

ASTES

Advances in Science, Technology & Engineering Systems Journal

Special Issue

Advancement in Engineering and Computer Science

2019

www.astesj.com
ISSN: 2415-6698

EDITORIAL BOARD
(Special Issue)

Editor-in-Chief

Prof. Passerini Kazmerski
Pritzker School of Molecular Engineering, University of Chicago, USA

Guest Editors

Prof. Wang Xiu Ying
Chongqing University, China

Prof. Yu Xiao Yan
Chongqing Normal University, China

Prof. María Jesús Espinosa Trujillo
Universidad Tecnológica Metropolitana, Mexico

Prof. Ahmad Yusairi Bani Hashim
Universiti Teknikal Malaysia Melaka, Malaysia

Prof. Mohamed Abdelaziz Hassan Eleiwa
University of Hail, KSA

Prof. Nicolae Tudoroiu
John Abbott College, Canada

Editorial

The Special Issue on *Advancement in Engineering and Computer Science (2019)* in the *Advances in Science, Technology and Engineering Systems Journal (ASTES Journal)* presents a comprehensive and contemporary collection of research that underscores the deepening synergy between engineering disciplines and computer science. The year 2019 marked a period of accelerated integration, where computational intelligence, data-driven methodologies, and software-centric approaches became central to innovation across engineering domains. This special issue captures these developments by bringing together contributions that reflect both the theoretical foundations and practical implementations shaping the future of engineering and computing.

A defining theme of this issue is the pervasive role of computational technologies in transforming engineering practice. Advances in artificial intelligence, machine learning, and data analytics have enabled engineers to design systems that are more adaptive, efficient, and capable of handling complex, real-time challenges. The research presented explores applications of these technologies in areas such as predictive modeling, system optimization, image and signal processing, and intelligent control systems. These approaches have significantly enhanced the ability to analyze large datasets, automate decision-making processes, and improve overall system performance.

The issue also highlights the rapid evolution of software engineering and its critical role in supporting modern technological systems. Contributions examine innovative software architectures, algorithmic developments, and programming frameworks that facilitate the development of scalable, secure, and high-performance applications. From cloud computing and distributed systems to cybersecurity and software reliability, the studies demonstrate how advancements in computer science are enabling more robust and flexible engineering solutions.

Another important aspect addressed in this special issue is the integration of hardware and software systems. The convergence of embedded systems, Internet of Things (IoT) technologies, and cyber-physical systems has led to the development of intelligent environments where physical processes are seamlessly connected with computational platforms. The research showcases how such integration enables real-time monitoring, control, and optimization in applications ranging from smart infrastructure and industrial automation to healthcare and environmental monitoring.

Innovation in computational modeling and simulation is also prominently featured. The use of advanced numerical methods, simulation tools, and virtual prototyping techniques has become essential for analyzing complex engineering systems and reducing development time and cost. The contributions illustrate how these tools are being applied to improve design accuracy, enhance system reliability, and support informed decision-making across various engineering fields.

The editorial team expresses its sincere gratitude to the authors, reviewers, and contributors for their valuable efforts in making this special issue possible. Their dedication has ensured the publication of high-quality research that advances both engineering and computer science. The works presented in this issue reflect a continuing shift toward intelligent, interconnected, and software-driven systems, offering meaningful insights and inspiration for future research and technological advancement.

Guest Editor

Prof. Wang Xiu Ying

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Special Issue **March 2019**

CONTENTS

Detection the Places of the Heat Energy Leak on the Underground Thermal Pipelines Using the Computer System
by Iryna Zhuravska, Dmytro Lernatovych and Oleksandr Burenko

Logic Error Detection System based on Structure Pattern and Error Degree
by Yuto Yoshizawa and Yutaka Watanobe

Q2 YouTube: Quantitative and Qualitative Information Analysis based Influencer-aware YouTube Channel Ranking Scheme
by Ji Hyeon Lee and Hayoung Oh

Microstrip Patch Antenna for Ultra-Wideband Applications
by QADDI Mohamed Hamza, SRIFI Mohamed Nabil and MHARZI Hassan

Productify News Article Classification Model with Sagemaker
by Johannes Linden, Xutao Wang, Stefan Forsstrom and Tingting Zhang

Melanoma detection using color and texture features in computer vision systems
by Antonio Fuduli, Pierangelo Veltri, Eugenio Vocaturo and Ester Zumpano

Emulation of Bio-Inspired Networks
by Zdenek Kolka, Viera Biolkova, Dalibor Biolek and Zdenek Biolek

Application of Open-Source Optimization Library "Extremum" to the Synthesis of Feedback Control of a Satellite
by Andrei Panteleev and Valentin Panovskiy

Fully Homomorphic Encryption Scheme Based On Complex Numbers
by Khalil Hariss, Maroun Chamoun and Abed Ellatif Samhat

Low-Cost and Accurate Computational System for Efficiency Measures over Photovoltaic Arrays
by Héctor L. Núñez-Ramírez, Gloria G. Carvalho-Kassar, Freddy C. Brito-Maestre, Yaremi I. Gamboa-Maldonado, Luis A. Santos-Avendaño, Carlelines Gavidia-Toro, Orlando Villarroel-Ramos and Dino Di Rosa-Ulloa

Detecting Malicious Assembly using Convolutional, Recurrent Neural Networks
by Michael Santacroce, Daniel Koranek and Rashmi Jha

The Study of PSi & PSo Algorithm for Reducing Power of the Mobile Communication Network
by Jongsin Kim, Yonggil Choi and Younghoon Oh
Walker Ergonomic Design for Visually Impaired Elderly
by Fang-Lin Chao

Frequency-Tunable Narrow-Band Ladder-Shape Microstrip Patch Antenna for TV Applications
by Maria Moussa, Mervat Madi and Karim Kabalan

Creating a Digital Twin: Simulation of a Business Model Design Tool
by Kira Rambow-Hoeschele, Nick Giani Rambow, Matthias Michael Hampel, David Keith Harrison and Bruce MacLeod Wood

Flame Distribution and Attenuation in Narrow Channels Using Mathematical Software
by Bayan Sapargaliyeva, Aigul Naukenova, Bakhyt Alipova and Javier Rodrigo Ilarri

An Expanded Derivation of Suppressing Anisotropic Birefringence in a Rectangular Waveguide
by Gregory Mitchell

Performance Analysis of Thyristors Switched Capacitors used for Reactive Power Compensation of Induction Motor
by Dmitry Ivanovich Panfilov, Ahmed Elsayed ELGebaly, Michael Georgievich Astashev and Alexander Nikolaevich Rozhkov

Fast Determination of Tsunami Source Parameters
by Mikhail Lavrentiev, Dmitry Kuzakov and Andrey Marchuk

Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Case Study
by Renato Mio, Midori Sanchez, Quino Valverde, José Lara and Francisco Rumiche

All-Pass RC-Filters Architecture with Independent Adjustment of the Main Parameters Based on Differential Difference Amplifiers
by Darya Denisenko, Nikolay Prokopenko and Nikolay Butyrlagin

Assessment of the Quality and Sustainability Implications of FIFO and LIFO Inventory Policies through System Dynamics
by Phoebe Lim Ching, Jose Edgar Mutuc and John Anthony Jose

Multi-Step Iteration Algorithm of Total Asymptotically Quasi-Nonexpansive Maps
by Salwa Salman Abed and Zahra Mahmood Mohamed Hasan

Intelligent Foundations for Knowledge Based Systems
by Mukundan Kandadai Agaram

Development of Smart Technology for Complex Objects Prediction and Control on the Basis of a Distributed Control System and an Artificial Immune Systems Approach
by Samigulina Galina and Samigulina Zarina

A Study on Development of Evaluation Metrics for Learners in Physical Computing
by Woochun Jun
Analysis of Fuzzy Data for the Selection of Ingredients of the Polymer Composition by Technical Functions in Quality Management
by Ilya Vasilyevich Germashev, Tatyana Petrovna Mashihina, Evgeniya Vyacheslavovna Derbisher, Vyacheslav Evgenyevich Derbisher and Mikhail Alexeevich Kharitonov

A Systematic Mapping Study of Cloud, Fog, and Edge/Mobile Devices Management, Hierarchy Models and Business Models
by Isaac Odun-Ayo, Rowland Goddy-Worlu, Victor Geteloma and Emanuel Grant

An Extension of Throughput Drop Estimation Model for Three-Link Concurrent Communications under Partially Overlapping Channels and Channel Bonding in IEEE 802.11n WLAN
by Kwenga Ismael Munene, Nobuo Funabiki, Md. Manowarul Islam, Minoru Kuribayashi, Md. Selim Al Mamun and Wen-Chung Kao

Toward a Smart Campus Using IoT: Framework for Safety and Security System on a University Campus
by Alghamdi Abdullah, Mohammed Thanoon and Anwar Alsulami

Multimedia Application Development With Islamic Critical Reflection Through 3-2-1 Technique for Novice Teacher Internship Program
by Evi Fatimatur Rusydiyah, Rakhmawati Rakhmawati, Eni Purwati, Moh. Hafiyusholeh and Ahmad Hanif Asyhar

pH Neutralization Process Control based on Active Disturbance Rejection Control
by Mohanad Hamad Eljack Elameen, Mona Bahaeldin Musa Hago and Wen Tan

Deep Feature Representation for Face Sketch Recognition
by Weiguo Wan and Hyo Jong Lee

Forecasting Bio-economic Effects in the Milk Production based on the Potential of Animals for Productivity and Viability
by Gennadii Georgievich Cherepanov, Anatolii Ivanovich Mikhalskii and Zhanna Anatolievna Novosrltseva

Survey on Semantic Similarity Based on Document Clustering
by Rowaida Khalil Ibrahim, Subhi Rafeeq Mohammed Zeebaree and Karwan Fahmi Sami Jacksi

Aggrandized Random Forest to Detect the Credit Card Frauds
by Jisha. Mulanjur Vadakara and Dhanasekaran Vimal Kumar

Automatic Stitching of Medical Images Using Feature Based Approach
by Kyi Pyar Win, Yuttana Kitjaidure and Kazuhiko Hamamoto

Experimental Evaluation of Transmission between Two XBee Modules Using Radio-over-Fiber Technique
by Luis Alejandro González Mondragón, Leidy Johana Quintero Rodríguez, Ana Gabriela Correa Mena, Jorge Rodríguez Asomoza, Alejandro García Juárez and Ignacio Enrique Zaldívar Huerta

Sensor Based on-the-go Detection of Macro Nutrients for Agricultural Crops
by Shraddha Balayya Hiremath, Raghavendra Shet, Nikita Patil and Nalini Iyer

Optimization Method of Wideband Multilayer Meander-Line Polarizer using Semi-Analytical approach and Application to 6-18GHz Polarizer including test with Horn Antenna
by Wafa Abdouni-Abdallah, Muhammad Saeed Khan, Athanasios Konstantinidis, Anne-Claude Tarot and Aziz Ouacha

Effects of Using Fuzzy Material Handling Inputs in the Genetic Algorithm for Machine Layout
by Wanwanut Boongsood and Chiranuwat Jadram

Development of Teaching Aids Selection Standards in Maker Education
by Woochun Jun

Educational Domain Management Knowledge Content Identification and Knowledge Updating Method, Based on Enterprise Management Information Interactions
by Jurij Tekutov, Saulius Gudas, Vitalijus Denisovas and Julija Smirnova

Optimal Discrete-time Sliding Mode Control for Nonlinear Systems Subject to Input Constraints
by Olfa Jedda and Ali Douik

Multi-Restricted Area Avoidance Scenario Using Hybrid Dynamical Model and Its Predictive Controller
by Sutrisno, Widowati, R. Heru Tjahjana, Sunarsih and Kartono

Inference Rule of Collision Risk Index based on Ship Near-Collision via Adaptive Neuro Fuzzy Inference System
by Ho Namgung

Linearity Improvement of VCSELs based Radio over Fiber Systems utilizing Digital Predistortion
by Muhammad Usman Hadi, Jacopo Nanni, Olivier Venard, Geneviève Baudoin, Jean-Luc Polleux, Pier Andrea Traverso and Giovanni Tartarini

Smart Meter Data Analysis for Electricity Theft Detection using Neural Networks
by Priyamvada Chandel and Tripta Thakur

Feature Selection for Musical Genre Classification Using a Genetic Algorithm
by Abba Suganda Girsang, Andi Setiadi Manalu and Ko-Wei Huang

An Adaptive Fuzzy-Sliding Mode Controller for The Bridge Crane
by Trinh Luong Mien
An Efficient Automotive Paint Defect Detection System
by Sohail Akhtar, Adarsh Tandiya, Medhat Moussa and Cole Tarry

Model-Driven Engineering Infrastructure and Tool Support for Petrochemical Industry Automation
by Thaise Poerschke Damo, Leandro Buss Becker and Fabio Paulo Basso

Effects of Educational Support Robots using Sympathy Expression Method with Body Movement and Facial Expression on the Learners in Short and Long-term Experiments

by Yuhei Tanizaki, Felix Jimenez, Tomohiro Yoshikawa, Takeshi Furuhashi and Masayoshi Kanoh

Quantitative Traffic Congestion Analysis Approach in Ahmedabad
by Tsutomu Tsuboi

The Impact of Using Upgraded Biogas on Generator Performance
by Wilson Babu Musinguzi and Ibrahim Luqman Mpungu

Well Balanced Multi-value Sequence and its Properties Over Odd Characteristic Field
by Md. Arshad Ali, Yuta Kodera, Md. Fazle Rabbi, Takuya Kusaka, Yasuyuki Nogami, Satoshi Uehara and Robert H. Morelos-Zaragoza

Photoluminescence Investigation of Inhomogeneous Porous P-type Si
by Róbert Brunner, Pavel Vojtek, Zuzana Zábudlá and Emil Pinčík

Prediction of Non-Communicable Diseases Using Class Comparison Data Mining
by Ammar Al-Dallal and Amina Al-Moosa

Dysphoria Detection using EEG Signals
by Norhaslinda Kamaruddin, Mohd Hafiz Mohd Nasir and Abdul Wahab

Integrating Diacritics Restoration and Question Classification into Vietnamese Question Answering System
by Bui Thanh Hung

Study of Performance of Bio- Inspired Strategies Applied to Pursuit Evasion Game Under Feedback Laws
by Lairenjam Obiroy Singh and Rajagopalan Devanathan

Eye Feature Extraction with Calibration Model using Viola-Jones and Neural Network Algorithms
by Farah Nadia Ibrahim, Zalhan Mohd Zin and Norazlin Ibrahim

A Critical Analysis of Topics in Software Architecture and Design
by Janet Bishung, Ooreofe Koyejo, Adaugo Okezie, Boma Edosomwan, Sylvester Ani, Abisola Ibrahim, Austin Olushola and Isaac Odun-Ayo

Providing Underlying Process Mining in Gamified Applications – An Intelligent Knowledge Tool for Analyzing Game Player's Actions
by Anna Tatsiopoulou, Christos Tatsiopoulos and Basillis Boutsinas

Improvement opportunities of a Simulation/Expert System Approach for Manufacturing System Sizing: A review and proposal
by Lajmi Imen, Masmoudi Wassim, Elleuch Mounir and Chtourou Hedi

Brain-inspired IoT Controlled Walking Robot – Big-Foot
by Anna Lekova, Ivan Chavdarov, Bozhidar Naydenov, Aleksandar Krastev and Snezhanka Kostova

Optical Braille Recognition Software Prototype for the Sinhala Language
by Shanmuganathan Vasanthapriyan and Malith De Silva

Virtualization in Cloud Environment: Bandwidth Management
by Isaac Odun-Ayo, Blessing Idoko and Temidayo Abayomi-Zannu

Fuzzy Simulation of Historical Associative Thesaurus
by Olena Basalkevych and Olexandr Basalkevych

Multi-Band Hand of God Antenna with Functionality Variation with Use of Slots and Feed Position
by Mervat Akram Madi and Karim Youssef Kabalan

Cloud Service Level Agreements and Resource Management
by Isaac Odun-Ayo, Blessing Udemezue and Abiodun Kilanko

Transformation of Solar Energy to Electricity with Photovoltaic Systems, Reduction of Electrical Consumption and Installation of LED Lamps in the ITSAO
by Jeremias Bravo Tapia, Juan José Rivera Toledo, Francisco Ramos Guzmán and Yoxkin Estévez Martínez

Application of The Half-Sweep Egsor Iteration for Two-Point Boundary Value Problems of Fractional Order
by Rostang Rahman, Nur Afza Mat Ali, Jumat Sulaiman and Fatihah Anas Muhiddin

Automatic Service Orchestration for e-Health Application
by Anatolii Petrenko and Bogdan Bulakh

Exchange Rate Modeling: Medium-Term Equilibrium Dynamics
by Anton Kuzmin

Event Monitoring using Distributed Pattern Recognition Approach on Integrated IoT-Blockchain Network
by Anang Hudaya Muhamad Amin, Nazrul Muhaimin Ahmad and Subarmaniam Kannan

Effects of Different Activation Functions for Unsupervised Convolutional LSTM Spatiotemporal Learning
by Nelly Elsayed, Anthony Maida and Magdy Bayoumi

A Novel Strategy For Prompt Small Cell Deployment In Heterogeneous Networks
by Dorathy Abonyi

Localization of Emerging Leakages in Water Distribution Systems: A Complex Networks Approach
by Matteo Nicolini

DSTATCOM-Fuel Cell System on Radial Low-voltage Distribution Network for Mitigating Voltage Rise Caused by High Penetration of Photovoltaic Systems

by Dylon Hao Cheng Lam, Jianhui Wong, Yun Seng Lim and Jimmy Chin Yong Hee

Conversion of 2D to 3D Technique for Monocular Images using Papilio One
by Priya Arokia and Anupama Patil

Investigating The Detection of Intention Signal During Different Exercise Protocols in Robot-Assisted Hand Movement of Stroke Patients and Healthy Subjects Using EEG-BCI System
by Maryam Butt, Golshah Naghdy, Fazel Naghdy, Geoffrey Murray and Haiping Du

Resonator Influence Simulation of Designed Close-Open Standing Wave Thermoacoustic Engine
by Sugiyanto, Samsul Kamal, Joko Waluyo and Adhika Widyaparaga

Noise Cancellation Algorithm Based on Air- and Bone-Conducted Speech Signals by Considering an Unscented Transformation Method
by Hisako Orimoto and Akira ikuta

The Role of Information Technology Human Capability in the Implementation of Information Technology Governance (ITG): A Systematic Literature Review on Malaysian Organizations
by Nor Aziati Abdul Hamid, Chin Wei Liew, Nor Hazana Abdullah and Siti Sarah Omar

Estimating Academic results from Trainees' Activities in Programming Exercises Using Four Types of Machine Learning
by Toshiyasu Kato, Yuki Terawaki, Yasushi Kodama, Teruhiko Unoki and Yasushi Kambayashi

Improving the Performance of Hadoop Framework Using Optimization Process in the Information Management
by Ramachandran Ravi Sowmiyasree, Nachimuthu Maheswari and Manickam Sivagami

Overview on 5G Radio Frequency Energy Harvesting
by Sanae El Hassani, Hind El Hassani and Noureddine Boutammachte

A Modular Multipurpose, Parameter Centered Electronic Health Record Architecture
by Norbert Maggi, Luca Douglas Magnoni, Carmelina Ruggiero, Roberta Gazzarata and Mauro Giacomini

Hybrid Technique for Enhancing Underwater Image in blurry conditions
by Chrispin Jiji and Nagaraj Ramrao

Using Privacy Enhancing and Fine-Grained Access Controlled eKYC to implement Privacy Aware eSign
by Puneet Bakshi and Sukumar Nandi

Location Prediction based on Variable-order Markov Model with Time Feature and User's Spatio-temporal Rule
by Ying Xia, Yu Gong, Xu Zhang and Hae-young Bae

Computer Security as an Engineering Practice: A System Engineering Discussion
by Robert M. Beswick

Hidden Surface Removal for Interaction between Hand and Virtual Objects in Augmented Reality
by Takahiro Ishizu, Makoto Sakamoto, Masamichi Hori, Takahiro Shinoda, Takaaki Toyota, Amane Takei and Takao Ito

Evaluation of the Effects of Bidding Strategy with Customized Pricing on the Individual Prosumer in a Local Energy Market
by Borislava Spasova, Daisuke Kawamoto and Yoshiyasu Takefuji

Distribution of Bit Patterns in Binary Sequence Generated Over Sub Extension Field
by Md. Arshad Ali, Yuta Kodera, Takuya Kusaka, Yasuyuki Nogami, Satoshi Uehara and Robert H. Morelos-Zaragoza

An Analysis of Social Media Usage in Winery Businesses
by Constantina Costopoulou, Maria Ntaliani and Filotheos Ntalianis

A Formal Ontology-based Framework Towards an Intelligent Temporal Scheduler
by Patience Usoro Usip, Udoinyang Godwin Inyang, Daniel Ekpenyong Asuquo and Enobong Umoren

A Proposal of Control Method Considering the Path Switching Time of SDN and Its Evaluation
by Kosuke Gotani, Hiroyuki Takahira, Misumi Hata, Luis Guillen, Satoru Izumi, Toru Abe and Takuo Suganuma

Transmission Line Restoration Using ERS Structure
by Sushri Mukherjee, Sumana Chattaraj, Md. Irfan Khan, Dharmbir Prasad, Pradip Barua and Harish Agarwal

Supporting Better Physical Activity in a Smart City: a Framework for Suggesting and Supervising Walking Paths
by Mário Rodrigues, Rita Santos, Alexandra Queirós, Anabela Silva, João Amaral, Patrícia Simões, Jorge Gonçalves, Ciro Martins, António Pereira and Nelson Pacheco da Rocha

Development of Wavelet-Based Tools for Event Related Potentials' N400 Detection: Application to Visual and Auditory Vowelling and Semantic Priming in Arabic Language
by Nadia Mountaj, El-Mehdi Hamzaoui, Mohamed Majid Himmi and Mireille Besson

A Balanced RF Power Amplifier using the Improvement 3-dB Quadrature Hybrid Couplers for Mitigation of the Reverse Intermodulation in FM Broadcasting Systems
by Panya Hantula and Rangsan Tongta

Effectiveness and Comparison of Digital Substations Over Conventional Substations
by Aashir Waleed, Umar Siddique Virk, Muhammad Tanveer Riaz, Shaikh Bilal Mehmood, Saeed Ahmad, Muhammad Rameez Javed and Ali Raza

Optimized Design of PM Halbach Array Linear Generator for Sea Wave Energy Converters Operate at Maximum Power Transfer
by Ahmed Elsayed ELGebaly and Mohamed Kamal El-Nemr

A Fuzzy-Based Approach and Adaptive Genetic Algorithm in Multi-Criteria Recommender Systems
by Mohamed Hamada, Abdulsalam Latifat Ometere, Odu Nkiruka Bridget, Mohammed Hassan and Saratu Yusuf Ilu

Marginalizing Last Mile Logistics Cost through 4th Party Milk Run
by Robert de Souza, Linda William and Cher Kian Lee

Estimation of Target Maneuvers from Tracked Behavior Using Fuzzy Evidence Accrual
by Stephen Craig Stubberud, Kathleen Ann Kramer and Allen Roger Stubberud

Conceptual Architecture for the Continuity if Workflow in Activities on Multi-Devices: Case Study Co-Kitchen
by Mirian Janeth Avalos-Viveros, Luis Gerardo Montané-Jiménez, Gabriela Sánchez Morales, Carmen Mezura-Godoy and Edgard Benítez-Guerrero

Material, Structural Optimization and Analysis of Visible-Range Back-Illuminated OPFET photodetector
by Jaya V. Gaitonde and Rajesh B. Lohani

The Fusing Framework Between Lifting Carrier and Tractor-Trailer for Modern Transportation
by Ha Quang Thinh Ngo, Thanh Phuong Nguyen and Hung Nguyen

Study and Implementation of Various Image De-Noising Methods for Traffic Sign Board Recognition
by M. Monica Subashini, Abhinav Deshpande and Ramani Kannan

A Lightweight, Hardware-Based Support for Isolation in Mixed-Criticality Network-on-Chip Architectures
by Giacomo Valente, Paolo Giammatteo, Vittoriano Muttillo, Luigi Pomante and Tania Di Mascio

Multiple-Optimization based-design of RF Integrated Inductors
by Houcine Marouani, Amin Sallem, Mondher Chaoui, Pedro Pereira and Nouri Masmoudi

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 6, 1-9 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Q^{2}YouTube: Quantitative and Qualitative Information Analysis based Influencer-aware YouTube Channel Ranking Scheme

Ji Hyeon Lee[1], Hayoung Oh[*2]

[1]*Ajou University, Department of English Language and Literature, Suwon, Korea*

[2]*Ajou University, DASAN University College, Suwon, Korea*

ARTICLE INFO

Article history:
Received: 14 June, 2019
Accepted: 08 October, 2019
Online: 20 November, 2019

Keywords:
Quality aware YouTube Ranking scheme
Flipped learning maximization
Influencer Analysis
Data analysis

ABSTRACT

With the development of big data, artificial intelligence and deep learning, various social information networks are becoming exponentially intelligent. Of all the various social networks, YouTube is so popular that it is called the YouTube era. Not only video viewers, but also actual video producers, influencer youtubers, are increasing, allowing individuals as well as operators to use and market contents on various topics. As the number of users and contents increases, the choice of information increases, but it is more difficult for individual and business owners to select information that meets their needs. Therefore, this study aims to analyze the channel of the subject that the user needs from various angles and provide the ranking of the channel to individuals and businesses. We crawl the channels and measure the average awareness and influencer youtuber of channels and channel videos by analyzing the qualitative data of quantitative data and comments which are statistical information. As a final case study, we recommend the English learning channel to individual users based on numerical data statistics and emotional analysis results to show the maximum of flipped learning effect regardless of time and space. Plus, we prove the maximum of the effectiveness of marketing by influencer channel which is influential to operators in the medium and long term.

1 Introduction

There are about 3.4 billion people using social media (SNS) among 7.6 billion people worldwide. Compared with the year 2018, about 280 million people started the SNS within a year. Especially with the development of big data, artificial intelligence and deep learning, various social information networks are becoming exponentially intelligent. One of them, YouTube is the most famous social platform for satisfying various objects such as users as well as businesses on top of the recommender system. However, as the development of technology and the emergence of big data have caused the exponential growth of YouTube channels, users are having difficulties in selecting personalized videos. In other words, users have difficulty in selecting a channel on top of the basic recommendation of YouTube, and they spend additional time searching for a good channel or comparing contents directly [1]–[4]. In aspects of the business operator, the new influencer marketing on top of YouTube data analysis is also emerging as a new and successful marketing strategy. Influencers are influential individuals, which can include the general public as well as entertainers and SNS stars. Using the recommedation on top of the influencer aware YouTube data analysis, the general public can easily consume the contents and at the same time be satisfied with the content.

As the application case studies based on influencer aware YouTube data analysis, in this paper we introduce the possibility of the flipped learning maximization of user education as well as business marketing satisfaction.

Regardless of age, interest in lifelong education has increased in various fields such as English and Chinese, and various contents based on smartphone such as internet lecture, radio, and learning application are attracting attention. As the number of smartphone users increases, demand for online content is growing more than offline. However, since the above-mentioned online service requires the motivation of the user to voluntarily learn, the more the time passes, the less motivated and the less interested in learning.

The popularity of the YouTube-based beyond time and space learning is increasing due to the characteristics of these

[*]Corresponding Author: Hayoung Oh, Address, Contact No & hyoh79@gmail.com

online learners. Internet lectures are similar to flipping learning in pedagogy, which is not an existing concept of 'off-line class', but can be learned anytime and anywhere by transcending time and space. In particular, Youtube provides big data that can be analyzed statistically of videos that are composed by influencer Youtuber in a free environment and have lots of choices in social network, so it can maximize the effect of flipped learning education. In other words, influencer aware YouTube video ranking based education does not require as much concentration as text-oriented education, and users can use it for free time very effectively. In particular, influencer aware YouTube video ranking based online education can maximize the effectiveness of education because it shows the candid reviews of ordinary users and the cases that are natural but helpful in real life on top of big data analysis quantitatively and qualitatively. Similar to the general online education methodologies, the influencer YouTube aware flipped learning education can solve the physical limitations and anyone can be educated in real time or streaming (pre-recorded streaming).

In aspects of the business operators, influencer aware YouTube data analysis based the marketing has become a new and solid strategy. As of February 2017, people spent 25.7 billion minutes on YouTube, a video platform. This shows that YouTube spending time has more than tripled in the past two years compared to 2015. In other words, research in video and photo-based YouTube influencers analysis techniques is critical in aspects of users and businesses, as it indicates that people are more interested in video content than text content over time. Specifically, if the business operators recommend the personalized video to the users by referring the features of influencer YouTuber with high quantitative and qualitative data analysis, the satisfactions of the users as well as businesses can be maximized. Different from the previous related schemes of YouTube big data anlysis, we consider velocity, veracity(Truth), value as well as volume and variety among five characteristics of big data.

As a result, in this paper, we propose an influencer aware YouTube ranking scheme for the channel recommendation technique to maximize the satisfaction of users as well as the businesses. As mentioned before, we show the flipped learning effect through explicit and implicit analysis of various data of the influencer provided by YouTube centering on the YouTube English learning channel. Based on numerical data such as the number of views and likes of channels collected by using Google and R's tuber package for four channels with the same ranking of the social channel. The primary statistic is given and those of positive and negative are calculated through emotional analysis of the comment. Finally, based on the numerical data statistics and emotional analysis results, the English learning channel is recommended to the user to maximize the flipped learning effect and the business satisfaction regardless of time and space. In future, the proposed method can be used for flipped learning of various contents because it can be used for analysis of big data of influencer aware YouTube channel in various fields as well as language field.

2 Related Work

Recently, in an environment where big data has grown exponentially, personalized recommendations are a great way to efficiently search for information and discover content with high satisfaction.

2.1 Recommender system

Based on how designing recommender systems, the recommendation systems are usually classified into the following categories [5]–[8]: Collaborative recommendation systems; Content-based recommendation systems; and Hybrid recommendation systems.

Collaborative filtering is currently the most wise and popular approach because of focusing more on rating-based recommendations. The collaborative methods predict the utility, preference or taste of items per a particular user using the items previously rated by other users as shown in figure 1.

On the other hand, content-based approaches are generally based on classical information retrieval such as keywords as content descriptors. Therefore, the content-based methods are very popular for recommending documents, news articles or web pages. Lastly, the hybrid systems [9, 10] is to combine two or more different approaches to maximize desirable properties.

	$Item_1$	$Item_2$	$Item_3$	...	$Item_n$
$User_1$	0	3	4	...	5
$User_2$	1	1	2	...	?
:	:	:	:	...	:
$User_m$	4	?	5	...	4

Figure 1: Collaborative recommendation systems

2.2 Social Network aware Recommender system

Youtube, which efficiently provides dynamic data such as video among various contents, is gaining popularity around the world. Although a basic recommendation system on the YouTube platform is installed to recommend high-profile videos, personal satisfaction is not high. Three factors can be used to determine ranking of recommendation systems. These are each of video quality, user specificity and diversity. The quality of the video includes quantitative information such as views, comments, favorites, upload time, and so on. User specificity is related to individual taste and preference. The way to check this is to consider the number of views or watch time of the watch record. Thereafter, in order to maintain the diversity of the video, there is a method of limiting the number of videos of the same channel [11]. In [12], the authors designed the recommendation system in consideration of the factors. In [13]–[16], different studies have attempted to verify the significance of influencer decision through R-based data analysis rather than heuristic

judgment. In this case, the data commonly used are nine characteristics (i.e., features) such as total number of views of the channel, number of videos of the channel, average number of videos views of the channel, number of channel subscribers, likes and dislikes of each video. In [13], the authors tried to normalize covariance of each variable and analyzed the correlation between each variable with views of the subordination variable. Plus, they proved that the average number of views on the channel and the number of subscribers on the channel had a positive relationship with the number of views. However, the degree of correlation was insufficient, and the statistical data alone have limitations in producing significant results, such as the determination of video recognition, etc. In [14], the authors first focused on the influencer YouTube data analysis with accompanying the 50 channels made using the blind date app 'swipe'. With the data, the release date of the video, likes , dislikes, the number, video length, hash tag count, total number of views of channels, number of videos of channels, average number of video views of channels, number of channel subscribers were analyzed as variables, and Principal Component Analysis and Multiple Regression Analysis were performed as well as R-based data analysis. The cumulative contribution rate of PCA1 and PCA2 was more than 80% of the total to compared these two factors to determine the importance of each variable. For the first component, the average number of views on the channel, comments on swipe videos, views on swipe videos, and likes of videos were -1.25, -2.6, -2.49, and -2.83, respectively, showing a strong negative value, while the video length of swipe video and the hashtags of swipe video were 3.31 and 1.95, respectively, showing a strong positive value. The second component showed a negative value with the number of views and video subscribers, and a weak positive value with the length of the video and tag count. In addition, author concluded that the number of videos of the channel is most influential on the influencer attribute decision model among the variables when the two components are judged synthetically. In Multiple Regression Analysis, only the number of videos reveals significance. However, the limitation of the Principal Component Analysis was that the differences among the variables were large, and for Multiple Regression Analysis, there is a mismatch with the Principal Component Analysis results, therefore the study also has a limit. In [15], the authors analyzed the comment of each video with time to demonstrate the continuing influence of YouTube Influencer. The group is classified as "Large group" and "Small group" based on the total number of comments. In the case of "Large group", the number of comments per an hour is similar and the decreasing width is constant over time. In the case of "small group" is assumed that the number of comments per an hour is irregular, and that the decreasing width is not constant over time. Accordingly, the paper selected the "Big3" of the group and used the decreasing width and consistency obtained through the existing "Large group" and "Small group" as an evaluation indicator. In the case of decreasing width, the smallest percentage was 3% and the largest percentage 2%. For consistency, the standard deviation was used, and the results were derived that the standard deviation of the "Large group" was relatively small compared to the standard deviation of the "Small group." This has proven that "Large group"'s influencer continues to receive people's attention in a stable manner. However, the limitations still exist because of the lack of connectivity with the recommendation system of YouTube through high-profile video analysis.

In [16], the authors proposed the importance of the analysis of the comments quantified through the questionnaire and suggested the link between the analysis of the comments and the generation of profits. According to the survey, 12 percent of users post their opinions on a regular basis, which means about 96 million people are actively writing their opinions, considering the size of the YouTube community is estimated at 800 million. Also, 34 % of the respondent of investigation read comments frequently and read comments 2-3 times after watching 53 % of the video. In the case of comments, those received a lot of likes if those were reasonable, practical and content-customized, and eventually showed that it could be used to increase the popularity of the videos and make profits. However, the existing study also has limitations due to the lack of connectivity with the recommendation system of YouTube through high-profile video analysis, and there is a limit to analyzing human contradictory behavior. In [17] the authors used YouTube comment analysis, TF-IDF (term frequency inverse document frequency) and Association Rule Analysis to select influencers. Comment analysis used the B-rated YouTube channel when searching for "English study", "English education" and "English learn" on the social blades that provide the YouTube ranking system. The comparison of TF (Terminal Frequency) and TF-IDF analysis then demonstrated that the TF-IDF analysis elicited more significant words. For example, when word analysis was conducted with the frequency of words (TF), common words related to learning such as "learn" and "lesson" or words that had nothing to do with the video content such as "love," "video" and "channel" were heavily weighted. In contrast, for TF-IDF analysis, words related to video content such as "resort," "plan," "proposal," "vegan," and "vegetarian" are expected to be talking about travel first and vegetarian second. Compared to TF analysis, TF-IDF extracts features of each video content and shows them well. Finally, the association among words was analyzed through the Association Rule Analysis. As a result of the Association Rule Analysis, the authors can see that words with high TF are generally generated as rules. In case of using TF-IDF, there is a limit to judge the association among words heuristically. In the case of Association Rule Analysis, the rules are generated based on the frequency of words in general, so that only meaningless words are repeated in the YouTube environment. However, the paper is considered to provide a cornerstone of the proposed technique because it found that the two methods can be only applied to select the influencers. Recommender systems of [18] collects information about users and businesses and how they are related. The recommender systems let users share their reviews about products, places, establishments or services. If a product receives more votes or reviews, it

becomes more popular among customers.

It is expected that higher number of positive reviews is beneficial for marketing's success of the product. Interactions based on user's followers are usually allowed by recommender systems. A user is considered a fan of another user when there's no personal relationship between them but still the former wants to see the latter's posts in his or her own main page. This type of social link is based on interests and affinities with celebrities, special products, places or common interests. When a review is posted, it is seen by the submitter's friends and followers, increasing its chances of readings about the reviewed product or place. The more social links a user has, the more persons will likely read the review. This social component has become an interesting aspect of recommender systems because it allows an increase of popularity of a product or place. In the case of the recommendation system mentioned in the thesis, the user-to-user social network is formed through the follow, and is also distinguished by the user-user fan through the interaction relationship. In general, reviews by users with many social networks (users with many fans) are more likely to be accessed by more users. These characteristics are an interesting part of the recommendation system, and many businesses approach these users for satisfaction. This proves that the existing research direction of recommending to companies for profit generation is reasonable. The current model excludes text analysis, but acknowledges that the analysis of comments exerts an influence on Influencer's numerical data.

3 System Model

The data used for the proposed scheme was based on 54 channels of influence, which produced promotional videos on the "Enligh Education" related channel. The videos are easy to use influencer attribute analysis because general public influencer can directly produce video and observe other users' reactions in various formats such as numerical information and comments on YouTube. YouTube's visible numerical data shows the date of creation of each video, likes, dislikes, length of the videos, number of hash tags, total number of views on the channel, the number of videos on channel, average number of views on the channel's video, and number of channel subscribers. The primary analysis of suggested techniques is to look at their correlation using various numerical data provided on YouTube quantitatively. At this time, the set hypothesis is that "if the correlation of each numerical data is high, can it be judged unconditionally that the recognition of the corresponding video is high?" In general, it is easy for general users to judge the quality of channel and video by trusting only the numerical data based ranking system of YouTube. However, since the average number of views of a channel is increased with only single click of a YouTube video, the actual playback time is unknown, such as whether the user actually viewed the videos, which parts of the videos were particularly popular, or whether they stayed for a certain period of time, rather than simply clicking them. In addition, YouTube does not provide access to statistics and visualization information (e.g., demographics information) of the audience, such as age, gender, etc., which videos are usually viewed highly by which user class. Therefore, we simultaneously analyze influencer YouTube data quantitatively and qualitatively as shown in figure 2

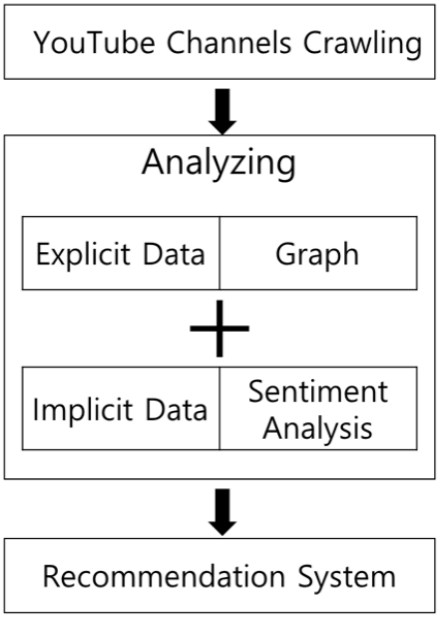


Figure 2: Recommendation System analysis flowchart

In detail, the English learning channels corresponding to B level, which is the self-rating of the social blade, were filtered based on the influencer criteria of the social network data. Those are the number of subscribers, English learning channels, which have a much higher number of views and likes, generally have a higher level awareness of videos as well as the number of videos with the long days of channel opening. We chose channels that were opened between 2016 and 2018 to allow for the latest videos of channel. The YouTube channel has its own concept, and English learning channels are largely divided into lecture type and talker type. One of the reasons that users prefer YouTube over Internet lectures is because they can learn English through content that is not formal and has a free atmosphere. Therefore, focusing on the talker-type channel except for the lecture type, we conducted research on ETK English, Interactive English, JForrest English and Real English With Real Teachers.

The study method used the statistical analysis software R tuber package. There are research [10] that used the chrome extension tool SelectorGadget. However, due to the limitations of the crawlable area due to the manual selection of the user and lack of real time property, considering the characteristics of R in which various packages exist, a tuber package for YouTube is used. We crawled the data on YouTube and explicitly extracted a statistical graph for the primary numerical social network data. Afterwards, we crawled the comment big data for each video on each channel through the package

for YouTube data access only and the Google open API, and performed the YouTube data qualitative analysis.

The overall research process is shown in figure 2

After crawling YouTube channel data in the Tuber package, we set up different research methods for each data type quantitatively and qualitatively. The social numeric data such as the number of likes and subscribers extracts the graph based on the mutual relationship and applies the qualitative approach through the emotional analysis to the comment data in which the opinions of actual users are reflected on the channel.

4 Proposed scheme

4.1 Research technique

YouTube provides statistics on channel and channel video. In this study, channel information and channel video information are distinguished from each other because channel and channel video are different. We use the number of subscribers to measure channel itself awareness. The recognition of channel videos is based on qualitative data analysis using comment data analysis as well as quantitative data such as number of views, likes, and comments. Therefore, the recognition of one channel, that is, the number of subscribers, is not popular and highly rated. After analyzing the quantitative and qualitative data of the channel and the channel video, the channels are ranked by synthesizing them.

4.1.1 Crawling

After selecting the influencer channels to be ranked, call the channel ID from the YouTube url address as follows and use the R tuber package to crawl the statistical information like followings. https://www.youtube.com/channel/ (Channel ID). When we enter the YouTube channel main screen, the above url address is generated, and the last part is the channel ID. The channel information is coded by entering the channel ID into the get_channel_stats function of the tuber package. The yt_search function calls up the video list of the channel, and the video statistics information is also retrieved using the ID of each video in the same manner as above. In order to organize the crawled statistical information by video and channel, the title and channel name are added to the same row and stored as a data frame. Each channel is refined in the same manner as described above, and then organized into one data frame. Then, column names such as "Title", "Likes", "Views", "Subscriptions", "Channel" The video title defines character type, number of likes, number of views, number of subscribers as integer type, and channel name as factor type.

The comment data corresponding to the qualitative data also uses the tuber package. A comment on the video is called up to generate a comment data frame for each channel, which consists of a comment and an video ID in one row. This identifies channel names as well as quantitative data.

4.1.2 Quantitative Data Analysis

The results of quantitative data analysis are graphically displayed for visualization. Quantitative data is a measure of channel awareness and channel video awareness, and uses statistics provided by YouTube. The awareness of the channel is the number of subscribers, and the video recognition per channel is confirmed by the number of views, likes, and comments. Since the number of subscribers and the number of comments are fixed, they can be displayed as a bar graph. The reason why the number of comments in a video is classified as a fixed value is because the comment does not need to consider the deviation for each video fairly. Since the number of comments is a measure of how much viewers are interested in the channel, we consider only fixed values equally. On the other hand, the number of views and likes are taken into account for distribution and the deviation is taken into account by video.

4.1.3 Qualitative data analysis

In this study, the comment data is analyzed at the same time as the statistical data provided by YouTube. Likes, in addition to the number of subscribers, we can rate those channels and videos with the feedback from people who actually watched the video. It is difficult to judge that the number of views is high because it can be climbed even if the user clicks the video once even if the viewer does not watch the video to the end. On the contrary, it is possible to determine whether the video is good or not by checking whether the user has stayed in the video for a longer time, that is, whether the video has been watched until the end. Comments on video, except spam, are mainly left after watching video, mainly video related content. The higher the ratio of the result, the comment and the video content match, the more the commenter will see the video. Therefore, the comment analysis can be used to determine whether the video is suitable for ranking highly. To understand the correlation between video and comment, we invoke video script and comment file and refine data. Sort videos and comment data classified by corpus by frequency. Then, the comment data is compared with the video data, and the matching word is denoted by 1, and the non-matching word is denoted by 0. Scoring function is used to score differently for each frequency, divided by the total number of corpus and expressed as a ratio.

That is, the proposed scheme used Term Frequency (TF) to measure the frequency of each video based on script and comments. If the comment data matches the script, a list of the words is generated. The more words and duplicates in scripts and comments, the more the subject and content match. As the number of words in the list increases, it means that the viewer has a high degree of understanding and understanding of the video. Therefore, the video is recommendable to other users with the increasing recommendation reliability by weighting the degree of correspondence between video and comment. The response of the viewer, that is, the content of the comment, becomes the criteria.

The proposed scheme also consider Inverse Document Frequency (IDF) for TF-IDF (Term Frequency-Inverse Document Frequency). The TF-IDF (Term Frequency-Inverse Document Frequency) is often used for keyword extraction and similarity between documents. The higher the value of the TF indicating the appearance frequency of a word, the higher the importance of the word is, but it is not an absolute measurement method. It is more likely that the word is often used in other documents, and the frequency and importance are inversely proportional. Thus, in many document clusters, we multiply the frequency of occurrence of the word IDF with TF to assign weights to the only frequently appearing words in a particular document. Therefore, we also considered TF-IDF per each channel video for the proposed scheme.

4.1.4 *Normalization*

Since quantitative data and qualitative data analysis can not be accessed from the same point of view, we use a normalization function. As shown in figure 3, we use the normalization function to integrate the quantitative and qualitative data analysis of each channel to provide the final channel ranking. As shown in figure 3, the reasons for putting the weight on the 20th word list are as follows. As a result of checking the list of words based on the experimental data, word matching ratios up to 20th were distributed on average regardless of the total number of words, and other words showed the same or subtle differences. Therefore, based on the TF-IDF, weights are assigned to the words from the first to the 20th.

Function 1 Scoring

```
1: if Words in MatchingList >=20
2:   for (i in 1:3)
3:     MatchingList[i] = 1.8
4:   for (i in 4:10)
5:     MatchingList[i] = 1.5
6:   for (i in 11:20)
7:     MatchingList[i] = 1.2
8: else
9:    for (i in 1:3)
10:     MatchingList[i] = 1.8
```

Figure 3: Normalization scoring function to integrate the quantitative and qualitative data analysis of each channel

The detailed rules for weighting criteria are as follows. The numbers 1 and 0 are assigned to the words that match each other. Since we place weight on words with a high percentage of match between video and comment content, they should be basically higher than the number 1 already granted. Therefore, based on various experiments as shown in figure 4, we have weighed the combinations of various numbers from 1.1 to 2 and confirmed the number that best matches the existing match ratio.

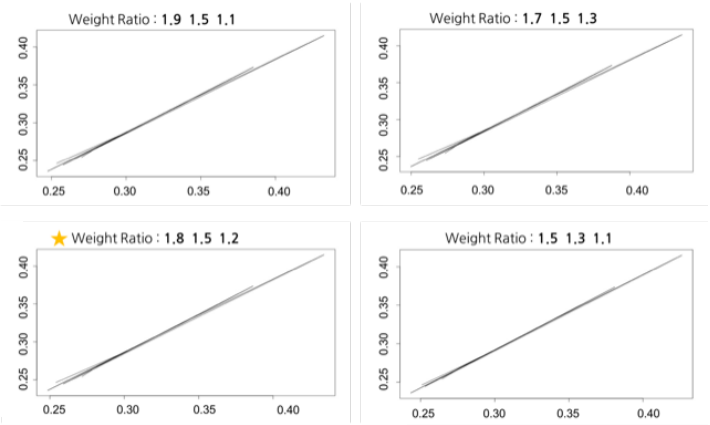


Figure 4: Weight ratio

4.2 *Experiment Data*

All the performance evaluation codes were written in the R environment suitable for data analysis, and the influencer channels with the subject of English learning were selected as the experimental subject. Four English language channels (English Coach, Interactive English, JForrest English, and Real English with Real Teachers) were selected based on the ratings of social blades that provide social rankings. In order to experiment with the channels that can be compared, we chose channels that can be compared with each other considering the channel information such as the total number of views, the number of subscribers, etc., except for the old channel. Quantitative data and qualitative data were analyzed for each channel. First, quantitative data analysis results are shown in figures figure 5–8.

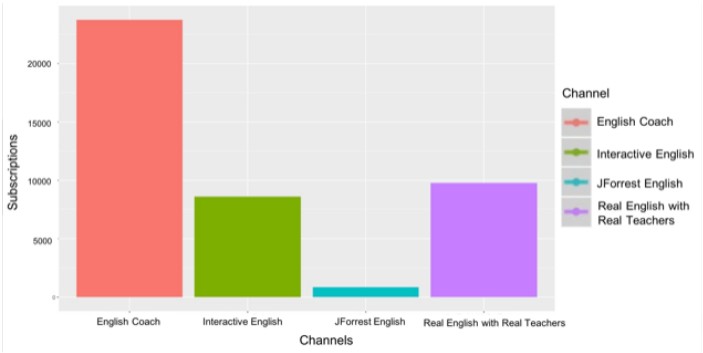


Figure 5: # of subscriber per channel

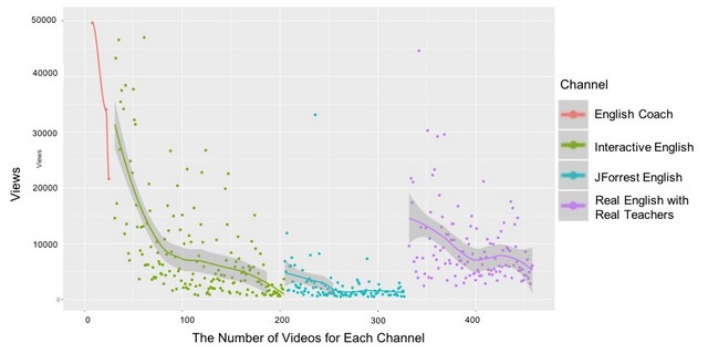


Figure 6: # of video views per channel

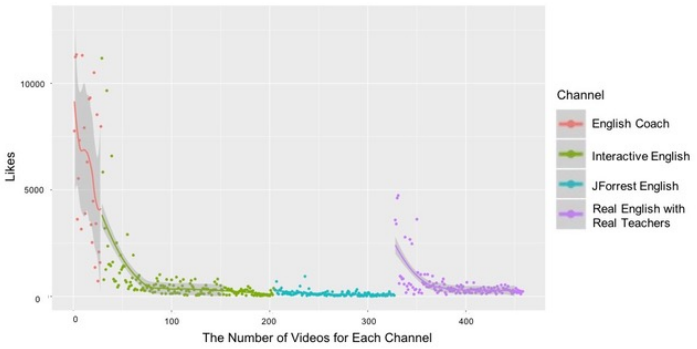


Figure 7: # of video likes per channel

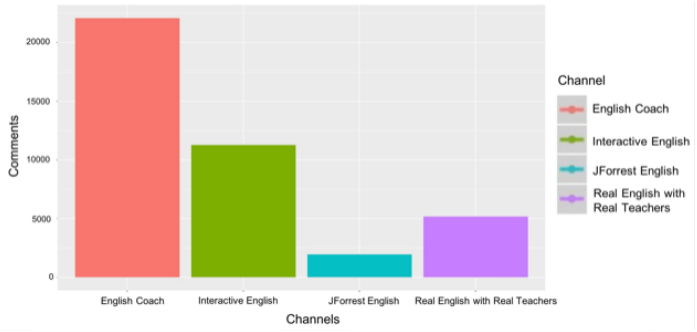


Figure 8: # of video comments per channel

Figure 9 and 10 show additional result graphs when considering different video data sets on the same channel.

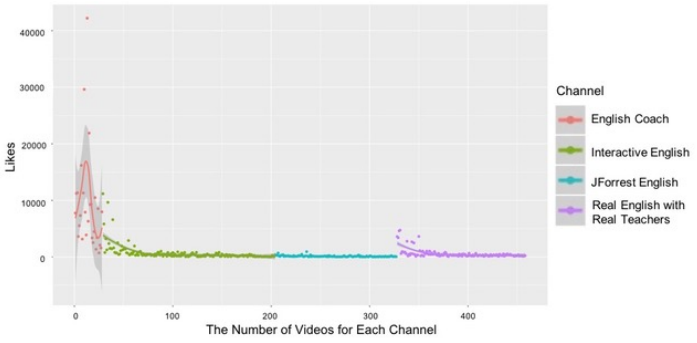


Figure 9: # of video likes per channel

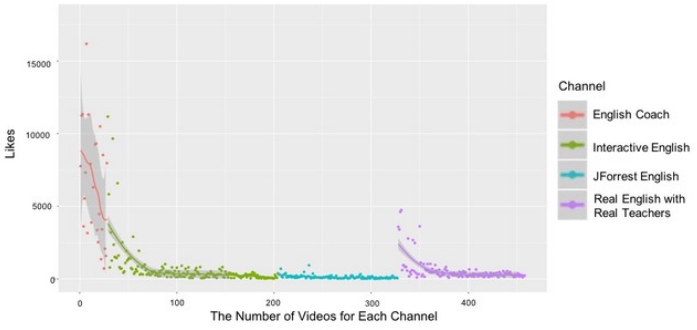


Figure 10: # of video likes per channel

Qualitative data analysis is shown in figure 11. The correlation between videos and comments is compared and scores are given.

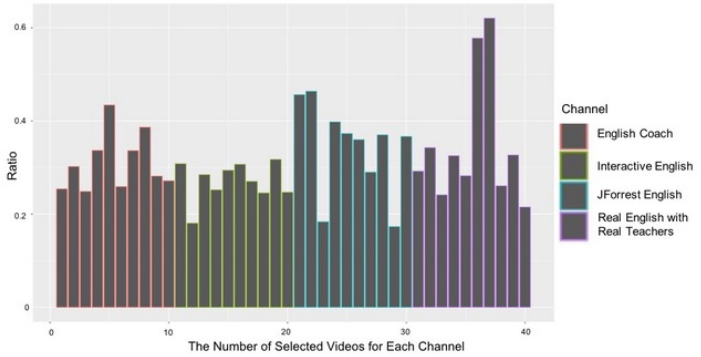


Figure 11: Correlation between video and comments per channel (i.e., Graph Analysis)

The above analyzes are different from each other of x coordinates because the fixed value graph and the distribution graph are mixed. To compare each graph, the analysis standard should be unified. As a result, all analysis results were adjusted through the normalization formula to integrate quantitative and qualitative data.

Figures 12, 13, 14 and 15 show cases where the quantitative data is normalized and re-expressed. Finally, considering both quantitative and qualitative information, it can be expressed as shown in figure 16. As the graph of quantitative and qualitative integrated normalization can be seen, satisfaction with other users can be increased at recommendation (ranking) when video is recommended considering qualitative data at the same time than when considering only quantitative data. Since the user actually watches the video for a long time and has a high probability of making a comment based on the contents of the video. In other words, the proposed scheme is more meaningful than the previous studies, the explicit analysis using the social network data alone, because it does not know the recognition of the video implicitly.

In addition, we found that the recommended videos of the proposed scheme were tracked for a certain period of time based on the same criteria. As a result, the amount of quantitative information such as the number of likes, comments, likes, and subscribers were increases and the quality correlation between video contents and comments was much higher, the performance was better than the existing methods.

In aspects of the business operators, their marketing strategies based on influencer YouTube data analysis were more successful than traditional approaches. Since the number of users and consistency degree between the contents and comments on top of big data analysis quantitatively and qualitatively were more higher with the trace analysis regardless of the time and space.

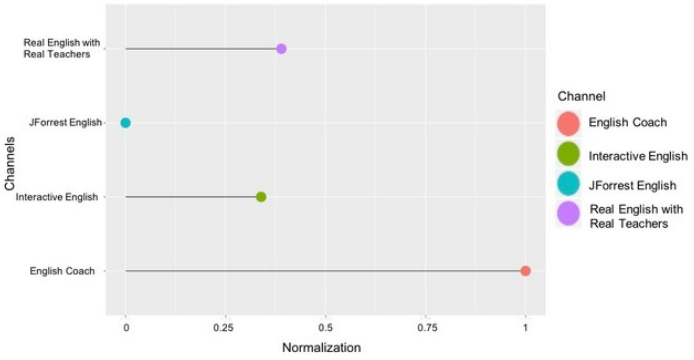


Figure 12: (Normalization) # of subscriber per channel

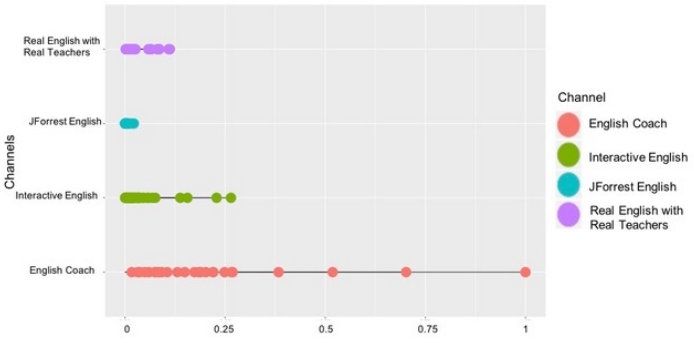


Figure 13: (Normalization) # of video likes per channel

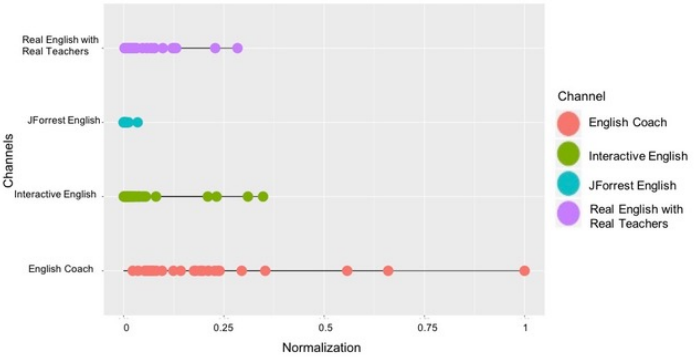


Figure 14: (Normalization) # of video views per channel

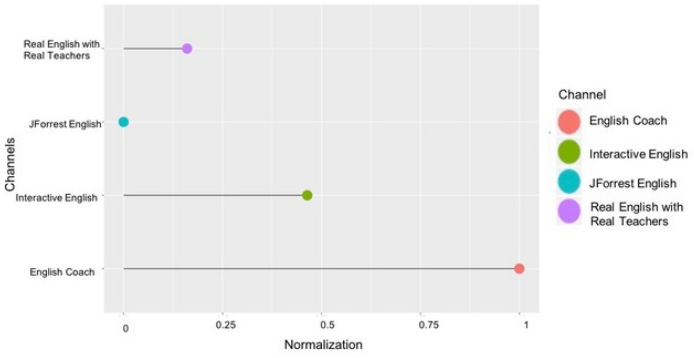


Figure 15: (Normalization) # of video comments per channel

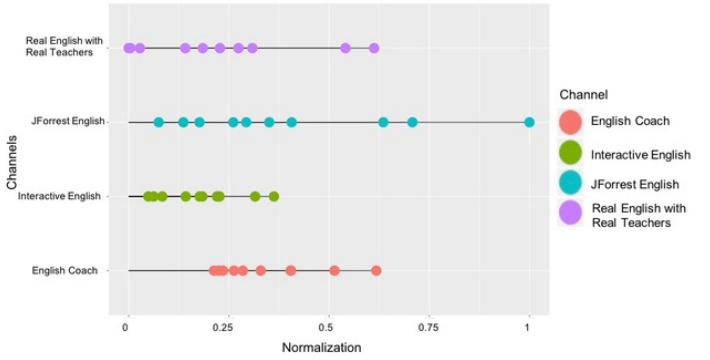


Figure 16: (Normalization) Correlation between video and comments per channel

5 Conclusion

With the development of social networks and wired and wireless networks, global users are constantly using YouTube, a huge platform from multiple angles. Although it provides a basic personalized ranking system in YouTube, it is based on simple statistical information. Therefore, it does not support sophisticated ranking and recommendation system by utilizing hidden secret information qualitatively at the same time. In this study, we utilize two factors at the same time, and in particular, we prove that not only highly visible channels but videos are visible. This study is very meaningful because it can recommend the personalized recommendation system and the customized learning channel to the user in the future, and it can lay the cornerstone to lead to the application field that maximizes the flipped learning effect regardless of time and space as well as novel marketing strategy to the business operator.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This research was supported by the MIST(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute for Information & communications Technology Promotion)"(20150009080031001), Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2017R1D1A1B03035557) and Ajou University.

References

[1] E. Çano, M. Morisio, "Hybrid Recommender Systems: A Systematic Literature Review" Intelligent Data Analysis, vol. 21, no. 6, pp. 1487-1524, 2017. https://arxiv.org/abs/1901.03888

[2] A. M. Elkahky, Y. Song, and X. He, "A multi-view deep learning approach for cross domain user modeling in recommendation systems" In Proceedings of the 24th International Conference on World Wide Web, New York, NY, USA, 2015. ACM. https://doi.org/10.1145/2736277.2741667

[3] K.J. Oh, W.J. Lee, C.G. Lim, H.J. Choi, "Personalized news recommendation using classified keywords to capture user preference" 16th International Conference on Advanced Communication Technology. https://doi.org/10.1109/ICACT.2014.6779166

[4] P. Chiliguano, G. Fazekas, "Hybrid music recommender using content-based and social information" 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/ICASSP.2016.7472151

[5] I. Munemasa, Y. Tomomatsu, K. Hayashi, T. Takagi, "Deep reinforcement learning for recommender systems" 2018 International Conference on Information and Communications Technology (ICOIACT). https://doi.org/10.1109/ICOIACT.2018.8350761

[6] X. Amatriain, L. Gatos, "Building industrial-scale real-world recommender systems" RecSys '12 Proceedings of the sixth ACM conference on Recommender systems. https://doi.org/10.1145/2365952.2365958

[7] J. Carbonell and J. Goldstein. "Diversity-based Reranking for Reordering Documents and Producing Summaries" In Conference on Research and Development in Information Retreival (SIGIR), 1998. http://doi.acm.org/10.1145/290941.291025

[8] L. Chen, G. Zhang, and H. Zhou. "Improving the Diversity of Top-N Recommendation via Determinantal Point Process" In Large Scale Recommendation Systems Workshop at the Conference on Recommender Systems (RecSys), 2017. http://arxiv.org/abs/1709.05135

[9] Y. Koren, R. Bell, and C. Volinsky. "Matrix Factorization Techniques for Recommender Systems" Computer 42, 8 (2009), 30–37. http://dx.doi.org/10.1109/MC.2009.263

[10] N. Lathia, S. Hailes, L. Capra, and X. "Temporal Diversity in Recommender Systems" In Conference on Research and Development in Information Retreival (SIGIR), 2010. http://doi.acm.org/10.1145/1835449.1835486

[11] Young Il Chang, Jung You-Soo, "A Study on YouTube Product Review Channel Subscribers' Product Attitude Formation Process," e-business research, 20(2), 78, 2019.

[12] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy and Taylor Van Vleet, "The YouTube Video Recommendation System" in Proceedings of the fourth ACM conference on Recommender systems Pages 293-296, 2010.

[13] Minwoo Kim, JeongRyeon Park, Jiwon Park and Hayoung Oh, "Channel Attribute Analysis Scheme for Trustworthy Youtube Influencer Detection", in Proceeding of The 29th Joint Conference on Communications and Information on Big Data and Social Network, 2019.

[14] Minwoo Kim, JeongRyeon Park, Jiwon Park and Hayoung Oh, "Influencer Attribute Decision-Making based on Principal Component Analysis" in Proceeding of KIPS Conference on Web Science, Korea, pp 672-674, 2019.

[15] Ji-Won Park, Min-Woo Kim, Jeong-Ryeon Park and Hayoung Oh, "Stable Influencer Selection Criteria Scheme through Youtube Analysis of Hourly Comments" " in Proceeding of Korea Institute of Next Generation Coputing, 2019.

[16] JeongRyeon Park, Minwoo Kim, Jiwon Park, Hayoung Oh "Influencer Identification based on TF-IDF and association rule analysis" in Proceeding of KIPS Conference on information system, Korea, pp 293-295, 2019.

[17] Peter Schultes, Verena Dorner, and Franz Lehner, "Leave a Comment! An In-Depth Analysis of User Comments on YouTube" in Proceeding Wirtschaftsinformatik Proceedings 2013. 42, pp659-673, 2013.

[18] Andres Bejarano, Agrima Jindal, Bharat Bhargava "Measuring user's influence in the Yelp recommender system" in Poceeding of PSU Reasearch review, Vol. 1 No. 2, pp. 91-104, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3,01-09 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Detection the Places of the Heat Energy Leak on the Underground Thermal Pipelines Using the Computer System

Iryna Zhuravska*, Dmytro Lernatovych, Oleksandr Burenko

Petro Mohyla Black Sea State University, Mykolaiv, 54003, Ukraine

A R T I C L E I N F O

Article history:
Received: 28 March, 2019
Accepted: 16 April, 2019
Online: 04 May, 2019

Keywords:
Thermal energy loss
Underground heat pipelines
Pyrometer
GPS coordinates
Mobile application
Heat map
Google FireBase

A B S T R A C T

The purpose of this paper is to improve the thermographic method for detection the places of heat energy losses on the urban underground thermal pipelines using the computer system. This goal was achieved by building a heat map of the surveyed heating main with data layering directly on a computer map of the area, which can be accessed from a mobile device by authorization via the Internet (profile on the social network Google, Facebook, Twitter, etc.). An analytical model was proposed for converting the temperature values measured over the underground heating main by a non-contact digital pyrometer to an image of the RGB color model using the parameters of the HSV color model. The implementation of the integrated Google technologies complex (Google Firebase, Firebase Authentication, and Google Maps) in the technological process of urban housing and public services was realized. The information-measuring computer system (IMCS) was developed. The proposed improvements in the thermographic method provided independently from each other work of different service teams due to the exchange of data between structural units through the cloud data storage. The results of the practical research allowed detecting the place of the underground breakthrough of the heat pipe quickly, to position the repair equipment over the accident section precisely using the GPS module on the mobile device. Thus, the size of the dug road section was minimized, material resources were saved. The experimental studies carried out confirmed the correctness of the theoretical assumptions.

1. Introduction

The urgency of the searching problem and the reduction of heat loss via the construction elements is due to the need for a rigorous energy saving. This problem is exacerbated by the work of heat power engineering in many countries because of limited natural fuel resources. Therefore, significant attention is paid to reducing operating costs, and timely detection and elimination of emergencies on the heat mains. The detection of heat energy losses on underground thermal pipelines is the most problematic.

In the conditions of the modern urban economy, losses on heating lines are enormous. Increased heat losses along the pipelines and massive coolant leaks, in addition to economic damages, cause excessive thermal pollution.

In order to comply with environmental requirements, it is necessary to implement a set of measures to eliminate the causes of heat leaks on heating mains. It is necessary to improve the methods allowing receiving thermal energy on the main highways.

*Corresponding Author: Iryna Zhuravska, 68 Desantnykiv str., Mykolaiv, 54003, Ukraine. Email: iryna.zhuravska@chmnu.edu.ua

2. Related Works

2.1. Overview of methods for determining losses in heat pipes

Methods for detecting and eliminating defects in main pipelines have been described by many researchers and scientists. [1, 2]. All over the world, common attention is paid to non-destructive techniques and technologies like acoustic leak detection, thermal imaging, tracer gas, radiotracers, etc. [1–4].

Scientifically based methods for non-destructive testing of leakages in pipelines are methods based on analysis of transient processes in pipes, in particular, the impedance method [5]. However, these methods have substantial limitations in the representation of system characteristics such as pipeline length (no more than 90 m) and location of leakage.

The detection and recovery time of the carrying capacity of pipeline sections with local damage can be significantly reduced through the measuring and statistical analytical procedures with using of the information-measuring computer system (IMCS).

The method of the negative pressure wave (NPW) is one of the most modern. Using this method, it is possible to track the appearance of a leak on the heating main without leaving the pipeline zone, while remaining at the central control station [6, 7]. At the time of occurrence of leakage of fluid in the pipeline, dilution waves appear. The pressure wave reaches the measuring points on either side of the leak and is recorded at the central dispatch post. The computing system processes the results of the incoming information and promptly calculates the leak location (within 1 minute). However, the accuracy of this method does not exceed ± 200 m.

Currently, sophisticated methods that combine acoustic and correlation principles of leak detection with remote registration on servers over the air are used on heating mains. However, the authors themselves have determined that such practices are ineffective when the heating main is located deeper than 2.5 m [8]. Also, the instruments used in the implementation of these methods do not work in the presence of increased noise (for example, moving vehicles). Considering that the majority of urban heating pipelines are laid under highways, these factors have a negative impact on the research results. The harmful properties of these methods also include the fact that to obtain the measurement results it is necessary to insert the peak of the sensor a good swift kick in the road ground (thus, destroying the highway). Also, there is a danger of the sensor damage by a stone, metal object, the ice, etc. on its way.

2.2. *The advantages and features of using the thermographic method*

From the analysis of the above methods, we can conclude that the thermographic method is the most promising. Infrared is a proven noninvasive technology, which locates leaks in the systems of heat pipes without any destruction owing to the digging of highways, under which heating mains in cities are usually laid.

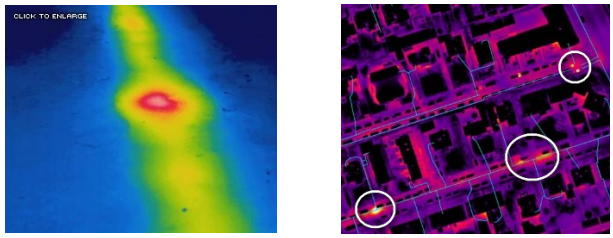
Thermal infrared cameras can see leaking heating pipes, water/steam leaks, etc. If used properly, infrared technology is one of the most highly useful diagnostic tools available today.

It's available to detect places of the local heat leak on the underground thermal pipelines using the IMCS. Such a system should use infrared thermographic devices which can transmit data through wireless ports. This solution helps to determine thermal energy loss location to eliminate unnecessary digging, destruction and expense.

Figure 1, a shows how infrared technology is used to detect a heat loss from steam pipes located under the streets of New York and attributed as a breakdown [9]. These thermal mains demand immediate repair.

Infrared technology in combination with automatic analysis of aerial thermal images allows localizing leaks on district heating pipes immediately in several neighborhoods as shown in Figure 1, b.

In this case, it is essential that the diagnostic equipment moves strictly within the spatial corridor (street) along the heating main [10]. For example, such methods as remote sensing or airborne thermography are widely used in Sweden and Norway [11].



a. b.

Figure 1. Infrared detection the local (a) and large-scale (b) places of heat loss following an accident

Figure 2 shows the location of a non-leaking, but un-insulated unions on the heat pipelines (Mykolaiv, Ukraine). Thanks to the thermal imager, it is determined that this is a non-emergency situation and, in this case, a planned repair is needed. In addition, the thermal imaging method has helped not to unearth the entire section of the roadway over the heating main, but to carry out repairs through point pits.



Figure 2. Detection of the places technological heat energy losses through un-insulated unions

Results of heat pipeline leak detection have statistically analyzed using the methods of variance analysis and the pairwise comparison methods. The researches have shown that a higher going speed of IR camera would reduce the effect of noise on the collected thermal contrast and, therefore, would give better results of prediction the leak location [12].

2.3. *On lineaments and permits for the use of UAVs for moving thermographic equipment*

Usually, the leak imaging system is placed on rotor-wing and fixed-wing UAVs [13, 14]. One of the best modern hardware solutions for studying energy losses on the thermal line is the thermal imaging cameras mounted on a quadrocopter. By applying a UAV, it is possible to specify its operation area accurately [15]. But this solution is relatively expensive and costs about USD 20,000.

Moreover, most often such systems don't determine accurate GPS coordinates of the accident and are therefore useless for the repair team. At present, the leak detection team should wait for repairers at the identified accident section or create written and photo records.

It is positive that modern leak detection methods locate heat leak accurately to 1 sq. m, so any excavation reduces to a minimum when the pipes are repaired. However, such systems also don't have an accurate reference to the location by GPS coordinates. After detection, the place of thermal energy leaks, the written report with photos of thermogram locations and recommendations are made [3, 16, 17].

Due to the lack of the identification of technological objects in these methods, a manual inspection with accordance to the route was visualized on the map is done, after which a heat map is created [18].

In the Czech Republic, the system Workswell WIRIS has been developed, which can be fitted with a GPS sensor for storing location information on the position of the drone when producing a record about the detection of hot water leaks [14]. This system is designed to be mounted on a drone (UAV) too.

But there are two problems for the implementation of this method of study of energy losses. First, the above means of the thermal measurements are too expensive. Second, on June 11, 2018, the State Aviation Service of Ukraine banned flights of remotely manned aircraft (drones or UAVs) over roads of state importance, streets of cities and villages, etc. [19]. Similar prohibitions on the use of drones on the territory of settlements exist in other countries – Italy, Austria, Czech Republic, Great Britain, etc. [20, 21]. In USA the FAA rule will not allow operation of a small unmanned aircraft over a moving vehicle because impact of a small unmanned aircraft may distract the driver of a moving vehicle and result in an accident [22].

Thus, the problem under consideration is relevant not only for Ukraine but also for the majority of countries of the world using underground heat mains.

3. Problem Formulation

The purpose of the work is to improve the thermographic method for detection the places of heat energy losses on the urban underground thermal pipelines using the IMCS.

The object of the work is the data receiving and handling process for detection heat loss using a combination of terrain maps and results of heat measurements.

The subject of the work is a system for creating heat maps of energy leak on thermal pipelines according to the data of pyrometric non-contact measurements over the surface of underground heat mains.

The purpose of the work is achieved by solving the following research tasks:

- Analysis of existing algorithms and methods implementing heat map construction on the basis of collected data.
- Analysis of the existing base of devices for research and analysis the current temperature of objects.
- Development of an analytical model for forming a heat map and calculating its statistical indicators.
- Development of a software application for detection of heat energy leaks on thermal pipelines and visualizing the results of measurements on heat maps of the area.

In this article, it is proposed to collect thermal data using an on-duty car equipped the IMCS based on tablet and manual pyrometer with wireless transmission of measured temperatures to achieve the goal.

Unlike the above examples, this measuring device (that costs about $100 USD) has several advantages such as:

- Measuring the temperature of the object with the help of non-contact infrared beam.
- The ability to transmit data through Bluetooth port 4.0 to the computer equipment.
- Laser sight in the form of a circle with a point in the middle, that well visible at a sufficiently large distance (from the on-duty car).
- It is compact, convenient and not harmful for the operator health, unlike the equipment described in [5–9].

The practical value of the work is:

- Detection and localization of the breakthroughs of the heating main for the planning of repairing works, maintenance, etc.
- The possibility of identifying objects which emit heat.

4. Methods and techniques used for the solution of the problem

4.1. Development of the information-measuring computer system for detecting heat leaks

Thus, due to the inability to use drones for the inspection of heating mains in urban areas, it is proposed to analyze the thermal pipe state using a non-contact pyrometer. For example, the Testo 805i model can be used. The pyrometer and the tablet are the part of the IMCS on board the on-duty machine moving the diagnostic team.

Bluetooth infrared thermometer Testo 805i allows transmitting the measurement data to the mobile device (smartphone, tablet, etc.) via Bluetooth port 4.0 at a distance of 15–20 m. It requires iOS 8.3 (or later) / Android 4.3 (or later).

Unfortunately, the Testo 805i doesn't measure automatically, so, to receive data, it is necessary to press the button on the device each time. Doing every 10 seconds, the temperature will be measured every 100 m at a speed of 40 km/h (10 m/s).

The measured temperature is recorded into the database of the mobile device and uploaded to the cloud storage.

The structure of such a database should include separate fields for displaying different types of data in the processed request (dates, ambient temperature, temperature above the surface of the heating main, geolocation data of the break section, etc.) [23].

Then measured data can be plotted on the terrain map as values or as a heat map.

The structural scheme of the improved thermographic method of detecting leak places on heat pipes is shown in Figure 3.

Black color indicates the main blocks, hardware and software components and the primary connections between them, used in currently applied thermographic methods for detecting the breakthroughs of the heating main. The first proposed improvements to existing methods are highlighted in red.

It should be noted that thanks to the proposed modifications, the repair team can operate autonomously, regardless of the diagnostic team, and receives information about the place of the

necessary repairs directly from the cloud storage. Also, the presence of the geo-coordinates of the accident section allows you to accurately position the repair equipment over the place of the heating main breakthrough. Replacing equipment from thermal imaging to pyrometric not only significantly reduces the cost of the solution, but also allows you to obtain digitized data associated with GPS coordinates.

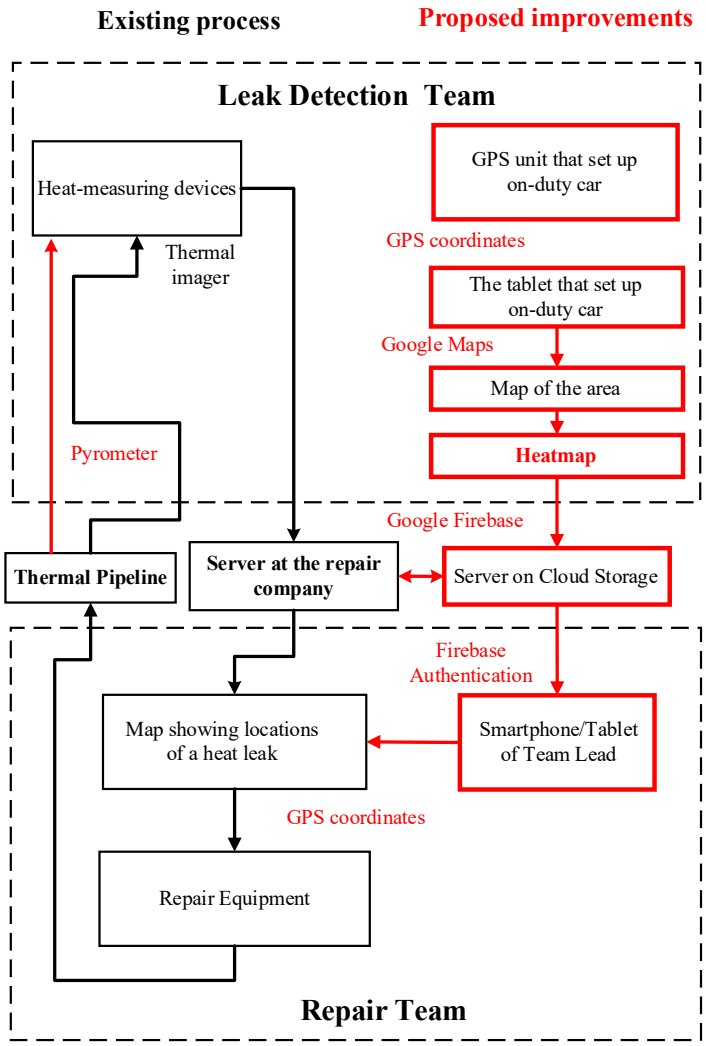


Figure 3. Structural scheme of the improved method for detecting leak places on heat pipes

Figure 4 shows a flowchart of the heat map formation.

Initially, temperature data is measured with a pyrometer, and if these data have been measured, they are recorded in the database, if not – the heat map will not be built, further measurements are crucial.

Further, when the temperature data have been received to the phone (or tablet), there is a process of tracking the coordinates, namely latitude, and longitude, using a GPS module of mobile phone or tablet. If the data is received – it is also entered in the database, and in the case of its absence – the heat map will not be built. The on-duty car equipped with IMCS moves further, to a point with new geolocation, where subsequent measurements would be made.

The final stage is the very construction of a heat map. To do this, you need to add GPS coordinates with measured temperature values into Google Maps [24]. To do this, it is necessary to use the library *com.google.maps.android:android-maps-utils.*

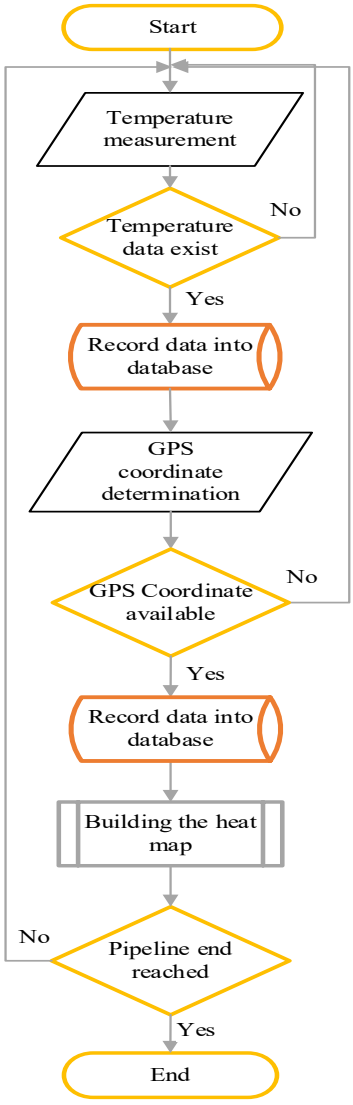


Figure 4. Flowchart of the heatmap building algorithm

The distinctive feature of the proposed method is the application of cloud technologies when obtained data (temperature and GPS coordinates) is uploaded to Google Firebase [25]. In this case, we propose an extended technological process shown in Figure 3.

4.2. Results of measurements and input data for research

During the practical research, measurements of the temperature over the heating main were carried out in preparation for the heating season (in October) on the segment of the thermal pipeline limited by the geocoordinates given in Table 1, along one street.

As a result of the temperature measurements, a breakthrough of the heat pipe was detected. The temperature over the heat main that differs significantly from others evidenced that.

Table 1. Temperature measurements on the thermal pipe in preparation for the heating season (in October)

Temperature (°C)	East longitude (degree)	North latitude (degree)	Conclusion on the state of the thermal pipe
+13.8	32.011009	46.970312	holistic
+14.2	32.012510	46.967368	holistic
+14.1	32.013154	46.965597	holistic
+22.6	32.013060	46.962030	breakthrough
+25.2	32.012781	46.961381	breakthrough
+7.2	32.012688	46.958510	holistic
+7.0	32.011994	46.953942	holistic
+7.1	32.012019	46.949474	holistic

Table 2. Temperature measurements on the thermal pipe during the heating season (in January)

Temperature (°C)	East longitude (degree)	North latitude (degree)	Conclusion on the state of the thermal pipe
+3.6	32.011009	46.970312	holistic
+4.1	32.012510	46.967368	holistic
+4.4	32.013154	46.965597	holistic
+5.2	32.013060	46.962030	holistic
+5.2	32.012781	46.961381	holistic
+3.4	32.012688	46.958510	holistic
+3.0	32.011994	46.953942	holistic
+4.1	32.012019	46.949474	holistic

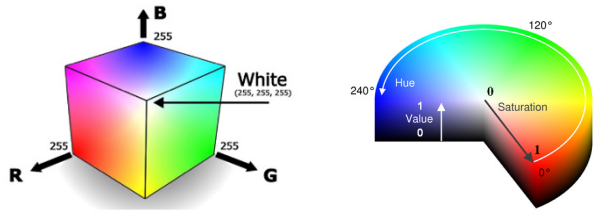
After carrying out repairs, another temperature measurement of the heating main was made at the same points. Measurements were made during the heating season (in January), then the environment temperature was lower (Table 2). Therefore, the average temperature over the surface of the thermal pipelines has decreased compared to the previous measurements made in October.

However, as it can be seen, the temperature along the heating main does not differ significantly from one another, which is why we conclude that the breakthrough of the heating main in this place has been eliminated.

4.3. An analytical model for the formation of a heat map

It should be noted that tabular data is very inconvenient for perception and rapid analysis. Therefore, it was decided to build the heat map with visualizing of found heating main emergency sections on the area map on the basis of the measurements. For this purpose, an analytical model for the formation of a heat map was developed.

Two color models, RGB and HSV, are selected for conversions [26]. The RGB model describes the color space based on Red, Green and Blue colors (Figure 5, a). The HSV model operates with parameters Hue, Saturation and Value (Figure 5, b).



a. b.

Figure 5. RGB (a) and HSV (b) color models

However, if the colors are supposed to be interpreted using the human visual system, it is advisable to integrate the RGB space and the HSV space [27]. Besides, diagnostic and repair work on heating mains can be carried out at different times of the day, and the HSV color space is more resistant to changes in lighting. Therefore, it is advisable to formulate the mathematical equivalent of the projection of the HSV model vectors in the RGB color space.

In contrary to the studies about the usage of color models mentioned above [26–27], in this article, we propose using parameters Hue and Value of the HSV color model to construct the color gradient in the RGB color model. In this case, parameter Saturation equals a constant value. This approach reduces the computing power requirements concerning an IMCS set in an on-duty car. Thanks to that building a heat map for the thermal mains and combining it with a terrain map can be performed even on a tablet instead of a laptop with higher performance, but also more cumbersome and inconvenient to use in a moving car.

The pyrometer readings, corresponding to the seasonal temperature value of the section of the pipeline, are successively added to the array of temperatures $\{t_{i,j}\}$, where $i = \overline{0,N}$ is the number of all N rows of temperature values for each from N seasons, $j = \overline{0,25}$ – the number of columns of each temperature value in the matrix of the input file data to build a heat map.

To recount the parameters of the HSV color model in the value of each spectrum of the RGB color model, the temperature coefficient k of the HSV model should be calculated according to (1). This will allow scaling the temperature values to the color of a certain tone. To switch to the parameters of the color model, it is also necessary to calculate the basic statistical parameters, namely the maximum and minimum values of the temperature (2) – (3) from the set of temperatures (4):

$$k = \frac{(sp_{max} - sp_{min})}{(t_{max} - t_{min})}, \quad (1)$$

$$t_{max} = max\{t_{i,j}\}, \quad (2)$$

$$t_{min} = min\{t_{i,j}\}, \quad (3)$$

$$i = \overline{0,N}; j = \overline{0,25}, \quad (4)$$

where $sp_{max} = 360°$ – selected the maximum value of the color tone (Hue) of the HSV model, which corresponds to the red color; $sp_{min} = 120°$ – selected the minimum value of Hue of the HSV model, which corresponds to the green color. On the basis of the set minimum (green) and maximum (red) Hue values, a range is formed in the middle of which the value of 240° corresponds to the blue color, which in turn is the spectrum of the RGB color model.

Then the color tone parameter $H_{i,j}$ for each heat map segment using the temperature coefficient for the conversion the temperature values to the color values are calculated on the basis of the formula (5):

$$H_{i,j} = sp_{min} + (t_{i,j} - t_{min})\cdot k. \quad (5)$$

In a heat map, to draw a small segment of a certain color of the RGB palette, you need to convert the temperature values using the HSV color parameters. In the proposed analytical model (6) – (9) the saturation "S" is taken as a constant ($S = 1$). This parameter does not appear in further calculations to reduce the computing costs of a mobile device with limited computing capabilities, which is used in the described IMCS for thermal pipelines inspection. $V_{min} = 0.1$ – the minimal digit of the parameter "Value".

Therefore, the next step in forming a thermal image in the proposed analytical model is the conversion of temperature data into a color RGB model based on the rated values of HSV model parameters.

The values of the red, green, and blue components of the RGB spectrum are labeled $R_{i,j}$, $G_{i,j}$, and $B_{i,j}$, respectively and calculated as follows:

$$R_{i,j} = \begin{cases} V_{min}, H_{i,j} = 0 \; or \; H_{i,j} = 1; \\ |V_{min} - ad_{i,j}|, H_{i,j} = 2; \\ V, H_{i,j} = 3; \\ |V_{min} + ad_{i,j}|, H_{i,j} = 4. \end{cases} \quad (6)$$

$$G_{i,j} = \begin{cases} V_{min}, H_{i,j} = 0 \; or \; H_{i,j} = 1; \\ |V_{min} - ad_{i,j}|, H_{i,j} = 2; \\ V, H_{i,j} = 3; \\ |V_{min} + ad_{i,j}|, H_{i,j} = 4. \end{cases} \quad (7)$$

$$B_{i,j} = \begin{cases} V_{min}, H_{i,j} = 0 \; or \; H_{i,j} = 1; \\ |V_{min} - ad_{di,j}|, H_{i,j} = 2; \\ V, H_{i,j} = 3; \\ |V_{min} + ad_{i,j}|, H_{i,j} = 4, \end{cases} \quad (8)$$

$$ad_{i,j} = (V - V_{min})\cdot(H_{i,j}). \quad (9)$$

In this case, the higher the temperature the more the color is close to the red, and vice versa – the colder the temperature the more the color is closer to the blue. As shown in Figure 6, a, in the described model, the temperature scale is implemented in the temperature range from 0 to +25°C, which corresponds to the season when the heat pipelines in the region under consideration are prepared for the heating season.



Figure 6. Temperature scale for the season

It should be noted that thermal measurements can be carried out in different seasons when the temperature of the environment varies considerably over the seasons. When researching at another time of year, the visual temperature scale should be replaced by another scale reflecting temperature values that coincide with the seasonal ones.

When constructing temperature scales, it should be noted that the operating temperature range for the pyrometer is from minus 10°C to +50°C. But with negative values of the ambient temperature, the measurement error is significantly increased. Therefore, further temperature scales were built only for ranges in which the pyrometer measurement error did not exceed 1.5 %.

4.4. The software development

The software was developed in Android Studio 3.3.1, an application created for the Android platform with JDK ver. 11.0.

Let's consider the life cycle of the application. On the initial phase of the application, the task should be sent to diagnostic teams for inspection of the area.

Using the Firebase Cloud Messaging (FCM), the system server sends messages to clients about current tasks. There is an ability to send notifications (up to 2K bytes) and data messages (up to 4K bytes).

Upon receiving the assignment, the diagnostic teams on a duty car are headed to places of information gathering. After their arrival, the system reads its GPS coordinates and sends a message to the server about the diagnostic team location.

The diagnostic team of specialists measures the temperature values using pyrometers. Data via Bluetooth 4.0 technology are transmitted to a tablet/notebook storing a local database in SQLite. When the tablet is connected to the Internet, information is sent to the cloud database.

After completing the inspections by the diagnostic team, other services (mobile clients) will receive a message through the FCM for further works on the territory.

Using the web version of the application helps to create reports based on the data collected. The *JasperReports* library is used to generate reports on the server.

The UML Class Diagram of the developed application is given in "the help file" in the code repository. The shared code that would be good for further researching by other scientists is also placed there.

The following classes were developed:

CurrentApplication – the class responsible for the life cycle of the application. It performs one or another action depending on the state of the application.

PointModel is the class responsible for representing the single point information that is displayed on the map.

PointPackModel – a class serving as a container for the PointModel list.

BaseActivity is the base class (parent) for all Activity entity. This class is responsible for the elements with visual content.

BaseMapActivity is a class with the basic functionality of Google Map.

SplashActivity is the class responsible for displaying the application loading screen.

MainActivity is the class responsible for displaying a map with functional points.

5. Experimental Results and Analysis

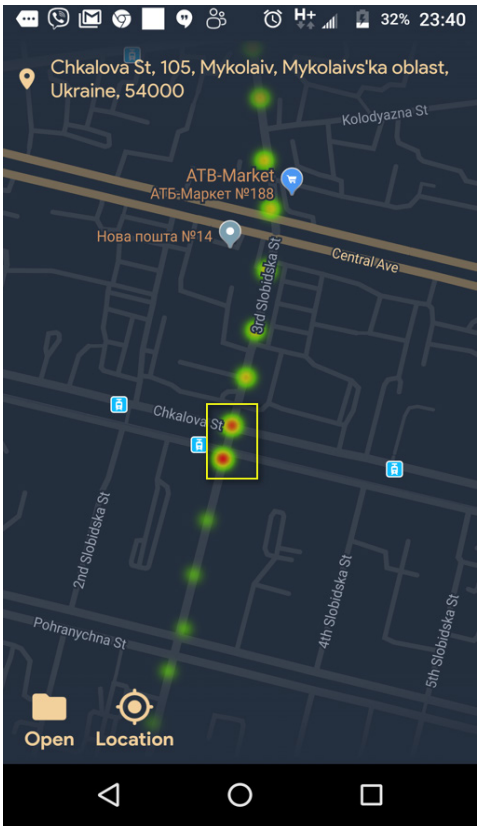
Figure 7 shows screenshots of the interface of the developed Android application.

In Figure 7 you can see the map and two buttons. The button "Location" is responsible for the current location of your mobile device (or tablet). The button "Open" is responsible for opening a file from the database needed to build a heat map.

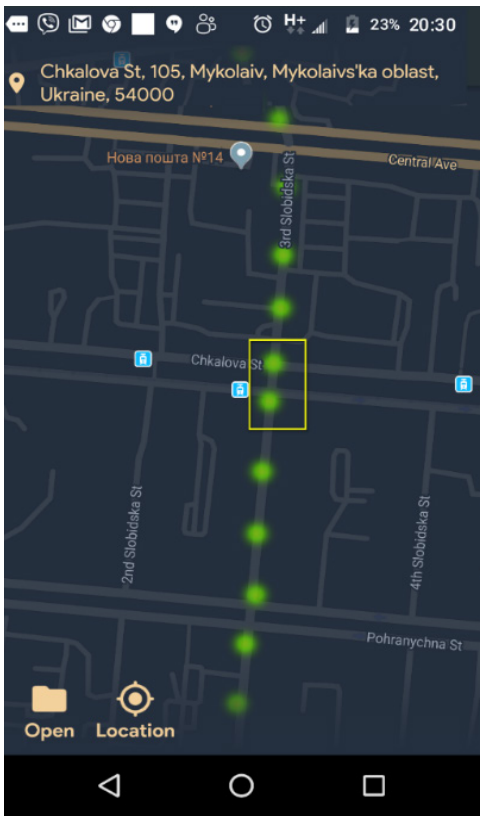
The file was filled via a Bluetooth-port with results of temperature measurements made a pyrometer Testo 805i.

From the analysis of the heat map in Figure 7, the place of heat energy leak from the underground pipeline can be seen (in the area highlighted by a yellow square). It is possible to speak about the presence of heat leaks based on the temperature gradient constructed on the scale in fig. 6. The red color inside the selected two spots (in the color spots above on the map there is no such color at all, or it is represented in an insignificant proportion to the area of the whole spot) indicates the presence of heat leaks in these places of the heating main. In the developed system, accurate geo-coordinates (latitude and longitude) of the section of the break of the underground heating main are registered and available to the repair team (see Table 1). These coordinates can be refined by more frequent pyrometric measurements at a previously diagnosed area and marked on a Google map of a larger scale. Then it will be possible, with higher sampling, to perform image segmentation of the spot of the color scale for sensors of the automated system of technological control [28]. The maximum spot area, indicated in the red inside, will show a more accurate accident section. This will allow the positioning of repair equipment directly above the area of the heating main break. The accuracy of diagnosis is limited only by the measurement interval of the device (3 seconds) and the speed of movement of the operator of the diagnostic team.

Automated fixation of the geo-coordinates of the accident site with their plotting on the terrain map is performed in none of the methods described in the literature.



a.



b.

Figure 7. The heatmap showing the breakdown (a) on the thermal pipeline and its eliminated (b)

With the data of the developed IMCS, it is also possible to conclude that after the detected leakage point along the pipe main, the temperature over the surface of the underground heating pipe has been reduced to the ambient temperature on the date of measurement (green labels in Fig. 7, but without a color gradient according to the developed temperature scale for the studied season).

After the computer diagnostics were carried out, the received data was transferred to the corresponding repair company. The unearthing of the heating pipeline by the found coordinates of the breakthrough has confirmed the fact of the accident precisely in the detected place of the heating main. It was discovered as a result of temperature measurements and reflected in Figure 7, a. The red color in the center of the spot corresponds to the raised temperature. And Figure 7, b shows the same place, but already with the eliminated breakthrough (see the yellow rectangle).

The advantages of the proposed method are that the repair team can operate independently of the work schedule and the location of the diagnostic team specialists. When the repair team arrives at the place of heat energy leakage, the brigadier is authorized on Firebase by any mobile device (smartphone, tablet, etc.). In this case, his profile in any social network can be used – Google, Facebook, Twitter, etc.

When the access to the heat map of the thermal pipeline has given, brigadier of the repair team can compare the own GPS coordinates with the coordinates received from the cloud service. Then it is able to combine the repair equipment accurately (excavator, pump, etc.) with detected places of the heat energy leak from the heat map.

6. Conclusion

Based on the analysis of the methods the detection of the thermal energy leak places on thermal pipelines, the relevance and necessity of developing new analytical models and software for the search and mapping leaks on heating mains with the help of heat maps have been established.

An analytical model for forming a heat map and calculating its statistical parameters was proposed. The convert of the temperature values of a non-contact digital pyrometer from an underground heating main to the image of the RGB color model using the parameters of the HSV color model was substantiated.

A software application for building heat maps based on measured temperature data for the Android platform with JDK 11.0 version was developed. Introduced Firebase Authentication, integrated with popular social services. Due to the heat map layered on Google Maps, the repair team got the opportunity access to detected places of the heat energy leak regardless of the work schedule and the location of the diagnostic team.

Experimental researches with the developed IMCS for the needs of urban housing and communal services for detection places of heat energy leak in the city underground thermal mains due to pipe accident have been carried out. The conducted studies confirmed the sufficient accuracy of the results for the optimal location of repair equipment on detected places of the pipe breakthrough.

The perspective of further development of the thermographic method to detect the places of heat energy loss using the IMCS and cloud services has shown.

As a development of the method, it can be supplemented, besides the technology of processing surface temperatures, also with information technologies of visual information processing on purpose more accurate geo position of heat leaks [29].

In the future, the proposed method for the diagnosis of thermal outflows can be used for both technical and biological objects. For each of the goals, it will be necessary to build separate temperature scales in the appropriate temperature range based on the proposed analytical model. Also, the pyrometric equipment should also be selected with the accuracy of measurement and the price of dividing the scale corresponding to the constructed temperature scale.

The developed system has a lower cost (up to 50 times) compared with similar thermographic systems based on thermal imagers. Also, the proposed system is safe for the operator compared to diagnostic systems based on ultrasound, radiation and electromagnetic non-invasive methods for monitoring heat leaks described in the analyzed literature.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The research was carried out with the support of the Ministry of Education and Science of Ukraine within the framework of the state-funded scientific research works at the Petro Mohyla Black Sea National University.

References

[1] Non-destructive Testing and Repair of Pipelines. E. N. Barkanov, A. Dumitrescu, I. A. Parinov (eds), Springer, Cham, 2018. doi: 10.1007/978-3-319-56579-8.

[2] M. Fahmy, O. Moselhi, “Detecting and locating leaks in Underground Water Mains Using Thermography” in 26th International Symposium on Automation and Robotics in Construction (ISARC 2009), 61–67, 2009. https://www.irbnet.de/daten/iconda/CIB14819.pdf.

[3] “Central heating leaks” [L.D.S Leak Detection Specialists Ltd.]. https://www.leakdetectionspecialists.co.uk/central-heating-leaks.htm.

[4] “Leak Detection in Heat Exchangers and Underground Pipelines Using Radiotracers” [Material for education and on-the-job training for practitioners of radiotracer technology], Vienna, International Atomic Energy Agency, 2009. https://www-pub.iaea.org/MTCD/Publications/PDF/TCS-38_web.pdf.

[5] S. Kim, “Holistic diagnosis of pipeline system with impedance method” PROCEDIA ENGINEER, 70, 924–933, 2014. doi: 10.1016/j.proeng.2014.02.103.

[6] J. Liua, H. Sua, Ya. Maa, G. Wanga, Yu. Wangb, K. Zhang, “Chaos characteristics and least squares support vector machines based online pipeline small leakages detection” CHAOS SOLITON FRACT, 91, 656–669, Oct. 2016. doi: 10.1016/j.chaos.2016.09.002.

[7] M. Valincius, M. Vaisnoras, A. Kaliatka, “Study and demonstration of pressure wave-based leak detection in a district heating network” STRUCT INFRASTRUCT E, 14 (2), 151–162, 2018. doi: 10.1080/15732479.2017.1330892.

[8] X. H. Tang, Y. B. Liu, L. J. Zheng, C. B. Ma, H. Wang, “Leak Detection of Water Pipeline Using Wavelet Transform Method” in International Conference on Environmental Science and Information Application Technology (ESIAT 2009), Wuhan, PEOPLES R CHINA, 2009, 2, 217–220. doi: 10.1109/ESIAT.2009.57.

[9] “Infrared Inspection Services – Thermography and Thermal Imaging in New York, NYC, Long Island and New Jersey for Commercial Industry and

Property Management" [web-site of Infrared Imaging Services LLC]. http://www.infraredimagingservices.com/commercial.

[10] M. P. Musiyenko, I. M. Zhuravska, I. V. Kulakovska, A. V. Kulakovska, "Simulation the behavior of robot sub-swarm in spatial corridors" in 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO–2016), Kyiv, Ukraine, 2016, 382–387. doi: 10.1109/ELNANO.2016.7493090.

[11] O. Friman, P. Follo, J. Ahlberg, S. Sjokvist, "Methods for large-scale monitoring of district heating systems using airborne thermography" IEEE T GEOSCI REMOTE, 52 (8), 5175–5182, Aug. 2014. doi: 10.1109/TGRS.2013.2287238.

[12] A. Hawari, M. Khader, W. Hirzallah, T. Zayed, O. Moselhi, "Integrated sensing technologies for detection and location of leaks in water distribution networks" WATER SCI TECH-W SUP, 17 (6), 1589–1601, Dec. 2017. doi: 10.2166/ws.2017.044.

[13] G. R. Stockton, "Leak Detection – Aerial Thermal Mapping Reveals Leaks in All Types of Systems" in CampusEnergy2017 [International District Energy Association]. Miami, FL, Feb. 20–24, 2017. https://www.districtenergy.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=5ae4b471-ee47-9f2f-4443-94a8fe505273.

[14] "Inspection of hot-water and warm-water pipes" [web-site of the Workswell s.r.o.]. https://www.workswell-thermal-camera.com/inspection-hot-water-warm-water-pipes/.

[15] I. Zhuravska, I. Kulakovska, M. Musiyenko "Development of a method for determining the area of operation of unmanned vehicles formation by using the graph theory" Eastern-European Journal of Enterprise Technologies, 2, 3 (92), 4–12, 2018. doi: 10.15587/1729-4061.2018.128745.

[16] "Thermal Mapping – Inspection of District Heating" in Case Study for Drone Systems – with Thermal Capture Recorder 2.0 640, Copenhagen, Denmark. https://thermalcapture.com/thermal-mapping-inspection-of-district-heating/.

[17] P. M. Bach, J. K. Kodikara, "Reliability of infrared thermography in detecting leaks in buried water reticulation pipes" IEEE J SEL TOP APPL, 10 (9), 4210–4224, 2017. doi: 10.1109/JSTARS.2017.2708817.

[18] E. Cornelis, F. Meinke-Hubeny, "The STRATEGO Project No: IEE/13/650 (Multi-level actions for enhanced Heating & Cooling plans)", European Union, May 29 2015. http://stratego-project.eu/wp-content/uploads/2014/09/D3_7a-STRATEGO-WP3-Mapping-methodology-and-data-sources.pdf.

[19] "Rules for Use of Airspace of Ukraine", State Aviation Administration of Ukraine, 430/210, May 11 2018. https://avia.gov.ua/wp-content/uploads/2018/05/TYMCHASOVYJ-poryadok-VPPU_31.05.18-2.doc (In Ukrainian).

[20] "Regulation of the European Parliament and of the Council on common rules in the field of civil aviation", Brussels, July 12 015. http://ec.europa.eu/smart-regulation/impact/ia_carried_out/docs/ia_2015/swd_2015_0262_en.pdf.

[21] "Airspace restrictions for unmanned aircraft and drones" [Civil Aviation Authority – CAA in the United Kingdom]. https://www.caa.co.uk/Consumers/Unmanned-aircraft/Our-role/Airspace-restrictions-for-unmanned-aircraft-and-drones/.

[22] "Unmanned Aircraft Systems (UAS)" [U.S. Department of Transportation – Federal Aviation Administration, FAA]. https://www.faa.gov/uas/.

[23] M. Fisun, M. Dvoretskyi, A. Shved, Y. Davydenko, "Query parsing in order to optimize distributed DB structure" in 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2017), Bucharest, Romania, 2017, 1, 172–178. doi: 10.1109/IDAACS.2017.8095071.

[24] E. Price, "Create a heat map of your Google location history with this tool", Apr. 11, 2018. https://lifehacker.com/create-a-heat-map-of-your-google-location-history-with-1830215410.

[25] K. S. Ashok, Mastering Firebase for Android Development: Build real-time, scalable, and cloud-enabled Android apps with Firebase, Packt Publ., 2018.

[26] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, J. Jatakia, "Human skin detection using RGB, HSV and YCbCr color models" [ICCASP/ICMMD-2016] ADV INTEL SYS RES, Atlantis Press, 137, 324–332, 2017.

[27] M. Loesdau, S. Chabrier, A. Gabillon, "Hue and saturation in the RGB color space" [Lecture Notes in Computer Science], 8509, 203–212, Springer, Cham, 2014. doi: 10.1007/978-3-319-07998-1.

[28] M. Fisun, W. Smith, A. Trunov, "The vector rotor as instrument of image segmentation for sensors of automated system of technological control" in the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2017), Lviv, Ukraine, 2017, 1, 458–463. doi: 10.1109/STC-CSIT.2017.8098828.

[29] G. Shcherbakova, S. Antoshchuk, A. Sachenko, M. Gerganov, M. Polyakova, V. Krylov, " " in the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing (DSMP 2018), Lviv, Ukraine, 2018, 1, 605–608. doi: 10.1109/DSMP.2018.8478551.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 1-15 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Logic Error Detection System based on Structure Pattern and Error Degree

Yuto Yoshizawa*, Yutaka Watanobe

The University of Aizu, Computer Science and Engineering, Fukushima 965-8580, Japan

ARTICLE INFO

Article history:
Received: 12 May, 2019
Accepted: 14 August, 2019
Online: 06 September, 2019

Keywords:
Logic Error Detection
Programming Education
e-Learning
Online Judge

ABSTRACT

The importance of programming skills has increased with advances in information and communication technology (ICT). However, the difficulty of learning programming is a major problem for novices. Therefore, we propose a logic error detection algorithm based on structure patterns, which are an index of similarity based on abstract syntax trees, and error degree, which is a measure of appropriateness for feedback. We define structure patterns and error degree and present the proposed algorithm. In addition, we develop a Logic Error Detector (LED) Application Programming Interface (API) based on the proposed algorithm. An implementation of the proposed algorithm is used in experiments using actual data from an e-learning system. The results show that the proposed algorithm can accurately detect logic errors in many programs solving problems in the Introduction to Programming set.

1 Introduction

The Fourth Industrial Revolution has increased the demand for engineers [1] [2] [3]. Until recently, computer science was learned by university students, often in engineering fields. Because of a shortage of engineers in many countries, attempts have been undertaken to introduce computer science into secondary school education.

A study suggested that the high school ICT curriculum should focus on computer science [4]. A report by the Royal Society of the United Kingdom emphasized the importance of computer science skills in primary education [5]. The report pointed out problems in ICT education, such as those in class content and the loss of opportunities for students. In addition, it proposed to redefine ICT into three areas, namely computer science, information technology, and digital literacy, among which computer science is a particularly important area. The movement to implement programming education is accelerating globally.

For introducing programming in school education, many problems have been identified and various teaching methods have been proposed [6] [7]. A report summarizing three years of compulsory programming education in the UK published in 2014 found that it was patchy and fragile [8]. The report briefly discussed the difficulty of programming education.

A student's inherent interest in computer programming affects learning. Learners with an interest in programming can effectively learn in large groups with a single teacher, as is done in the university. Such learners can solve problems by investigating them or asking questions. However, traditional teaching methods may be unsuitable for novices and learners who are not interested in programming. One of the most effective ways for novices to learn is one-on-one learning with a teacher. This is especially true for programming because programs that accomplish a given task can be written in various ways. Teachers must thus strictly evaluate the program written by a learner.

Programming includes complex language constructs and requires logical thinking, similar to that used in mathematics. Programmers must consider algorithms, memory space, execution time, and code size. However, exhaustive and quantitative evaluation of programs by a human teacher is difficult. In addition, programs may not work as expecting due to various bugs. It is difficult even for experienced teachers to deal with every bug.

To improve this situation, it is necessary to assign a knowledgeable teacher to several novice programmers. However, this is not possible in classes that introduce programming due to a lack of human resources. To solve this problem, methods for estimating the difficulty of programming problems using cluster analysis and problem recommendation algorithms have been proposed [9] [10]. In addition, attempts have been made to identify early crisis situations experienced by novice programmers using clustering [11]. However, many studies considered only optimizing human resources and

*Corresponding Author: Yuto Yoshizawa, The University of Aizu, Computer Science and Engineering, Fukushima 965-8580, Japan

presenting a course to users. The quality of teaching still depends on the actual teacher. To solve this problem comprehensively, an intellectual tutoring system is required.

Online judge systems are environments in which programming can be learned independently [12]. An online judge is an online execution environment that executes and verifies source code submitted by users. Users can improve their programming skills by solving problems in various categories (e.g., syntax, semantics, algorithms, data structures). In addition to automatic verification, an online judge provides a quantitative performance evaluation (execution time, memory usage, etc.) of the program. Although online judges satisfy certain requirements of an autonomous learning environment, it is extremely difficult for novice programmers to continue learning using online judges because they do not provide concrete feedback. Originally, online judges were created for programming contests, in which a solution is either accepted or rejected. Thus, programmers can only recognize that their program has an error, not which type of error (e.g., logic error, compile error, run-time error). For such a program, which is "executable for any input", but "does not satisfy the specifications of the problem", the judgment does not provide useful support for the user. Therefore, novice programmers tend to give up if their logic errors cannot be debugged.

There are various approaches to programming education, including methods that use program execution traces and program dependence graphs. Here, we propose a logic error detection algorithm based on structure patterns, error degree, and abstract syntax trees. A structure pattern is a list generated from the corresponding abstract syntax tree and is an index of similarity at the structure level. An error degree is a value generated from a result of the detection algorithm and is an index of similarity at the characteristic level. The goal of the proposed algorithm is to provide optimal and personalized feedback to individuals. The proposed algorithm provides personalized feedback based on filtering using structure patterns, error degree, and the characteristics of abstract syntax trees.

In this paper, an algorithm for detecting logic errors and the corresponding API, which can be used to construct intelligent coding editors, are presented. The results of quantitative and qualitative experiments are shown. The quantitative experiments are used to verify the detection range of logic errors. The results show that the proposed algorithm can detect logic errors in many cases. The qualitative experiments are used to determine the detection accuracy and evaluate the appropriateness of feedback based on the detection results. The proposed algorithm is shown to detect logic errors and provide appropriate feedback in many cases. In some cases, generated feedback was inappropriate for learning support. This paper is an extension of the work originally presented in "2018 9th International Conference on Awareness Science and Technology (iCAST)" titled "Logic Error Detection Algorithm for Novice Programmers based on Structure Pattern and Error Degree" [13].

The remainder of this paper is organized as follows. Section 2 presents related research. Section 3 describes the logic error detection algorithm. Section 4 describes the LED API. Section 5 presents experiments that use data from an online judge system. Section 6 shows and discusses the experimental results and suggests possible improvements. Finally, Section 7 summarizes the main conclusions.

2 Related Research

The goal of the present study is to construct an autonomous learning environment for novice programmers. However, autonomous learning of programming has not been clearly defined.

Learner autonomy has been mainly considered in the field of foreign language education by researchers such as Holec [14] [15]. Dickinson proposed the concept of autonomous learning, where learners do not rely on teachers, but are themselves responsible for the results, and make all decisions related to learning [16]. Therefore, autonomous learning can be broadly defined as self-learning controlled by the learner.

This definition of autonomous learning assumes that the learning environment is appropriate. However, there is a shortage of teachers in programming education. In the present study, autonomous learning of programming is defined as autonomous learning in an environment without an actual teacher.

Thus far, a number of approaches to support novice programmers have been proposed. In this paper, we present methods that have different targets and structures, and require different educational resources.

Wu et al. designed practice teaching and assessment methods around an online judge system and proposed resolution methods for existing problems in hybrid learning [17]. Semi-automatic evaluation methods that focus on reducing the burden on teachers have also been proposed [18]. In this approach, students are grouped according to the structure of their source code submitted to the online judge, and guidance is provided for each group. Using a network of online judge users, a hybrid method that allows other users to be asked for guidance has been proposed [19]. This approach uses a clustering result based on the similarity of source code. In addition, a method that uses solutions and a language model based on long short-term memory (LSTM) networks for bug detection has been proposed [20].

Fault localization with program execution traces is a well-known technique for detecting logic errors. Using this concept, Su et al. constructed a Moodle-based online learning environment to provide feedback for logic errors [21]. In a study related to this technique, the accuracy of logic error detection using only 20% of the source code and the effectiveness of feedback using a lightweight spectrum-based algorithm were verified [22].

There are approaches that target specific logic errors to support novices. These approaches focus primarily on conditional branches and detect common logic errors, such as careless mistakes in conditional expressions, priority violations, and infinite loops [23] [24].

Targeted at advanced novice programmers, FrenchPress [25] is an eclipse plugin that focuses on programming style and supports programmers who have not yet assimilated the object-oriented paradigm.

In the present study, logic errors are detected by comparing a source code that may include logic errors with correct source codes stored in an online judge system. A comparison of source code is generally performed to detect code clones (i.e., similar or identical fragments of code) [26] [27] [28]. The following four types of method are commonly used to detect code clones.

Text-based methods directly compare source code as text. This type of comparison includes a lot of unique information, such as

variable names. Token-based methods compare tokens obtained by a lexical analysis of source code. Since this type of comparison is not affected by unique information, such as variable names, it has relatively high accuracy. Tree-based methods compare code based on abstract syntax trees obtained by a syntax analyzer. Since parts unrelated to the meaning of the program are omitted from the syntax tree, the structural information of the source code is used for comparison. Semantic-based methods compare code using various information obtained from a parser. A program dependence graph is primarily used. It is possible to detect discontinuous code clones, which are difficult to find using other methods.

The results obtained from text or token comparison are not necessarily consistent with the structure of the source code. Therefore, it is difficult to generate accurate feedback when using text- and token-based methods for detecting logic errors. Since the program dependency graph used in semantic-based approaches has information on data dependency and control dependency, it is possible to perform a comparison based on semantics rather than the description of source code. However, in the isomorphism judgment of graphs, although the semantics of the programs are the same, it is considered difficult to detect logic errors that differ only in internal values. A comparison using abstract syntax trees is based on the structural information of source code and thus can be performed at the syntax level, considering for example *if* statements or *for* statements. Therefore, this method is considered to be suitable for detecting logic errors by a feedback algorithm.

3 Logic Error Detection Algorithm

In the present study, two kinds of source code, namely "wrong code" and "correct code", are used to demonstrate the proposed algorithm. The wrong code is code submitted to an online judge and judged as "Wrong Answer". It is the target for detecting logic errors by the proposed algorithm. The correct code is code judged as "Accepted" and stored in the online judge system. It is the comparison target for detecting logic errors by the proposed algorithm.

3.1 Logic Error

A logic error is a bug in a program that triggers erroneous behavior but does not cause abnormal termination. In other words, logic errors induce behavior not intended by the programmer. Programs that contain logic errors are valid programs that can be compiled and executed. To debug a logic error, the user must determine the cause of the error by tracing variable information during execution using a number of test cases. Table 1 shows some common examples of logic error.

3.2 General Approach of Proposed Algorithm

In the present study, we propose a logic error detection algorithm based on a comparison between wrong and correct code. The proposed algorithm uses a comparison method based on an abstract syntax tree, which is a data structure in which the grammatical structure of source code is represented by finite labeled oriented trees. An abstract syntax tree is recursively defined by a minimum unit, called a syntax, generated by a parser. For example, *if* and *for* statements are recursively defined by a number of constructs.

Table 1: Common examples of logic error

Type of logic error	Cause
Loop condition	Incorrect number of iterations of *for* statement
Conditional branch	Incorrect conditional expression in *if* statement
Output format	Incorrect rounding
Data type	Incorrect data type
Calculation	Incorrect operator in binary expression
Indexing	Wrong index accessed

In the present study, a construct that is recursively defined by other constructs is denoted as a nonterminal construct, and a construct that is not recursively defined by other constructs is denoted as a terminal construct. Examples of nonterminal constructs are *if* statements and conditional expressions, and examples of terminal constructs are character strings and numeric values. Specifically, the algorithm compares abstract syntax trees of the target source code by applying the following steps recursively.

1. If the information on terminal constructs of the same kind is not different, it is not judged as a logic error.
2. If the information on terminal constructs of the same kind is different, it is judged as a logic error.
3. If the constructs are the same kind of nonterminal construct, they are compared to each other.
4. In other cases, a logic error is judged to have occurred.

The algorithm detects constructs that are logic errors. Algorithm 1 shows the pseudocode of the logic error detection algorithm.

The proposed algorithm can compare compilable source code. However, because of the characteristics of abstract syntax trees, a comparison of source code with different tree shapes generates detection results that are unsuitable for detecting logic errors. In addition, even if the shapes of the trees are the same, a comparison of code with different node meanings will generate incorrect detection results. Therefore, the proposed algorithm has two phases.

In the first phase, an appropriate comparison target is selected. The proposed algorithm compares only two source codes and detects logic errors. However, the user may attempt to solve a problem using various approaches. For example, a loop can be implemented using a *for* statement or a *while* statement. If the detection algorithm is applied to the source code of programs that have quite different structures, incorrect results will be obtained. Figures 1 and 2 show examples of a comparison of source code that uses different algorithms. A suitable comparison target is required to detect logic errors.

Algorithm 1 Logic Error Detection

Require:
x = expression in wrong code, x_n = n-th expression of x
y = expression in correct code, y_n = n-th expression of y
Ensure:
E = an array of Logic Error : Array

```
1:  function DETECTION(x, y)
2:      E ⇐ empty array
3:      if x and y are terminal constructs of the same kind then
4:          if x and y are the same information then
5:              return E                                          ▷ Step 1
6:          else
7:              return E ⇐ x                                      ▷ Step 2
8:          end if
9:      else if x and y are nonterminal constructs of the same kind then
10:         if the number of expressions of x and y are the same then
11:             for i = 1 to the number of expressions in x do
12:                 Connect E and DETECTION(x_i, y_i) together    ▷ Step 3
13:             end for
14:             return E
15:         end if
16:     end if
17:     return E ⇐ x                                              ▷ Step 4
18: end function
```



Figure 1: Example of detection with a *for* statement approach

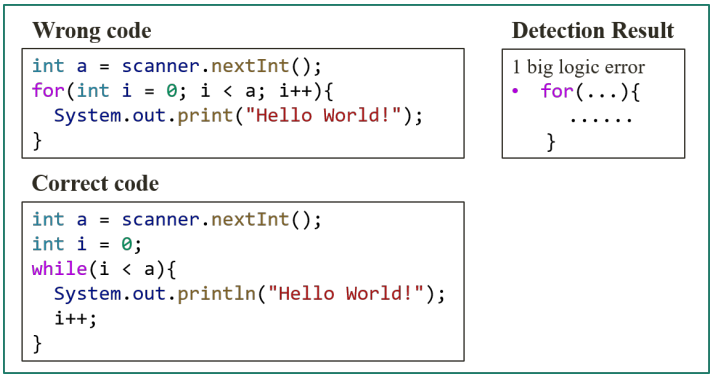


Figure 2: Example of detection with a *while* statement approach

In the second phase, the best detection result is selected. The proposed algorithm generates feedback for the user based on a comparison of the source code. However, the algorithm that the user employs to solve a problem is independent of the source code structure. For example, how conditional branches and libraries are used depends on the user's characteristics. If the detection algorithm is applied to the source code of programs with different algorithms, we may provide feedback while ignoring the characteristics of the corresponding user.

Figures 1 and 3 show examples of a comparison of source code that uses different loop constructions. Therefore, we need to give feedback based on correct codes that has the same characteristics as those of the target codes.

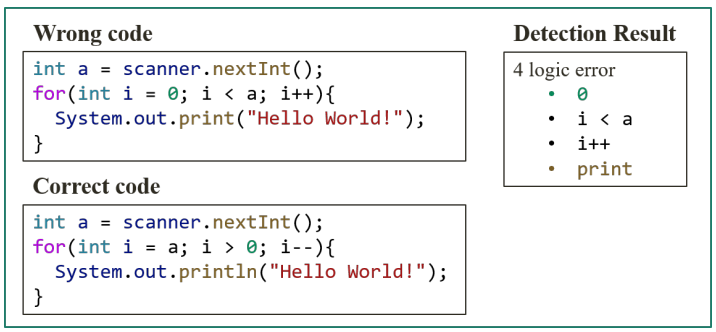


Figure 3: Example of detection with a decrement algorithm

3.3 Structure Pattern

In the present study, structure patterns are used in the first phase. A structure pattern is a list generated from the corresponding abstract syntax tree and is an index of similarity at the structure level. Moreover, it is core data obtained by a depth-first search of the abstract syntax tree and includes a list of expressions and block statements. The generated list contains expressions such as a variable declaration and assignment, and control structures such as *for* and *if* statements. Therefore, a structure pattern can be considered to express the general form of source code.

Figure 4 shows a sample source code. Figure 5 shows an example of an abstract syntax tree parsed from the sample source code. Table 2 shows an
example of a structure pattern generated from the sample source code.

```
int a = sc.nextInt();
int b = sc.nextInt();
if(a < b) {
  System.out.println("a < b");
}else if(a > b) {
  System.out.println("a > b");
}else {
  System.out.println("a == b");
}
```

Figure 4: Sample source code

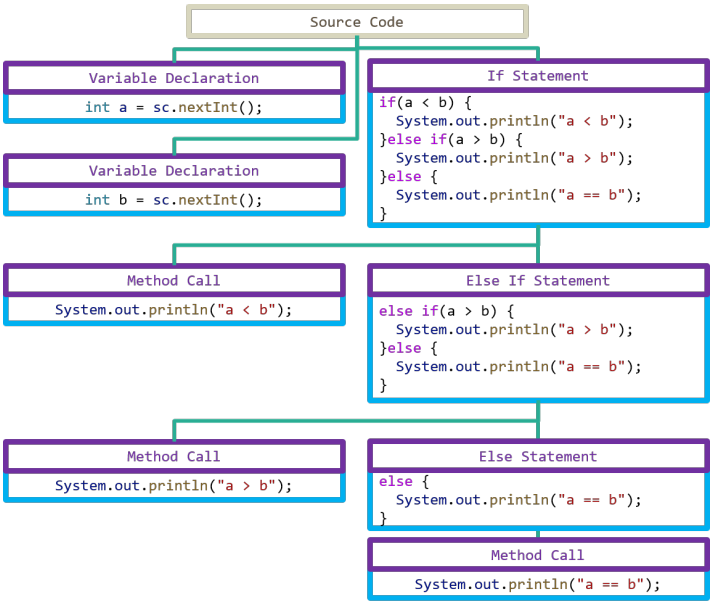


Figure 5: Abstract syntax tree for the source code in Figure 4

Programs with the same structure patterns have the same order and kind of constructs. Therefore, we can obtain appropriate comparison targets by filtering correct code using the structure pattern. For example, the structure patterns for the source code in Figures 1 and 2 are shown in Tables 3 and 4, respectively. As shown, appropriate correct codes for comparison can be obtained by filtering according to the structure pattern.

Table 2: Structure pattern for the source code in Figure 4

Index No.	Syntax
1	Variable Declaration
2	Variable Declaration
3	*If* Statement
4	Method Call
5	*Else* Statement
6	*If* Statement
7	Method Call
8	*Else* Statement
9	Method Call

Table 3: Structure pattern of wrong code and correct code shown in Figure 1

Index No.	Syntax
1	Variable Declaration
2	*For* Statement
3	Method Call

Table 4: Structure pattern of correct code shown in Figure 2

Index No.	Syntax
1	Variable Declaration
2	Variable Declaration
3	*While* Statement
4	Method Call
5	Binary Expression

3.4 Error Degree

In the present study, the error degree is used in the second phase. An error degree is a value generated from a result of the detection algorithm and is an index of similarity at the characteristic level. Moreover, it is a total value obtained by assigning a weight according to the type of detected logic error and can be expressed as (1).

x = the number of logic errors
w = weight of logic error
n = the number of logic error types

$$E_{rrordegree} = \sum_{i=1}^{n} x_i w_i \tag{1}$$

The error degree is zero when the wrong code is compared with itself. An increase in error degree indicates that the source code has characteristics different from those of the wrong code. Therefore, we can obtain optimal feedback by comparing the wrong code with all versions of correct code and by filtering the detection results using the error degree. Table 5 shows the weighting rules. The weight depends on the logic error type in the source code. For example, errors at the element level, such as terminal constructs

and operators, have lower weight, whereas errors at the syntax level, such as the type of syntax and its presence in the code, have higher weight. Errors related to the number of constructs are weighted in proportion to this number.

Table 5: Weighting rules for error degree

Error type	Weight
Terminal construct	1
Operators in expressions	3
Number of constructs	2*constructs
Existence of component	5
Construct that exists only on one side	10
Types of construct for comparison	10

The error degree compensates for the weaknesses of filtering by structure pattern. For example, Figures 6 and 7 show the results of logic error detection for the source code of programs with the same structure patterns presented in Figure 1 and Figure 3. As shown in these figures, more accurate feedback can be obtained by calculating the error degree from the detection result.

Detection Result

Logic Error	Error Type	Weight
print	Terminal construct	1

Error Degree = 1

Figure 6: Detailed detection results for source code in Figure 1

Detection Result

Logic Error	Error Type	Weight
0	Types of construct for comparison	10
i < a	Operators in expressions	3
i ++	Operators in expressions	3
print	Terminal construct	1

Error Degree = 17

Figure 7: Detailed detection result for source code in Figure 3

3.5 Extended Detection Algorithm

The purpose of the proposed algorithm is to support autonomous learning by detecting logic errors in a wrong code and providing the corresponding feedback. The detection algorithm provides feedback through the following steps.

1. Analyze the judged source code and generate a structure pattern.
2. From the data stored in the online judge system, extract correct codes with the same structure pattern.
3. Compare all of the extracted correct codes with the wrong code, and calculate the error degree based on the detection results.
4. Provide information obtained based on the detection result with the lowest error degree as feedback for the user.

Algorithm 2 shows the pseudocode of the extended logic error detection algorithm.

4 Implementation of Algorithm API

Section 3 described the proposed logic error detection algorithm for general online judge systems. In the present study, we also developed an LED web API for the proposed algorithm that is compatible with Aizu Online Judge (AOJ) [29].

4.1 Aizu Online Judge

AOJ is an online judge system developed and operated by the University of Aizu. An online judge system is a service that compiles and executes programs submitted by users and automatically evaluates their accuracy and performance using various test cases and verification machines. At present, AOJ contains source code from 60,000 users and 4 million judgments. AOJ contains various problems, such as those used in programming contests (e.g., ACM-ICPC) and those specific to a certain topic, including Introduction to Programming (ITP), “Algorithm and Data Structures”, and related libraries. For these problems, limit values are set for CPU time and memory usage, and the user can practice by solving the problem under these limitations.

The user obtains a verdict by submitting the answer code for each question to the judge system. The judge system instantly provides feedback (accepted/rejected) based on various test cases for the problem, together with the CPU time and memory usage. The type of failure (Time Limit Exceeded, Memory Limit Exceeded, Runtime Error, etc.) is also given in the feedback. Users can debug their code based on the feedback and resubmit code as many times as desired. In the present study, the aim is to detect logic errors, so we only cover programs corresponding to "Accepted" and "Wrong Answer".

AOJ has an API for accessing a part of its database. The API can be used for developing various tools. For example, there are submitted source code, submission history including log of trial and error.

4.2 Logic Error Detector API

In the present study, we developed an LED that uses source codes stored in AOJ. Furthermore, by implementing peripheral functions, we developed as a web API for the LED.

Algorithm 2 Extended Logic Error Detection

Require:
X = submitted code
Y = array of accepted programs, Y_n = n-th accepted program
Ensure:
F = Feedback
function EXTENDED DETECTION(X) ▷ Step 1
  R = empty array
  $p_{attern} \Leftarrow$ structure pattern of X ▷ Step 2
  $Y \Leftarrow$ all the correct codes submitted for a given problem as X
  $Y \Leftarrow Y$ that have the same structure pattern as p_{attern} ▷ Step 3
  for i = 1 to number of Y **do**
    add detection result based on ALGORITHM 1(X, Y_n) to R ▷ Step 4
  end for
  return $F \Leftarrow$ detection result with the lowest error degree in R ▷ Step 5
end function

From the viewpoint of data availability and practicality, source codes in Java language are analyzed and utilized for the LED API. There are three main reasons for selecting Java. Firstly, it is often adopted as an introduction to object-oriented programming languages for novice programmers. Secondly, it is the second most frequently used language in AOJ. The use of stored data makes it possible to detect logic errors more accurately, so there must be a sufficient number of comparison objects. Thirdly, an open-source library called JavaParser can be used to convert source code written in Java into an abstract syntax tree.

The LED API is implemented using docker containers that serve as a web server, an application server, and a database server. Figure 8 shows an architecture diagram of the LED API.

The web container, based on Nginx, is responsible for front-end processing and acts as a reverse proxy. At the front-end, we implemented tools for research and simple editors using Vue.js and trial of API is possible. The reverse proxy controls access to the application container. Only the web container is exposed to the outside world via the API to prevent attacks on the application and database containers.

The application container is responsible for the back-end processing by the LED, which is the application that implements the proposed algorithm. Since only the host can connect to the database container, it can be used together with the reverse proxy of the web container to prevent attacks on the database container.

LED is a Java application developed using Spring Boot based on Spring Framework. MySQL and Hibernate, as an object-relational mapper, are adopted. JavaParser is used for parsing source code written in Java.

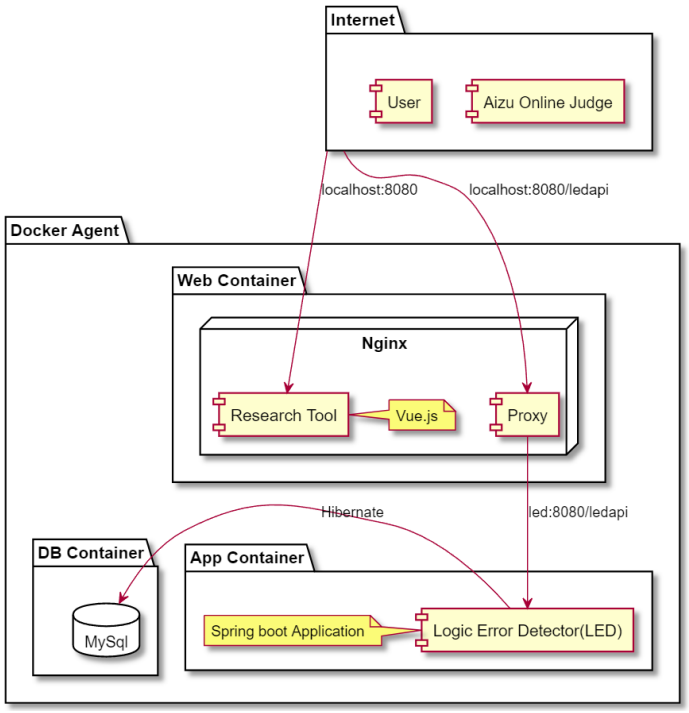


Figure 8: LED architecture diagram

4.3 Class Composition

Figure 9 shows a class diagram of LED. The SourceCode class is used to represent source codes submitted to and stored in AOJ. It has a many-one relation with the Problem, Status, and StructurePattern classes. It includes information on the ID of the user who submitted the code, the ID of the judge, and the submitted code.

The StructurePattern class represents the structure pattern of source code for each program. It has an one-to-many relationship with the SourceCode and PatternComponent classes.

The PatternComponent class is used to represent the type and order of syntax in the structure pattern. It has many-to-one and composition relationships with the StructurePattern class. This is because the structure pattern includes the type and order of syntax, and there are cases where the source code of multiple programs has the same structure patterns, but the pattern component completely depends on the structure pattern.

The ErrorReport class is used to represent the comparison result

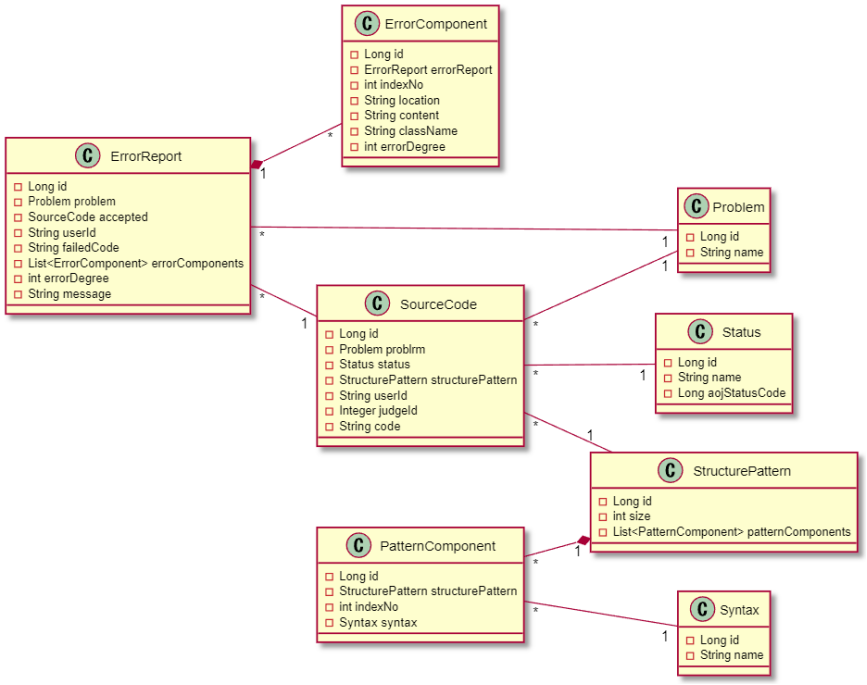


Figure 9: LED class diagram

of the logic error detection algorithm and is also a feedback class of the LED API. It has a many-to-one relationship with the SourceCode and Problem classes. It has many-to-one and composition relationships with the ErrorComponent class. It includes information on wrong codes in which logic errors were detected and correct codes used to generate the detection result selected by filtering using the error degree.

The ErrorComponent class is used to represent details of the detected logic error. It has the location, content, class of syntax, and error degree information of the logic error detected in the wrong code.

4.4 Behavior

Figure 10 shows the relationship between classes and objects. The flow of logic error detection using the LED API is as follows.

1. The user submits a source code that is judged as "Wrong Answer".
2. The wrong code is analyzed and its structure pattern is generated.
3. The structure pattern of the wrong code is searched for in the database.
4. Correct codes with the same structure pattern as that of the wrong code are searched for in the database and used as candidate codes for detection.
5. Using LED, all candidate codes are respectively compared to the wrong code and an error report is generated.
6. The error report with the lowest error degree is provided to the user as feedback.

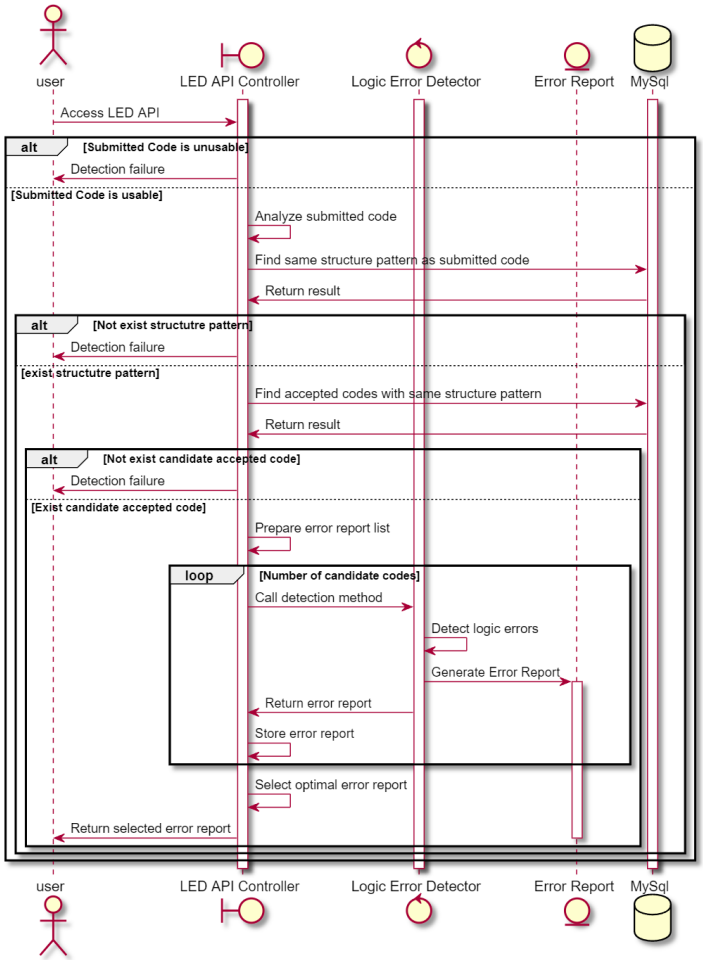


Figure 10: LED sequence diagram

5 Experiment

We conducted quantitative and qualitative experiments using the LED API and AOJ. The LED API is first used to detect logic errors in wrong code submitted by a user who has never solved the problem. Therefore, in this experiment, source codes judged to be "Wrong answer" in AOJ were used as wrong codes.

5.1 Basic Dataset from AOJ

AOJ has a problem set for novices called ITP. In this experiment, all problems in ITP were used. The dataset consists of source code that satisfies the following conditions.

- Submitted to solve the target problems.
- Judged as “Accepted” or “Wrong Answer”.
- Written in Java.
- Consists only of the main method.
- Considered not to be created by the same user.

Tables 6 and 7 show the details of the dataset. “Problem Id” is the problem id of the experimental target. “Usable Programs” is the number of available programs. “Structure Pattern” is the number of structure patterns. “Common Pattern Programs” is the number of programs with the most common structure pattern.

5.2 Experiment 1: Quantitative Method

This subsection describes experiments conducted using quantitative methods.

AOJ stores all the source codes submitted by users. Therefore, there are correct and incorrect answers to certain problems submitted by a given user. By comparing correct codes and the wrong code submitted by a given user, it is possible to obtain reliable feedback *F* necessary for debugging. From the inclusion relation of feedback *F* and feedback *F'* obtained using the LED API, we measure what portion of logic errors can be detected by the proposed algorithm.

In this experiment, we evaluated whether the following statements are true:

- The LED API can be used to detect all logic errors.
- The LED API can be used to detect some logic errors.

For evaluating the first statement, we measured whether *F'* completely includes *F*. The code debugged based on information provided by feedback *F* has already been accepted by AOJ. Therefore, when feedback *F'* completely includes feedback *F*, it can be regarded that all the logic errors were detected.

For evaluating the second statement, we measured whether *F'* contains some of *F*. During debugging, a user may modify not only the algorithm, but also variable names, the order of processing, etc. Therefore, when some of feedback *F* are included in feedback *F'*, it can be regarded that some logic errors were detected.

The following procedures were applied to each target problem.

1. Apply the proposed algorithm to correct code and wrong code submitted by a given user and obtain feedback *F*.
2. Generate feedback *F'* from the wrong code using the LED API.
3. Determine the inclusion relation of *F* and *F'*.

The dataset consists of source code that satisfies the following conditions.

- Correct code and wrong code that submitted by the same user.
- These two programs have the same structure patterns.

5.3 Experiment 2: Qualitative method

This subsection describes experiments conducted using qualitative methods.

In Experiment 1, the accuracy of logic error detection was investigated using a quantitative method. However, the experiment only measured the detection range, and it is not known whether the feedback is sufficient for debugging. Therefore, we manually debugged a code using feedback from the LED API and verified whether the wrong code was acceptable by AOJ. In this experiment, we evaluated whether the following statements are true:

- The LED API can be used to correctly detect logic errors.
- The provided feedback is appropriate for supporting learning.

The following procedures were applied to each target problem.

1. Randomly select a wrong code in AOJ.
2. Detect logic errors using the LED API.
3. Debug the wrong code manually using feedback and verify whether the debugged code is acceptable by AOJ.

Table 6: Problem Information

Problem ID	Topic	Summary of Problem
ITP1_1_A	Getting Started	Print "Hello World" to standard output.
ITP1_1_B		Calculate the cube of a given integer x
ITP1_1_C		Calculate the area and perimeter of a given rectangle.
ITP1_1_D		Read an integer S [second] and convert it to h:m:s.
ITP1_2_A	Branch on Condition	Print small/large/equal relation of given two integers a and b.
ITP1_2_B		Read three integers a, b and c, and print "Yes" if $a < b < c$, otherwise "No".
ITP1_2_C		Read integers and print them in ascending order.
ITP1_2_D		Read a rectangle and a circle, and determine whether the circle is arranged inside the rectangle.
ITP1_3_A	Repetitive Processing	Print "Hello World" 1000 times.
ITP1_3_B		Read an integer x and print it as is for multiple test cases.
ITP1_3_C		Read two integers x and y, and print them in ascending order for multiple test cases.
ITP1_3_D		Read three integers a, b, and c, and print the number of divisors of c in [a, b].
ITP1_4_A	Computation	Read two integers a and b, and calculate a / b in different data types.
ITP1_4_B		Calculate the area and circumference of a circle for given radius r.
ITP1_4_C		Read two integers, a and b, and an operator op, and then print the value of a op b.
ITP1_4_D		Read a sequence of n integers $a_i(i=1,2,...n)$, and print the minimum value, maximum value, and sum of the sequence.
ITP1_5_A	Structured Program I	Draw a rectangle whose height is H cm and width is W cm.
ITP1_5_B		Draw a frame whose height is H cm and width is W cm.
ITP1_5_C		Draw a chessboard whose height is H cm and width is W cm.
ITP1_5_D		Structured programming without *goto* statement.
ITP1_6_A	Array	Read a sequence and print it in reverse order.
ITP1_6_B		For given a deck of cards, find missing cards.
ITP1_6_C		Count the number of elements in a three-dimensional array.
ITP1_6_D		Read a n×m matrix A and a m×1 vector b, and print their product Ab.
ITP1_7_A	Structured Program II	Read a list of student scores and evaluate the grade for each student.
ITP1_7_B		Find the number of combinations of integers that satisfy 1) three distinct integers from 1 to n, and 2) sum of the integers is x.
ITP1_7_C		Read the number of rows r, columns c, and a table of r×c elements, and print a new table that includes the total sum for each row and column.
ITP1_7_D		Read a n×m matrix A and a m×l matrix B, and print their product.
ITP1_8_A	Character	Convert uppercase/lowercase letters to lowercase/uppercase.
ITP1_8_B		Read an integer and print the sum of its digits.
ITP1_8_C		Count the number of appearances for each letter (ignoring case).
ITP1_8_D		Find a pattern p in ring-shaped text s.
ITP1_9_A	String	Read a word W and text T, and print the number of times that word W appears in text T.
ITP1_9_B		Read a deck (a string) and shuffle it, and print the final state (a string).
ITP1_9_C		Read a sequence of Taro and Hanako cards and report the final scores of the game.
ITP1_9_D		Perform a sequence of commands including reverse and replace operations.
ITP1_10_A	Math Functions	Calculate the distance between two points $P1$($x1$, $y1$) and $P2$($x2$, $y2$).
ITP1_10_B		For the given two sides of a triangle a and b and the angle C between them, calculate the area, circumference, and height of the triangle.
ITP1_10_C		Calculate the standard deviation.
ITP1_10_D		Read two n-dimensional vectors x and y, and calculate Minkowski's distance, where $p=1,2,3$, ∞.
ITP1_11_A	Structure and Class	Simulate rolling a dice.
ITP1_11_B		Print the integer on the right side face of the rolled die.
ITP1_11_C		Read two dice and determine whether they are identical.
ITP1_11_D		Read n dice and determine whether they are all different.

6 Results and Discussion

In this section, we present and discuss the results of each experiment.

6.1 Results of Experiment 1

Table 8 shows the results of Experiment 1. "Problem Id" is the problem id of the experimental target. "All Users" is the number of users who submitted source code. "Usable Users" is the number

Table 7: Statistics of the Dataset Used in the Experiment

Problem ID	Number of Programs	Usable Programs	Structure Pattern	Common Pattern Programs
ITP1_1_A	5132	5094	5	5030
ITP1_1_B	4408	3713	50	1274
ITP1_1_C	3292	2743	56	744
ITP1_1_D	2365	1832	68	504
ITP1_2_A	2836	2231	82	629
ITP1_2_B	2283	1683	68	619
ITP1_2_C	2779	1099	71	234
ITP1_2_D	2919	1682	60	628
ITP1_3_A	2290	2194	15	1436
ITP1_3_B	3180	1303	90	71
ITP1_3_C	2538	727	75	67
ITP1_3_D	1622	983	58	261
ITP1_4_A	3715	2605	66	541
ITP1_4_B	2263	1825	56	316
ITP1_4_C	1506	347	49	35
ITP1_4_D	3027	735	48	54
ITP1_5_A	1477	701	65	107
ITP1_5_B	1404	354	46	57
ITP1_5_C	1236	421	53	58
ITP1_5_D	3120	717	58	147
ITP1_6_A	1345	447	46	53
ITP1_6_B	1027	181	25	23
ITP1_6_C	1340	212	19	55
ITP1_6_D	743	295	32	36
ITP1_7_A	1683	196	29	28
ITP1_7_B	1287	230	28	33
ITP1_7_C	839	109	19	14
ITP1_7_D	1207	216	24	37
ITP1_8_A	927	329	44	23
ITP1_8_B	797	227	30	32
ITP1_8_C	1444	152	23	19
ITP1_8_D	721	233	23	44
ITP1_9_A	1409	270	27	35
ITP1_9_B	642	151	17	28
ITP1_9_C	710	174	23	29
ITP1_9_D	613	74	11	17
ITP1_10_A	594	387	28	98
ITP1_10_B	670	339	24	84
ITP1_10_C	935	78	11	17
ITP1_10_D	420	61	8	14
ITP1_11_A	135	20	4	11
ITP1_11_B	109	18	4	9
ITP1_11_C	108	11	2	8
ITP1_11_D	19	0	0	0

of users who submitted both correct and wrong code with the same structure patterns (and thus logic errors can be detected using the LED API). "Whole Success Users" is the number of users for whom all logic errors were detected using the LED API. In other words, this is the result of the first statement. "Partial Success Users" is the number of users for whom some logic errors were detected using the LED API. In other words, this is the result of the second statement. "Whole Success Ratio" is the ratio of Whole Success Users to Usable Users. "Partial Success Ratio" is the ratio of Partial Success Users to Usable Users. "Difference Success Ratio" is the difference between Whole Success Rate and Partial Success Rate.

The first item is the complete detection rate of logic errors. The results of the experiment show that all logic errors were detected for many problems. Overall, we succeeded in detecting logic errors in more than 70% of programs.

The second item is the partial detection rate of logic errors. The results of the experiments show that a higher detection ratio for partial logic errors was achieved for many problems. Overall, we succeeded in detecting logic errors in more than 80% of programs.

6.2 Results of Experiment 2

Table 9 shows the results of Experiment 2. The first item is the detection accuracy of logic errors. Considering the features of the evaluation method, it was considered that logic errors could not be correctly detected when the source code could be modified only by factors outside the feedback range. The results of the experiment show that we successfully detected logic errors in 100% of test cases. Therefore, the proposed detection algorithm has sufficient ability to detect logic errors in wrong code.

The second item is related to the appropriateness of feedback. The experimental results show that we succeeded in providing appropriate feedback in 80.4% of the test cases. This evaluation item is based on the qualitative method used to verify the appropriateness for supporting learning. Therefore, we should discuss the appropriateness of the judgment.

6.3 Discussion

In this subsection, we discuss the experimental results. Some ideas for improvement based on the experimental results are also given.

6.3.1 Discussion for Experiment 1

The results of Experiment 1 confirm that the proposed detection algorithm can support users who attempt to solve problems. We also confirmed that the detection algorithm can detect all the logic errors in a wrong code in many cases. However, results for statement 2 indicate that feedback may contain syntax that is unrelated to the algorithm. Therefore, it is necessary to improve detection accuracy by improving the detection algorithm.

6.3.2 Discussion for Experiment 2

The results of Experiment 2 confirm that logic errors can be detected by the proposed detection algorithm. In addition, we confirmed that the detection algorithm can provide appropriate feedback in many cases. Therefore, it is necessary to generalize some specific cases and to improve the algorithm.

6.3.3 Discussion for All Experiments

The results of the two experiments confirmed that the proposed algorithm can accurately detect logic errors and provide appropriate feedback. The proposed algorithm detects logic errors using source codes stored in the online judge system. To increase accuracy, we filter the codes using structure patterns and error degree. Therefore, detection accuracy depends on the diversity of source code stored in the online judge system. In the present study, although we only considered codes written in Java, the proposed algorithm can be applied to all context-free languages. There is more source code written in languages such as C and C ++, and thus logic errors can be detected with higher accuracy.

6.3.4 Improvements

In this subsubsection, we suggest some improvements based on the experimental results.

Based on the results of Experiment 1, we suggest the following two improvements. One improvement is the removal of elements not related to logic errors detected by the detection algorithm. For example, elements with the same meaning in different formats, such as variable names and logical expressions, can be removed because they are not related to logic errors. Another improvement is the use of a comparison method that is independent of the structure pattern (i.e., generalization of the algorithm). In Experiment 1, a rule regarding the structure pattern as a condition of the dataset was used. This rule was used because the LED API does not have a function for comparing
source codes with different structure patterns. Therefore, we suggest generalizing the algorithm by unifying the variable declaration part and determining the format of the calculation formula. This generalization would make it possible to compare source codes with different structure patterns without changing the meaning of the algorithm.

Based on the results of Experiment 2, we suggest the following three improvements. The first improvement is a comparison method of conditional branching. The conditional branching algorithm strongly depends on the logical expression that is employed by the user. In particular, filtering by structure pattern is not effective for problems that require multiple standard outputs using conditional branching. Therefore, we suggest a comparison based on a logical expression and feedback regarding the processing order. The second improvement is a comparison method of binary expressions. Because binary expressions are parsed using certain rules, a comparison between binary expressions from different structures are not effective. Therefore, we suggest making the error degree more accurate by implementing special logic to compare binary expressions. The third improvement is a review of the weighting when the type of syntax is the same but the elements are different. As an example, method call expressions can be very different, even using the same kind of syntax.

Table 8: Results of Experiment 1

	Users		Success Users		Success Rate		
Problem ID	All	Usable	Whole	Partial	Whole	Partial	Difference
ITP1_1_A	3502	358	335	358	93.58%	100.00%	+6.42%
ITP1_1_B	2425	144	110	121	76.39%	84.03%	+7.64%
ITP1_1_C	2006	187	114	142	60.96%	75.94%	+14.97%
ITP1_1_D	1618	130	86	102	66.15%	78.46%	+12.31%
ITP1_2_A	1692	291	249	263	85.57%	90.38%	+4.81%
ITP1_2_B	1525	126	106	112	84.13%	88.89%	+4.76%
ITP1_2_C	1505	87	49	74	56.32%	85.06%	+28.74%
ITP1_2_D	1194	261	93	111	35.63%	42.53%	+6.90%
ITP1_3_A	1542	270	201	208	74.44%	77.04%	+2.59%
ITP1_3_B	1405	129	108	114	83.72%	88.37%	+4.65%
ITP1_3_C	1199	63	55	57	87.30%	90.48%	+3.17%
ITP1_3_D	1059	105	92	97	87.62%	92.38%	+4.76%
ITP1_4_A	1107	315	180	252	57.14%	80.00%	+22.86%
ITP1_4_B	1134	250	157	184	62.80%	73.60%	+10.80%
ITP1_4_C	1022	14	14	14	100.00%	100.00%	0%
ITP1_4_D	909	176	107	147	60.80%	83.52%	+22.73%
ITP1_5_A	977	24	21	22	87.50%	91.67%	+4.17%
ITP1_5_B	911	33	22	26	66.67%	78.79%	+12.12%
ITP1_5_C	880	20	14	16	70.00%	80.00%	+10.00%
ITP1_5_D	968	58	51	53	87.93%	91.38%	+3.45%
ITP1_6_A	831	17	15	16	88.24%	94.12%	+5.88%
ITP1_6_B	656	15	12	14	80.00%	93.33%	+13.33%
ITP1_6_C	623	33	26	27	78.79%	81.82%	+3.03%
ITP1_6_D	566	11	9	11	81.82%	100.00%	+18.18%
ITP1_7_A	719	21	20	20	95.24%	95.24%	0%
ITP1_7_B	643	24	21	23	87.50%	95.83%	+8.33%
ITP1_7_C	595	3	3	3	100.00%	100.00%	0%
ITP1_7_D	484	66	49	64	74.24%	96.97%	+22.73%
ITP1_8_A	636	14	13	14	92.86%	100.00%	+7.14%
ITP1_8_B	591	2	0	1	0.00%	50.00%	+50.00%
ITP1_8_C	525	5	5	5	100.00%	100.00%	0%
ITP1_8_D	448	11	7	8	63.64%	72.73%	+9.09%
ITP1_9_A	518	52	35	42	67.31%	80.77%	+13.46%
ITP1_9_B	456	7	7	7	100.00%	100.00%	0%
ITP1_9_C	441	6	2	5	33.33%	83.33%	+50.00%
ITP1_9_D	323	4	4	4	100.00%	100.00%	0%
ITP1_10_A	432	32	15	22	46.88%	68.75%	+21.88%
ITP1_10_B	392	38	20	31	52.63%	81.58%	+28.95%
ITP1_10_C	458	14	9	10	64.29%	71.43%	+7.14%
ITP1_10_D	279	2	2	2	100.00%	100.00%	0%
ITP1_11_A	84	1	1	1	100.00%	100.00%	0%
ITP1_11_B	64	1	1	1	100.00%	100.00%	0%
ITP1_11_C	32	1	1	1	100.00%	100.00%	0%
ITP1_11_D	13	0	0	0	0.00%	0.00%	0%

Table 9: Results of Experiment 2

Feedback Type	Reason for Judgment	Cause of Logic Error	Rate	Total
Appropriate	Can be debugged	Literal	6.5%	80.43%
		Output format	19.6%	
		Binary expression	15.2%	
		Loop condition	8.7%	
		Type and output format	2.2%	
		Conditional branch	19.6%	
		Used type	4.3%	
		Used method	4.3%	
Inappropriate	Direct answer	Position of variables	17.4%	19.57%
	Ignored the algorithm	Non-optimal feedback	2.2%	
Wrong	Cannot be debugged		0.0%	0.0%

However, even with such a comparison, the error degree is a very small numerical value. In particular, standard output methods and various library methods are sometimes compared, so that optimal feedback cannot be generated in some cases. Therefore, we suggest measuring to what extent the syntax has the same meaning by using the namespace and arguments. We also suggest detecting the parent syntax itself as a logic error, such as when only variable names are detected as logic errors.

7 Conclusion

This study proposed a logic error detection algorithm based on structure patterns and error degree to support novice programmers. The proposed algorithm uses data in an online judge system to detect logic errors in wrong code and provides feedback for supporting problem solving. In addition, we developed the LED web API based on the proposed algorithm. The LED API is compatible with the online judge system, so it can promote autonomous learning with the corresponding environment.

Two experiments were conducted to verified the accuracy of the proposed algorithm. The experimental results show that the proposed algorithm can accurately detect logic errors and provide appropriate feedback. The proposed algorithm detected logic errors in programs for all the problems in the Introduction to Programming 1 set with high accuracy. In addition, the detection accuracy depends on the amount of data in the online judge system.

Although we constructed the autonomous learning environment and detection algorithm in Java, the proposed algorithm can be applied to various programming languages, such as C/C++, Python, and Ruby. Although the presented LED API is oriented to detec logic errors, the algorithm is capable of suggesting possible solutions. So, our future works include the extension of the API after conducting further experiments.

References

[1] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" 2013. [Online]. Available: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

[2] I. Koichi and T. Yuta, "Digitization, Computerization, Networking, Automation, and the Future of Jobs in Japan (article in Japanese with an abstract in English)," Research Institute of Economy, Trade and Industry (RIETI), Policy Discussion Papers (Japanese) 18009, May 2018. [Online]. Available: https://ideas.repec.org/p/eti/rpdpjp/18009.html

[3] W. E. Forum, "The future of jobs report 2018," 2018. [Online]. Available: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

[4] J. Gal-Ezer, C. Beeri, D. Harel, and A. Yehudai, "A high school program in computer science," *Computer*, vol. 28, no. 10, pp. 73–80, Oct. 1995. [Online]. Available: http://dx.doi.org/10.1109/2.467599

[5] S. Furber *et al.*, "Shut down or restart? the way forward for computing in uk schools," *The Royal Society, London*, 2012. [Online]. Available: https://royalsociety.org/~/media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf

[6] Y. Okajima, "Research on the impact of compulsory programming education," *Chuo University Policy Research Institute Annual Report*, no. 21/2017, pp. 203 – 218, 2018-08-23. [Online]. Available: http://id.nii.ac.jp/0341/00005893

[7] K. Komatsu, "Problems and countermeasures of programming education (article in japanese)," *Business review, Faculty of Business Administration, Bunkyo Gakuin University*, vol. 25, no. 1, pp. 83–104, dec 2015. [Online]. Available: https://ci.nii.ac.jp/naid/40020853834/en/

[8] S. Furber *et al.*, "After the reboot: computing education in uk schools," *The Royal Society, London*, 2017. [Online]. Available: https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf

[9] C. M. Intisar and Y. Watanobe, "Classification of online judge programmers based on rule extraction from self organizing feature map," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 313–318.

[10] T. Saito and Y. Watanobe, "Learning path recommender system based on recurrent neural network," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 324–329.

[11] C. M. Intisar and Y. Watanobe, "Cluster analysis to estimate the difficulty of programming problems," in *Proceedings of the 3rd International Conference on Applications in Information Technology(ICAIT)*, 2018.

[12] S. Wasik, M. Antczak, J. Badura, A. Laskowski, and T. Sternal, "A survey on online judge systems and their applications," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 3:1–3:34, Jan. 2018. [Online]. Available: http://doi.acm.org/10.1145/3143560

[13] Y. Yoshizawa and Y. Watanobe, "Logic error detection algorithm for novice programmers based on structure pattern and error degree," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 297–301.

[14] H. Holec, *Autonomy and foreign language learning.* ERIC, 1979.

[15] D. G. Little, *Learner autonomy: Definitions, issues and problems.* Authentik Language Learning Resources, 1991.

[16] L. Dickinson, "Autonomy and motivation a literature review," *System*, vol. 23, no. 2, pp. 165–174, 1995.

[17] H. Wu, Y. Liu, L. Qiu, and Y. Liu, "Online judge system and its applications in c language teaching," in *2016 International Symposium on Educational Technology (ISET)*, July 2016, pp. 57–60.

[18] S. Buyrukoglu, F. Batmaz, and R. Lock, "Increasing the similarity of programming code structures to accelerate the marking process in a new semi-automated assessment approach," in *2016 11th International Conference on Computer Science Education (ICCSE)*, Aug 2016, pp. 371–376.

[19] E. Stankov, M. Jovanov, B. Kostadinov, and A. M. Bogdanova, "A new model for collaborative learning of programming using source code similarity detection," in *2015 IEEE Global Engineering Education Conference (EDUCON)*, March 2015, pp. 709–715.

[20] Y. Teshima and Y. Watanobe, "Bug detection based on lstm networks and solution codes," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2018, pp. 3541–3546.

[21] X. Su, J. Qiu, T. Wang, and L. Zhao, "Optimization and improvements of a moodle-based online learning system for c programming," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–8.

[22] E. Araujo, M. Gaudencio, D. Serey, and J. Figueiredo, "Applying spectrum-based fault localization on novice's programs," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–8.

[23] T. Nguyen and C. Chua, "A logical error detector for novice php programmers," in *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, July 2014, pp. 215–216.

[24] K. K. Sharma, K. Banerjee, I. Vikas, and C. Mandal, "Automated checking of the violation of precedence of conditions in else-if constructs in students' programs," in *2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)*, Dec 2014, pp. 201–204.

[25] H. Blau, S. Kolovson, W. R. Adrion, and R. Moll, "Automated style feedback for advanced beginner java programmers," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–9.

[26] J. Feng, B. Cui, and K. Xia, "A code comparison algorithm based on ast for plagiarism detection," in *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, Sept 2013, pp. 393–397.

[27] G. Tao, D. Guowei, Q. Hu, and C. Baojiang, "Improved plagiarism detection algorithm based on abstract syntax tree," in *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, Sept 2013, pp. 714–719.

[28] Y. Zhang, X. Gao, C. Bian, D. Ma, and B. Cui, "Homologous detection based on text, token and abstract syntax tree comparison," in *2010 IEEE International Conference on Information Theory and Information Security*, Dec 2010, pp. 70–75.

[29] Y. Watanobe, "Development and operation of an online judge system," *Information Processing*, vol. 56, no. 10, pp. 998–1005, sep 2015. [Online]. Available: https://ci.nii.ac.jp/naid/40020591038/

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3,10-15 (2019)

www.astesj.com

Special Issue on Advanced Electrical and Communication Technologies

ASTESJ

ISSN: 2415-6698

Microstrip Patch Antenna for Ultra-Wideband Applications

QADDI Mohamed Hamza*[1], SRIFI Mohamed Nabil[1] and MHARZI Hassan[2]

[1]*Telecommunication Systems Research Group, National School of Applied Sciences, Ibn Tofail University Kenitra, Morocco*

[2]*Laboratory of Electric Engineering, Computing and Mathematical Sciences National School of Applied Sciences, Ibn Tofail University Kenitra, Morocco*

ARTICLE INFO

Article history:
Received: 31 January, 2019
Accepted: 17 April, 2019
Online: 04 May, 2019

Keywords:
Ultra-wideband
Micro-strip antenna
Directivity
Gain
Slots

ABSTRACT

A new microstrip patch antenna for ultra-wideband applications is presented in this paper. The design and performance of the antenna component are discussed. The propounded antenna is mounted on a compact FR-4 substrate having dimensions 20 x 30 x 1.6 mm^3 with relative permittivity εr=4.3. The rectangular patch antenna is slotted with two types; rectangular and semi-circle slots so as to ensure a broad bandwidth. The results reveal that the antenna covers the frequency range of 3.1–7.5 GHz with a reflection coefficient reduced to -55 dB and a maximum gain of 5.9 dB. The details of the simulated and measured results for reflection coefficient are presented, showing a good agreement between them. To analyze the effect of the slots, the surface current distribution is investigated. The performances of good impedance matching are achieved within the operating band. Simulations are performed using CST Microwave Studio. The propounded antenna can be deployed for UWB applications and other radio communication services systems such as high-resolution radar, military communication, communications and sensors, position location and tracking.

1. Introduction

UWB technology has recently attracted significant attention in wireless communications thanks to its main features in particular high bandwidth and data rate, ensuring secure communication in military applications. The world of ultra-wideband (UWB) has changed in very recent history. In February 2002, the FCC (Federal communications commission) issued a ruling that UWB could be used for commercial applications on unlicensed basis as well as for radar and safety applications. The decision authorized very low power spectral density emission in a bandwidth ranging from 3.1 to 10.6 GHz [1-3]. UWB appears to be a technology very promising for wireless communications at very high levels data-rate, high-precision radar and imaging systems. UWB systems use short pulses (of the order of the picosecond), repeated at a certain rate that can be up to several giga-impulses per second, offering thus a very high bandwidth with a very low transmission power level. This offers UWB systems the possibility of coexisting with other systems.

*Corresponding Author: QADDI Mohamed Hamza, 27 Bir Rami EST, Kenitra, Morocco, Cell No; +212 661634417, qaddihamza@gmail.com

To fulfill the demand in this type of communications, many research on UWB antenna design have been published. Most of them focused on patch antennas because of their simple structure, small size, low cost, low complexity, light weight, and high-speed data rate. They are considered as good candidates for UWB applications among many versions of UWB antennas. Nevertheless, micro-strip patch antenna has narrow bandwidth.

To resolve this problem and obtain the desired performances, many methods are used. The typical technique focus on cutting slots of several forms in the radiating element, the ground plane, and the feed line. Nowadays, many slot shapes are available and discussed by various researchers [4-6]. Slot antennas have proven to be useful in that context. However design and placement of these slots on the radiating element is challenging as compared to the other traditional techniques. The purpose of this project is to design a micro-strip patch antenna for UWB applications that covers the frequency range of 3.1–7.5 GHz.

2. Design and Configuration of the Antenna

The propounded antenna is a rectangular patch antenna which is fed by a micro-strip line and mounted on a dielectric substrate named FR4-lossy, with thickness h=1.6mm and relative

permittivity εr=4.3, having dimensions of 20 x 30 x 1.6 mm^3. These specifications for the substrate are a part of material parameters window on CST Microwave Studio. The material used in the patch and the ground is lossy copper with a thickness of Mt=0.035mm as given in Tab .1. The width of the line feeding is 3 mm and the length is 14.5 mm. The partial ground plane is also used with the propounded antenna geometry as shown in Figure 1. To investigate the impact of various antenna's characteristics (different slots), a meticulous parametric study was performed. The parametric values of the propounded antenna are shown in Tab.1. The antenna is excited by a 50-Ω micro-strip feed line. The parameters measurement of the patch are calculated using the following formulas [7]:

Width formula of rectangular:

$$W_p = \left(\frac{c}{2 f_r}\right)\left(\frac{\varepsilon_r + 1}{2}\right)^{-0.5}$$

Where c = 3×108 ms-1 and $\varepsilon_r = 4.3$

Formula of effective dielectric constant:

$$\varepsilon_{eff} = \left(\frac{\varepsilon_r + 1}{2}\right) + \left(\frac{\varepsilon_r - 1}{2}\right)\left(1 + \frac{12\ h}{W_p}\right)^{-0.5}$$

At h=1.6mm

Formula of length extension:

$$\Delta L = 0.412\ h\left(\frac{\varepsilon_{eff} + .3}{\varepsilon_{eff} - 0.258}\right)\left(\frac{\left(\frac{W_p}{h}\right) + 0.264}{\left(\frac{W_p}{h}\right) + 0.8}\right)$$

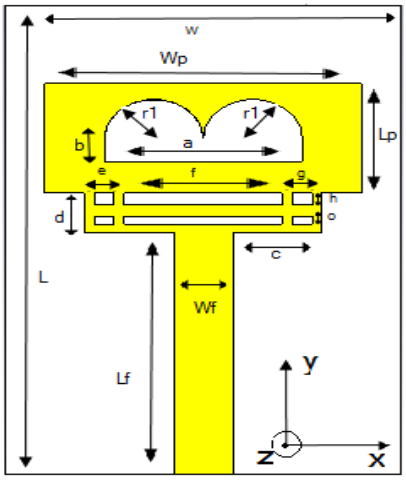
Length formula of rectangular:

$$L_p = \left(\frac{c}{2 f_r \sqrt{\varepsilon_{eff}}}\right) - 2\Delta L$$

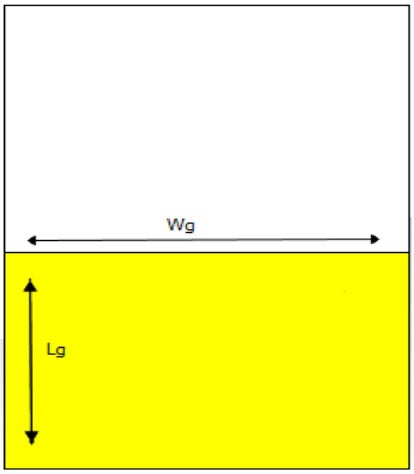
Table 1: Parameter values of the propounded antenna

Parameters	Values (mm)
L	30
W	20
Lp	7
Wp	16
Lg	14
Wg	20
h	1.6
Mt	0.035
Wf	3
Lf	14.5
a	10
b	1.5
c	4.5
d	2.5
e	1
f	8
g	1
h	0.75
o	0.5
r1	2.5

To amplify the coupling between the feed line and the slots, multiple slots are introduced in order to broaden the operating bandwidth of the antenna. Having a quite large slot is logical transition to obtain wide bandwidth from an aperture-coupled patch. The first two slots are designed by combining rectangular and semi-circle slots placed in a juxtapose way [8-10]. In addition, different other rectangular slots are included in the radiating element. The structure of the antenna design is illustrated in Fig .1.



(a)



(b)

Figure 1 Design of the propounded UWB antenna (a) Top (b) Bottom

To improve the performance of micro-strip patch antenna, it is advisable to design a combination of two or more slots on the radiating element. This can be observed from the discussed literature survey that a micro-strip patch antenna with slots can be designed for wireless applications [11]. The technique employed in this paper retains the same elementary structure of the antenna with different slots included in the radiating element. In order to obtain a wide bandwidth, a rigorous investigation on different parameters in terms of feed width (Wf), feed length (Lf) and different slots width and length was performed. From the analysis we found the optimum dimensions of above parameters to be 3 mm, 14.5 mm respectively for the feed line, to meet the

requirement. The slots dimensions and other antenna's parameters are listed in Tab .1 on the basis of undertaken parametric analysis.

Figure 2 presents the simulated reflection coefficient of the antenna, the antenna without rectangular slots and the antenna without semi-circle slots. When different slots are included, a wide bandwidth is achieved from 3.1 to 7.5 GHz with three resonant frequencies at 3.9, 5.95 and 7.3 GHz. Effect of using different slots is shown. We notice that the propounded antenna is the best candidate to meet the requirements in terms of UWB bandwidth. The flow chart of the propounded antenna's design methodology is presented in figure 3.

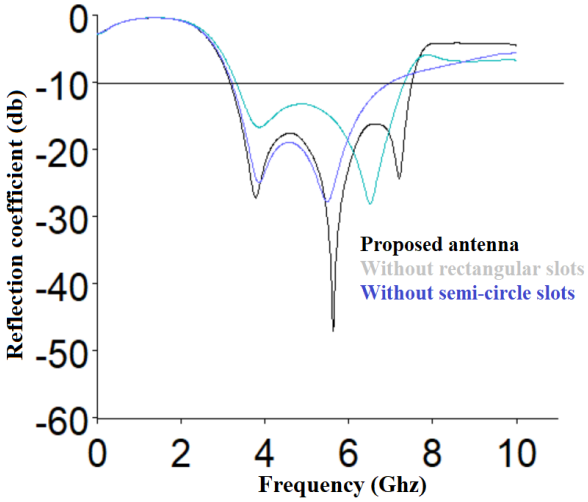


Figure 2 Simulated reflection coefficient of the propounded antenna, the antenna without rectangular slot and the antenna without semi-circle slots

3. Results and Discussions

The propounded UWB antenna is designed by opting for the optimal values of above parameters in order to cover the frequency range 3.1- 7.5 GHz. Both numerical and experimental tests have been achieved to assess the effectiveness of the propounded antenna and the results are presented in this section. Toward this end, the antenna was fabricated on FR4-lossy substrate (30 x 20 x 1.6 mm^3) with dielectric constant of Ɛr = 4.3. Figure 4 depicts the fabricated antenna for experimental verification.

The reflection coefficient of the antenna has been analyzed using CST Microwave STUDIO simulator and verified using another electromagnetic simulator HFSS as shown in Figure 5. A 3.1–7.5 GHz frequency range below -10 dB of return loss S11 is obtained. The reflection coefficient (S11), measured in decibel (dB), can be calculated using the formula:

$$S_{11} = -20log_{10}\left|\frac{Z_{in} - Z_0}{Z_{in} + Z_0}\right|$$

Where Z0 is the characteristic impedance of the 50 Ω SMA port and Zin is the driving point impedance of the antenna. Figure 5 shows a comparison of the simulated reflection coefficient using both CST and HFSS. The results corroborate that there is a good agreement between the two simulations with a small variation as the two softwares use different numerical techniques, finite integration technique, a relative of FDTD for CST Microwave Studio and Finite Element Method (FEM) for HFSS [12].

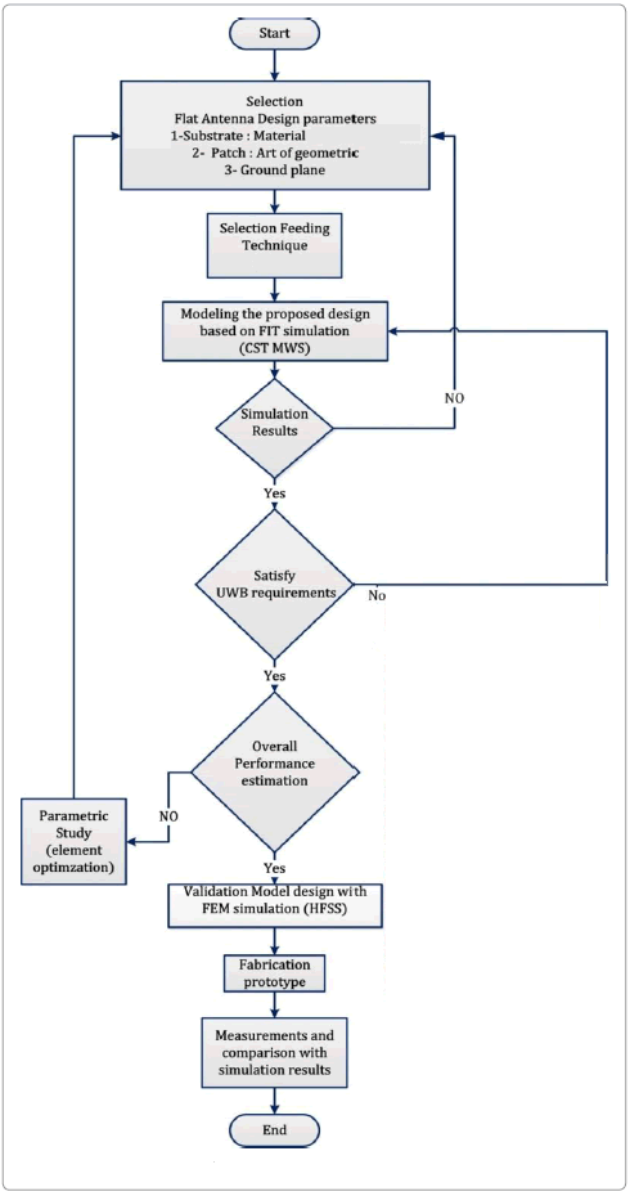


Figure 3 The flow chart of the propounded antenna's design methodology.

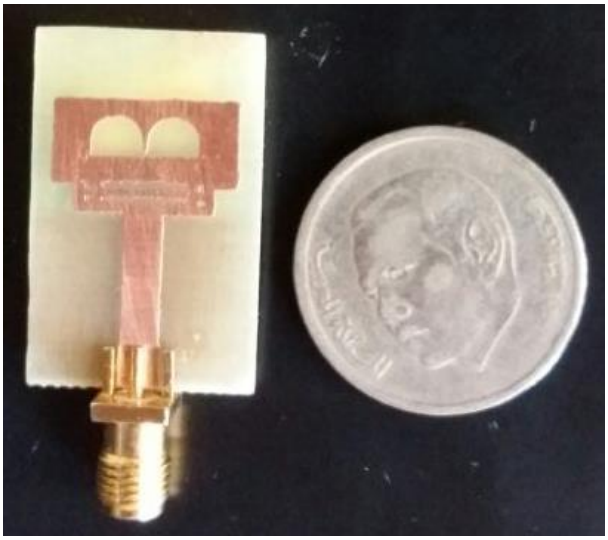


Figure 4 Photograph of the fabricated UWB antenna

The UWB antenna was measured after fabrication to examine the performance of the proposed approach. Measured reflection coefficient is achieved by vector network analyzer. The measured and simulated reflection coefficients of antenna are compared in Figure 6. From measured results, we can notice that the antenna is operating with a -10dB bandwidth from 3.1 to 7.5 GHz. A good agreement between simulation and measurement results is observed with a little discrepancy which is mainly owing to the fabrication margin. It could also be because to the impact of the feeding cable, feeding connector and the antenna fixation support as the structure is small.

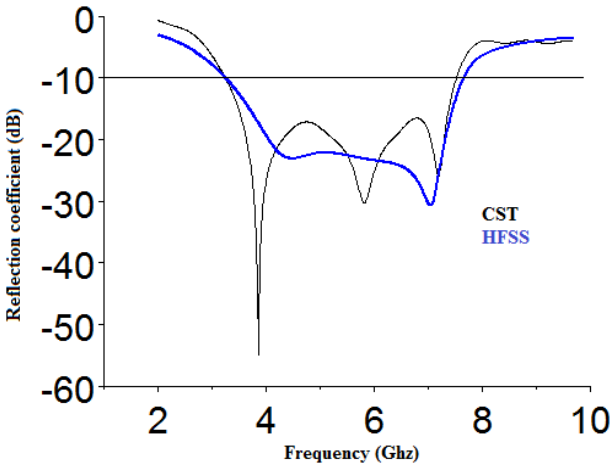


Figure 5 Simulated reflection coefficient of the propounded antenna

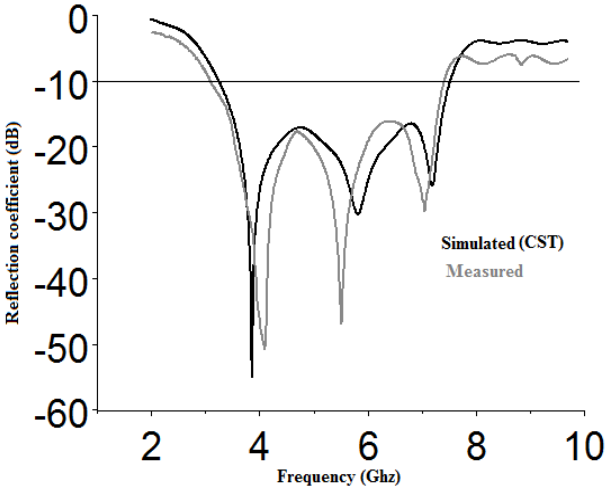


Figure 6 Measured and simulated reflection coefficient of the UWB antenna

The electromagnetic solver, CST Microwave STUDIO, is used to numerically analyze and improve the antenna's configuration. Figure 7 presents the simulated results of the propounded antenna with the rectangular slot combined with semi-circle slots width, from 0.5 to 1.5 mm. It should be noticed that one variable at the time was varied, the others being constant. It is apparent that the bandwidth for UWB band decreases as the width decreases from 1.5 to 0.5 mm. Consequently, it is concluded that b= 1.5 mm is the optimal value for the bandwidth from 3.1 to 7.5 GHz, covering the whole UWB range with three resonant frequencies at 3.9, 5.95 and 7.3 GHz.

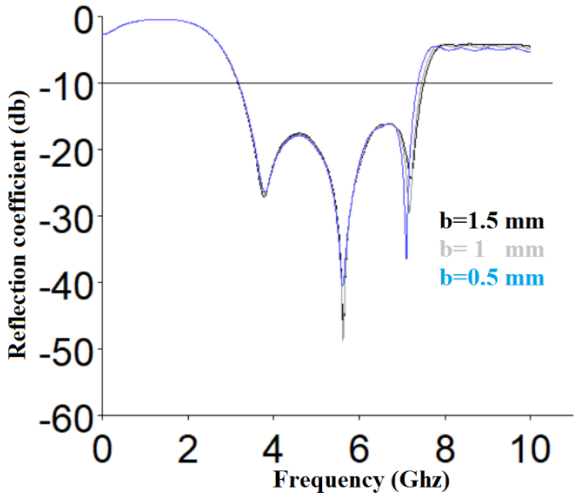
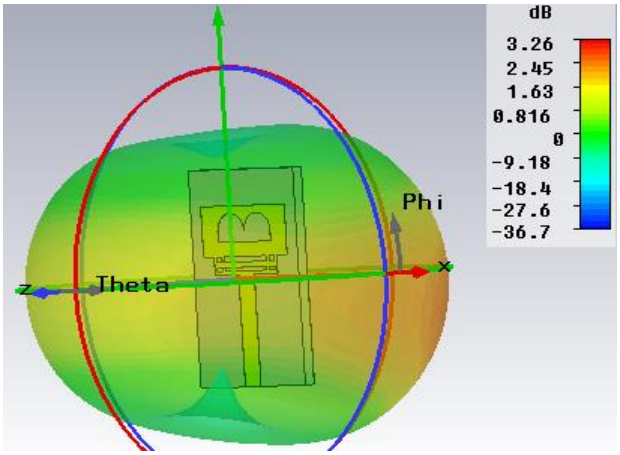
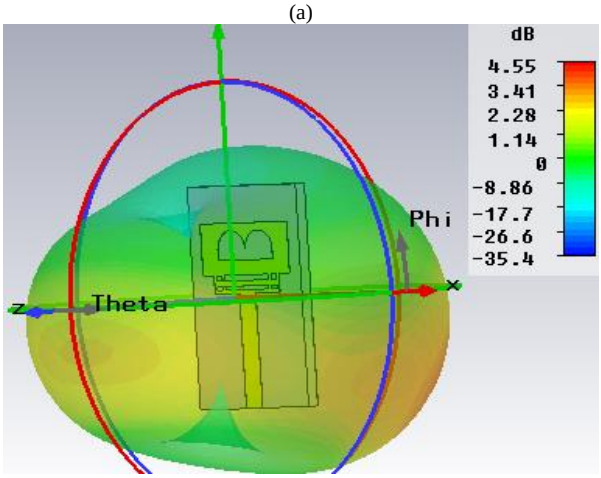


Figure 7 Simulated reflection coefficient response of the antenna as a function of "b". All other parameters are the same as listed in Table 1

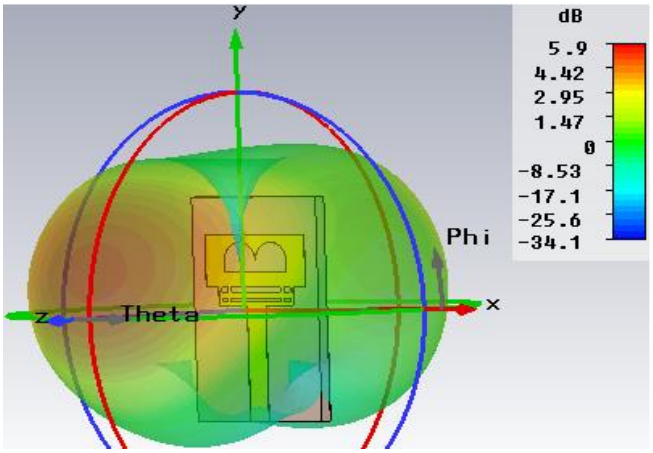
The propounded UWB antenna has an acceptable quasi omnidirectional radiation pattern needed to receive information signals from all directions. Figure 8 presents three-dimensional radiation patterns at three frequencies 3.9, 5.95 and 7.3 GHz. The gain of the antenna attain 3.26 dB at 3.9 GHz, 4.55 dB at 5.95 GHz and 5.90 dB at 7.3 GHz.



(a)

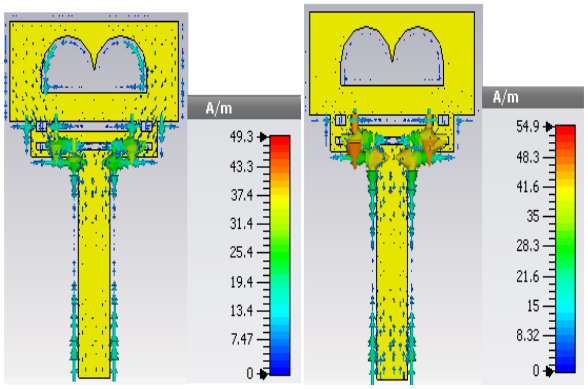


(b)

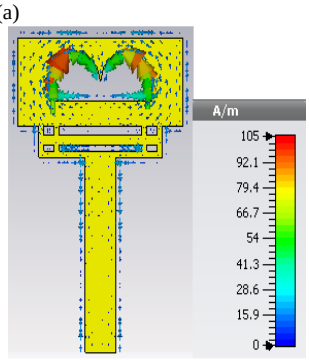


(c)

Figure 8. 3D radiation patterns of the propounded antenna at (a) 3.9, (b) 5.95 and (c) 7.3 GHz.



(a) (b)



(c)

Figure 9 Surface Current distributions at (a) 3.9 GHz, (b) 5.95 GHz and (c) 7.3 GHz

Surface current distribution has been depicted in Figure 9 and analyzed at discrete frequencies (a) 3.9 GHz, (b) 5.95 GHz and (c) 7.3 GHz, to give a physical perception and have a deeper understanding on the resonant behavior of the antenna. Figure 9 (a) and (b) shows that the surface current distributions are mainly flow and distributed though along the rectangular slots and the feed part. From Figure 9 (c), it can be observed that most of surface currents are concentrated on the edges of the interior and exterior of semi-circle slots. Effective coupling between different slots affords wide-band matched impedance bandwidth, therefore, the surface current is distributed evenly over the radiating patch. The maximum gain of the propounded antenna is presented in Figure 10. As expected, the antenna has a good gain over the whole operating frequency range with a maximum gain up to 5.9 dB.

The simulated radiation efficiency of the antenna is presented in Figure 11. It can be noticed that the antenna reaches a maximum radiation efficiency of 81% and more than 60% over the whole UWB frequency range.

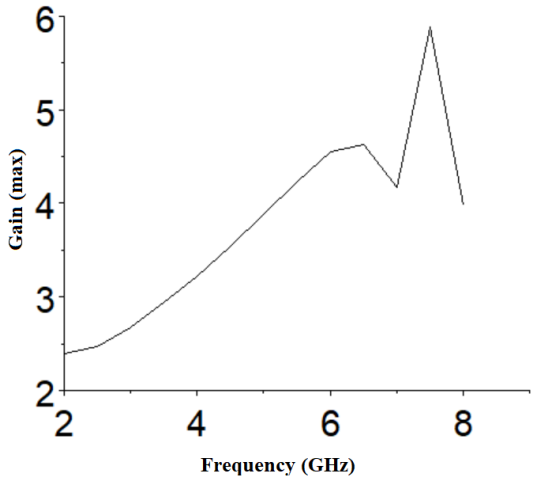


Figure 10 The gain of the propounded antenna

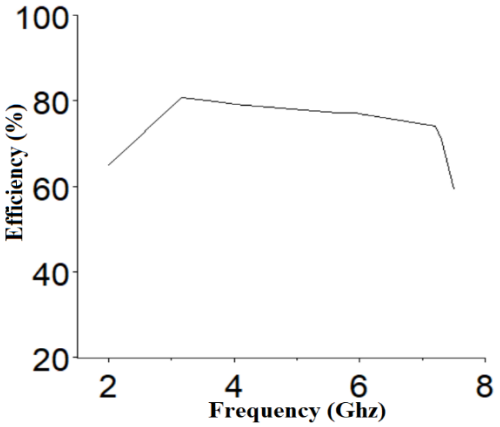


Figure 11 simulated radiation efficiency of the antenna

A comparison between our antenna and other published research is demonstrated in Table 2 in terms of their dimensions, bandwidth, gain and applications, in order to reinforce the concept of design. The almost stable radiation pattern with gain up to 5.9 dB makes our antenna suitable for being used for UWB applications.

Table 2 A comparison between the propounded antenna and other published work

References	Dimensions (in mm)	Bandwidth UWB	Gain at resonant frequency (dB)	Applica-tions
[13]	13x15	3.07-11.64	5.49	UWB applications
[14]	22x24	2.8-11.4	3.6 5.4	UWB applications
[15]	27x32.42	3.1-5.2 5.8-10.6	2.33 5.49	UWB applications
Our work	20x30	3.1- 7.5	3.2 4.5 5.9	UWB applications

4. Conclusion

The micro-strip patch antenna is easy to manufacture, replaceable, low profile and highly efficient. Such antennas are strongly recommended in satellite and wireless communication. In this paper, a new UWB patch antenna is presented with a wide bandwidth from 3.1 to 7.5 GHz with three resonant frequencies at 3.9, 5.95 and 7.3 GHz. The optimized structural parameters have been reached after many optimization and parameter sweeps on antenna performance. The propounded antenna can be used in multiple UWB applications that requires a wide bandwidth and reduced return loss at the operating frequency of the frequency range. The simulated results provide that, the maximum bandwidth attained numerically is 4.4 GHz due to the multiple slots used in the antenna. The results have revealed that our new fabricated antenna is definitely highly convenient for UWB applications such as high-resolution radar, military communication, communications and sensors, position location and tracking.

References

[1] Federal Communications Commission (FCC) "New Public Safety Applications and Broadband Internet Access among Uses Envisioned by FCC Authorization of Ultra-Wideband Technology", February 2002.

[2] Federal Communications Commission (FCC), "Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems", First Report & Order, Washington DC, February 2002.

[3] Chen, D. and C. H. Cheng, "A novel compact ultra-wideband (UWB) wide slot antenna with via holes," Progress In Electromagnetics Research, Vol. 94, pp. 343–349, 2009. doi:10.2528/PIER09062306

[4] A. Ibrahim, M. A. Abdalla, and A. Boutejdar, "A Printed Compact Band-Notched Antenna Using Octagonal Radiating Patch and Meander Slot Technique for UWB Applications", Progress In Electromagnetics Research M, Vol. 54, pp. 153–162, 2017. doi:10.2528/PIERM16122805

[5] Changzhou Hua, Yunlong Lu, and Taijun Liu, "UWB Heart-Shaped Planar Monopole Antenna with a Reconfigurable Notched Band", Progress In Electromagnetics Research Letters, Vol. 65, pp. 123–130, 2017. doi:10.2528/PIERL16120203

[6] A. Pirooj, M. Naser Moghadasi, F. B. Zarrabi, and A Sharifi, "A Dual Band Slot Antenna for Wireless Applications with Circular Polarization", Progress In Electromagnetics Research C, Vol. 71, pp. 69–77, 2017. doi:10.2528/PIERC16111401

[7] C. A. Balanis. Antenna Theory analysis and design. New York, 2007

[8] Kharakhili, F. G., M. Fardis, G. Dadashzadeh, A. Ahmadi, and N. Hojjat, "Circular slot with a novel circular microstrip open ended microstrip feed for UWB applications," Progress in Electromagnetics Research, Vol 68, pp. 161–167, 2007. doi:10.2528/PIER06071901

[9] Marqués, R.; Mesa, F.; Martel J.; Medina, F. "Comparative analysis of edge and broadside coupled split ring resonators for metamaterial design: Theory and experiment". IEEE Transactions on Antennas and Propagation, Vol. 51, No. 10, pp. 2572-258, October 2003. DOI: 10.1109/TAP.2003.817562

[10] I. Latif, L. Shafai and S. K. Sharma, "Bandwidth Enhancement and Size Reduction of Microstrip Slot Antennas". IEEE Transactions on Antennas and Propagation, Vol. 53, No. 3, pp. 994-1003, March 2005. DOI: 10.1109/TAP.2004.842674

[11] M. Hamza QADDI, Marko SONKKI, Sami MYLLYMAKI, Hassan MHARZI, M. Nabil SRIFI, and Heli JANTUNEN, "Novel compact patch antenna for Ultra-wideband (UWB) applications", International Journal Of Microwave And Optical Technology, Vol. 13, pp. 343-350, July 2018.

[12] M. Särestöniemi, T. Tuovinen, M. Hämäläinen, K. Y. Yazdandoost, E. Kaivanto, and J. Iinatti, ''Applicability of Finite Integration Technique for the modelling of UWB channel characterization,'' in Proc. ISMICT, San Diego, CA, USA, 2012, pp. 1–4. DOI: 10.1109/ISMICT.2012.6203045

[13] M. Z. Mahmud et al "A triangular coupled-resonator antenna for ultra-wide applications" Applied physics A, 2017. https://doi.org/10.1007/s00339-016-0639-x

[14] Rezaul Azim, Mohammad Tariqul Islam, and Norbahiah Misran. "Compact Tapered-Shape Slot Antenna for UWB Applications". IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, Vol. 10, pp.1190-1193, October 2011. DOI: 10.1109/LAWP.2011.2172181

[15] Arun Kumar. "A microstrip UWB antenna for next generation communication system". Int. J. Wireless and Mobile Computing, Vol. 15, No. 3, pp.270–278, 2018. doi.org/10.1504/IJWMC.2018.096010

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 5, No. 2, 13-18 (2020)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Productify News Article Classification Model with Sagemaker

Johannes Lindén*, Xutao Wang, Stefan Forsström, Tingting Zhang

Mid. Sweden University, Department of Information Systems and Technology (IST), 851 70, Sweden

ARTICLE INFO

Article history:
Received: 29 May, 2019
Accepted: 09 July, 2019
Online: 09 March, 2020

Keywords:
Machine learning
Data mining
Big data
News events
Journalists
Editors
Text analysis
Natural language processing
NLP
Paragraph vectors

ABSTRACT

News companies have a need to automate and make the process of writing about popular and new events more effective. Current technologies involve robotic programs that fill in values in templates and website listeners that notify editors when changes are made so that the editor can read up on the source change on the actual website. Editors can provide news faster and better if directly provided with abstracts of the external sources and categorical meta-data that supports what the text is about. To make categorical meta-data a reality an auto-categorization model was created and optimized for Swedish articles written by local news journalists. The problem was that it was not scale-able enough to use out of the box. Instead of having this local model that could make good predictions of the text documents, the model is to be deployed in the cloud and an API interface is created. The API can be accessed from the tools where the articles is being written and therefore these services can automatically assign categories to the articles once the journalist is done writing it. To allow scale-ability to several thousands of simultaneously categorized articles and at the same time improving the workflow of deploying new models easier the API is uploaded to Sagemaker where several models are trained and once an improved model is found that model will be used in production in such a way that the system organically adapts to new written articles. An evaluation of Sagemaker API was done and it was concluded that the complexity of this solution was polynomial.

1 Introuction

Modelling data with machine learning algorithms has shown to give promising results in various of areas, for example image processing and robotics. The areas are growing and machine learning becomes more advanced for every day. Natural language processing is one area that is difficult and requires larger and deeper networks to perform. The larger a network is the more computational power is required to compute a prediction. For a system in production the speed and memory usage could be fatal to the incoming requests of the application. This article addresses the difficulties to take such a system from the model produced by the algorithm to a productified version utilized by a product or users in real time. The response time, memory consumption and computational power are of importance when these types of models will be available with minimal waiting- and downtime. This article addresses one way to scale the application to handle all requests from the user base of a media company distributed over 22 regions in Sweden and reasons behind possibilities to scale up to the entire Sweden user base of local news production line.

In a longer run the focus is also to use the categories metadata point in other products such as improving article personalization algorithms, data analytic of supply and demand of article categories. This article is an extension from Linden et al presented at Fedcsis conference 2018 [1]. The auto-categorization prediction accuracy are improved by parameter optimization since the last article and are now at 80 % and that is the underlying reason why the system is ready for an production environment. The classification model will categorize articles into different news topics and reply with a given confidence for each category. The while the journalists writes their articles in they would like to have categories in or close to real time.

2 Related Work

The novelty of this research, that concerns productification, is the idea to take the auto-categorization neural network

*Johannes Lindén, Holmgatan 10 Sundsall Sweden, 010-142 80 69 & johannes.linden@miun.se

model from its experimental phases all the way to a use case where it brings value. Several research projects has been done for managing and supporting big data pipelines, where live data needs to be processed by several organizations (for example analysis and business intelligence) [2]-[3]. The auto-categorization model perdictions needs a similar automatic data pipeline to uphold the load of each prediction and at the same time allow analysis of the result that will be brought back to the model in a feedback loop. Blaiszik et al made a scalable service that manages machine learning models through a structured API [4]. Blaiszik is using docker containers to wrap the model logic into an environment with controlled variables. This research approach is similar to Blaiszik but instead of using containers for only the models the idea is to have one container for the service as well. Cloud based services such as AWS Sagemaker or IBM's Watson has benefits as well as flaws explained in detail by Hummer et al that builds their own cloud service for training machine learning models [5]. According to Hummer's research it is essential for the AI expert to focus on the model itself and care less about the pipeline and how to set it up properly but Hummer also mentions the importance of having the settings available to tweak the service for special needs. For simplicity this research are using Sagemaker to deploy models in production since there is knowledge about other AWS services available and having all systems in the same cloud has other benefits of latency and authorization flexibility. An evaluation investigation in signal processing was issued that was written by Massimo Ravasi and Marco Mattavelli where they motivated the importance of the system complexity and the partitioning of hardware and software evaluation [6].

As for the auto-categorization model itself Payne et al have made an evaluation of their SVM classification algorithm which yields promising results in performance [7]. The model used in this research utilizes a bigger neural network to comprehend a larger problem domain. Rather than only focusing on a domain concerning digital archives it is focused on local news which could be very specific to each region of the country but could also be very general where regions have common information needs of local news.

3 Approach and Model

The proposed four step model in previous research that predicts categories of arbitrary text paragraphs are going to be the main model in this research that will conduct scalability experiments and propose a system structure of this machine learning model. See Figure 1 for an overview of previous implementation.

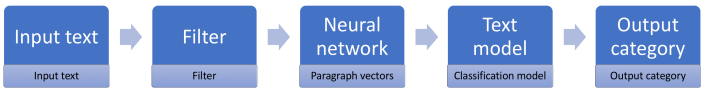


Figure 1: An overview of the model

The input is the algorithm parameters θ and a single document D, which is interpreted as a sequence of words $w_1, w_2, \cdots, w_n$. The output of the model is a set of category probabilities $c_1, c_2, \cdots, c_i, \cdots, c_m$ computed by 1.

$$c_i = P(\text{ith category}|D,\theta) \approx P(\text{ith category}|D) \quad (1)$$

In this research the model are used for example as shown in Figure 2, the article shown to the left in a web-browser will request a category from the trained auto-categorization model and retrieve a list of probable categories.

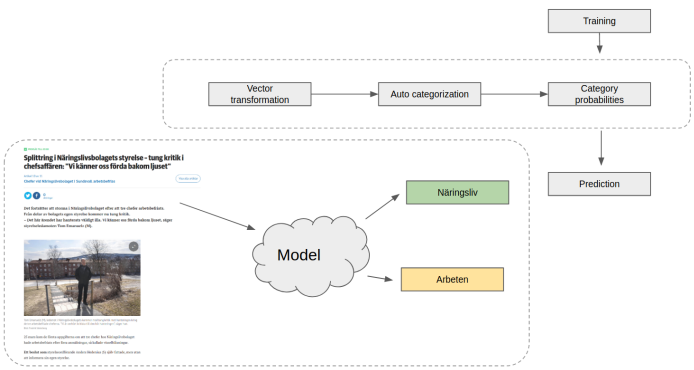


Figure 2: Auto-categorization in a real case scenario. While a user navigates to a article page the category of that particular article is determined based on the body/text within it.

The algorithm takes an unstructured sequence of words forming a text paragraph as input. Language of this model is Swedish and even though the text length could be of any length the articles that has been evaluated has the length between 5 to 600 words in the first step of Figure 1. Text paragraphs are filtered before training to not include articles with only links or empty texts since these articles are not describing what the article is about. Within an text paragraph some additional modifications is done which is step two in Figure 1. Non-alphabetic characters, exclamation and question marks are replaced by full stops. XML tags and its attributes are removed. Some additional optional modifications are left in the settings file to be turned on or off while using the system. Such modifications include Named Entity Recognition (NER) substitution of locations, organizations, names and time units, another modification are Part-of-speech (POS) tagger and Dependency parser which will filter out certain types of tagged words and leaf nodes [8]. POS and Dependency parser chosen in this research where Google SyntaxNet [9, 10]. Out of the resources mentioned in Nilson et al, we selected a treebank made by Jan Einarsson's project, which is well documented [11]-[12]. The other steps in Figure 1 are described in the following sections.

3.1 *Paragraph Vectors*

The third step in Figure 1 is a constructed one layered neural network model that will transform the filtered text paragraphs from previous step into vectors. The paragraph vectors are unique vectors that mapps the paragraph to a high dimensional vector space. A softmax activation function

makes sure that the elements are within the interval -1 to 1. The vectors are constructed in such a way that it is possible to relate a paragraph from the distance between another paragraph. The relation of two paragraphs can be obtained by computing the cosine similarity which will give a positive value when the documents are sharing similar contexts, a value close to zero when no relatedness could be found, and a negative number when the paragraphs appears in in opposite contexts [13]. The vectors can also use common subtraction and additions operations in case other related context will be retrieved, as shown in 2.

$$king - man + woman = queen \tag{2}$$

The paragraph vectors context awareness contains information about what makes a document category as shown in previous research [1]. The paragraph vectors are as before computed using the PV-DM algorithm which is an extension of the known word2vec algorithm bag of words (WV-BOW) [1, 14].

Experiments previously conducted showed that the PV-DBOW paragraph algorithm is reliable when implemented by the distributed memory vector concatenated with the distributed bag of words vector described by Mikolov et al [13].

3.2 Text Model

The paragraph representation vectors is input to the fourth step where the auto-categorization takes place. The categorization algorithms we proposed in previous research experiments where decision trees, random forest, multi layer perception and long-short term memory (LSTM). The one outperformed the others where the LSTM model which are used in this extended paper with the possibility to exchange to another model at any time through a settings parameter in the system. The output is the category belonging probabilities for each considered category described in 1. The LSTM model has a time parameter that each time sequence are used for a vector transformed sentence in the paragraph. Figure 3 shows the LSTM input format, each colored line is a sentence which are fed in per time-slot in the model for training .

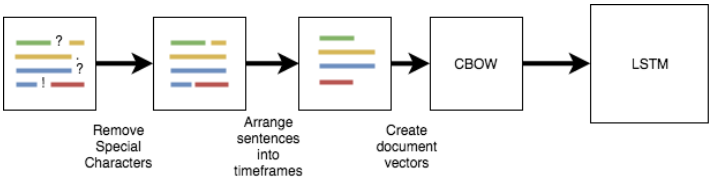


Figure 3: The input vector of the LSTM algorithm is shown above. It first applies the filter layer conditions, then divides the document into sentences to be used as input data for the CBOW algorithm that produces the document vectors.

LSTM models are based on the principles of recurrent neural networks (RNN). A RNN is constructed like a neural network with an input, hidden layers and a output layer. The RNN-cell can be visualised as shown in Figure 4. The activation function of an RNN is usually the *tanh* function. For each iteration, the model is trained by backpropagation through the network. The purpose of RNN is to have a short-term memory that remembers previous neurons. One of the first and simple constructions of RNN is the recurrent neural network language Model (RNN-LM). The hidden layer of an RNN-LM algorithm remembers the neurons one time-step back in the training history [15].

For different use-cases there are different variations of RNNs. Andrej Karpathy summarises the different networks that are used into different mappings [16]. One-to-one mapping is the original algorithm, for example the RNN-LM algorithm. One-to-many mapping when there is one input and several RNNs connecting to several outputs. This mapping can, for example be, used for image prediction with one image and several words that predict the image. Many-to-one mapping is where there are several inputs mapping to one output, this mapping can for example be used for classification. Many-to-many mapping is what Karpathy describes as two different mappings: one mapping that maps to an equal number of input and output RNNs (N-N), and another mapping that maps to a different number of inputs and outputs (N-M). The N-N mapping can, for example, be used to predict video sequences over time, while N-M mapping can be used for translation problems.

LSTM networks are a special case of RNN that tends to solve a problem in the original RNN. The long-term memory in RNN (the gradient descent) exponentially diverges to infinity or converges to zero in many cases and LSTM introduces an additional memory cell and forget parts of the information and therefore avoiding more cases of vanishing gradients. LSTM networks introduce three sigmoid layers and certain gates that only let parts of the information through to compensate for the vanishing gradient. The first sigmoid layer determines what information that is important from the previous LSTM-cell, the second sigmoid layer determines what information is important from the *tanh* layer in the current cell and the third sigmoid layer determines what information will be passed to the next LSTM-cell. The gates that open or close based on the input from the previous LSTM-cell either remove or add information to a cell state that is also passed through to the next LSTM-cell. The third sigmoid layer extracts a piece of information from the cell state to the output value [17, 18].

3.3 Output Category

Output category is the last, fifth, step of the model which determines the output categories. This step interprets the output of the categorization model and limits the categories to suggest via a threshold of the category probability. If the probability is higher than the average probability of seen paragraphs the category are recommended and sent to output of the model. Before the categories that will be replied is returned from the auto-categorization model the probabilities are normalized according to 3.

$$\frac{value - min(value)}{max(value) - min(value)} \tag{3}$$

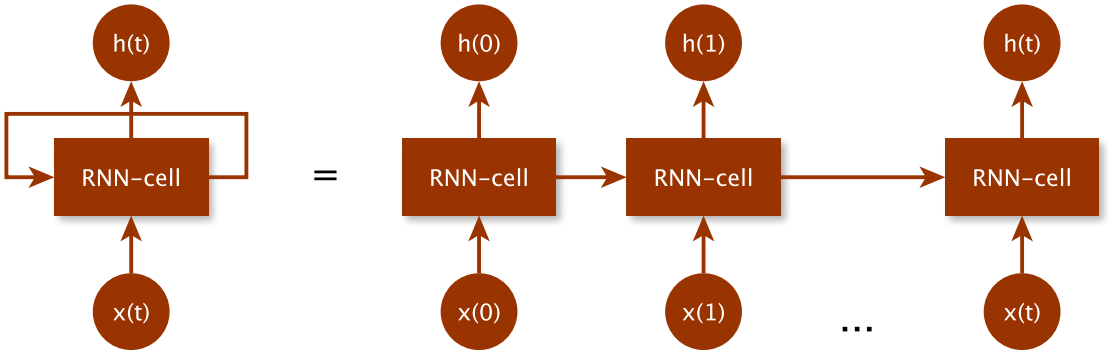


Figure 4: A recurrent neural network cell. The cell to the left is the general notation of the hidden layer in an RNN model. The right unfolded version is the representative of the RNN with a short-term length of t.

4 Architecture

The architectural approach consists of the system, the data pipeline and the neural network model. For making the auto-categorization available to use on texts written by the journalists all three pieces has to be in place. The procedure should be close to real time and this was accomplished by a system that could easily be scaled up to bigger machines with load distributed support for the user prediction requests. Figure 5 shows an overview of the system.

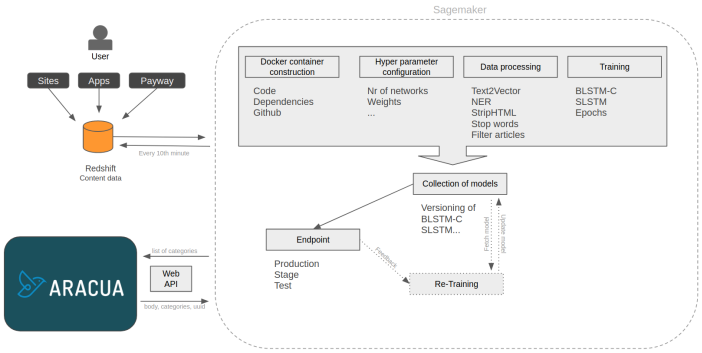


Figure 5: Architectural overview of production environment.

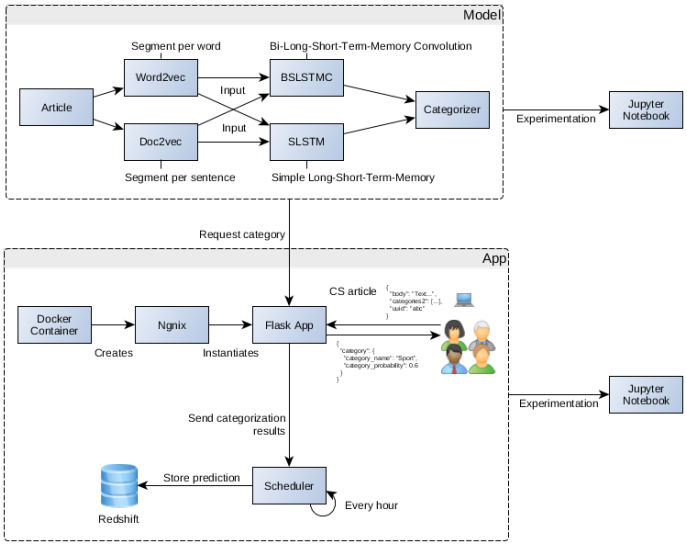


Figure 6: Architecture of the server system.

Aracua is the name of the system the journalists are writing their articles in, this system will request categories given the text of the article. Each categorization request is stored together with the previous category for further live evaluation purposes. The requests are stored in a AWS Redshift cluster to compare the final categories of each article with the suggested ones along the way while the journalists are writing the article [19]. In Redshift other services will have access to the auto-categorized categories to be used for example in personalization purposes. This section will further describe the deployment process of the model.

4.1 System

A Ngnix server was used for the system that can handle requests and reply the the model prediction of categories. The server instantiates python flask applications on demand when the load is high and one instance cannot keep up the clients requests as illustrated in Figure 6.

All runs inside a docker container to make it easy to increase the memory and processing power on an AWS EC2 instance if needed [20]. Each time a client requests categories for an article the user also sends the current categories on the article so that in a later step it is possible to evaluate an article over time as it is being worked on.

The predictions and articles are temporarily stored in memory to speed up the replies for the same article. A scheduled task is triggered every 10 minute or when number of articles in memory are larger than 1000 documents. The scheduler dumps these articles in memory to a persistent database where evaluation will be performed. The model is parameterized so that depending on the settings different models are being used for the prediction see Section 4.3.

4.2 Data pipeline

The system gets article contents that are being written by journalists from Aracua and stores them in a database. The articles are distributed on various of different places depending on the product that will use them. The auto-categorization product retrieves all articles and trains its models on these articles as described in Section 3. Each model that is trained is stored separately so that it is easy to switch models in production. A retraining process is started in production to dynamically improve the model for the articles. There are not only models for each environment (developement, staging and production) but also models for different versions and improved models as the time progresses.

4.3 Neural network model

The parameters that is used to train the model are stored in a configuration file. The configuration file contains what model structures that are going to be used, how long the training will take in terms of number of epochs, layer sizes pre-processing steps to be done, where the data is taken from and model stored to. Currently the following setup shown in Table 1 is used since it was shown to be the best one in previous experiments.

Table 1: Neural network model parameters

Param	Value
Layer size	3
LSTM timestep size	200
Document vector size	1000
Pre-processing steps	NER, StripHTML, Filter stop words, Part-of-speech

4.4 Hardware specifics

The evaluation conducted in this research is performed on an AWS Graviton Processor with two 64-bit Arm cores. The processor is caped at 2.3 GHz clock frequency, the memory is 16 GB and high network speed according to AWS specifications [21]. The computer used corresponds to the specification of ”ml.m4.xlarge” in Amazon Instance specifications[21].

5 Results

Figure 5 and 6 shows the resulting production system and how it works. The main result is that this system is used by thousands of journalists all over Sweden. Additional results of this extended article is to provide evaluation metrics of the system in terms of responsiveness to the journalists. Figure 7 shows how long the response time is in average given the number of clients trying to access the system at the same time.

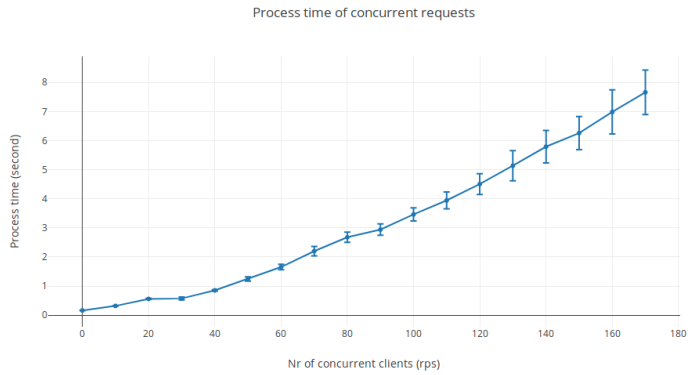


Figure 7: Process time of concurrent requests with standard deviation.

By analysing the number of characters an article body consists of when categorizing it is possible to say how much implications long articles have on the system, the resulting analysis can be seen in Figure 8.

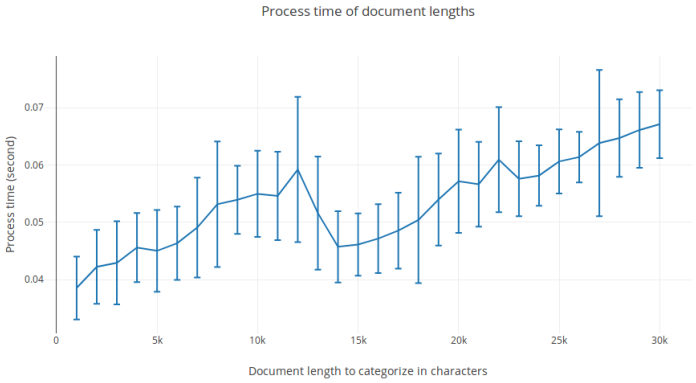


Figure 8: Process time of document lengths with standard deviation.

The training process is the process that initially takes the longest time of all in this research. Training a model with all articles of the categories in question takes up to 2.2 hours using a document vector embedding size of 600 elements as shown in Figure 9.

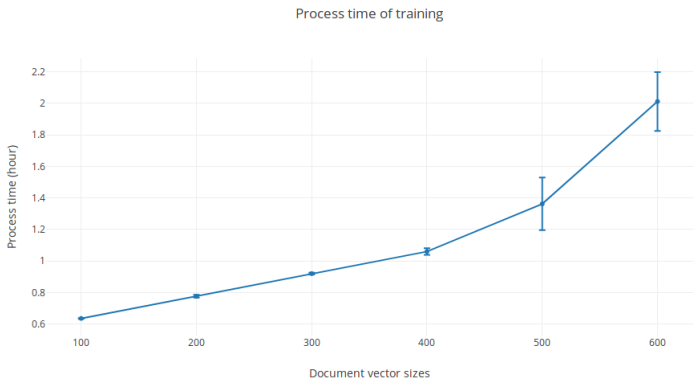


Figure 9: Process time of training with standard deviation.

6 Conclusion

In this article a proposition for a productification of a categorization algorithm that through a web API serves journalists with automatically categorized texts in their writing tool. The main question was if the performance of such model could be serving a large amount of users in their daily work and in that case how many users can it handle. It can be concluded that the number of concurrent requests have somewhat a linear time complexity but could potentially be exponential of higher orders for larger number of clients. Analysing the article length it can be interpreted to have an impact of the process time but would not be a problem for journalists since it is rare that they write long articles. Around 15 000 characters seem to be a good number of characters to have in an article. Training performance is growing with polynomial complexity and could be a risk as more categories are introduced and larger vector embedding needs to be considered. Future work for this project will be to include more categories with more complex algorithms which will use a wisdom of the many approach.

References

[1] Johannes Lindén, Stefan Forsström, and Tingting Zhang. Evaluating combinations of classification algorithms and paragraph vectors for news article classification, 2018.

[2] Ryan Chard, Kyle Chard, Jason Alt, Dilworth Y Parkinson, Steve Tuecke, and Ian Foster. Ripple: Home automation for research data management. In *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 389–394. IEEE, 2017.

[3] Jack Deslippe, Abdelilah Essiari, Simon J Patton, Taghrid Samak, Craig E Tull, Alexander Hexemer, Dinesh Kumar, Dilworth Parkinson, and Polite Stewart. Workflow management for real-time analysis of lightsource experiments. In *2014 9th Workshop on Workflows in Support of Large-Scale Science*, pages 31–40. IEEE, 2014.

[4] Ben Blaiszik, Kyle Chard, Ryan Chard, Ian Foster, and Logan Ward. Data automation at light sources. In *AIP Conference Proceedings*, volume 2054, page 020003. AIP Publishing, 2019.

[5] Waldemar Hummer, Vinod Muthusamy, Thomas Rausch, Parijat Dube, and Kaoutar El Maghraoui. Modelops: Cloud-based lifecycle management for reliable and trusted ai.

[6] Massimo Ravasi and Marco Mattavelli. High-level algorithmic complexity evaluation for system design. *Journal of Systems Architecture*, 48(13-15):403–427, 2003.

[7] Nathaniel Payne and Jason R Baron. Auto-categorization methods for digital archives. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 2288–2298. IEEE, 2017.

[8] Atro Voutilainen. Part-of-speech tagging. *The Oxford handbook of computational linguistics*, pages 219–232, 2003.

[9] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. Globally normalized transition-based neural networks. *arXiv preprint arXiv:1603.06042*, 2016.

[10] Chris Alberti, Daniel Andor, Ivan Bogatyy, Michael Collins, Dan Gillick, Lingpeng Kong, Terry Koo, Ji Ma, Mark Omernick, Slav Petrov, et al. Syntaxnet models for the conll 2017 shared task. *arXiv preprint arXiv:1703.04929*, 2017.

[11] Jens Nilsson and Johan Hall. *Reconstruction of the Swedish Treebank Talbanken.* Matematiska och systemtekniska institutionen, 2005.

[12] Jan Einarsson. Talbankens talsprkskonkordans. 1976.

[13] Quoc V Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. In *ICML*, volume 14, pages 1188–1196, 2014.

[14] Johannes Lindén. Understand and Utilise Unformatted Text Documents by Natural Language Processing algorithm. 46(0), 2017.

[15] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černocky, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5528–5531. IEEE, 2011.

[16] Andrej Karpathy. The unreasonable effectiveness of recurrent neural networks. *Andrej Karpathy blog*, 2015.

[17] Christopher Olah. Understanding lstm networks. *GITHUB blog, posted on August*, 27:2015, 2015.

[18] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. LSTM Neural Networks for Language Modeling. In *Interspeech*, pages 194–197, 2012.

[19] Joseph Baron and Sanjay Kotecha. Storage options in the aws cloud. *Amazon Web Services, Washington DC, Tech. Rep*, 2013.

[20] Gideon Juve, Ewa Deelman, Karan Vahi, Gaurang Mehta, Bruce Berriman, Benjamin P Berman, and Phil Maechling. Scientific workflow applications on amazon ec2. In *2009 5th IEEE international conference on e-science workshops*, pages 59–66. IEEE, 2009.

[21] Amazon Co. Amazon SageMaker Instance Types - Amazon Web Services (AWS). https://aws.amazon.com/sagemaker/pricing/instance-types/.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 16-22 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Melanoma detection using color and texture features in computer vision systems

Antonio Fuduli[1], Pierangelo Veltri[2], Eugenio Vocaturo[3], Ester Zumpano[*,3]

[1]*Department of Mathematics and Computer Science, University of Calabria, 87036, Italy*

[2]*DSCM - Bioinformatics Laboratory Surgical and Medical Science Department, University Magna Graecia, 88100, Italy*

[3]*DIMES - Department of Computer Science, Modelling, Electronic and Systems Engineering, University of Calabria, 87036, Italy*

ARTICLE INFO

Article history:
Received: 14 June, 2019
Accepted: 22 August, 2019
Online: 06 September, 2019

Keywords:
Melanoma detection
Feature selection
Classification

ABSTRACT

All forms of skin cancer are becoming widespread. These forms of cancer, and melanoma in particular, are insidious and aggressive and if not treated promptly can be lethal to humans. Effective treatment of skin lesions depends strongly on the timeliness of the diagnosis: for this reason, artificial vision systems are required to play a crucial role in supporting the diagnosis of skin lesions. This work offers insights into the state of the art in the field of melanoma image classification. We include a numerical section where a preliminary analysis of some classification techniques is performed, using color and texture features on a data set constituted by plain photographies, to which no pre-processing technique has been applied. This is motivated by the necessity to open new horizons in creating self-diagnosis systems for accessible skin lesions, due also to a huge innovation of cameras, smartphones technology and wearable devices.

1 Introduction

The World Health Organization recognizes skin cancer as one of the most lethal cancers that is spreading widely in many parts of the world in recent decades. Globally, each year between 2 and 3 million of non-melanoma skin cancers and 132,000 of melanoma skin cancers occur [1]. Cutaneous melanoma is still incurable. A successful treatment of skin lesion is heavily dependent on the timeliness of the diagnosis: it is crucial to detect and remove the skin lesion in the early stage. One of the major allert of this tumor is that, in the early stages, melanoma remains hidden, showing itself to be similar to other benign lesions. This aspect, combined with the aggressiveness and rapidity with which it expands, creates difficulties even for expert dermatologists; in addition, in its early phase, it is often underestimated [2]. Dermatologists, in clinical evaluation, in order to identify the tumor lesions of the skin, adhere to some reference protocols. Generally, they refer to the ABCDE protocol [3], the 7-point checklist [4] and the Menzies method [5]. If the investigations conducted leave doubts about the nature of the lesion, the specialist prescribes an analysis of the affected tissue by biopsy. Even today, the reliability of the diagnosis of melanoma can be improved with the support of latest-generation automated analysis tools. Although the human factor continues to be the main discriminating element for the diagnostic accuracy [2], on the other hand such tools could play a relevant role to support the specialist in the diagnosis procedure.
Therefore if, on one hand, it is evident the need to design sophisticated tools supporting the specialists in providing early and accurate diagnosis, on the other hand it is crucial also the necessity to develop low cost solutions alerting in case of suspicious skin lesions. Technology is driving a huge innovation in the healthcare domain and artificial intelligence is playing a key role in providing tools and applications supporting physicians in diagnosis and patient treatment. In particular, computer vision systems are revealing to be fundamental in medical image analysis, since they allow to have non-invasive diagnoses. The specific frameworks for the automatic analysis of skin lesions are constituted by the following fundamental steps: image acquisition, image processing and analysis, features extraction and classification.

This paper is an extension of work originally presented at the 9th International Conference on Information, Intelligence, Systems and Applications (IISA 2018). The goal of the original paper [6] was to offer insights into the state of the art in the field of melanoma image classification by both providing a review of the various features used in the elaboration of computer vision systems and focusing on those ones which appear to be the mostly used on the basis of the recent literature. Here a preliminary comparison between two different approaches (Support Vector Machine and Multiple Instance Learning)

[*]Corresponding Author:Ester Zumpano, DIMES - Department of Computer Science, Modelling, Electronic and Systems Engineering, University of Calabria, 87036, Italy & zumpano@dimes.unical.it

is proposed, pointing out the key role played by the feature selection (color and texture). In particular, this study was inspired by the good results obtained in [7], where the authors applied Multiple Instance Learning techniques for classifying some medical dermoscopic images drawn from the Ph^2 database [8], by using only color features. Our numerical experimentations, reported in this paper, showed that using only color features is not at all high-performance when the classification process is performed on a data base constituted by common plain photographies. On the other hand using in addition texture features provides reasonable classification results, especially if we consider that no pre-processing techniques were used.

This paper is organized as follows. In Section 2 we present the pathogenic mechanisms of the skin cancer, while in Section 3 we recall the main steps adopted in biomedical image processing. In Section 4, a preliminary numerical study is performed for classifying a set of 200 plain photographies (100 melanomas and 100 common nevi), using color and color/texture features. Finally some conclusions are drawn in Section 5, looking at possible future developments.

2 Skin lesions

The first layer of the skin is the epidermis, which is constituted by different types of cells, each of them characterized by specific biological chemical functions: *keratinocytes* have a mechanical function of protection from the external environment, *basal cells* have a reproductional function of the cells of the most superficial layer and *melanocytes* provide protection against ultraviolet rays throught the production of melanin.

Dermis is the second layer of the skin, and is constituted by dermal papillary and reticular dermis. It is located below the epidermis and has a thicker consistency (Figure 1). The third layer of the skin is the subcutaneous layer or hypodermis, and, as the dermis, contains many blood and lymphatic vessels and protects the underlying tissues and the internal organs [9].

The different biological structure of the various layers of the skin implies their specific optical properties. When a beam of white light hits the skin, one part is reflected, another part penetrates into the superficial layers and the last part is absorbed. The specific amount of light that is reflected or absorbed depends on genetic factors of the population including the specific phototype to which the individual belongs. Exposure to light triggers melanocytes that produce melanin which is activated as a particularly effective protective filter on shades of blue light.

Normally the melanin deposits are placed in the epidermis (nevi); in the case of malignant lesions, such as the feared melanoma, melanocytes produce excess melanin and these deposits tend to invade the underlying levels of the skin. The presence of melanin in the dermis is one of the most indicative factors in the presence of melanoma (Figure 2).

Further factors indicative of a malignant lesion are the increased blood flow at the periphery of the lesion, evidence of epithelial desquation and the low flow of blood within the lesion which appears with a bluish red hue. Precisely the color features provide clinical evidence for a suspicion of skin cancer. At this point the specialist prescribes a histology to make an explicit diagnosis. Figure 3 shows a common nevus (a), a dysplastic nevus (b) (which is a benign lesion) and finally a melanoma (c) [8].

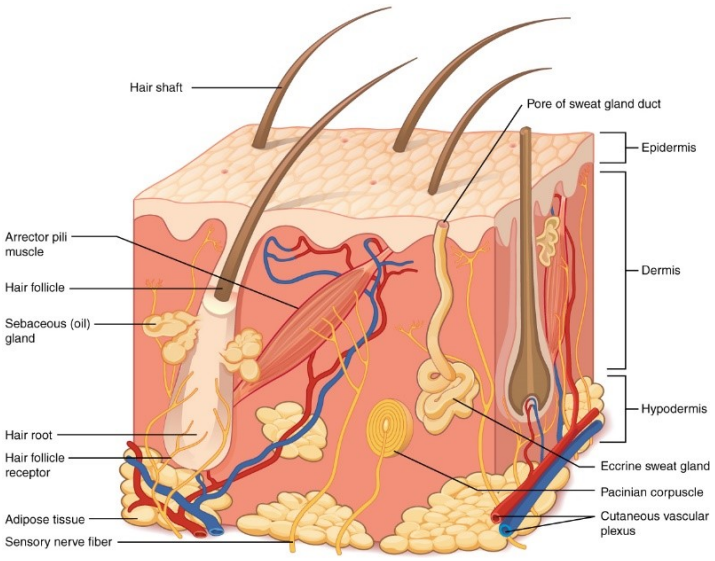


Figure 1: Skin layers

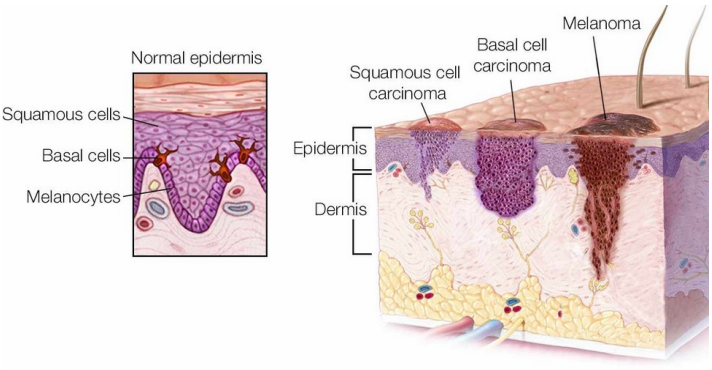


Figure 2: Types of skin cancer

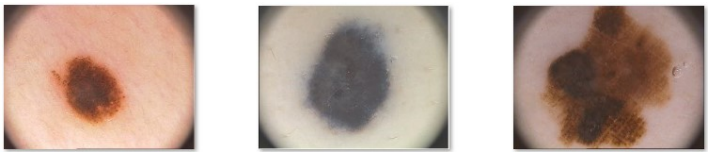


Figure 3: Dermoscopic images of common nevus (a), dysplastic nevus (b) and melanoma (c)

3 Main steps in computer vision systems

In this section we will focus on the fundamental phases that characterize the automatic frameworks for image analysis. We will also refer to the most widespread techniques that characterize the various steps of an artificial vision system: acquisition, pre-processing, segmentation, extraction and selection of features (Figure 4). The classification would be the final step, which is not the main object

of this document, apart from some brief discussion in the numerical section.



Figure 4: Main steps in biomedical image processing

3.1 Image acquisition

The first step of computer vision systems is the acquisition of the digital image. The proliferation of techniques as well as the development of new software paradigms has found its way into the numerous techniques that have been presented by researchers and industrial stakeholders. Based on the corresponding pros and cons, each techniques results more suitable for specific cases and less for other ones. Among the mainly widespread techniques we report total cutaneous photography, dermoscopy, ultrasound, magnetic resonance imaging (MRI), confocal scanning laser microscopy (CSLM), positron emission tomography (PET), optical coherence tomography (OCT), multispectral imaging, computed tomography (CT), multifrequency electrical impedance and Raman spectra. The epiluminescence microscopy (ELM), also known as dermoscopy, allows an in-depth analysis of the surface texture of skin lesions as the technique makes the epidermis translucent. Transmission electron microscopy (TEM) is generally used to monitor the evolution of melanoma. Through an appropriate light source the layers of the skin become translucent and this makes possible to appreciate the vascularization of the lesion also analyzing the colorimetric variations of the pigmentation. Computed tomography (CT) is also used to assess the patient's subjective response, which also allows diagnosis to be made on suspicious lesions.

Positron emission tomography (PET) is a technique that involves the use of contrast fluids through the different absorption of cutaneous melanoma metastases [10]. In the presence of a tumor, the absorption of fluoro-deoxyglucose has a direct correlation with the rate of its proliferation and therefore with the degree of its malignancy. Alternative screening methods are the multifrequency electrical impedance [11] and the Raman spectra [12].

The biological characteristics of a healthy tissue are different from those ones of a diseased tissue. Using the Raman spectra, this different way of diffracting the light coming from a laser beam is exploited, obtaining useful information on the physical structure of the lesion. The spread of wearable devices as well as the enhancement of the smatphones cameras create the conditions for being able to realize applications for auto-screening potentially integrable in platform solutions, through which the patient can interact with specialists aiming at rapid and punctual diagnoses [13]. The increasing costs of health suggest to Governance bodies the adoption of dematerialized health models that can provide decision and responsibility processes, encouraging predictive and preventive analysis [14].

3.2 Image pre-processing

The presence of potential noise linked both to the instrumentation with which the image was acquired, and to the presence of undesired artefacts such as hair and dermatoscopic gel may imply inaccuracies in the diagnosis phase. The objective of the pre-processing phase is to optimize image quality: this phase has a strategic role and is decisive for the entire classification process. The primary task consists in separating the lesion area from the healthy skin. This operation results to be quite difficult as dermoscopic images often present irregular lighting, frames, gelatinous material and intrinsic skin features that make the identification of a lesion, for example trought edges detection methods, a difficult task. Therefore, anything that could affect the image needs to be localized and then removed, masked or replaced. Many techniques have been used in literature in order to perform this task, like image resizing, masking, removal of artefacts, transformation of images into various color spaces. Another important element that commonly penalizes the analysis of dermatoscopic images is the presence of hairs. Recent literature proposes various approaches to optimize images, which involve the use of morphological operators [15], or the detection of curvilinear structures [16] or envisaging a phase of image restructuration through painting [17].

Unfortunately, most of these techniques often leave behind undesirable blurring, invalidating an effective use of the color image. Zhou et al. proposed an algorithm able to carry out an accurate detection of curvilinear artefacts while preserving the lesion structure [18]. The presented approach is very effective for removing the dermatoscopic misuring stick and the hair, but on the other hand it requires high computational requirements. The pre-processing phase of an image can require several operations such as: enhancement, restoration and hair removal (Figure 5).

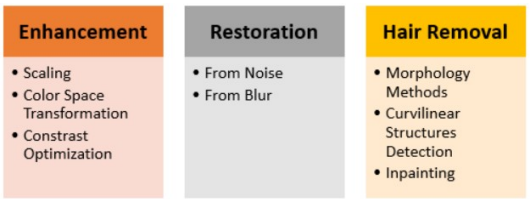


Figure 5: Fundamental steps and techniques for pre-processing

Among the various tools to remove unwanted artefacts, the filters are the most immediate ones. Several Computer Vision Systems dedicated to dermatology adopt filters by applying them directly to grayscale photos or on individual color channels in the case of color images [19, 20, 21].

Generally, for each type of unwanted artefacts, a specific strategy involving the use of dedicated solutions is adopted. Many effective methods use color space transformation to optimize both the quantization of color and of contrast, passing from RGB images to different color spaces for enhancing the variation of the color on the edges and inside the lesion.

The pre-processing phase follows antithetical goals. On the one hand, in order to contain the computational complexity, the objective is to identify a minimal number of colors to be used in image analysis. On the other hand, insufficient contrast would make edge detection very difficult. A solution adopted in many circumstances involves the use of histograms, with which it is possible to perform stretching operations to expand the pixel values in the interval [0, 255], or equalization to obtain a uniform distribution of the pixel values. The important task of appropriately enhancing the contrast between the image area where the lesion is contained and the background consisting of healthy skin can be obtained by compensating for irregular lighting through the use of monomorphic filters and high-pass filters. Adaptive and recursive weighted median filter are used to remove air bubbles and dermoscopic gels [22]. An interesting solution involves local processing carried out on a portion of pixels (generally 3x3), assigning the median to the central pixel. Using these techniques, it is possible to eliminate the noise by containing the alterations that may arise in the original image. The strategies referred to, as well as the other solutions presented in the literature, have as their final objective that of being able to favor the segmentation of the image and the extraction of functional features to the classification process. The strategies to be adopted depend on the nature of the considered images and of the impurities that must be managed [23].

3.3 Image segmentation

The analysis and identification of a lesion can be performed directly by specialists who rely on their experience or by using software solutions for image segmentation, such as IMAGEJ or EDISON (Edge Detection and Image Segmentation system). In fact, in melanoma detection, the main task of the segmentation process is to separate the skin lesion from the background consisting of healthy skin (Figure 6), in order to analyze color, structure, shape, texture and information related to vascularization of the lesion.



Figure 6: (a) dermatoscopic image (b) segmented lesion

The choice of the color space in which it is best to represent images is important, as it determines the rule of the various color channels. Consequently, it will be important to choose the appropriate segmentation algorithm that takes into account the modes of representation chosen for the images. Without the pretension to be exhaustive, some types of segmentation methods in this field are the following:

1. threshold-based, including the Otsu method, those that use the maximum value of the entropy and those that use histograms, such as the independent histogram search algorithm (IHP) [24];

2. region-based, including region growing and watershed segmentation;

3. pixel-based (aimed at labeling the pixel), both supervised methods such as the Support Vector Machine, unsupervised methods such as fuzzy c-means, and those with reinforcement such as artificial neural networks;

4. methods based on customized models, both adaptive and representable through deep multilevel networks.

3.4 Feature extraction

The goal of the segmentation process is to identify the region of the image that contains the lesion by discarding the pixels which are part of the background consisting of healthy skin. In this way the extraction and selection of the most appropriate features to be used for the melanoma detection by artificial vision systems will be favored. In particular, feature extraction implies the enhancement of some specific measures on portions of images: this fundamental step is not often sufficiently discussed and detailed in many researches. On the other hand, a recent study of the relevant features for artificial vision systems in melanoma detection can be found in [6]. As for the extraction of features, many different approaches have been proposed of the statistical and/or filtering type. Among the adopted methods to be mentioned, we cite the principal component analysis, useful for determining the geometric and the shape features, the co-occurrenze matrix [25], useful for the evaluation of texture features, the Fourier spectrum [18], the Gaussian kernels [26], and the transformation of the wavelet package [27], useful for performing image optimization and the extraction of boundaries aimed at reducing data redundancy [28]. In addition, many automated decision support systems implement the feature extraction by exploiting the traditional diagnostic analysis protocols such as the ABCDE system [25], the 7-point checklist [29], the 3-point checklist [30], the model analysis [31] and the Menzies method [32].

4 Numerical experimentations

In this section a preliminary numerical comparison between two different approaches (Support Vector Machine and Multiple Instance Learning) is reported, using color and color/texture features.

This study starts from some considerations related to the works [7, 33, 34]. In [33] the authors analyzed the role played by the color and the texture features, showing empirically that using only the color features outperforms the use of the texture features: very good results were obtained by means of different type of classifiers on a image data set drawn from the Ph^2 database [8], containing 200 melanocytic lesions images (80 common nevi, 80 atypical nevi and 40 melanomas). All the images of the database were selected on the basis of their quality, resolution and dermoscopic features. Good results were obtained also in [7] on the same data set Ph^2,

Table 1: 5-fold cross-validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	52.50	57.50	71.00	60.50	69.50	74.00
Sensitivity (%)	39.28	60.37	59.21	63.52	71.07	70.59
Specificity (%)	67.29	56.09	83.46	57.81	68.67	77.21
F-score (%)	45.74	58.53	65.99	61.11	69.54	72.10
CPU time (secs)	3.00	0.08	2.52	2.67	0.42	2.98

Table 2: 10-fold cross-validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	47.50	56.50	71.00	54.50	68.00	72.00
Sensitivity (%)	37.19	61.55	59.63	59.91	70.98	67.30
Specificity (%)	61.75	54.23	83.97	49.62	65.30	78.16
F-score (%)	41.83	57.93	65.14	56.27	68.22	69.54
CPU time (secs)	2.93	0.02	3.26	3.31	0.01	3.82

Table 3: Leave-One-Out validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	48.50	57.50	71.50	55.00	67.50	69.00
Sensitivity (%)	33.00	64.00	57.00	59.00	69.00	63.00
Specificity (%)	64.00	51.00	86.00	51.00	66.00	75.00
F-score (%)	39.05	60.09	66.67	56.73	67.98	67.02
CPU time (secs)	4.19	0.02	4.44	3.49	0.01	5.50

by applying Multiple Instance Learning techniques and using only color features.

On the other hand, the objective in [34] was to select the most important image features usable in melanoma detection. In this case, differently from [33] where a high quality data set was adopted, the authors performed their experimentations on a data set constituted by plain photographs publicly available from two online databases https://www.dermquest.coml and http://www.dermins.netl, obtaining very good results in terms of sensitivity (96.77%) and, consequently, in terms of F-score (90.91%). The key point of this work consisted in a very accurate pre-processing and segmentation steps.

On the contrary, the objective of our experimentation consists in simply evaluating the classification results, obtainable on 200 images (100 melanomas and 100 common nevi) drawn from the same database used in [34], but without performing any pre-processing step aimed at cleaning up the images from possible noises [23, 35]. This choice is motivated by the necessity to investigate on the possibility to create fast self-diagnosis systems for accessible skin lesions.

In particular, our experiments were performed by using three configurations of the data set, corresponding respectively to the 5-fold cross-validation, the 10-fold cross-validation and the leave-one-out validation.

For each configuration, initially only the color (RGB) features were used, adding, only successively, those ones related to the texture by means of the co-occurence matrix.

The following classification methods, each of them implemented in Matlab (version R2017b), were adopted:

- Support Vector Machine with linear kernel (SVM)[36];
- Support Vector Machine with RBF kernel (SVM-RBF) [36];
- MIL-RL [37].

Some words are to be spent about the MIL-RL method, proposed in [37]. Such a method belongs the class of the Multiple Instance Learning (MIL) algorithms. The MIL paradigm (see for example [38, 39]) is a relatively new approach for classification problems which fits very well on image classification, as shown in the recent works [40, 41] and, in particular, also for medical images and video analysis [42]. It differs from the classical supervised classification approaches, since it is aimed at classifying sets of items: such sets are called *bags* and the items inside the sets are called *instances*. The main characteristic of a MIL algorithm is that, in the learning phase, only the labels of the bags are known, while the labels of the

instances inside the bags are unknown. The MIL paradigm applies very well to image classification since in general an image (bag) is classified on the basis of some portions (instances) of it.

About the MIL-RL code, the same Matlab implementation adopted in [7] were used, while, for the SVM approaches with both linear and RBF kernels, the *fitcsvm* and *predict* subroutines for training the classifier and for computing the testing correctness, respectively, were utilized. Such subroutines are included in the Matlab optimization package.

The results of our experimentations are reported in Tables 1, 2 and 3, in terms of accuracy (testing correctness), sensitivity, specificity, F-score and CPU time. For each of these parameters and for each table, the best result was underlined. It is worth noting that the sensitivity is in general more important than the specificity because it measures the capability to identify positive patients. This is in fact taken into account in the F-score value.

Looking at the tables, it is possible to observe that using both color and texture features improves the results, with respect to using only color features. Moreover the better results are generally obtained in corresponding to the MIL-RL algorithm, whose F-score values are however comparable with those ones obtained by using the SVM with RBF kernel. On the other hand using the SVM-RBF is really faster than using the SVM with linear kernel or the MIL-RL approach.

5 Conclusions

In this work some numerical experimentations were performed to classify a data set constituted by plain photos of different sizes, some of which are low resolution images and others are blurred or with hairs. Although in [34] it has been shown that using some pre-processing methodologies on the images improves the classification performances, in our experiments we preferred do not use such techniques. In this way it was intended to operate in the worst conditions in order to compare the classification performances obtainable by using, on one hand, only color features (as proposed in [40, 41, 43]) and, on the other hand, color and texture features. The fact that the used data set consists of plain photographs, instead of high-quality dermatoscopic images such as those ones constituting the Ph^2 database, justifies the obtained classification performance.

Adopting additional sets of features together with more sophisticated algorithms providing nonlinear separation surfaces, such as in [44], opens up new possible scenarios for the implementation of efficient solutions, useful to provide a further support to specialists and, on the other hand, to have preliminary self diagnoses, since the diffusion of smartphones allows the possibility to take photos of one's own skin injuries. The use of medical solutions [45], as well as the possibility of creating Apps, BOTs and Educational Games [46], designed for different classes of population, are possible scenarios of interest.

References

[1] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 3, pp. 177–193, 2017.

[2] R. Pariser and D. Pariser, "Primary care physicians' errors in handling cutaneous disorders: A prospective survey," *Journal of the American Academy of Dermatology*, vol. 17, no. 2, pp. 239–245, 1987.

[3] R. Sanghera and P. S. Grewal, "Dermatological symptom assessment," in *Patient Assessment in Clinical Pharmacy*, pp. 133–154, Springer, 2019.

[4] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.

[5] S. Menzies, C. Ingvar, and W. McCarthy, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma," *Melanoma Research*, vol. 6, no. 1, pp. 55–62, 1996.

[6] E. Vocaturo, E. Zumpano, and P. Veltri, "Features for melanoma lesions characterization in computer vision systems," in *2018 9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018*, 2019.

[7] A. Astorino, A. Fuduli, P. Veltri, and E. Vocaturo, "Melanoma detection by means of multiple instance learning," *Interdisciplinary Sciences: Computational Life Sciences, DOI: 10.1007/s12539-019-00341-y*, in press, 2019.

[8] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "Ph^2 - A dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437–5440, 2013.

[9] G. Anastasi, S. Capitani, M. Carnazza, S. Cinti, O. Cremona, R. De Caro, R. Donato, V. Ferrario, L. Fonzi, A. Franzi, *et al.*, "Trattato di anatomia umana," 2010.

[10] C. Pleiss, J. Risse, H.-J. Biersack, and H. Bender, "Role of fdg-pet in the assessment of survival prognosis in melanoma," *Cancer Biotherapy and Radiopharmaceuticals*, vol. 22, no. 6, pp. 740–747, 2007.

[11] P. ÅAberg, I. Nicander, J. Hansson, P. Geladi, U. Holmgren, and S. Ollmar, "Skin cancer identification using multifrequency electrical impedance - a potential screening tool," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2097–2102, 2004.

[12] X. Feng, A. Moy, H. Nguyen, Y. Zhang, M. Fox, K. Sebastian, J. Reichenberg, M. Markey, and J. Tunnell, "Raman spectroscopy reveals biophysical markers in skin cancer surgical margins," vol. 10490, 2018.

[13] T. Maier, D. Kulichova, K. Schotten, R. Astrid, T. Ruzicka, C. Berking, and A. Udrea, "Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result," *Journal of the European Academy of Dermatology and Venereology*, vol. 29, no. 4, pp. 663–667, 2015.

[14] E. Vocaturo and P. Veltri, "On the use of networks in biomedicine," *Procedia Computer Science*, vol. 110, pp. 498–503, 2017.

[15] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[16] M. Fleming, C. Steger, J. Zhang, J. Gao, A. Cognetta, L. Pollak, and C. Dyer, "Techniques for a structural analysis of dermatoscopic imagery," *Computerized Medical Imaging and Graphics*, vol. 22, no. 5, pp. 375–389, 1998.

[17] P. Wighton, T. Lee, and M. Atkins, "Dermascopic hair disocclusion using inpainting," vol. 6914, 2008.

[18] H. Zhou, G. Schaefer, A. Sadka, and M. Celebi, "Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 26–34, 2009.

[19] H. W. Lim, H. Honigsmann, and J. L. Hawk, *Photodermatology*. CRC Press, 2007.

[20] I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris, "An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 86–98, 2005.

[21] M. Anantha, R. Moss, and W. Stoecker, "Detection of pigment network in dermatoscopy images using texture analysis," *Computerized Medical Imaging and Graphics*, vol. 28, no. 5, pp. 225–234, 2004.

[22] A. Dehghani Tafti and E. Mirsadeghi, "A novel adaptive recursive median filter in image noise reduction based on using the entropy," pp. 520–523, 2013.

[23] E. Vocaturo, E. Zumpano, and P. Veltri, "Image pre-processing in computer vision systems for melanoma detection," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018*, pp. 2117–2124, 2018.

[24] D. Gómez, C. Butakoff, B. Ersbøll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, 2008.

[25] N. Abbasi, H. Shaw, D. Rigel, R. Friedman, W. McCarthy, I. Osman, A. Kopf, and D. Polsky, "Early diagnosis of cutaneous melanoma: Revisiting the abcd criteria," *Journal of the American Medical Association*, vol. 292, no. 22, pp. 2771–2776, 2004.

[26] G. Surówka and K. Grzesiak-Kopeć, "Different learning paradigms for the classification of melanoid skin lesions using wavelets," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pp. 3136–3139, 2007.

[27] C. Lee and D. Landgrebe, "Decision boundary feature extraction for neural networks," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 75–83, 1997.

[28] I. Maglogiannis, E. Zafiropoulos, and C. Kyranoudis, "Intelligent segmentation and classification of pigmented skin lesions in dermatological images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3955 LNAI, pp. 214–223, 2006.

[29] M. Healsmith, J. Bourke, J. Osborne, and R. Graham-Brown, "An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 130, no. 1, pp. 48–50, 1994.

[30] H. Soyer, G. Argenziano, I. Zalaudek, R. Corona, F. Sera, R. Talamini, F. Barbato, A. Baroni, L. Cicale, A. Di Stefani, P. Farro, L. Rossiello, E. Ruocco, and S. Chimenti, "Three-point checklist of dermoscopy: A new screening method for early detection of melanoma," *Dermatology*, vol. 208, no. 1, pp. 27–31, 2004.

[31] A. Steiner, H. Pehamberger, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. ii. diagnosis of small pigmented skin lesions and early detection of malignant melanoma," *Journal of the American Academy of Dermatology*, vol. 17, no. 4, pp. 584–591, 1987.

[32] S. W. Menzies, K. A. Crotty, C. Ingvar, and W. H. McCarthy, *An atlas of surface microscopy of pigmented skin lesions: dermoscopy*. McGraw-Hill Sydney, Australia, 2003.

[33] C. Barata, M. Ruela, M. Francisco, T. Mendonca, and J. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.

[34] S. Mustafa and A. Kimura, "A svm-based diagnosis of melanoma using only useful image features," in *2018 International Workshop on Advanced Image Technology, IWAIT 2018*, pp. 1–4, 2018.

[35] E. Vocaturo, E. Zumpano, and P. Veltri, "On the usefulness of pre-processing step in melanoma detection using multiple instance learning," in *13th International Conference on Flexible Query Answering Systems (FQAS) 2019, in press*, 2019.

[36] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[37] A. Astorino, A. Fuduli, and M. Gaudioso, "A Lagrangian relaxation approach for binary multiple instance classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 2662–2671, 2019.

[38] J. Amores, "Multiple instance classification: review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 2013.

[39] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, 2018.

[40] A. Astorino, A. Fuduli, M. Gaudioso, and E. Vocaturo, "A multiple instance learning algorithm for color images classification," in *Proceedings of the 22Nd International Database Engineering & Applications Symposium*, IDEAS 2018, (New York, NY, USA), pp. 262–266, ACM, 2018.

[41] A. Astorino, A. Fuduli, M. Gaudioso, and E. Vocaturo, "Multiple instance learning algorithm for medical image classification," in *Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16-19, 2019*, 2019.

[42] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 213–234, 2017.

[43] A. Astorino, A. Fuduli, P. Veltri, and E. Vocaturo, "On a recent algorithm for multiple instance learning. Preliminary applications in image classification," in *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, vol. 2017-January, pp. 1615–1619, 2017.

[44] M. Gaudioso, G. Giallombardo, G. Miglionico, and E. Vocaturo, "Classification in the multiple instance learning framework via spherical separation," *Soft Computing, DOI: 10.1007/s00500-019-04255-1*, in press, 2019.

[45] E. Zumpano, P. Iaquinta, L. Caroprese, G. Cascini, F. Dattola, P. Franco, M. Iusi, P. Veltri, and E. Vocaturo, "Simpatico 3d: A medical information system for diagnostic procedures," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018*, pp. 2125–2128.

[46] E. Vocaturo, E. Zumpano, L. Caroprese, S. M. Pagliuso, and D. Lappano, "Educational games for cultural heritage," in *Proceedings of 1st International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding co-located with 15th Italian Research Conference on Digital Libraries (IRCDL 2019), CNR Area in Pisa, Italy, January 30, 2019*, pp. 95–106, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 21-28 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Emulation of Bio-Inspired Networks

Zdenek Kolka[*,1], Viera Biolkova[1], Dalibor Biolek[1,2], Zdenek Biolek[1,2]

[1]*Brno University of Technology, 616 00 Brno, Czech Republic*

[2]*University of Defence, 662 10 Brno, Czech Republic*

ARTICLE INFO

Article history:
Received:17 May, 2019
Accepted:10 June, 2019
Online: 09 July, 2019

Keywords:
Continuous-Time System
Discrete-Time System
Emulation
Neuron Modeling

ABSTRACT

The paper deals with hardware emulation of bio-inspired devices and nonlinear dynamic processes of complex nature by means of mixed-mode analog-digital emulators. The discretized state model of the emulated system serves for real-time calculation of dependent quantities. In contrast to input-output emulation known in control systems, the proposed approach emulates the ports of an electrical multiport network. The paper discusses the stability of the emulation process and the possibility of partitioning the system into two parts, one being emulated digitally and the other via an analog circuitry. The procedure is illustrated on the example of emulating the Fitzhugh-Nagumo model of neuron and the model of amoeba adaptation. The paper is an extension of our paper presented at the NGCAS 2018 conference in Valletta, Malta. The extended version deals newly with the choice of the integration method and provides a deeper stability analysis and more examples of emulation of biological models.

1. Introduction

Emulation consists in replacing a part of the electrical system with another system that has similar characteristics but is more convenient to implement. This technique has become popular in the field of electronic circuits with the (re)introduction of mem elements [1, 2]. For example, the memristors and other promising nanodevices for digital computational systems, massively parallel analog computations or elegant modeling of the neuron cells, are still in the experimental phase and are not available as off-the-shelf components. Emulation allows performing circuit experiments with equivalents of these novel elements and is also useful for demonstration and educational purposes [3].

The first emulators were proposed as analog circuits. Single-purpose emulators such as [4–6] are simple and elegant, but their disadvantage is the inability to change easily their characteristics. Emulators of memristive, memcapacitive and meminductive devices based on mutators provide large universality because they transform a nonlinear resistor, which can be easily modified, to the respective constitution relation of the mem element [7–8].

Another problem is the emulation of blocks with floating ports, which is difficult for purely analog emulators. Several two-terminal floating emulators were proposed [5, 9, 10]. However, all the emulators are based on grounded devices and are "floating" only in the case of neglecting parasitic parameters. In [11], a genuinely floating memcapacitor was proposed on the principle of switched capacitors. The first implementation of a floating resistive port using a mixed-mode system was proposed in [12] with the use of a digital potentiometer, whose resistance was controlled by a microcontroller by means of pre-programmed algorithms.

On the other hand, there are dynamic systems that cannot be emulated via the above single-purpose emulators. For example, the well-known Hodgkin-Huxley model of the cell membranes in a neuron is represented by a set of nonlinear differential equations [13]. The equivalent electrical model is a two-terminal device, consisting of a linear capacitor, two nonlinear memristive devices, and biasing sources [14]. Usually, these models contain large-value inductors and capacitors, which are not useful for practical laboratory experiments.

The demand for emulating general dynamic systems resulted in developing mixed-mode analog-digital emulators [15, 16]. They consist of a central digital unit, which controls one or more floating analog ports of three possible types: controlled voltage source, controlled current source, and digital potentiometer [17]. The independent port currents and voltages are digitally processed

[*]Corresponding Author: Zdenek Kolka, Brno University of Technology, Email: kolka@feec.vutbr.cz

according to the mathematical model of the emulated device, and the computed dependent quantities are used for controlling the analog ports. A great advantage of such an approach is the ability to change easily the emulated system by means of changing the software and the possibility of designing truly floating ports with the use of digital isolators without compromising the precision.

The paper summarizes our experience of the emulation of two complex nonlinear dynamic systems via the mixed-mode approach [15]. The problem of discretization analog system and its (in)stability, which occurs when connecting a digital emulator to an analog circuitry, is analyzed. The paper is an extension of our paper [1] presented at the NGCAS 2018 conference in Valletta, Malta. The extended version deals newly with the choice of the integration method and provides a deeper stability analysis and more examples of emulation of biological models.

2. Emulation Principle

2.1. Basic Setup

In general, the emulator represents a mixed-mode analog-digital system with m floating electrical ports. For the purpose of emulation, let us constrain ourselves to such systems where each port has an independent input quantity (voltage or current) and the other quantity (current or voltage) is computed as a response.

Thus, the emulator consists of a digital core, implemented in a microcontroller (MCU) or FPGA, and ports in the form of controlled current or voltage sources, Figure 1. Using A/D converters with a supporting circuitry, the core measures independent quantities, computes the response, and sets the dependent quantities using D/A converters, followed by reconstruction filters and corresponding controlled sources.

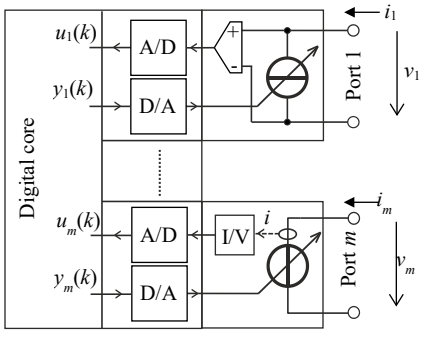


Figure 1: Multiport emulator shown with controlled current source at Port 1 and controlled voltage source at Port m.

The emulated system is generally a nonlinear discrete-time system of the n-th order with m inputs and m outputs, whose mathematical representation can be written as

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)), \quad (1)$$

$$\mathbf{y}(k) = \mathbf{g}(\mathbf{x}(k), \mathbf{u}(k)), \quad (2)$$

where $\mathbf{x}(k) \in \mathrm{R}^n$ is the state vector, $\mathbf{u}(k) \in \mathrm{R}^m$ are the samples of independent (input) quantities, and $\mathbf{y}(k) \in \mathrm{R}^m$ are the dependent (output) quantities. The continuous functions $\mathbf{f}:(\mathrm{R}^n,\mathrm{R}^m) \to \mathrm{R}^n$ and $\mathbf{g}:(\mathrm{R}^n,\mathrm{R}^m) \to \mathrm{R}^m$ are generally nonlinear.

Figure 2 shows the time characteristics of a discrete-time system (1), (2) implemented in a real logical device and connected to a continuous-time circuit. The inputs $\mathbf{u}(k)$ are sampled with a period T and the outputs $\mathbf{y}(k)$ are available after the computation and transfer delay $t_c < T$. The analog circuit the emulator is connected to, which formally includes the anti-aliasing and reconstruction filters, responds to the new outputs $\mathbf{y}(k)$, and the response is sampled as the next input $\mathbf{u}(k+1)$.

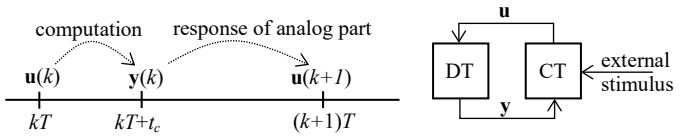


Figure 2: Timing of the emulation process (DT – discrete-time emulator, CT – continuous-time analog circuit).

2.2. Stability of Emulation

As shown in Figure 2, connecting the emulator to an application circuit creates a feedback system, which combines discrete-time and continuous-time parts. Let us consider the one-port case ($m = 1$). The stability of some fixed point of the system (1), (2) can be examined by linearizing the state description

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k), \quad (3)$$

$$y(k) = \mathbf{C}\mathbf{x}(k) + Du(k), \quad (4)$$

where $\mathbf{A} \in \mathrm{R}^{n\times n}$, $\mathbf{B} \in \mathrm{R}^{n\times 1}$, $\mathbf{C} \in \mathrm{R}^{1\times n}$, and $D \in \mathrm{R}$ are real matrices.

Equations (3), (4) represent a single-input single-output linear system, which can be characterized by the z-domain input-output transfer function [18]

$$H(z) = \frac{Y(z)}{U(z)} = \mathbf{C}\frac{\mathrm{adj}[z\mathbf{I} - \mathbf{A}]}{\det[z\mathbf{I} - \mathbf{A}]}\mathbf{B} + D, \quad (5)$$

where "adj" denotes the adjoint matrix, $\mathbf{I}$ is the unity matrix, and $U(z)$ and $Y(z)$ are the z-transforms of the sequences $u(k)$ and $y(k)$.

The D/A converter generates a continuous-time signal $y(k) \to y(t)$, which stimulates the analog circuit, and the circuit response is sampled by the D/A converter as $u(t) \to u(k)$. The continuous-time circuit, which also includes the anti-aliasing and reconstruction filters, can be linearized around an operating point, which corresponds to the fixed point of the discrete system. The input-output small-signal transfer function will be

$$G_C(s) = \frac{U(s)}{Y(s)}, \quad (6)$$

where $U(s)$ and $Y(s)$ are the Laplace transforms of the continuous-time signals $u(t)$ and $y(t)$.

Considering the signal conversion $y(k) \to y(t) \to G_c \to u(t) \to u(k)$, the properties of the analog circuit can also be characterized in the discrete domain by means of the pulse transfer function [18]. For the zero-order hold (ZOH) D/A converter the transfer function will be

$$G(z)=\frac{U(z)}{Y(z)}=Z\left\{\frac{1-e^{-Ts}}{s}G_C(s)e^{-t_c s}\right\}, \qquad (7)$$

where $Z\{\bullet\}$ denotes the z-transform of the sampled impulse response of the continuous-time transfer function, $(1\text{-}e^{-Ts})/s$ is the transfer function of ZOH, and $e^{-t_c s}$ represents the processing delay t_c, which was formally added to the response of the analog circuitry. Note that (7) neglects the quantization introduced by real D/A and A/D converters.

Then the characteristic closed-loop equation will be

$$H(z)G(z)=1. \qquad (8)$$

The roots of (8) determine the stability of the emulation process.

The existence of two types of port (controlled current or voltage source, see Figure 1) is also dictated by the requirement of emulation stability. Let us consider a simple emulation of the resistor R_E by a controlled current source. Then the independent quantity will be the voltage ($u := v$) and the dependent quantity will be the current ($y := i$). The emulator transfer function (5) will be simply

$$H(z)=\frac{Y(z)}{U(z)}=\frac{I(z)}{V(z)}=\frac{1}{R_E}, \qquad (9)$$

i.e. $i(k) = v(k)/R_E$.

Let the emulator port be loaded with a physical resistor R. Considering a first-order reconstruction filter, but no anti-aliasing filter and no processing delay the small-signal transfer function (6) will be

$$G_C(s)=-R\frac{1}{1+s\tau_r}, \qquad (10)$$

where τ_r is the time constant of the reconstruction filter. The positive current i flows into the positive terminal of the emulator port, i.e. (10) has the negative sign.

The pulse transfer function (7) will be

$$G(z)=Z\left\{-\frac{1-e^{-Ts}}{s}\frac{R}{1+s\tau_r}\right\}=R\frac{1-c}{z-c}, \qquad (11)$$

where $c = exp(-T/\tau_r)$.

Substituting (9) and (11) into the characteristic equation (8) leads to

$$-\frac{R}{R_E}\frac{1-c}{z-c}=1. \qquad (12)$$

Considering the stability condition $|z| < 1$ for the root of (12) and with respect to the typical choice $\tau_r \approx T$, which gives $0 < c < 1$, we obtain the condition

$$R < R_E\frac{1+c}{1-c} \qquad (13)$$

for stable operation of the emulating process. Thus, if the Thévenin-equivalent resistance of the analog circuitry connected to the emulator port is higher than the limit (13), the system will oscillate, although the emulated device is a positive-value resistor. A more detailed analysis for other types of devices can be found in [19].

The example underlines the necessity for an emulation-stability analysis. A circuit that would be perfectly stable if realized from physical components may become unstable if some parts are emulated by a discrete-time system.

3. Demonstrations

3.1. Emulator Hardware

The emulator used for demonstrations is a two-port modification [17] of our emulator [15]. The system consists of the main board STM NUCLEO-F722ZE with a 32-bit MCU ARM Cortex-M7 with a single-precision floating-point unit connected to PC via USB. An add-on card implements two isolated ports, each with 16-bit A/D and D/A converters. The ports are fully floating, which is achieved by the use of integrated DC-DC converters and digital SPI isolators for converter control. The total parasitic capacitance of the floating part to the ground is about 11 pF.

A pluggable controlled voltage or current source can be connected to the ports as shown in Figure 3. The operating area of the four-quadrant voltage and current modules is ±3 V and ±3 mA. The emulator can easily achieve a sampling rate of 100 kHz. Depending on the software, it can emulate one or two independent floating devices or a general two-port network.

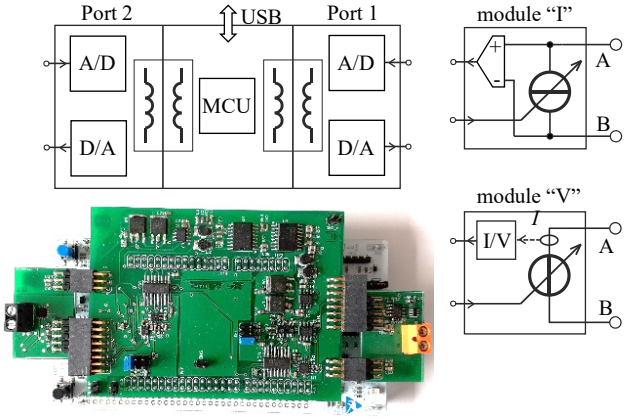


Figure 3: Emulator with pluggable IO modules.

3.2. FitzHugh-Nagumo Model

The FitzHugh-Nagumo (FHN) model of neuronal excitability [20] is a simplification of the well-known Hodgkin-Huxley model [13]. The model represents the prototype of an excitable system. When the input quantity exceeds some certain threshold, the system will generate a pulse. The two-dimensional FHN model is given in its normalized form as

$$\frac{d\upsilon}{d\tau} = \upsilon - \frac{\upsilon^3}{3} - w + j\,, \tag{14}$$

$$\frac{dw}{d\tau} = \frac{1}{\sigma}\left(\upsilon + a - bw\right), \tag{15}$$

where υ represents the normalized membrane potential, w is an auxiliary variable, j is a stimulus current, and τ is the model time. The model dynamics is determined by a set of three parameters with the usual values $a = 0.7$, $b = 0.8$, and $\sigma = 12.5$ [21].

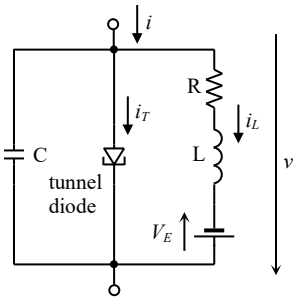


Figure 4: Equivalent circuit to (14) and (15) proposed by Nagumo [20].

The system (14), (15) can be represented by an equivalent electrical circuit shown in Figure 4 with a nonlinearity similar to that of the tunnel diode, which was proposed by Nagumo [20]. Let us consider the following mapping of FHN model quantities to the physical circuit in Figure 4:

$$v = \kappa_v\,\upsilon\,,\ i = \kappa_i\,j\,,\ i_L = \kappa_w\,w\,,\ t = \kappa_t\,\tau\,. \tag{16}$$

The transformation coefficients were chosen such that the variable υ will represent voltage in volts ($\kappa_v = 1$ V), j and w will be currents in milliamps ($\kappa_i = \kappa_w = 1$ mA), and one unit of model time τ will correspond to one millisecond of the physical time ($\kappa_t = 1$ ms), i.e. the dynamics will be speeded up a thousand times to get a convenient duration of generated pulses.

Considering (16), the parameters of the circuit elements will be

$$C = \frac{\kappa_t\,\kappa_i}{\kappa_v},\ R = \frac{b\,\kappa_v}{\kappa_w},\ L = \frac{\kappa_t\,\kappa_v\,\sigma}{\kappa_w},\ V_E = a\,\kappa_v\,. \tag{17}$$

For the chosen transformation coefficients we have $C = 1$ µF, $L = 12.5$ H, $R = 800$ Ω, and $V_E = 0.7$ V. The i-v characteristic $i_T(v)$ of the "tunnel diode" will be

$$i_T(v) = \kappa_i\left(\frac{(v/\kappa_v)^3}{3} - v/\kappa_v\right), \tag{18}$$

which corresponds to the first two terms of RHS of (9).

1) Direct Digital Emulation

Equations (14), (15) represent a dynamical system with the input current j and output potential υ. In accordance with the electrical model in Figure 4 the system can be emulated by a controlled voltage source. The output voltage $\upsilon(k)$ will be computed from the samples $j(k)$ of the measured current.

The real-time operation of the emulator requires the use of explicit integration methods with low numerical complexity [22]. Let us consider an ordinary differential equation of the n-th order in the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, u)\,, \tag{19}$$

where $\mathbf{x} \in \mathrm{R}^n$ and u is a scalar input. Then the explicit multistep linear integration scheme can be formulated as

$$\mathbf{x}(k+1) = \mathbf{x}(k) + h\sum_{i=0}^{r-1}\gamma_i\,\mathbf{f}(\mathbf{x}(k-i), u(k-1))\,, \tag{20}$$

where r is the order of the method, h is the integration step, and γ_i are coefficients. For $r = 1$ we get the classical forward Euler method with $\gamma_1 = 1$, and for $r = 2$ the Adams-Bashforth method of the 2nd order (AB2) with $\gamma = \{3/2, -1/2\}$ [22].

The practical implementation of the multistep scheme (20) has a relative low numerical complexity as it requires just storing r-1 past values $\mathbf{f}(k\text{-}i) := \mathbf{f}(\mathbf{x}(k\text{-}i), u(k\text{-}i))$.

The computation of each step starts with the evaluation of RHS of (14), (15) for current values of the quantities

$$\mathbf{f}(k) = \begin{array}{|c|}\hline \upsilon(k) - \upsilon(k)^3/3 - w(k) + j(k) \\ \hline \left(\upsilon(k) + a - b\,w(k)\right)/\sigma \\ \hline \end{array}\,. \tag{21}$$

In the case of the Euler method the next step is

$$\begin{array}{|c|}\hline \upsilon(k+1) \\ \hline w(k+1) \\ \hline \end{array} = \begin{array}{|c|}\hline \upsilon(k) \\ \hline w(k) \\ \hline \end{array} + h\,\mathbf{f}(k) \tag{22}$$

and for AB2 we have

$$\begin{array}{|c|}\hline \upsilon(k+1) \\ \hline w(k+1) \\ \hline \end{array} = \begin{array}{|c|}\hline \upsilon(k) \\ \hline w(k) \\ \hline \end{array} + \frac{h}{2}\left[3\mathbf{f}(k) - \mathbf{f}(k-1)\right], \tag{23}$$

where $h = T/\kappa_t$. The coefficient $\kappa_t = 1$ ms scales the time axis so that the time unit in the original system (14), (15) corresponds to one millisecond in (22) and (23).

The emulator was configured as a current-controlled voltage source using the module "V", where the output quantity is the transformed state variable $v(k) = \upsilon(k)\ \kappa_v$ and the control quantity is $j(k) = i(k)/\kappa_i$. The sampling rate was set to 100 kHz.

Figure 5 shows the results of an experiment where the digital FHN model was "excited" by short pulses with an amplitude of 3 V, a width of 8 ms, and a period of 100 ms applied through a 10 kΩ resistor R_d. It can be seen that each pulse triggers a characteristic excursion, after which the system relaxes back to the equilibrium state.

The discretization and emulation of the continuous-time system bring two problems with the stability: the stability of the

integration method itself and the stability of the feedback emulation process as introduced in Section 2.2.

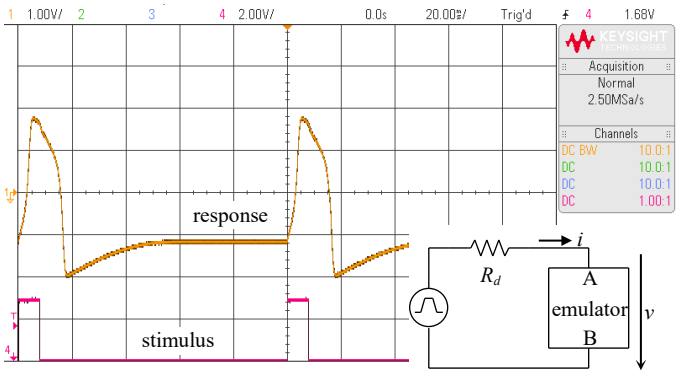


Figure 5: Response (channel 1 - orange) to excitation pulses (channel 4 - red).

In particular, it is known for explicit methods that an inappropriate choice of the integration step can lead to instability and an excessive truncation error [22]. Figure 6 shows a qualitative study comparing the performance of the methods (22) and (23) for the same setup as in Figure 5. For the 10 μs sampling period used, the results are indistinguishable from the nominal solution within the uncertainty introduced by the quantization by A/D and D/A converters. When the sampling period was increased to 500 μs, the waveform computed by AB2 showed numerical oscillations. Therefore, the first-order method was preferred in the experiments.

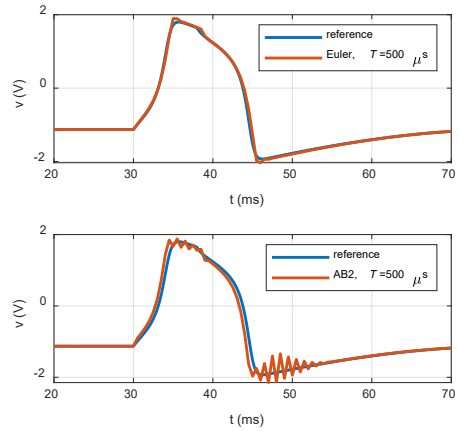


Figure 6: Comparison of forward Euler and 2nd–order Adams-Bashforth methods for large integration steps.

The stability of emulation as introduced in Section 2.2 can be qualitatively assessed based on the equilibrium stability analysis for $v = 0$, which corresponds to the OFF period of the signal source. The equilibrium of the FHN model can be obtained via solving the system (14), (15) for $d\upsilon/d\tau = 0$ and $dw/d\tau = 0$, and also by considering the relation $v = -R_d\, i$ determined by the driving circuit in Figure 5 during the OFF periods. After the transformation (16) the relation becomes $\upsilon\, \kappa_v = -R_d\, j\, \kappa_i$ and we obtain the system

$$0 = \tilde{\upsilon} - \frac{\tilde{\upsilon}^3}{3} - \tilde{w} - \frac{\tilde{\upsilon}\,\kappa_v}{R_d\,\kappa_i}, \tag{24}$$

$$0 = \tilde{\upsilon} + a - b\,\tilde{w}, \tag{25}$$

whose solution is

$$\tilde{\upsilon} = -1.129,\ \tilde{w} = -0.536\,. \tag{26}$$

Let us start with the forward Euler method. The equilibrium corresponds to a fixed point of the discretized system (22), because $\mathbf{f}_k = \mathbf{0}$ and thus $\upsilon(k+1) = \upsilon(k)$, $w(k+1) = w(k)$. The structure of the system (22) corresponds to (1), (2) and its linearization at the fixed point ($\tilde{\upsilon}$, $\tilde{w}$) leads to the following matrices (3) and (4):

$$\mathbf{A} = \begin{bmatrix} 1+h(1-\tilde{\upsilon}^2) & -h \\ h/\sigma & 1-hb/\sigma \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} h \\ 0 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix},\ D = 0\,. \tag{27}$$

Using (5), the emulator transfer function is

$$H(z) = \frac{V(z)}{I(z)} = \frac{\kappa_v}{\kappa_i}\,\frac{zh\sigma - h(\sigma - bh)}{z^2\sigma + z[h\sigma(\tilde{\upsilon}^2 - 1) - 2\sigma + bh] + (bh - \sigma)[h(\tilde{\upsilon}^2 - 1) - 1] + h^2}\,. \tag{28}$$

The simulator uses a 1st-order reconstruction filter with the time constant τ_r and no anti-aliasing filter. The Laplace-domain transfer function between the output of the D/A converter and the input of the A/D converter corresponding to the application schematics in Figure 5 will be

$$G_C(s) = \frac{I(s)}{V(s)} = -\frac{1}{R_d}\,\frac{1}{1 + s\,\tau_r}, \tag{29}$$

where the negative sign stems from the chosen orientation of v and i.

Using the standard s- and z-transforms, the pulse transfer function can be easily obtained as [23]

$$G(z) = Z\left\{\frac{1 - e^{-Ts}}{s}\,G_C(s)\,e^{-t_c s}\right\} = -\frac{1}{R_d}\,\frac{(1 - cd)z^{-1} + c(d-1)z^{-2}}{1 - c\,z^{-1}}, \tag{30}$$

where $c = exp(-T/\tau_r)$ and $d = exp(t_c/\tau_r)$.

The time constant of the reconstruction filter was $\tau_r = T = 10$ μs, which smoothed reasonably its staircase output, and the processing delay was $t_c = 3$ μs. Then the roots of the characteristic equation (8) are as follows:

$$\lambda_{1,2} = 0.9978 \pm j\ 0.002289,\ \lambda_3 = 0.3692,\ \lambda_4 = -3.4988 \times 10^{-4},$$

which indicates a stable operation. It can be shown that the doublet $\lambda_{1,2}$ corresponds to the poles of (28) because the FHN model is driven by a high-resistance source, which behaves like a stimulation by a current source (see Figure 5). The doublet is close to the unity circle, reflecting the fact that the sampling rate is orders of magnitudes faster than the modeled dynamics, i.e. the update in each step is relatively low ($v(k+1) \approx v(k)$). The roots λ_3 and λ_4 are predominantly influenced by the chosen time constant of the

reconstruction filter and MCU processing delay, and represent fast-decaying artifacts of the emulation process.

To compute the z-domain transfer function (5) for higher-order integration methods, the difference equation (20) should be transformed to an equation of the 1st order.

In the case of the Adams-Bashforth method of the 2nd order the scheme (23) can be transformed to

$$\begin{bmatrix} v(k+1) \\ w(k+1) \end{bmatrix} = \begin{bmatrix} v(k) \\ w(k) \end{bmatrix} + \frac{h}{2}\left(3\mathbf{f}(k) - \tilde{\mathbf{f}}(k)\right), \tag{31}$$

$$\tilde{\mathbf{f}}(k+1) = \mathbf{f}(k), \tag{32}$$

where $\tilde{\mathbf{f}}(k)$ is the vector of RHS of (14) and (15) delayed by one period. Then the linearization of (31) and (32) leads to

$$\mathbf{A} = \begin{bmatrix} 1+\frac{3h}{2}(1-v^2) & -\frac{3h}{2} & -\frac{h}{2} & 0 \\ \frac{3h}{2\sigma} & 1-\frac{3hb}{2\sigma} & 0 & -\frac{h}{2} \\ 1-v^2 & -1 & 0 & 0 \\ 1/\sigma & -b/\sigma & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3h/2 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, \; D = 0\,. \tag{33}$$

Repeating the same procedure as for the Euler method, we obtain the roots:

$\lambda_{1,2} = 0.9978 \pm j\,0.003594$,

$\lambda_3 = 0.3681$, $\lambda_4 = -0.0136$, $\lambda_5 = 0.0130$, $\lambda_4 = -3.0454 \times 10^{-4}$.

The doublet characterizes the emulated system dynamics and the other roots represent the artifacts of the discretization. Also in this case, the emulation is stable.

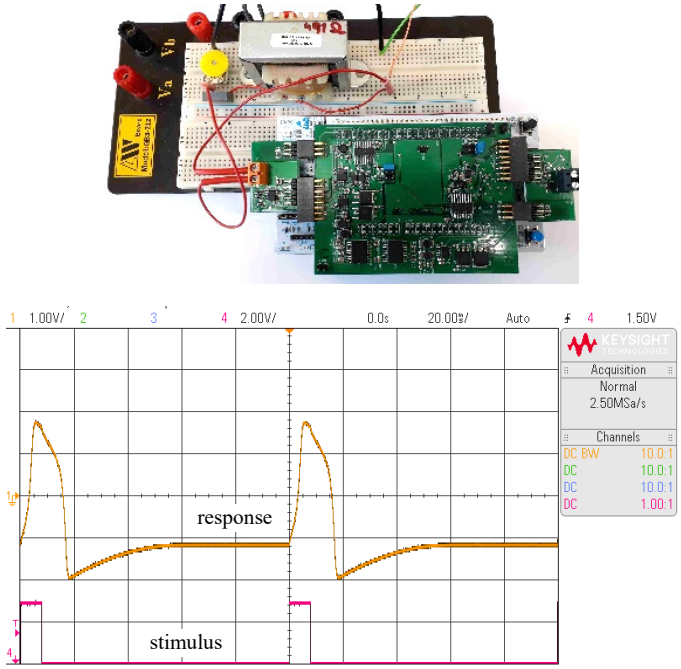


Figure 7: Hybrid circuit setup and response (orange) to excitation pulses (red).

2) Hybrid Circuit

In the hybrid approach the emulator is used to implement some parts of the model, while the rest of the model can be realized using standard electronic components as in Figure 4. In the case of the FHN equivalent circuit the emulator implements just the "tunnel diode". The cubic nonlinearity (18) was emulated as a voltage-controlled current source using the module "I".

Figure 7 shows the experiment setup on a breadboard and the response of the model for the same conditions as in Figure 5. The parameter L was changed to 16.2 H in this experiment to match the inductance of the available off-the-shelf choke.

Figure 8 compares the measured pulses with a PSpice simulation, the latter being regarded as a reference solution. All three waveforms overlap.

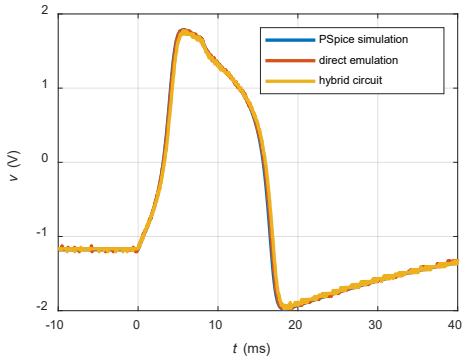


Figure 8: Comparison of the measured and simulated responses for L = 16.2 H (all waveforms overlap).

3.3. Model of Amoeba Adaptation

The amoeba can adapt to periodic environmental changes. An electrical model of the process was presented in [24], see Figure 9. The environmental conditions (temperature and humidity) are represented by a single quantity – the voltage $v(t)$, which is applied to the circuit. The response is the change in amoeba movement velocity represented by the voltage $v_c(t)$.

The memristor in Figure 9 is a voltage-controlled memristor with a threshold whose memristance R_M is governed by the following state equation [24]

$$\frac{dR_M}{dt} = f_A(v_M)\, f_W(R_M, v_M)\,, \tag{34}$$

where v_M is the voltage across memristor ($v_M = v_C$ here). The memristor activation function

$$f_A(v_M) = \frac{\beta - \alpha}{2}\left(|v_M + V_T| - |v_M - V_T|\right) - \beta v_M \tag{35}$$

is shown in Figure 9(b). The parameter α is the slope of the inner segment and β is the slope of outer segments.

The window function

$$f_W(R_M, v_M) = \theta(v_M)\theta(R_M - R_1) + \theta(-v_M)\theta(R_2 - R_M) \tag{36}$$

confines the memristance between the boundary values R_1 and R_2, and θ is the Heaviside step function.

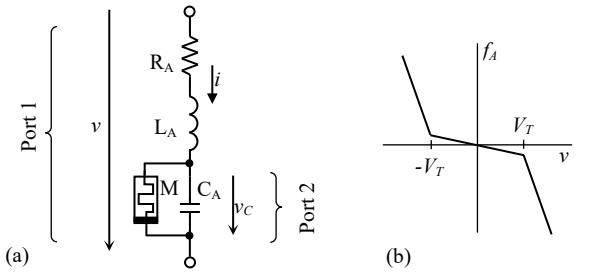


Figure 9: (a) Equivalent electrical model of amoeba's adaptation; (b) Activation function of memristor.

The model leads to a system of the 3rd order [25]

$$\frac{di}{dt} = \frac{1}{\kappa_t L_A}\left(-R_A i + v - v_c\right), \tag{37}$$

$$\frac{dv_c}{dt} = \frac{1}{\kappa_t C_A}\left(-\frac{v_c}{R_M} + i\right), \tag{38}$$

$$\frac{dR_M}{dt} = \frac{1}{\kappa_t} f_A(v_c) f_W(R_M, v_c), \tag{39}$$

where the parameters were identified as [24]: $R_A = 0.1\ \Omega$, $L_A = 2$ H, $C_A = 1$ F, $R_1 = 3\ \Omega$, $R_2 = 20\ \Omega$, $\alpha = 0.1\ \Omega$/Vs, $\beta = 100\ \Omega$/Vs, $V_T = 2.5$ V. The amoeba adaptation occurs on the time scale of ≈100 s, which is rather slow for laboratory experiments even with a digital memory oscilloscope. Therefore the time-axis scaling $\kappa_t = 10^{-3}$ was added to the equations. Now, one second of the model time corresponds to one millisecond of the real time.

The emulator was configured as a two-port network. The excitation voltage $v(t)$ is applied to Port 1 with the "I" module (controlled current source). Although the current i(t) is just an internal variable of the system (37)-(39), it can be generated at Port 1 to emulate fully the circuit from Figure 9(a). One ampere in the model corresponds to one milliampere in the emulated circuit. The model response based on $v_c(t)$ is generated at Port 2 with the "V" module. Alternatively, the output can be set as the memristance R_M to monitor the internal state of the model.

Equations (37)-(39) were discretized using the forward Euler method, similar to (22). The window function f_W was realized in the algorithm as a correction of the state variable R_M after each integration step. Whenever $R_M > R_2$, it is set back to R_2 and vice versa for the lower limit R_1.

The excitation voltage $v(t)$ represents environmental conditions of the amoeba. Favorable conditions correspond to a positive voltage and unfavorable conditions to a negative voltage. Long-term exposure to favorable conditions ($v > 0$) leads to $dR_M/dt < 0$ and after a sufficiently long interval the memristance R_M will be at its lower limit, i.e. $R_M = R_1$.

It has been observed that periodic intervals of unfavorable conditions make the amoeba adapt so that the organism can anticipate the intervals and decrease its velocity. The process can be demonstrated on an experiment from [24]. The "training" voltage waveform consists of three negative cosine pulses

$$v(t) = V_F + V_P \sum_{i=1}^{3} \left\{ \cos\left[\frac{2\pi(t - t_i)}{W_p}\right] - 1 \right\} \frac{w(t)}{2}, \tag{40}$$

where $w(t) = \theta(t - t_i)\theta(t_i + W_p - t)$ is a window function for masking the individual periods of the cosine function. The parameters used were $V_F = 0.1$ V, $V_p = 2$ V, and $W_p = 5$ ms. With respect to the time-axis scaling, the times of pulse starts for the irregular sequence were $t_i^{(irregular)} = \{10\text{ ms}, 19\text{ ms}, 34.5\text{ ms}\}$ and for the regular sequence $t_i^{(regular)} = \{10\text{ ms}, 19\text{ ms}, 28\text{ ms}\}$. The excitation signal was generated using the arbitrary waveform generator Keysight 33521B.

The output voltage on Port 2 was generated as [24]

$$v_{out}(t) = \min(v_C(t), V_F). \tag{41}$$

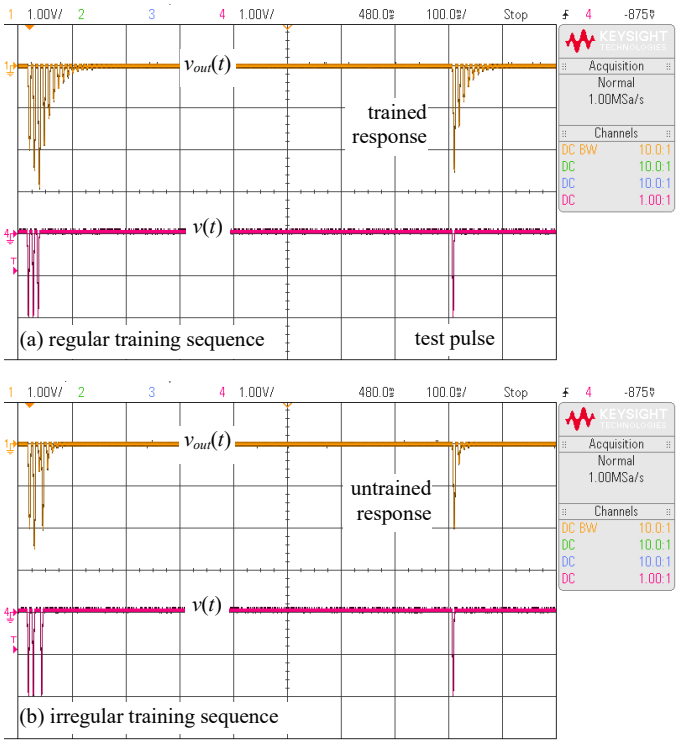


Figure 10: Excitation voltage (channel 4 – red) and amoeba response (channel 1 – orange) for (a) regular and (b) irregular training sequence.

Figure 10 shows the excitation voltage $v(t)$ and the response $v_{out}(t)$ generated by the emulator. The amoeba adaptation is represented in the model by the change of memristance R_M. The application of the irregular sequence of training pulses after a long period of favorable conditions (V_F) does not change the memristance significantly. On the other hand, the regular training sequence makes R_M increase to R_2 and the amoeba becomes "adapted". Both states can be tested by a single pulse at 800 ms after the training sequence start. The adapted response consists of several pulses of motion slowdown where the organism anticipates other unfavorable intervals. The response of the untrained organism is visibly smaller. Both results are in agreement with computer simulations in [24].

4. Conclusions

The results presented in the paper can be summarized as follows:

(1) Emulation provides the possibility of substituting an *m*-port network with a mixed-mode analog-digital system. The network modeled via multi-dimensional differential equations, which link terminal voltages and currents and internal state variables, can be emulated via digitally controlled current or voltage sources. The sources are controlled based on the port voltages or currents and the state variables via discrete-time equations of motion of the system being emulated.

(2) In some cases, it can be useful to emulate digitally only some key parts of the circuit, while the rest can be made up of (preferably passive) analog components.

(3) The combination of continuous-time and discrete-time blocks can bring stability problems, even in very simple circuits. The paper presents a methodology that can be used to assess the emulated system stability.

(4) Stability problems can be alleviated by an appropriate selection of the type of controlled source (voltage or current) which is used to realize each port. Details are given in our previous work [17].

(5) The paper presents a mixed-mode emulator with pluggable IO ports that has proven itself to emulate unconventional circuit elements (memristors, memcapacitors, meminductors, etc.) and also bio-inspired circuits. Especially in the case of models of biological systems it can avoid using bulky inductors and capacitors, which are common there.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by the Czech Science Foundation under grant No. 18-21608S. The research was also supported by the Project of Specific Research, K217 Department, UD Brno.

References

[1] Z. Kolka, V. Biolkova, D. Biolek, Z. Biolek, "Hardware Implementation of Bio-Inspired Models" in 2018 New Generation of CAS (NGCAS), Valletta, Malta, 102-105, 2018. https://doi.org/10.1109/NGCAS.2018.8572072

[2] D. Biolek, Memristor emulators. In Memristive networks. Springer book, New York, 2014.

[3] Y. V. Pershin, M. Di Ventra, "Teaching memory circuit elements via experiment-based learning" IEEE Circuits and Systems Magazine, **12**(1), 64—74, 2012. https://doi.org/10.1109/MCAS.2011.2181096

[4] Q. Zhaoa, C. Wangb, X. Zhangc, "A universal emulator for memristor, memcapacitor, and meminductor and its chaotic circuit" Chaos **29**, 013141, 2019. https://doi.org/10.1063/1.5081076

[5] D.S.Yu, H.Chen, H.H.C.Iu, "Design of a practical memcapacitor emulator without grounded restriction" IEEE Trans Circuits Syst II, **60**(6), 207–211, 2013. https://doi.org/10.1109/TCSII.2013.2240879

[6] M.Pd.Sah, R.K.Budhathoki, C.Yang, H.Kim, "Charge Controlled meminductor Emulator" J. of Semiconductor Technology and Science, **14**(6), 750–754, 2014. https://dx.doi.org/10.5573/JSTS.2014.14.6.750

[7] D. Biolek, V. Biolková, "Mutator for transforming memristor into memcapacitor" Electronics Letters, **46**(21), 1428—1429, 2010. https://doi.org/10.1049/el.2010.2309

[8] D.Yu, Y.Liang, H.H.C.Iu, L.O.Chua, "A universal mutator for transformations among memristor, memcapacitor, and meminductor" IEEE Trans. on Circ Syst Express Briefs, **61**(10), 758–762, 2014. https://doi.org/10.1109/TCSII.2014.2345305

[9] Y. V. Pershin, M. Di Ventra, "Emulation of floating memcapacitors and meminductors using current conveyors" Electronics Letters, **47**(4), 243—244, 2011. https://doi.org/10.1049/el.2010.7328

[10] M. Kumngern, "A floating memristor emulator circuit using operational transconductance amplifiers" in 2015 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC), Singapore, 679–682, 2015. https://doi.org/10.1109/EDSSC.2015.7285207

[11] D. Biolek, V. Biolkova, Z. Kolka, J. Dobes, "Analog Emulator of Genuinely Floating Memcapacitor with Piecewise-Linear Constitutive Relation" Circuits Syst. Signal Process, 35(1), 43-62, 2016. https://doi.org/10.1007/s00034-015-0067-8

[12] Y. V. Pershin, M. Di Ventra, "Practical approach to programmable analog circuits with memristors" IEEE Trans. Circ. Syst. I, 57(8), 1857–1864, 2010. https://doi.org/10.1109/TCSI.2009.2038539

[13] A. L. Hodgkin, A. F. Huxley, "A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve" J. Physiol., **117**(4), 500—544, 1952. https://doi.org/10.1113/jphysiol.1952.sp004764

[14] L. Chua, V. Sbitnev, H. Kim, "Hodgkin-Huxley Axon is Made of Memristors" Int. J. of Bifurcation and Chaos, **22**(3), 1230011-1-48, 2012. https://doi.org/10.1142/S021812741230011X

[15] D. Biolek, Z. Kolka, J. Vávra, S. Doan, "Universal Emulator of Memristive and Other Two-Terminal Devices" International J. of Unconventional Computing, **12**(4), 281—302, 2016.

[16] K. Ochs, E. Solan, S. Dirkmann, T. Mussenbrock, "Wave digital emulation of a double barrier memristive device" in Proceedings of 59th International Midwest Symposium on Circuits and Systems (MWSCAS), Abu Dhabi, 1-4, 2016. https://doi.org/10.1109/MWSCAS.2016.7869946

[17] J. Vavra, Z. Kolka, D. Biolek, "Two-Port Hybrid Emulator of Analog Devices and its Application in Emulation of Memistor" Journal of Advanced Manufacturing Technology (in print).

[18] W.Y. Yang, Signals and Systems with MATLAB. Springer-Verlag Berlin Heidelberg, 2009.

[19] Z. Kolka, V. Biolkova, D. Biolek, "Stability of Digitally Emulated Mem-Elements" in 2015 International Conference on Computing, Communication and Security (ICCCS), Pointe aux Piments, Mauritius, 415—419, 2015. https://doi.org/10.1109/CCCS.2015.7374184

[20] J. Nagumo, S. Arimoto, S. Yoshizawa, "An Active Pulse Transmission Line Simulating Nerve Axon" Proceedings of the IRE, **50**(10), 2061—2070, 1962. https://doi.org/10.1109/JRPROC.1962.288235

[21] A. Fuchs, Nonlinear Dynamics in Complex Systems: Theory and Applications for the Life-, Neuro- and Natural Sciences. Springer-Verlag Berlin Heidelberg, 2012.

[22] J. C. Butcher, Numerical Methods for Ordinary Differential Equations, 3rd edition. John Wiley & Sons, Chichester, UK, 2016.

[23] D. Zwillinger, CRC Standard Mathematical Tables and Formulae, 32nd Edition, CRC Press, Boca Raton, 2011.

[24] Y.V. Pershin, S. La Fontaine, M. Di Ventra, "Memristive model of amoeba learning" Physical Review E, **80**(2), 021926-1-6, 2010. https://doi.org/10.1103/PhysRevE.80.021926

[25] F. Caravelli, J. Carbajal, "Memristors for the Curious Outsiders" Technologies, **6**(4), 118, 2018. https://doi.org/10.3390/technologies6040118

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 23-29 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Application of Open-Source Optimization Library "Extremum" to the Synthesis of Feedback Control of a Satellite

Andrei Panteleev*, Valentin Panovskiy

Moscow Aviation Institute (National Research University), department "Mathematics and Cybernetics", Moscow, Russia

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 20 August, 2019
Online: 06 September, 2019

Keywords:
Open-source
Optimization algorithm
Optimal control

ABSTRACT

Current work demonstrates how open-source optimization library "Extremum" (OSOL Extremum) can be used to build feedback controller of a satellite. Proposed software was developed to as an attempt to eliminate current problems that are present in scientific area: black-box effect (i.e. there is no opportunity to explore source code, modify it, or simply verify), no code reuse (i.e. implemented procedures are accessible only within software that includes it), limitated application of modern optimization algorithms (i.e. number of optimization algorithms increases but most of them were verified only on synthetic tests). All of them lead to so-called reproducibility crisis.

1 Introduction

This paper is an extension of work originally presented in 2018 IV International Conference on Information Technologies in Engineering Education (Inforino) [1].

Optimization is a field of mathematics that is widely applied in engineering, e.g. optimal component design and control theory [2, 3, 4]. Traditionally, optimization algorithms are divided into two groups:

- analytic – these are determined algorithms which efficiency and convergence properties are proved; their main disadvantage is relatively poor performance which reduces the scope of applied problems that can be solved by them [5],
- (meta)heuristic – on the contrary, this group of algorithms has problems with the convergence of results: it is not guaranteed that an appropriate solution will be found, but these algorithms are much more computationally efficient and practically verified [6].

The current work consists of several parts:

- *Prerequisites – explains reasons which led to the OSOL Extremum development;*
- *Description of OSOL Extremum – provides information about software structure and technology stack;*
- *Synthesis of Feedback Control of a Satellite via the OSOL Extremum – describes mathematical task formulation;*
- *Application Results – demonstrates particular results of application of the developed software;*
- *Conclusion – summarizes the results and provides information about further research.*

2 Prerequisites

In order to understand why the development of a new software was performed it is necessary to describe its background. It will be explained in details in the following subsections.

2.1 Black-Boxing

This is the double-edged sword of software development. On the one hand, it allows one to use given tools without necessity to implement desired algorithm, but on the other - it deprives the opportunity to modify code the way you need it for the exact task. In other words, black-boxing makes all your software tools not as flexible as they can be. Also, you must be totally confident in the quality of the software that you use, because you have no opportunity to check confusing moments.

*Corresponding Author: Andrei Panteleev, Moscow Aviation Institute (National Research University), department "Mathematics and Cybernetics", Moscow, Russia, avpanteleev@inbox.ru

2.2 Code Reuse

Generally, academic scientists and software developers work differently. Mainly, scientists are more concerned to solve the given problem without any care about reuse of the scripts and software that they have developed. Programmers, conversely, think a bit differently: they use different programming concepts (e.g. object-oriented programming paradigm) in order to maximize the efficiency of written code in future. That gives an opportunity to create reusable tools speeding up further investigations.

2.3 Reproducibility Crisis

Last but not least, reproducibility crisis is general problem for the science. A lot of researches that are described in scientific articles cannot be verified or at least repeated. That happens because of the restrictions on the article size (i.e. there is no enough space to provide step-by-step description or to explain every moment in detail) and displaced informational focus (i.e. scientists are more concerned to present application results not the specification moments).

2.4 Section Summary

Mentioned problems lead to various limits in overall development speed. This coerces scientist to solve tasks that are out of his professional scope. One possible way to weaken effects of described problems is to develop open-source libraries that offer wide capabilities of API (application programming interface) including unified standard for algorithm creation, general testing environment, and huge spectrum of applied tasks.

Available open-source packages (e.g. python's scikit and scipy packages, R's specialized packages for optimization) partially solve described problems but mostly provide out-of-box tools. Also none of them fully support interval optimization algorithms and they are hardly connected to external programs. That is why it is important to develop library that can be used as set of building blocks for any desired algorithm.

3 Description of OSOL Extremum

This section provides information about OSOL Extremum presenting its main ideas and features. Source code of the library can be found on GitHub site [7].

3.1 Technology Stack

Software part of the developing library consists of modules that are developed using two program languages: Scala (main logic) and Python (for plotting, scripting, and non-math routines).

Scala belongs to multi-paradigm programming languages which support both object-oriented and functional features [8]. Object-oriented paradigm gives opportunity to develop inheritance hierarchy that simplifies code reuse a lot. Functional paradigm is often treated as more "native" to mathematicians because its main statements: treating functions as basic objects, data immutability, pure functions, and etc.). Mainly, these statements aimed to provide code stability decreasing places of potential code mistakes. Also. It should be noted that Scala runs on JVM platform, so it is easily maintained and distributed.

Python is a well-known language which is widely used by scientists of different specialization because of its relative simplicity and generous number of various packages (e.g. NumPy, SciPy, Matplotlib, and etc.) [9].

3.2 Library Structure

OSOL Extremum consists of several parts:

- implementation of basic required mathematical operations and objects, e.g. vector operations, function transformations, random number generators, and statistics calculation,
- general template for optimization algorithms that can be extended by other researches to implement desired method,
- block of implemented optimization algorithms,
- block of different tasks (including basic optimization tasks and complicated optimal control synthesis problems).

3.3 Current State and Future Goals

Current version of OSOL Extremum supports the following features:

- definition of basic optimization algorithm in a form of abstract template that supports wide range of running instructions, e.g. termination by max time or iterations and logging instructions,
- bunch of classical and modern optimization algorithms (including Random Search, Simulated Annealing, Cat Swarm Optimization, and Harmony Search) [5, 6],
- tools for the interaction with Simulink models.

In future it is planned to: implement more optimization algorithms, add support of interval optimization algorithms, add support of other modelling systems (e.g. Wolfram Mathematica), add support of other computational cores.

4 Synthesis of Feedback Control of a Satellite via the OSOL Extremum

This section will provide information how OSOL Extremum can be used to solve applied problem in the field of control theory.

4.1 Stage 1: Open-Loop Control Determination

Problem Definition. During the first stage it is required to solve the problem of optimal open-loop control determination.

Let's consider that the behavior of an object that is described by the following differential equation [10]:

$$\dot{x}(t) = f(t, x(t), u(t)), \tag{1}$$

where $t \in T = [t_0; t_1]$ is continuous time, T - system operational time, $x \in \mathbb{R}^n$ - system state vector, $u \in U \subset \mathbb{R}^q$ - control vector, U - set of admissible control values, $f(t,x,u) = (f_1(t,x,u), ..., f_n(t,x,u))^{\mathrm{T}}$ - continuous vector-function.

Initial state is given:

$$x(t_0) = x_0, \tag{2}$$

terminal state $x(t_1)$ should satisfy the following conditions:

$$\Gamma_i(x(t_1)) = 0, i = 1, ..., l, \tag{3}$$

where $0 \le l \le n$, $\Gamma_i(x), i = 1, ..., l$ are continuously differentiable, system of vectors $\left\{\frac{\partial \Gamma_i(x)}{\partial x_1}, ..., \frac{\partial \Gamma_i(x)}{\partial x_n}\right\}, i = 1, ..., l$ is linearly independent $\forall x \in \mathbb{R}^n$.

The only information which is used to form control is continuous time t. Hence, we consider open-loop control.

Set of admissible processes $D(t_0, x_0)$ is defined as a set of pairs $d = (x(\cdot), u(\cdot))$ which include trajectory $x(\cdot)$ and admissible control $u(\cdot)$ where $\forall t \in T : u(t) \in U$ that satisfy equations (1), (2), and (3).

On the set of admissible processes we consider cost functional

$$I(d) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt + F(x(t_1)), \tag{4}$$

where $f^0(t,x,u)$ and $F(x)$ are given continuous functions.

In order to transform the given problem into the problem without terminal constraints we can use penalty function:

$$\bar{I}(d) = I(d) + \sum_{i=1}^{l} E_i(x(t_1), \varepsilon_i, p_i), \tag{5}$$

where

$$E_i(x, \varepsilon_i, p_i) = \begin{cases} (p_i \cdot \Gamma_i(x))^2 \text{ if } |\Gamma_i(x)| > \varepsilon_i \\ |\Gamma_i(x)| \text{ otherwise,} \end{cases} \tag{6}$$

are error functions, p_i and ε_i, $i = 1, ..., l$ are penalty parameters and maximum allowed errors accordingly.

The task is to find such a pair $d^* = (x^*(\cdot), u^*(\cdot))$ that

$$d^* = \mathrm{Arg} \min_{d \in D(t_0, x_0)} I(d). \tag{7}$$

To solve task defined by equation (7) it is suggested to propose parametric form of control and convert initial problem to a problem of parametric optimization. The control, which is obtained by such a procedure, is suboptimal.

Application of Optimization Algorithms. To apply optimization algorithms it is required to parameterize controls in any suitable vector form. Proposed approach was tested using several parameterization schemes:

- polynomial control: $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}} \Rightarrow u(t) = a_0 + a_1 \cdot t + a_2 \cdot t^2 + ... + a_s \cdot t^s$,
- piecewise linear control: $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}} \Rightarrow u(t) = \frac{\tau_{i+1} - t}{\tau_{i+1} - \tau_i} \cdot a_i + \frac{t - \tau_i}{\tau_{i+1} - \tau_i} \cdot a_{i+1}$, where $\tau_i = t_0 + i \cdot \frac{t_1 - t_0}{s}$ and $\tau_i \le t < \tau_{i+1}$.

The value of the cost functional is calculated by the numerical methods using Euler, Euler-Cauchy or Runge-Kutta formulas [11, 12].

To obtain numerical results it is proposed to use two techniques:

- simply generate polynomial control (this technique will be called direct),
- start with determination of controls in polynomial form (with few degrees of freedom) and then use obtained solution as a seed for optimization procedure with piecewise linear controls (this technique will be called sequential).

Transformation, which is used during sequential technique, can be easily done by calculation of values of polynomial control in nodes of uniform grid (demonstrated on Figure 1).

Described techniques differ in this way: direct requires less time to obtain control, sequential requires more time (because of two parts) but potentially generates more efficient controls (due to bigger number of degrees of freedom).

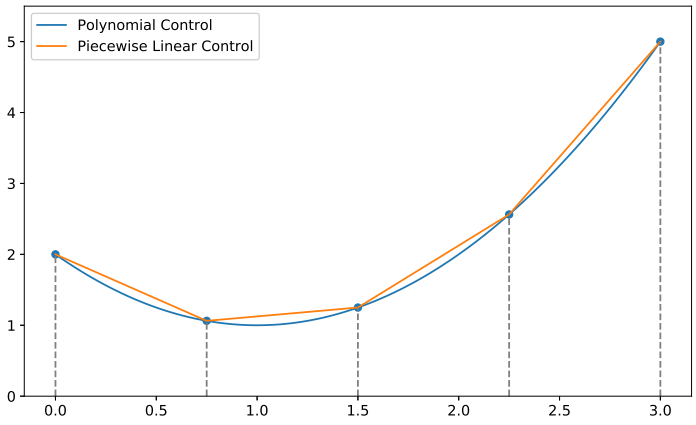


Figure 1: Conversion of polynomial control to piecewise linear one

Thus, initial task is transformed to the determination of optimal vector $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}}$. As the final result of the first stage set of suboptimal controls $\tilde{u}^{*,k}, k = 1, ..., K$ will be found, i.e. $\tilde{u}^{*,k}$ corresponds to initial state vector $x^k(t_0)$ (K is a number of totally observed initial conditions).

4.2 *Stage 2: Controller Construction*

Final form of the controller will be found during the final stage in a form of regressor. Current block consists of data preparation part and actually regressor construction.

Data Processing. During current phase initial data will be transformed and modified. There are various data processing procedures [13], e.g.

- **Cleaning**. This process ensures that only valid data will be used. In our case it is important to include this phase because on previous stage we use metaheuristic optimization algorithms which are not guaranteed to retrieve the best possible solution. Thus, we will eliminate from further consideration those controls which do not provide sufficient accuracy in the means of the final state error defined by equation (8).
- **Imputation** is widely used in machine learning. Mostly it is present in cases when there is a possibility of data loss. Considered case is determined so there is no need to include it.

- **Transformation**. During this phase data, which was previously cleaned, is transformed from raw format (in our case it is set of suboptimal controls) into format that is suitable for machine learning procedures. We will transform data to a table format.
- **Normalization** is required to standardize obtain data, e.g. nullify its mean values, normalize standard deviation, etc.

In terms of cleaning it is proposed to use only those controls which correspond to trajectories with relatively small terminal constraints errors, e.g.

$$\frac{1}{l} \cdot \sum_{i=1}^{l} |\Gamma_i (x(t_1))| \leq b, \tag{8}$$

where b is a bias value that can be treated as an acceptable level of average cumulative error.

Having set of suboptimal controls $\left\{\tilde{u}^{*,k}\right\}_{k=1}^{K}$ it is possible to find corresponding set of trajectories $\left\{\tilde{x}^{*,k}\right\}_{k=1}^{K}$ from equation (1). This information is sufficient to build initial dataset that will be considered as a base. Specifically, it is proposed to build dataset in a form of Table 1.

Table 1: Initial dataset template

j(index)	t	x_1	$\cdots$	x_n	u_1	$\cdots$	u_q
1	$t\Big\vert_{j=1}$	$x_1\Big\vert_{j=1}$	$\cdots$	$x_n\Big\vert_{j=1}$	$u_1\Big\vert_{j=1}$	$\cdots$	$u_q\Big\vert_{j=1}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$

Such a table can be built for any selected control and then all obtained tables can be concatenated. Thus, we have K different trajectories from each of which we sample in N discrete time moments. Finally, it results into the following bounds for j index: $j = 1, ..., K \cdot N$.

Regressor Construction. Regression is a classical example of so-called supervised learning [14]. In a considered scope of optimal control determination we can treat controls (derived from optimization procedures during the first stage) as an ideal output. Time, and state values can be treated as the input. In other words, desired feedback controller can be treated in this way:

$$\hat{u}_i \left(t\Big\vert_{j=j^*}, x_1\Big\vert_{j=j^*}, ..., x_n\Big\vert_{j=j^*} \right) = u_i\Big\vert_{j=j^*}, \tag{9}$$

where $i = 1, ..., q$ and j^* - possible index.

From equation (9) it is obvious how to generate control for previously observed values but it is not clear what to do with new values that will arise in future. That is why machine learning is proposed to be used here - because of its ability to generalize information basing on finite set of observations [15, 14].

In current work it is proposed to build regressor using well known and widely applied Random Forests [16], which are already implemented in Python open-source library Scikit-Learn [17].

4.3 Section Summary

The core component of the described approach is application of optimization algorithms which can be realized via the OSOL Extremum. Following sections will demonstrate how the proposed approach can be applied to a real control problem.

5 Application Results

5.1 Description of Dynamic Model

To demonstrate the proposed approach the problem of satellite rotational motion dampening is considered [18]. After the conversion to non-dimensional variables system of differential equations, which describes solid body motion relative to centre of mass, will have the following form:

$$\begin{cases} \dot{p} = \frac{u_1}{6}, \\ \dot{q} = u_2 - 0.2 \cdot p \cdot r, \\ \dot{r} = 0.2 \cdot (u_3 + p \cdot q). \end{cases} \tag{10}$$

Control vector constraints are represented via intervals $\forall t \in [0; 1]$: $U = [-500; 500] \times ... \times [-500; 500]$.

Cost functional is given by the formula:

$$I(d) = \int_0^1 (|u_1(t)| + |u_2(t)| + |u_3(t)|)\, dt. \tag{11}$$

Initial values for p_0, q_0, r_0 will be taken from sets $\{-25, -20, ..., -5, 0, 5, ..., 20, 25\}$ (for direct optimization procedure) and $\{21, 22, 23, 24, 25\}$ (for sequential optimization procedure) which results into 11^3 and 5^3 possible combinations accordingly.

Terminal constraints are the same for any starting configuration:

$$p(1) = q(1) = r(1) = 0. \tag{12}$$

Parameters of error functions are also the same for any starting configuration:

$$p_i = 10^3, \varepsilon_i = 0.1, i = 1, 2, 3. \tag{13}$$

Thus, modified functional cost will have the following from:

$$\begin{aligned} \bar{I}(d) = \int_0^1 (|u_1(t)| + |u_2(t)| + |u_3(t)|)dt + \\ + E_1\left(p(t_1), 0.1, 10^3\right) + \\ + E_2\left(q(t_1), 0.1, 10^3\right) + \\ + E_3\left(r(t_1), 0.1, 10^3\right). \end{aligned} \tag{14}$$

5.2 Intermediate Results

For All graphics in this and following section horizontal axis is associated with dimensionless time, vertical axis provides information for target variable (also in dimensionless units).

All results in the current section were provided assuming $s = 2$ for polynomial control and $s = 10$ for piecewise linear.

Figure 2 demonstrates the obtained results during the optimization stage (for sequential technique).

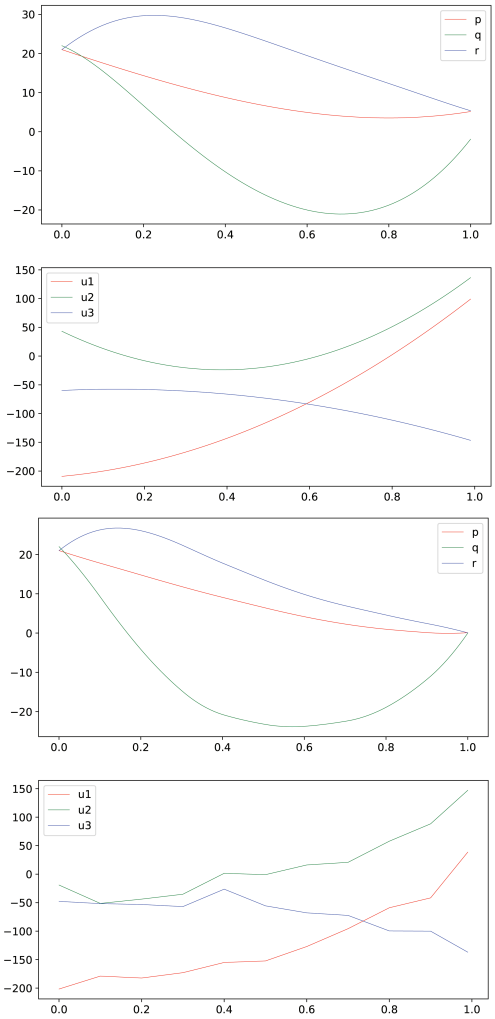


Figure 2: Results of sequential optimization procedure for initial conditions $p(0) = 21, q(0) = 22, r(0) = 21$

From the plot it is clear that during the first stage (generation of polynomial controls) some of them have corresponding trajectories with relatively big errors for terminal state. But it is successfully minimized during the construction of the piecewise polynomial control.

Better visual description of obtained controls can be obtained via bunch graphics demonstrated on Figure 3.

It should be noted that all columns will be scaled (except of the control columns), i.e. transformed to have maximum value equal to 1 and minimum - to 0.

5.3 *Final Results*

The following section will contain information about obtained regressors. Initial values are chosen in order not to duplicate previously used initial states. Each received trajectory was sampled $N = 101$ times in moments $t_i = 0.01 \cdot i$.

Direct technique. Example of application is demonstrated on Figure 4 and Figure 5.

Initial conditions: $p_0 = 24, q_0 = 16, r_0 = 16$.

Achieved terminal state: $p(t_1) = -0.067, q(t_1) = -0.065, r(t_1) = -0.094$.

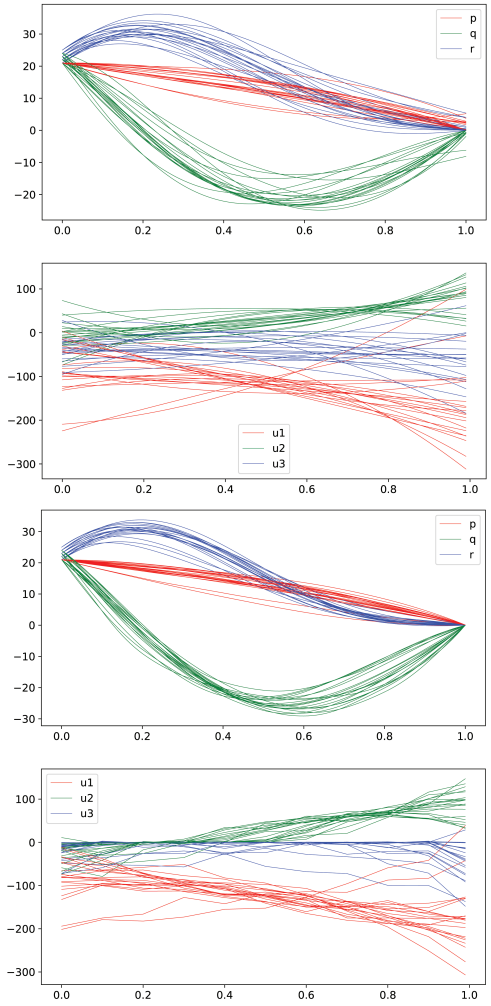


Figure 3: Bunches of trajectories and controls (for sequential optimization procedure)

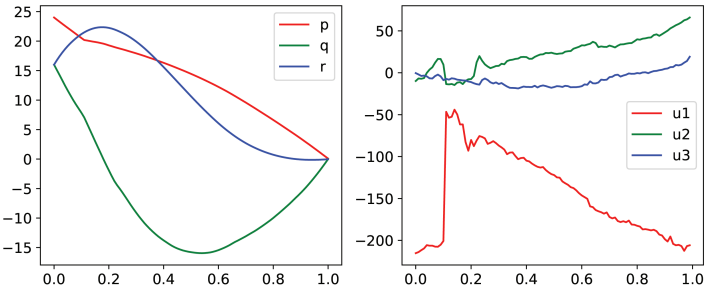


Figure 4: Example 1: trajectories and controls

Initial conditions: $p_0 = -9, q_0 = 13, r_0 = -24$.

Achieved terminal state: $p(t_1) = 0.096, q(t_1) = -0.008, r(t_1) = -0.019$.

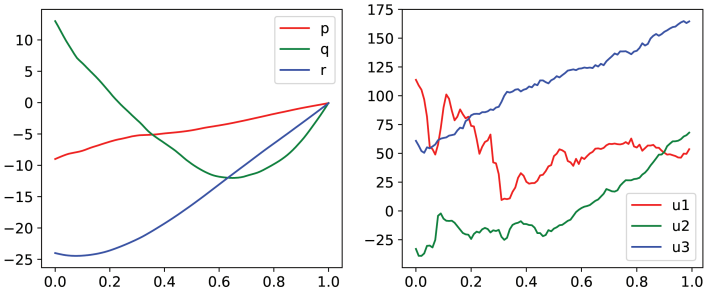


Figure 5: Example 2: trajectories and controls

Sequential technique. Example of application is demonstrated on Figure 6 and Figure 7.

Initial conditions: $p_0 = 24.7, q_0 = 23.3, r_0 = 21.1$.

Achieved terminal state: $p(t_1) = 0.101, q(t_1) = -0.063, r(t_1) = -0.002$.

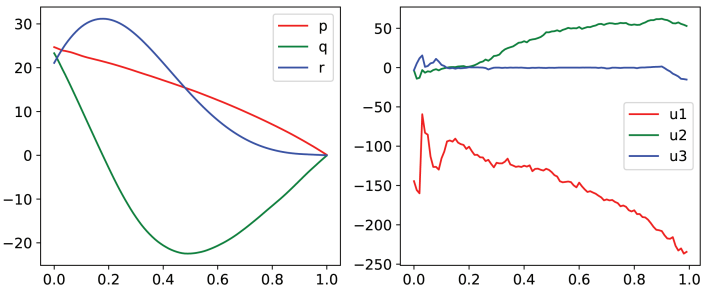


Figure 6: Example 3: trajectories and controls

Initial conditions: $p_0 = 21.1, q_0 = 21.2, r_0 = 21.3$.

Achieved terminal state: $p(t_1) = 0.09, q(t_1) = -0.067, r(t_1) = 0.058$.

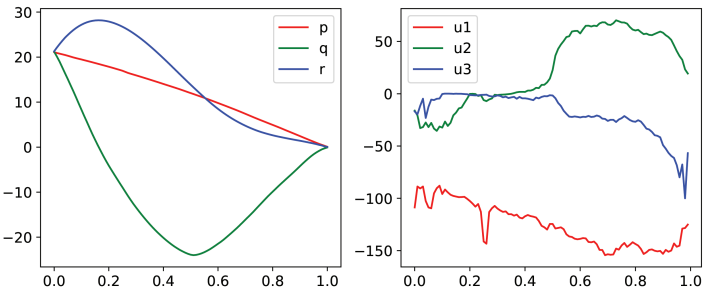


Figure 7: Example 4: trajectories and controls

From observed graphics it is evident that the proposed approach has a great potential to solve the problem of feedback control synthesis. Disadvantage of the proposed approach is that obtained controls are not smooth functions. Such particularity is treated as a disadvantage because it complicates the hardware implementation of the generated control and lowers durability of control mechanism due to the necessity of high-amplitude control adjustments.

6 Conclusion

Main results of the current work are the following:

- efficiency of the developed OSOL Extremum software complex was proved based on obtained solutions for tasks of open-loop control determination,
- proposed approach was tested on an applied task of satellite rotational motion dampening.

Further research is planned to be conducted in several different directions:

- find the way to make controls, that are generated by regressors, have smoother form,
- determine losses in efficiency in the terms of cost functional compared to other computational approaches (including analytic ones),
- find the way how to conduct proper comparative analysis and provide detailed comparison report with the existing software complexes in the terms of accuracy, computational efficiency, and algorithm coverage.

References

[1] Panteleev A.V., Panovskiy V. "2018 IV International Conference on Information Technologies in Engineering Education (Inforino)" in 2018 IV International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russia, 2018. https://doi.org/10.1109/INFORINO.2018.8581745

[2] Panteleev A.V., Letova T.A., Pomazueva E.A. "Parametric Design of Optimal in Average Fractional-Order PID Controller in Flight Control Problem". Automation and Remote Control, **79**(1): 153 – 166, 2018. https://doi.org/10.1134/S0005117918010137

[3] Kuo Y.-L., Wu T.-L. "Open-loop and closed-loop attitude dynamics and controls of miniature spacecraft using pseudowheels" Computers and Mathematics with Applications, **64**: 1282 – 1290, 2012. https://doi.org/10.1016/j.camwa.2012.03.072

[4] Panteleev A.V., Panovskiy V.N., Korotkova T.I. "Interval explosion search algorithm and its application to hypersonic aircraft modelling and motion optimization problems". Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software, **9**: 55 – 67, 2016. https://doi.org/10.14529/mmp160305

[5] Bazaraa M.S., Sherali H.D., Shetty C.M. Nonlinear programming. Theory and algorithms, New York, John Wiley & Sons, 2006.

[6] Gendreau M., and Potvin J.-S. Handbook of Metaheuristics, New York, Springer, 2010.

[7] GitHub: Open-Source Optimization Library - Extremum. https://github.com/wol4aravio/OSOL.Extremum

[8] Odersky M., Spoon L., Venners B. Programming in Scala, Walnut Creek, Artima Press, 2016.

[9] Lutz M. Learning Python, Sebastopol, O'Reilly Media, 2013.

[10] Panteleev A., Bortakovskiy A. Control Theory in Examples and Problems (Teorija upravlenija v primerah i zadachah), Moscow, Viskaya Shkola, 2003.

[11] Panteleev A., Metlitskaya D. "An application of genetic algorithms with binary and real coding for approximate synthesis of suboptimal control in deterministic systems", Autom. Remote Control, **72**:11, 2328 – 2338, 2011. https://doi.org/10.1134/S0005117911110075

[12] Panteleev A., Metlitskaya D. "Using the method of artificial immune systems to seek the suboptimal program control of deterministic systems", Autom. Remote Control, **75**:11, 1922 – 1935, 2014. https://doi.org/10.1134/S0005117914110034

[13] VanderPlas J. Python Data Science Handbook, Sebastopol, CA: O'Reilly, 2017.

[14] Herbrich R. Graepel T. Machine Learning: An Algorithmic Perspective, Boca Raton, FL, CRC Press, 2015.

[15] Goodfellow I., Bengio Y., Courville A. Deep Learning, Cambridge, MA, MIT Press, 2016.

[16] Breiman L. "Random Forests". Machine Learning, **45**(1): 5 – 32, 2001. https://doi.org/10.1023/A:1010933404324

[17] Hackeling G. Mastering Machine Learning with scikit-learn, Birmingham, Packt Publishing, 2014.

[18] Krylov I. "Numerical solution of the problem of the optimal stabilization of an artificial satellite". USSR Computational Mathematics and Mathematical Physics, **8**(1): 284 – 291, 1968. http://dx.doi.org/10.1016/0041-5553(68)90021-9

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 30-38 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Fully Homomorphic Encryption Scheme Based On Complex Numbers

Khalil Hariss*[1,2], Maroun Chamoun[1], Abed Ellatif Samhat[2]

[1] *Université Saint Joseph, ESIB, CIMTI, Mar Roukoz, Lebanon*

[2] *Lebanese University, Faculty of Engineering, CRSI, Hadath, Lebanon*

ARTICLE INFO

Article history:
Received: 15 June, 2019
Accepted: 20 August, 2019
Online: 06 September, 2019

Keywords:
Complex Numbers
Cloud Systems
Fully Homomorphic
Asymmetric Scheme
Bootstrapping
Approximate GCD

ABSTRACT

In this paper, we present a new Somewhat Homomorphic Encryption (SHE) scheme using computation over complex numbers. We then use Bootstrapping technique to make the scheme Fully Homomorphic (FH) and supports unbounded number of circuit depth. In addition to its homomorphic properties and security level, a main characteristic of the proposed new scheme is its simplicity as it is merely based on addition and multiplication operations over complex numbers. The new scheme is implemented under Python using SAGEMath library and evaluated. Then a crypt-analysis based on Approximate GCD problem is done. A comparison with the BGV, a well known Fully Homomorphic Encryption (FHE) scheme, shows that this new scheme is an efficient homomorphic encryption scheme.

1 Introduction

An encryption scheme is said to be FH, if it allows to perform addition and multiplication operations explicitly over the cipher-texts while performing implicitly the same operations over the plain-texts. This new type of ciphering is very required in cloud systems mainly when users treat the cloud as a third untrusted party. With Homomorphic Encryption (HE) the cloud is able to process over encrypted storage and the privacy of the user is preserved at the cloud side. In Figure. 1, we show the case of a mobile user sending an encrypted query to the cloud using HE, thus the cloud is able to process encrypted query over encrypted data to return an encrypted answer. The user can do the decryption.

The notion of homomorphism was first introduced by Rivest, Adleman and Dertouzo [1] as privacy homomorphism after the invention of RSA crypto-system [2]. The basic RSA crypto-system is a multiplicative homomorphic scheme that allows to perform multiplication operations over the cipher-texts. Give, for example, RSA public key, $p_k = (N, e)$ and cipher $c_i = (m_i)^e mod(N)$ for a plain-text m_i. It is simple to demonstrate that $\prod_i c_i = (\prod_i m_i)^e mod(N)$. Several works followed the RSA scheme. A state of art of HE is given in [3] – [6]. Some of the known homomorphic schemes are: Paillier crypto-system [7, 8], Domingo Ferrer crypto-system [9, 10] which is a HE scheme based on polynomial calculations. The Enhanced MORE [11] is another FHE based on matrix calculation and the NOHE [12] scheme is a lightweight FHE that profits from the simplicity and the homomorphic behavior of logic *NOT*. The most valued work in this domain was given by Craig Gentry in his Ph.D. thesis in 2009 [13, 14]. The work of Gentry was inspired from lattice based cryptography. Gentry first introduced a SHE scheme that supports a bounded number of operations over the cipher-texts. Then he developed a new refresh mechanism called Bootstrapping [15] – [19] to make the scheme FH and supports unbounded operations. An important FHE scheme is the BGV crypto-system developed by Brakerski, Gentry and Vaikuntanathan in [20, 21]. Since BGV scheme is also Somewhat Homomorphic (SH), Modulus Switching (MS) is a technique used with BGV to extend the circuit's depth during the evaluation procedure.

In this paper, our main motivation is to add a value in homomorphic encryption by designing a new simple and robust FHE based on simple complex numbers operations . We first build a new asymmetric SHE complex-based scheme, then we apply Gentry refresh mechanism (Bootstrapping). In comparison with BGV which is lattice based encryption, our scheme is simply based on complex addition and multiplication operations. In addition, Bootstrapping supports unbounded number of circuit depth in our scheme, while MS extends the BGV evaluation to a limited number.

The rest of the paper is organized as follows: in section 2, we present the basic concept of HE with the properties that can lead to a HE scheme. In section 3, we introduce our new FHE scheme built using complex numbers theory. In section 4 we consider our implementations for the new Complex-based scheme and the

*Corresponding Author: Khalil Hariss, Université Saint Joseph, khalil.hariss@net.usj.edu.lb

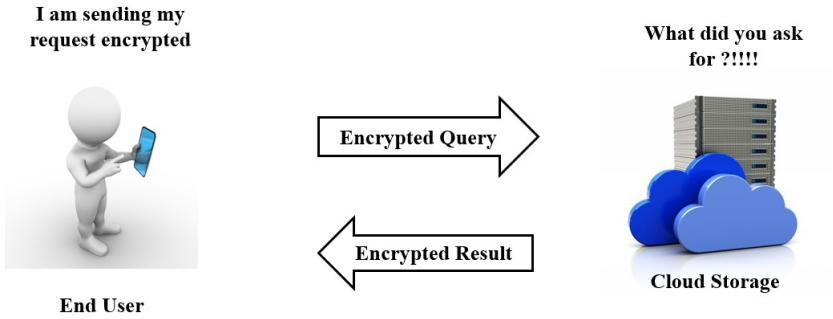


Figure 1: Cloud Scenarion Under Homomorphic Encryption

BGV scheme under Python using SAGEMath Library, followed by a crypt-analysis of our new scheme based on the Approximate GCD problem [22, 23, 24]. And finally the conclusion, comparison between the two schemes and future works are listed in section 5.

2 Homomorphic Encryption

In this work, we focus on asymmetric schemes. Consider an asymmetric scheme α, where p_k is the public key, s_k is the secret key, c_i is the ciphered text of a plain-text m_i. A homomorphic scheme has 3 different basic functions that are: $KeyGen_\alpha$ that generates (p_k, s_k), $Encrypt_\alpha(p_k, m_i)$ that outputs cipher-text c_i, $Decrypt_\alpha(s_k, c_i)$ that outputs the plain-text m_i.

2.1 Evaluation Function

In addition to the three basic functions listed above, a homomorphic scheme has also an evaluation function defined by $Evaluate_\alpha$ that takes as input the public key p_k, a circuit L and a tuple of cipher-texts $C = (c_1, c_2, c_3, ..., c_t)$, cipher of the input plain-texts vector $(m_1, m_2, ..., m_t)$. The evaluation function outputs a cipher-text Ψ given by:

$$\Psi = Evaluate_\alpha(p_k, L, C) \quad (1)$$

The scheme is homomorphic if we have: $\Psi = Encrypt_\alpha(p_k, L(m_1, m_2, ..., m_t))$.

2.2 Homomorphic Properties

When the circuit L performs a certain operation over the encrypted data as written in 1, the cloud is computing a predefined Boolean function f that can be written in a polynomial form. A polynomial form is a set of addition and multiplication gates. Thus to build a HE scheme we should satisfy these two basic properties:

1. Addition

$$Enc_\alpha(p_k, m_1) + Enc_\alpha(p_k, m_2) = Enc_\alpha(p_k, m_1 + m_2) \quad (2)$$

2. Multiplication

$$Enc_\alpha(p_k, m_1) \times Enc_\alpha(p_k, m_2) = Enc_\alpha(p_k, m_1 \times m_2) \quad (3)$$

where m_1 and m_2 are two plain-texts in $\{0, 1\}$ and $Enc_\alpha(p_k, m_i)$ is the encryption function.

3 Homomorphic Complex Scheme

In this section, we build our scheme using complex numbers. We first list the security parameters and then we detail our proposal.

3.1 Parameters

Based on a security parameter λ, different other parameters are generated as listed in [15] to build the new scheme:

1. γ: the bit length of the real and imaginary parts in the public key ($\gamma = \omega(\eta^2 log_2 \lambda)$).
2. ρ: the bit-length of the real and imaginary parts in the noise ($\rho = \omega(log_2(\lambda))$.
3. η: the bit length of the secret key ($\eta \geq \rho.\Theta(\lambda log_2^2(\lambda))$.
4. ρ': extra noise parameter.
5. τ: The number of public keys ($\tau \geq \gamma + \omega(log_2 \lambda)$).

3.2 Somewhat Homomorphic Complex Scheme Construction

The SH complex scheme construction is based on the following steps:

1. Secret key: The secret key s_k is defined by an η bit prime integer p.
2. Public keys: We build a public key bank (PK) formed of τ complex numbers having the following form:
   $PK = \{({p_{k1}}^h + i{p_{k2}}^h) = (({pq_1}^h + {\epsilon_1}^h) + i({pq_2}^h + {\epsilon_2}^h), 1 \leq h \leq \tau\}$.
   For each complex public key $({p_{k1}}^h + i{p_{k2}}^h)$, we build different parameters as follow:
   (a) random integer: $q_j^h \longleftarrow Z \cap [0, \frac{2^\eta}{p})$, random noise parameter: ${\epsilon_j}^h \longleftarrow Z \cap (-2^\rho, 2^\rho)$ where $j \in \{1, 2\}$.
   (b) $({p_{k1}}^h, {p_{k2}}^h)$ should have different parities.
   (c) $({\epsilon_1}^h, {\epsilon_2}^h)$ should have the same parity.

3. Encryption Function: The encryption of any bit m is given by the following equation:

$$c = Enc(p_k, m) = ({p_{k1}}^h + i{p_{k2}}^h) \times ({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m \quad (4)$$

where $({p_{k1}}^h + i{p_{k2}}^h)$ is chosen randomly from the public key bank (PK), ${R_1}^h + i{R_2}^h$ is a random complex chosen such that $|{R_1}^h| <<< \frac{p}{6|{\epsilon_1}^h - {\epsilon_2}^h|}$, $|{R_2}^h| <<< \frac{p}{6|{\epsilon_1}^h + {\epsilon_2}^h|}$, ${r_1}^h + i{r_2}^h$ is a random complex chosen between $(-2^{\rho'}, 2^{\rho'})$ such that ${r_1}^h$ and ${r_2}^h$ have the same parity and $|{r_1}^h - {r_2}^h| \ll \frac{p}{6}$(Choosing the intervals of R_1^h, R_2^h, r_1^h, r_2^h is explained in the upcoming decryption section).

4. Decryption Function:
Following 4, we can demonstrate that $c = Enc(p_k, m) = ({p_{k1}}^h + i{p_{k2}}^h) \times ({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m = (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + R_1{}^h \epsilon_1{}^h - R_2{}^h \epsilon_2{}^h + r_1{}^h + m) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + r_2{}^h)$.
The decryption of the cipher c is done as follows:

$$\begin{aligned} Intermediate = real(c) - imag(c) = \\ p(q_1{}^h R_1{}^h - q_2{}^h R_2{}^h - q_1{}^h R_2{}^h - q_2{}^h R_1{}^h) + \\ R_1{}^h(\epsilon_1{}^h - \epsilon_2{}^h) - R_2{}^h(\epsilon_1{}^h + \epsilon_2{}^h) + (r_1{}^h - r_2{}^h) + m. \\ m = mod(modNear(Intermediate, p), 2). \end{aligned} \quad (5)$$

Where $modNear(x, p) = y$, such that $y \in (-\frac{p}{2}, +\frac{p}{2})$.
Decryption works as long as:
$\frac{-p}{2} \leq {R_1}^h({\epsilon_1}^h - {\epsilon_2}^h) - {R_2}^h({\epsilon_1}^h + {\epsilon_2}^h) + ({r_1}^h - {r_2}^h) + m \leq \frac{p}{2}$.
In addition $({\epsilon_1}^h - {\epsilon_2}^h)$, $({\epsilon_1}^h + {\epsilon_2}^h)$ and $({r_1}^h - {r_2}^h)$ are even integers.
The decryption condition is satisfied as long as: $|{R_1}^h| <<< \frac{p}{6|{\epsilon_1}^h - {\epsilon_2}^h|}$, $|{R_2}^h| <<< \frac{p}{6|{\epsilon_1}^h + {\epsilon_2}^h|}$, $|{r_1}^h - {r_2}^h| << \frac{p}{6}$ and each pair of $({\epsilon_1}^h, {\epsilon_2}^h)$, $({r_1}^h, {r_2}^h)$ is formed of two integers having the same parity.

3.3 Complex Homomorphic Properties

We show in this section that the proposed scheme satisfies the homomorphic properties. Given two cipher-texts c_1, c_2 respectively for 2 different plain-texts m_1, m_2 encrypted using the complex scheme listed in section 3.2.
$c_1 = ({p_{k1}}^h + i{p_{k2}}^h)({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m_1 = (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + R_1{}^h \epsilon_1{}^h - R_2{}^h \epsilon_2{}^h + r_1{}^h + m_1) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + r_2{}^h)$. $c_2 = ({p_{k1}}^t + i{p_{k2}}^t)({R_1}^t + i{R_2}^t) + {r_1}^t + i{r_2}^t + m_2 = (pq_1{}^t R_1{}^t - pq_2{}^t R_2{}^t + R_1{}^t \epsilon_1{}^t - R_2{}^t \epsilon_2{}^t + r_1{}^t + m_2) + i(pq_1{}^t R_2{}^t + pq_2{}^t R_1{}^t + R_2{}^t \epsilon_1{}^t + R_1{}^t \epsilon_2{}^t + r_2{}^t)$.

1. Addition

$$\begin{aligned} c_1 + c_2 = \\ (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + pq_1{}^t R_1{}^t - pq_2{}^t R_2{}^t + R_1{}^h \epsilon_1{}^h \\ - R_2{}^h \epsilon_2{}^h + R_1{}^t \epsilon_1{}^t - R_2{}^t \epsilon_2{}^t + r_1{}^h \\ + r_1{}^t + m_1 + m_2) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + \\ pq_1{}^t R_2{}^t + pq_2{}^t R_1{}^t + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + R_2{}^t \epsilon_1{}^t \\ + R_1{}^t \epsilon_2{}^t + r_2{}^h + r_2{}^t). \end{aligned} \quad (6)$$

Decryption of $c_1 + c_2$ is obtained by calculating $mod(modNear(real(c_1 + c_2) - imag(c_1 + c_2), p), 2)$ having the following form:
$pQ_{add} + R_{add} + m_1 + m_2$ such that $R_{add} = {R_1}^h({\epsilon_1}^h - {\epsilon_2}^h) - {R_2}^h({\epsilon_1}^h + {\epsilon_2}^h) + {R_1}^t({\epsilon_1}^t - {\epsilon_2}^t) - {R_2}^t({\epsilon_1}^t + {\epsilon_2}^t) + ({r_1}^h - {r_2}^h) + ({r_1}^t - {r_2}^t)$.
Decryption works as long as R_{add} is even and $\frac{-p}{2} \leq R_{add} \leq \frac{p}{2}$.

2. Multiplication
Similar to addition, decryption of $c_1 \times c_2$ is obtained by calculating $mod(modNear(real(c_1 \times c_2) - imag(c_1 \times c_2), p), 2)$.
We can demonstrate that $real(c_1 \times c_2) - imag(c_1 \times c_2)$ have the following form:
$pQ_{mult} + R_{mult} + m_1 m_2$ such that $R_{mult} = ({r_1}^t {R_1}^h + m_2 {R_1}^h - {r_2}^t {R_2}^h)({\epsilon_1}^h - {\epsilon_2}^h) - ({r_1}^t {R_2}^h + {r_2}^t {R_1}^h + m_2 {R_2}^h)({\epsilon_1}^h + {\epsilon_2}^h) + ({r_1}^h {R_1}^t + m_1 {R_1}^t - {r_2}^h {R_2}^t)({\epsilon_1}^t - {\epsilon_2}^t) - ({R_1}^t {r_2}^h + {r_1}^h {R_2}^t + m_1 {R_2}^t)({\epsilon_1}^t + {\epsilon_2}^t) + ({R_1}^h {R_1}^t {\epsilon_1}^h - {R_2}^h {R_2}^t {\epsilon_1}^h)({\epsilon_1}^t - {\epsilon_2}^t) + ({R_2}^h {R_2}^t {\epsilon_2}^h - {R_1}^h {R_1}^t {\epsilon_2}^h)({\epsilon_1}^t + {\epsilon_2}^t) - ({R_1}^h {R_2}^t {\epsilon_1}^h + {R_2}^h {R_1}^t {\epsilon_1}^h)({\epsilon_1}^t + {\epsilon_2}^t) - ({R_2}^h {R_1}^t {\epsilon_2}^h + {R_1}^h {R_2}^t {\epsilon_2}^h)({\epsilon_1}^t - {\epsilon_2}^t) + {r_1}^h({r_1}^t - {r_2}^t) - {r_2}^h({r_1}^t + {r_2}^t) + m_2({r_1}^h - {r_2}^h) + m_1({r_1}^t - {r_2}^t)$.
Decryption also works as long as R_{mult} is even and $\frac{-p}{2} \leq R_{mult} \leq \frac{p}{2}$. We can simply demonstrate that R_{add} and R_{mult} are even since each couple of $({r_1}^h, {r_2}^h)$, $({r_1}^t, {r_2}^t)$, $({\epsilon_1}^h, {\epsilon_2}^h)$ and $({\epsilon_1}^t, {\epsilon_2}^t)$ is formed of two integers having the same parity, but having R_{add} and R_{mult} between $\frac{-p}{2}$ and $\frac{p}{2}$ is not always satisfied.
As a conclusion our new complex scheme is SH that supports a bounded number of addition and multiplication operations over the cipher-texts.

3.4 Making the Complex Scheme Fully Homomorphic

To make our complex scheme FH, we apply Bootstrapping [15, 17] in order to reduce the noise level after each operation. Starting from a complex cipher c, the plain-text m can be calculated following this equation $m \leftarrow [c^* - \lfloor \frac{c^*}{p} \rceil]_2$ where $c^* = real(c) - imag(c)$ since this decryption equation is much simpler than Equation. 5. First of all we can use Gentry's transformation to squash the decryption circuit. In this transformation we add to the public key some extra information about the secret key, and use this extra information to

post process the cipher-text c^*. The post processed cipher-text c^* can be decrypted more efficiently than the original cipher-text.

3.4.1 Bootstrapping:

Consider the evaluation procedure given in 1, we use instead of the arithmetic circuit L, the decryption circuit D_α of the SH scheme α. The main concept of Bootstrapping is that we have a cipher-text ψ_1 that encrypts m under p_{k1} that we want to refresh. s_{k1} is the secret key related to the public key p_{k1}. s_{k1} is encrypted under another public key p_{k2}. Let $\overline{s_{k1j}}$ be the encrypted secret bits j of s_{k1} under p_{k2}. Bootstrapping is given by this algorithm:

$$\begin{aligned} &Recrypt(p_{k2}, D_\alpha, < \overline{s_{k1j}} >, \psi_{1j}), \\ &Set \quad \overline{\psi_{1j}} \leftarrow Encrypt(p_{k2}, \psi_{1j}) \\ &Output \quad \psi_2 \leftarrow Evaluate_\alpha(p_{k2}, D_\alpha, \langle\langle \overline{s_{k1}} \rangle, \langle \overline{\psi_{1j}} \rangle\rangle) \end{aligned} \tag{7}$$

In 7, the function $Evaluate_\alpha$ takes as input the bits of s_{k1} and ψ_1, each encrypted under p_{k2} to evaluate homomorphically the decryption circuit D_α. The output ψ_2 is an encryption under p_{k2} of $Decrypt_\epsilon(s_{k1}, \psi_1) = m$. By applying 7, we are removing the error vector associated to the first cipher and adding another error vector. The progress is made as long as the second error vector in ψ_2 is shorter than the primitive in ψ_1.
In Figure. 2, we present the evaluation procedure of a circuit with high depth giving birth to a cipher c with a high noise level.
As given in Figure. 3, for a cipher c consider $D_c(sk) = Decrypt_{sk}(c)$, this operation is called squashing the decryption circuit (will be explained in the upcoming section) and $D_c(.)$ is a low depth polynomial in sk. Bootstrapping consists of evaluating $D_c(.)$ using the encryption of the secret key sk.

3.4.2 Squashing the Decryption Circuit:

Bootstrapping is possible as long as the decryption circuit is an arithmetical circuit of low depth. Squashing is the required transformation in making it possible as listed in[15, 16, 17, 18, 19]. Following Gentry's Squashing technique, we add three different extra parameters κ, θ, Θ that are function of the security parameter λ. $\kappa = \frac{\gamma\eta}{\rho'}$, $\theta = \lambda$, $\Theta = \omega(\kappa.log(\lambda))$. For a secret key $s_k^* = p$ and a public key p_k^* from the original complex homomorphic scheme α, we add to the public key a set $\overrightarrow{Y} = \{y_1, y_2, y_3, ..., y_\Theta\}$ of rational numbers in $[0, 2)$ with κ bits of precision, such that there is a sparse subset $S \subset \{1, 2, 3, ..., \Theta\}$ of size θ with $\sum_{i \in S} y_i \approx \frac{1}{p} (mod 2)$.
The secret key becomes the indicator of the subset S. The scheme of section 3.2 becomes FH complex by first applying the following modifications:

1. KeyGen:
   Generate $s_k^* = p$ and p_k^* as before. set $x_p \leftarrow \lfloor \frac{2^\kappa}{p} \rceil$. Choose randomly a Θ bit vector $\overrightarrow{s}$ with hamming weight θ, $\overrightarrow{s} = \langle s_1, s_2, ..., s_\Theta \rangle$ and let $S = \{i; s_i = 1 \text{ in vector } S\}$. Choose Θ random integers $u_i \in Z \cap [0, 2^{\kappa+1})$, $i = 1, 2, 3, ..., \Theta$ subject to the condition that $\sum_{i \in S} u_i = x_p mod(2^{\kappa+1})$. Set $y_i = \frac{u_i}{2^\kappa}$ and $\overrightarrow{Y} = \{y_1, y_2, ..., y_\Theta\}$. Hence each y_i is a positive number smaller than 2 with κ bits of precision after the binary point such that $[\sum_{i \in S} y_i]_2 = (\frac{1}{p}) - \Delta_p$ for some $\Delta_p < 2^{-\kappa}$.
2. Encrypt and Evaluate Algorithm:
   Generate the cipher-text $c^* = real(c) - imag(c)$ from a complex cipher c. Then for each $i \in \{1, 2, 3, ..., \Theta\}$, set $z_i \leftarrow [c^*.y_i]_2$ keeping only $n = \lceil log\theta \rceil + 3$ bits of precision after the binary point for each z_i. Output both c^* and $\overrightarrow{z} = \langle z_1, z_2, ..., z_\Theta \rangle$
3. Decrypt and Output Algorithm:

$$m \leftarrow [c^* - \lfloor \sum_{i \in S} s_i z_i \rceil]_2 \tag{8}$$

Lemma 1. *For every cipher-text $(c^*, \overrightarrow{z})$ that is generated by evaluating a circuit C, it holds that $\sum_i s_i z_i$ is within $\frac{1}{4}$ of an integer.*

Proof. Fix an arithmetic circuit L, public keys and secret keys generated with respect to the security parameter λ, with $\{y_i\}_{i=1}^\Theta$ the rational numbers in the public key and $\{s_i\}_{i=1}^\Theta$ the secret key bits. If we recall that y_i's were chosen that: $[\sum_i s_i y_i]_2 = (\frac{1}{p}) - \Delta_p$. We fix a circuit L, and t cipher-texts $\{c_i\}_{i=1}^t$ that encrypts the input to L and denote: $c = Evaluate_\epsilon(pk, L, c_1, c_2, .., ..., c_t)$, we need to establish this equation:

$$\lfloor \frac{c^*}{p} \rceil = \lfloor \sum_i s_i z_i \rceil (mod 2) \tag{9}$$

where $c^* = real(c) - imag(c)$ and the z_i's are computed as $[c^*.y_i]_2$ with only $\lceil log\theta + 3 \rceil$ bits of precision after the binary point, hence $[c^*.y_i]_2 = z_i - \Delta_i$ with $|\Delta_i| \leq \frac{1}{16\theta}$.
$[(\frac{c^*}{p}) - \sum s_i z_i]_2 = [(\frac{c^*}{p}) - \sum_i s_i [c^* y_i]_2 + \sum s_i \Delta_i]_2 = [(\frac{c^*}{p}) - c^* [\sum s_i y_i]_2 + \sum s_i \Delta_i]_2 = [(\frac{c^*}{p}) - c^*.(\frac{1}{p} - \Delta_p) + \sum s_i \Delta_i]_2$
$= [c^*.\Delta_p + \sum s_i \Delta_i]_2$. □

We claim that the final quantity inside the brackets has magnitude at most $\frac{1}{8}$. By definition, since c^* is a valid cipher-text output, the value $\frac{c^*}{p}$ is within $\frac{1}{8}$ of an integer. Together all these facts imply the lemma. To establish the claim, we first observe that $|\sum s_i \Delta_i| \leq \theta \times \frac{1}{16\theta}$. Regarding $c^* \Delta_p$, recall that the output c^* is obtained by evaluating the circuit L on the cipher-text c_i as listed in [15], for any polynomial P that implements the circuit L, for any $\alpha \geq 1$, if P's input has magnitude at most $2^{\alpha(\rho'+2)}$, its output magnitude at most $2^{\alpha(\eta-4)}$. In particular, when P's are fresh cipher-texts, which have magnitude at most 2^γ, P's output cipher-text c^* has magnitude at most $2^{\gamma(\eta-4)(\rho'+2)} < 2^{\kappa-4}$. Thus $|c^* \Delta_p| < \frac{1}{16}$.

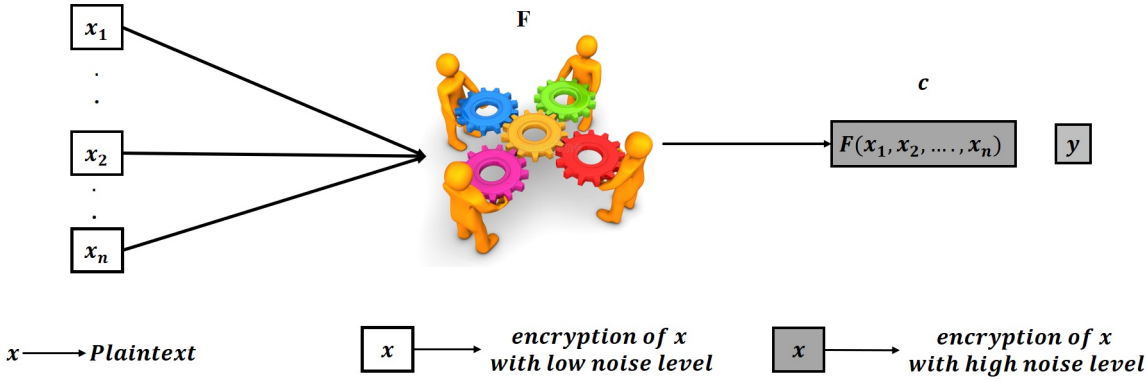


Figure 2: Circuit Evaluation with High Noise

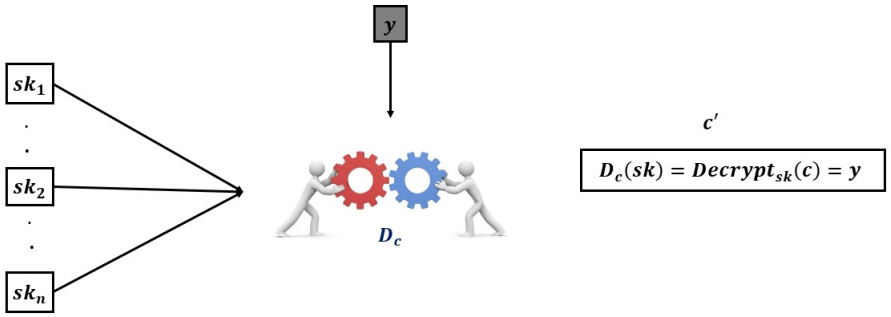


Figure 3: Fresh Cipher Generation

3.4.3 *Practical Implementation of Squashing and Bootstrapping:*

For the squashed decryption circuit given in 8, the implementation of $\sum_{i\in S} s_i z_i$ (such that $z_i = (z_{i1}, z_{i2}, ..., z_{in})$ is a vector of $n = \lceil log\theta \rceil + 3$ bits, where $1 \leq i \leq \Theta$), can be done with lower computational complexity as implemented in [16, 19]. This new implementation is based on dividing the new secret key $\overrightarrow{s}$ in θ boxes of $B = \dfrac{\Theta}{\theta}$ bits each, where each box has a single bit having the value 1. This will lead us to obtain a grade school addition algorithm that requires $O(\theta^2)$ multiplications instead of $O(\Theta.\theta)$. The secret key $\overrightarrow{s}$ is divided into $s_{k,i}$, the i^{th} secret key bit in box k, where $1 \leq k \leq \theta$ and $1 \leq i \leq B$. The resultant equation is:

$$m \leftarrow (c^* - \lfloor \sum_{k=1}^{\theta} (\sum_{i=1}^{B} s_{k,i} z_{k,i}) \rceil) mod(2) \tag{10}$$

We denote that the sum $q_k = \sum_{i=1}^{B} s_{k,i} z_{k,i}$ is obtained by adding B numbers, only one being nonzero. The decryption equation is now:

$$m \leftarrow (c^* - \lfloor \sum_{i=1}^{\theta} (q_k) \rceil) mod(2) \tag{11}$$

Where the q_k's are rational in $[0, 2)$ with n bits of precision after the binary point. Another form of the decryption equation is given by:

$$m = [c^*]_2 \oplus [\lfloor \sum_{i=1}^{\Theta} s_i z_i \rceil]_2 \tag{12}$$

Based on 12, we can deduce that the parity of the plain-text m is related to the parity of the primitive cipher-text c^* and the parity of $[\lfloor \sum_{i=1}^{\Theta} s_i z_i \rceil]_2$.

In order to make this concept much clearer, a simple implementation of bootstrapping, using low values, is explained in the following example. Giving randomly $\Theta = 4$, $\theta = 2$, $B = \dfrac{\Theta}{\theta} = 2$, and a precision bit $n = 1$. The secret key $\overrightarrow{s} = [b_1, b_2, b_3, b_4]$ such that ($b_1 = 1$ and $b_2 = 0$) or ($b_1 = 0$ and $b_2 = 1$), ($b_3 = 1$ and $b_4 = 0$) or ($b_3 = 0$ and $b_4 = 1$) based on the implementation of 10. Initial values for evaluating the decryption circuit using the secret key $\overrightarrow{s}$ without encryption are $s_1 = [s_{11} \quad s_{12}] = [b_1 \quad b_2]$, $s_2 = [s_{21} \quad s_{22}] = [b_3 \quad b_4]$, the z values are taken just as an example $z_1 = [z_{11} \quad z_{12}] = [[10] \quad [01]]$, $z_2 = [z_{21} \quad z_{22}] = [[11] \quad [01]]$. We start by applying the evaluation procedure over the plain-texts by calculating $Sum = \sum_{k=1}^{\theta} \sum_{i=1}^{B} s_{k,i} z_{k,i} = \sum_{k=1}^{2} \sum_{i=1}^{2} s_{k,i} z_{k,i} = (b_1 z_{11} + b_2 z_{12}) + (b_3 z_{21} + b_4 z_{22}) = [b_1 \quad b_2] + [b_3 \quad b_3 + b_4] =$

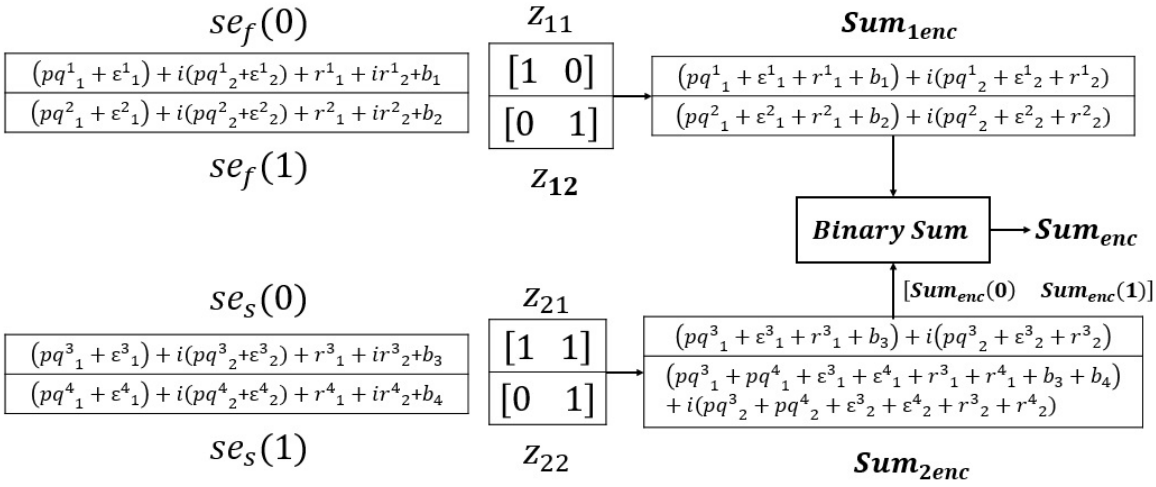


Figure 4: Circuit Evaluation over Cipher-texts.

$(Sum_1) + (Sum_2)$, then we apply the binary summation to calculate $Sum = Sum_1 + Sum_2 = [b_1 + b_3 \quad b_2 + b_3 + b_4 + b_1b_3] = [Sum(0) \quad Sum(1)]$ ($Sum(0) = Sum_1(0) + Sum_2(0)$ and $Sum(1) = Sum_1(1) + Sum_2(1) + Sum_1(0)Sum_2(0)$). As a result, the parity of $\lfloor Sum \rceil$ is given by $b_1 + b_3 + b_2 + b_3 + b_4 + b_1b_3$. Next step is to do the circuit evaluation over the complex cipher-texts. Let $se_f = [se_1 \quad se_2]$ be the cipher of the secret key $s_1 = [b_1 \quad b_2]$ and $se_s = [se_3 \quad se_4]$ be the cipher of the secret key $s_2 = [b_3 \quad b_4]$ using the encryption procedure listed in 4 with random values $(R_1^h + iR_2^h) = 1$.

$se_f = [se_f(0) \quad se_f(1)] = Enc(s_1) = [(pq_1^1 + \epsilon_1^1) + i(pq_2^1 + \epsilon_2^1) + r_1^1 + ir_2^1 + b_1 \quad (pq_1^2 + \epsilon_1^2) + i(pq_2^2 + \epsilon_2^2) + r_1^2 + ir_2^2 + b_2]$.

$$Sum_{enc} = \sum_{k=1}^{\theta}\sum_{i=1}^{B} se_{k,i}z_{k,i} = \sum_{k=1}^{2}\sum_{i=1}^{2} se_{k,i}z_{k,i} = Sum_{1enc} + Sum_{2enc},$$

where $Sum_{1enc} = se_f(0)z_{11} + se_f(1)z_{12}$ and $Sum_{2enc} = se_s(0)z_{21} + se_s(1)z_{22}$. An implementation of circuit evaluation over the cipher-texts is given in Figure.4.

As shown in Figure.4, after applying the binary summation over Sum_{1enc} and Sum_{2enc} we obtain $Sum_{enc} = [Sum_{enc}(0) \quad Sum_{enc}(1)]$ such that $Sum_{enc}(0) = p(Q_{0R} + iQ_{0I}) + (\epsilon_1^1 + \epsilon_1^3) + i(\epsilon_2^1 + \epsilon_2^3) + (r_1^1 + r_1^3) + i(r_2^1 + r_2^3) + \mathbf{b_1} + \mathbf{b_3}$.

$Sum_{enc}(1) = p(Q_{1R}+iQ_{1I})+(\epsilon_1^1\epsilon_1^3-\epsilon_2^1\epsilon_2^3)+(r_1^3\epsilon_1^1+r_1^1\epsilon_1^3-r_2^3\epsilon_2^1-r_2^1\epsilon_2^3)+(r_1^1r_1^3-r_2^1r_2^3)+b_1\epsilon_1^3+b_3\epsilon_1^1+b_1r_1^3+b_3r_1^1+(\epsilon_1^2+\epsilon_1^3+\epsilon_1^4)+(r_1^2+r_1^3+r_1^4)+\mathbf{b_1b_3}+\mathbf{b_2}+\mathbf{b_3}+\mathbf{b_4}+i((\epsilon_1^1\epsilon_2^3+\epsilon_2^1\epsilon_1^3)+(r_2^3\epsilon_1^1+r_1^1\epsilon_2^3+r_2^1\epsilon_1^3+r_1^3\epsilon_2^1)+(r_2^1r_1^3+r_1^1r_2^3)+b_1\epsilon_2^3+b_3\epsilon_2^1+b_1r_2^3+b_3r_2^1+(\epsilon_2^2+\epsilon_2^3+\epsilon_2^4)+(r_2^2+r_2^3+r_2^4))$.

Since $mod(modNear(real(Sum_{enc}(0)) - imag(Sum_{enc}(0)), p), 2) = b_1+b_3$ and $mod(modNear(real(Sum_{enc}(1))-imag(Sum_{enc}(1)), p), 2) = b_1b_3 + b_2 + b_3 + b_4$, as long as the encryption parameters are chosen such that the scheme is SH (i.e Sum_{1enc} and Sum_{2enc} are formed of complex ciphers that can be added and multiplied homomorphically as long as the scheme is SH). Based on all the calculations listed above, a fresh cipher-text of c^* can be written as:

$$fresh_{cipher} = \lfloor Sum_{enc}(0) + Sum_{enc}(1) \rceil + [c^*]_2 + \sum_j (p_{k_1}{}^j + ip_{k_2}{}^j)(R_1^j + iR_2^j) \quad (13)$$

4 Implementation and Security Analysis

In order to validate our work, first we did an implementation of our new Complex-based scheme, then we compared the evaluation of the logic circuit $C = (b_0 \oplus b_1) \bullet (b_2 \oplus b_3)$ using the two schemes: Complex and BGV.

As a brief overview of the BGV [20, 21], it is a FHE scheme that works over bit level and built using Lattice based cryptography. The security of this scheme depends on the hardness of Learning With Error (LWE) introduced by Oded Regev in [25], that relies on the complexity of solving a noisy linear system. Starting from a security parameter λ, we have a secret key $s \in Z_p^{[1,n]}$ and a cipher-text $c \in Z_p^{[n,1]}$. The hardness of LWE resides in taking the lattice dimension: $n \approx poly(\lambda)$ and the ring dimension: $p \approx poly(n)$ following [21, 25]. Homomorphic addition is achieved by simply adding the two cipher-texts, while homomorphic multiplication is done by calculating the Tensor product of the two different cipher-texts which increases the dimension exponentially. Key Switching (KS) is a new technique introduced to reduce the cipher dimension after each homomorphic multiplication. The basic BGV scheme is SH since the noise level will increase with circuit depth, therefore MS is another technique introduced to reduce the noise level after each arithmetic operation and extend to a higher circuit depth.

Finally a crypt-analysis of the new scheme is validated with the GACD attack. All simulations are done under Python with SAGE-Math Library using a machine having the following specifications (CPU: Intel Xeon, E5 – 2630, 2.40 GHZ, 8 CORES, 128 GB RAM).

4.1 Implementations and Results

1. **First Implementation** : First implementation with the new Complex-based scheme is done with three different security layers (**small**: $\lambda = 42$, $\rho = 26$, $\eta = 988$, $\gamma = 147456$, $\Theta = 150$, $\tau = 158$), (**medium**: $\lambda = 52$, $\rho = 41$, $\eta = 1558$, $\gamma = 843033$, $\Theta = 555$, $\tau = 572$), (**large**: $\lambda = 62$, $\rho = 56$, $\eta = 2128$, $\gamma = 4251886$, $\Theta = 2070$, $\tau = 2110$) with extra noise parameter $\rho' = \eta - \rho$, θ the hamming weight of vector $\overrightarrow{s}$ is equal to 15, n the precision after the binary point of each

Table 1: First Implementation Results

Parameters	*KeyBankGen*	*Encrypt*	*Decrypt*
small	0.136439000002 *s*	0.00102999999945 *s*	0.000683999998728 *s*
medium	2.732191 *s*	0.00130100000024 *s*	0.00339099999837 *s*
large	48.049091 *s*	0.00596199999927 *s*	0.0216390000023 *s*

Table 2: Second Implementation Results

Parameters	Complex mean execution time	BGV mean execution time
small	0.4102108 *s*	0.35041905 *s*
medium	2.8407822 *s*	2.42358273 *s*
large	27.40187732 *s*	38.52468052 *s*

z_i is taken as 4. Table. 1 shows in terms of execution time the generation of public key bank of τ complex public keys, 1 bit encryption and decryption.

2. **Second Implementation** : In the second implementation, we did the evaluation procedure of the logic circuit $C = Plain_{output} = (b_0 \oplus b_1) \bullet (b_2 \oplus b_3)$, (i.e. $Cipher_{Output} = ((c_0 + c_1) \times (c_2 + c_3)$, where $c_i = Enc(b_i)$, $for\ 0 \leq i \leq 3$) using the two schemes (Complex and BGV) with the same level of security λ. For BGV the three layers of implementation are (**small**: $\lambda = 42$, $n = 42$, $p \approx O(n^{20})$, $q \approx O(\frac{n^{20}}{2})$), (**medium**: $\lambda = 52$, $n = 149$, $p \approx O(n^6)$, $q \approx O(\frac{n^6}{2})$), (**large**: $\lambda = 62$, $n = 370$, $p \approx O(n^6)$, $q \approx O(\frac{n^6}{2})$). Table.2 shows a comparison in terms of mean execution time between the two implementations for 100 iterations (Evaluation procedure for the Compplex-based scheme is done with Bootstrapping mechanism while with BGV is achieved with KS and MS). In addition, results have shown that noise is efficiently reduced for the two schemes. One can see that for large case, the proposed Complex based-scheme performs better that BGV.

4.2 Security Analysis:

The crypt-analysis of the homomorphic encryption schemes will consider the techniques used to build such schemes. The security of the BGV-lattice based scheme relies on the hardness of solving LWE problems [25]. As for our new complex scheme, it uses mathematical operations and the typical attack in this case is General Approximate Common Divisor (GACD). Given a public key bank $PK = \{x_1, x_2, ..., x_\tau\}$ where $x_j = pq_j + r_j$ for $1 \leq j \leq \tau$, the basic idea of this attack is to reveal the value of the secret key p starting from PK. The secret key p can be revealed using different types of algorithms like the approximate GCD of two numbers discussed in [22], the approximate GCD of many numbers using the SDA algorithm [23] by applying a lattice based attack with *LLL* algorithm. Recently a new improved attack was introduced and implemented in [24], [19]. In our crypt-analysis, we tried to apply this new Approximate GCD attack. Given the public key bank $PK = \{x_j, 1 \leq j \leq \tau\}$, where $q_j \in [0, \frac{2^\gamma}{p})$ and $r_j \in [0, 2^\rho)$ are chosen uniformly and independently at random. The algorithm is as follows: For $j = 1, 2, ..., \tau$, let:

$$y_j = \prod_{i=0}^{2^\rho - 1} (x_j - i) \tag{14}$$

Equation 14 shows clearly that p divides the GCD $g = gcd(y_1, y_2, ..., y_\tau)$. To build this attack and depending on the choice of the (q_j, r_j), we will try to find a certain bound B not much larger than 2^ρ that with a high probability, all the prime factors of g except p are smaller than this bound B. The probability that all the prime factors of g except p are smaller than B is done based on [19]: "For every prime $p \geq B$ other than p, not all the x_j's are congruent to one of $(0, 1, ..., 2^\rho - 1) \quad mod p$". This happens with probability very close to $1 - (\frac{2^\rho}{p})^s$. Hence, the probability that all the prime factors of g except p are smaller than B is essentially given by the following Euler product:

$$P_{s,\rho}(B) = \prod_{p \geq B,\ p \neq p} (1 - \frac{2^{s\rho}}{p^s}) \tag{15}$$

Based on [19] and using the usual prime counting function $\pi(x)$ explained in [26], we can demonstrate that 15 converges to some positive number smaller than 1 and satisfies the following lemma:

Lemma 2. *For any $B > 2^{\rho + \frac{1}{s}}$, we have:*
$1 - P_{s,\rho}(B) < \frac{2s}{s-1} \times \frac{2^{s\rho}}{B^{s-1} log B}$

In our simulation, we picked $B = 2^{\frac{s+1}{s-1}\rho}$ and we got a success probability $P_{s,\rho}(B) > 1 - 2^{-\rho}$. The GACD attack is given by the pseudo code of Algorithm 1.

Algorithm 1 *GACD Attack*

procedure p=$GACD_(\eta,\gamma,\rho)$
$p \leftarrow random_prime(0, 2^{\eta})$
$s \leftarrow \rho \sqrt{\frac{s+1}{s-1}}$
$B \leftarrow \lfloor 2^{\rho} \rfloor$
$Fa \leftarrow Factorial(B)$
$x \leftarrow p \times random_integer(0, 2^{\gamma-\eta}) + random_integer(0, 2^{\rho})$
$i \leftarrow 0$
$g \leftarrow 1$
while $i < 2^{\rho}$ **do**
$g \leftarrow g \times (x - i)$
$i \leftarrow i + 1$
end while
$j \leftarrow 1$
while $j \leq s$ **do**
$x \leftarrow p \times random_integer(0, 2^{\gamma-\eta}) + random_integer(0, 2^{\rho})$
$i \leftarrow 0$
$z \leftarrow 1$
while $i < 2^{\rho}$ **do**
$z \leftarrow z \times (x - i)$
$i \leftarrow i + 1$
end while
$g \leftarrow Greatest_divisor_of_gcd(g, z)_prime_to_Fa$
if $\lfloor log_2(g) \rfloor \leq \eta$ **then**
Break
end if
$j \leftarrow j + 1$
end while
return g
end procedure

Due to the limited resources of our machine (CPU: Intel Xeon, E5-2630, 2.40 GHZ, 8 CORES, 128 GB RAM), the proposed GACD attack with the security levels related to λ is not feasible since the polynomial y_j given in 14 is of degree 2^{ρ} with coefficients of size γ and requires a memory of size $2^{\rho}\gamma$ bits. The required size of each level is: **small** : 1.125 Terra Byte, **medium** : 210759 Terra Byte, **large** : 34831286272 Terra Byte, while our machine is only 128 GB RAM.

5 Conclusion

In this paper, we profited from the simplicity of complex numbers properties by proposing a new SWE scheme based on complex numbers. We applied Gentry refresh mechanism to make our scheme FH. We then implemented our new scheme with the BGV using SAGEMath library. As a comparison with BGV, a well known FHE scheme, our new scheme is an efficient homomorphic scheme and performs better than BGV in terms of execution for large implementation. In addition, our scheme is simply based on simple complex operations rather than lattice based cryptography (homomorphic complex multiplication is done without dimension expansion rather than Tensor product) and Bootstrapping can support unbounded circuit depth, while MS used with BGV is limited to some circuit depth. Finally a crypt-analysis based on GACD attack is presented. Future work will consider the implementation of the GACD Attack given in section 4.2 with a more powerful machine in order to evaluate the approximate attack time.

Acknowledgment This work has been partially funded with support from the Lebanese University.

References

[1] R. Rivest, L. Adleman, and M. Dertouzos, "On data banks and privacy homomorphisms," in *Foundations on Secure Computation, Academia Press*, 1978, pp. 169–179. [Online]. Available: http://www.oalib.com/references/14708317

[2] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 26, no. 1, pp. 96–99, Jan. 1983. [Online]. Available: http://doi.acm.org/10.1145/357980.358017

[3] P. Martins, L. Sousa, and A. Mariano, "A survey on fully homomorphic encryption: An engineering perspective," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 83:1–83:33, Dec. 2017. [Online]. Available: http://doi.acm.org/10.1145/3124441

[4] L. Zhang, Y. Zheng, and R. Kantoa, "A review of homomorphic encryption and its applications," in *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*, ser. MobiMedia '16. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016, pp. 97–106. [Online]. Available: http://dl.acm.org/citation.cfm?id=3021385.3021405

[5] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, "Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 108–117, March 2013. [Online]. Available: http://doi.acm.org/10.1109/MSP.2012.2230219

[6] S. Fau, R. Sirdey, C. Fontaine, C. Aguilar Melchor, and G. Gogniat, "Towards practical program execution over fully homomorphic encryption schemes," in *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2013)*, Compiègne, France, Oct. 2013, pp. –. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00917061

[7] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology — EUROCRYPT '99*, J. Stern, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 223–238. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-48910-X_16

[8] M. Nassar, A. Erradi, and Q. M. Malluhi, "Paillier's encryption: Implementation and cloud applications," in *2015 International Conference on Applied Research in Computer Science and Engineering (ICAR)*, Oct 2015, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/7338149

[9] J. D. i Ferrer, "A new privacy homomorphism and applications," *Information Processing Letters*, vol. 60, no. 5, pp. 277 – 282, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020019096001706

[10] J. Domingo-Ferrer, "A provably secure additive and multiplicative privacy homomorphism*," in *Information Security*, A. H. Chan and V. Gligor, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 471–483. [Online]. Available: https://dl.acm.org/citation.cfm?id=744660

[11] K. Hariss, H. Noura, and A. E. Samhat, "Fully enhanced homomorphic encryption algorithm of more approach for real world applications," *Journal of Information Security and Applications*, vol. 34, no. Part 2, pp. 233 – 242, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2214212616303052

[12] K. Hariss, H. Noura, A. E. Samhat, and M. Chamoun, "Design and realization of a fully homomorphic encryption algorithm for cloud applications," in *Risks and Security of Internet and Systems*, N. Cuppens, F. Cuppens, J.-L. Lanet, A. Legay, and J. Garcia-Alfaro, Eds. Cham: Springer International Publishing, 2018, pp. 127–139. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-76687-4_9

[13] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford, CA, USA, 2009, aAI3382729. [Online]. Available: https://dl.acm.org/citation.cfm?id=1834954

[14] G. Craig, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: ACM, 2009, pp. 169–178. [Online]. Available: http://doi.acm.org/10.1145/1536414.1536440

[15] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully homomorphic encryption over the integers," in *Advances in Cryptology – EUROCRYPT 2010*, H. Gilbert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 24–43. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-13190-5_2

[16] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in *Advances in Cryptology – EUROCRYPT 2011*, K. G. Paterson, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 129–148. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-20465-4_9

[17] C. Gentry, S. Halevi, and N. P. Smart, "Better bootstrapping in fully homomorphic encryption," in *Public Key Cryptography – PKC 2012*, M. Fischlin, J. Buchmann, and M. Manulis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–16. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-30057-8_1

[18] J.-S. Coron, A. Mandal, D. Naccache, and M. Tibouchi, "Fully homomorphic encryption over the integers with shorter public keys," in *Advances in Cryptology – CRYPTO 2011*, P. Rogaway, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 487–504. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-22792-9_28

[19] J.-S. Coron, D. Naccache, and M. Tibouchi, "Public key compression and modulus switching for fully homomorphic encryption over the integers," in *Advances in Cryptology – EUROCRYPT 2012*, D. Pointcheval and T. Johansson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 446–464. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-29011-4_27

[20] Z. Brakerski and V. Vaikuntanathan, "Fully homomorphic encryption from ring-lwe and security for key dependent messages," in *Proceedings of the 31st Annual Conference on Advances in Cryptology*, ser. CRYPTO'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 505–524. [Online]. Available: http://dl.acm.org/citation.cfm?id=2033036.2033075

[21] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ser. ITCS '12. New York, NY, USA: ACM, 2012, pp. 309–325. [Online]. Available: http://doi.acm.org/10.1145/2090236.2090262

[22] N. Howgrave-Graham, "Approximate integer common divisors," in *Cryptography and Lattices*, J. H. Silverman, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 51–66. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-44670-2_6

[23] J. C. Lagarias, "The computational complexity of simultaneous diophantine approximation problems," *SIAM J. Comput.*, vol. 14, no. 1, pp. 196–209, Feb. 1985. [Online]. Available: http://dx.doi.org/10.1137/0214016

[24] Y. Chen and P. Q. Nguyen, "Faster Algorithms for Approximate Common Divisors: Breaking Fully-Homomorphic-Encryption Challenges over the Integers," in *EUROCRYPT 2012*, ser. Lecture Notes in Computer Science, D. Pointcheval and T. Johansson, Eds., vol. 7237, IACR. Cambridge, United Kingdom: Springer, Apr. 2012, pp. 502–519. [Online]. Available: https://hal.inria.fr/hal-00864374

[25] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," in *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '05. New York, NY, USA: ACM, 2005, pp. 84–93. [Online]. Available: http://doi.acm.org/10.1145/1060590.1060603

[26] E. Bach and J. Shallit, *Algorithmic Number Theory*. Cambridge, MA, USA: MIT Press, 1996. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-662-04053-9_2

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 39-45 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Low-Cost and Accurate Computational System for Efficiency Measures over Photovoltaic Arrays

Héctor L. Núñez-Ramírez[*,1], Gloria G. Carvalho-Kassar1[1], Freddy C. Brito-Maestre[2], Yaremi I. Gamboa-Maldonado[3], Luis A. Santos-Avendaño[1], Carlelines Gavidia-Toro[4], Orlando Villarroel-Ramos[3], Dino Di Rosa-Ulloa[4]

[1]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Fotónica, 01060A, Caracas, Venezuela.*

[2]*Universidad Central de Venezuela, Departamento de Comunicaciones, 01050, Caracas, Venezuela.*

[3]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Propagación y Antenas, 01060A, Caracas, Venezuela.*

[4]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Electrónica de Comunicaciones, 01060A, Caracas, Venezuela.*

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 14 July, 2019
Online: 06 September, 2019

Keywords:
Efficiency Measures
Photovoltaic Arrays
Software
Two-diode model

ABSTRACT

Great efforts have been shown in the literature to improve the efficiency of photovoltaic (PV) systems. In this workpaper, authors present the development of a computational system based on free software that allows to estimate with an accuracy superior to 99.9 percent the performance of the PV array connected to a current inverter block, from the physical models behind. For this purpose, a solar radiation measures was made in the laboratory, and also voltage, current and temperature tests over the whole system, adding a storage interface and serial data transmission port to drive and process the data from a computer. Once the system was tested and calibrated in laboratory, the I-V experimental curves of the PV system were obtained in real time, in order to define the most important parameters: short-circuit current, open-circuit voltage, maximum current and voltage, in a natural environment. Those parameters were the input data to the computational tool based on Scilab, which allow estimate the performance of PV system described, solving the non-linear equations of the physical models used. The performance results were compared with a similar computational tool developed in Matlab, as well as with a commercial solar simulator used by the manufacturer of PV module. A highly accurate computational system based in Scilab was obtained allows to evaluate the efficiency and the filling factor of an array for different temperature and solar irradiance levels in real time.

1 Introduction

Great efforts have been made to improve the efficiency in electric power generation using photovoltaic (PV) systems, since the advantages in comparison with others alternative energy sources, being generally fixed systems, which do not require inputs or fuels for operate, just sunlight, which geographically is available in most cases.

This workpaper is an extensión of work originally presented in IEEE 2018 Open Innovations Conference (OI) [1], as a contribution on developing software to determine the behavior of photovoltaic arrays, e.g. [2, 3], in order to be able to size correctly a deployment in any cases: residential, commercial, industrial and even farming environments. Most of these software require license payment or the source code can not be modified. Even, the most recent and interesting works based in two-diode model simulation [4] – [9] use Matlab to evaluate the performing of photovoltaic arrays. However, this work shows the results of an alternative computational system, based on two-diode model used for polycrystalline silicon PV cells [10]– [13], that determines the performance of a cell under sunlight irradiance variations and hence the relevance of knowing the efficiency of the device under study. To achieve this goal, the behavior of a semiconductor in sunlight conditions have been considered as well as temperature changes, to determine the current density between the terminals of the cell, generally consisting of a n-type and p-type semiconductor materials, which are brought

[*]Héctor L. Núñez-Ramírez, Fundación Cendit - Caracas 01060A, +582128088275, hectornunez2002@gmail.com

togheter in the well-known n-p junction. The current of carriers under the photovoltaic phenomenon are deduced by means of the equations that describe the physical model that lead us to a non-linear equations system, which will be solved by software using different Scilab functions. The coefficients of equations have been determined through the most important parameters of any photovoltaic array: short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), maximum current (I_{max}) and voltage (V_{max}), all of them determined experimentally in Photonic Laboratory of Cendit. Once the equations are solved, the I-V and P-V curves of the array under test are obtained, and they are compared with the results obtained using Matlab and the solar simulator facility from one manufacturer of the module under study. However, the results obtained include the performance of the DC-AC current inverter, which is part of the system, using a complete bridge configuration, PWM controller and LC filtering at the output. The computer system developed shows not only the I-V and P-V curves of the device under test but also shows the efficiency and the filling factor of whole set.

This workpaper was developed in five sections: Physical and Mathematical Modeling to focuses phenomena and equations according with the photovoltaic arrays used; Experimental Setup to describe the procedures performed in the laboratory to obtain the data required; Software Development with the scope of describe the tools used and the programming developed, as well as the use of the obtained experimental data; Results and Conclusions to analyze, compare, discuss and verify the operation of the computational tool presented.

2 Physical and Mathematical Modelling

The physical properties of a commercial photovoltaic array made of polycrystalline silicon was considered in this case, constituted by a n-p junction of semiconductor materials under the physical fundamentals of cells of this type described in [14, 15]. Isolated, each semiconductor material retains its neutral charge. The holes are the majority carriers in the p-type material and their positive charge is compensated by negatively-charged ionized acceptor atoms, while the electrons are the majority carriers in the n-type material where there is a large concentration of free electrons compensated by positively-charged ionized donor atoms. If both semiconductor materials are placed together, a very large difference in electron concentration occurs that causes an hole diffusion current from p-type material towards the n-type material as well as an electron diffusion current from the n-type material through the metallurgical junction to the p-type material, but the region close to the junction become free of mobile charge carriers, reason why this area is denoted as a depletion or space-charge region. The diffusion generates minority carriers in each region. The gradual depletion of the charge carriers generates a charge zone made up of acceptors and donor atoms, which is no longer compensated by the free mobile carriers. As quasi-neutral regions are denoted the regions around of the depletion zone, where the neutral charge is maintained (see figures 1 and 2).

In the depletion region, under equilibrium conditions and darkness, an internal electric field is formed which forces the charge carriers to move in the opposite direction to the concentration gradient, therefore, the diffusion current continues its flow, while the forces act on the charge carriers, compensate each other, then the diffusion and drift (due to the internal electric field created around the junction) currents that originate from the carriers are compensated, reason why the net current density is zero ($I_{drift} = I_{diffusion}$). When an increase in temperature occurs, electron-hole pairs are generated, provided that thermal energy can undo the bonds, increasing the minority carriers within the material, which causes a current density due to thermal-generation.

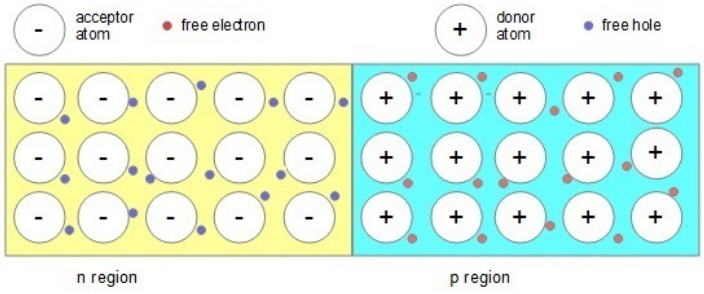


Figure 1: n-p junction (top: before junction).

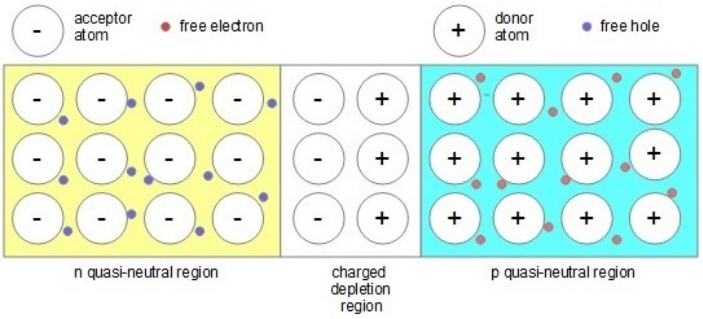


Figure 2: n-p junction (after junction).

The difussion currents (due to minority carriers) into n-type and p-type materials causes recombination currents since the minority carriers recombine with the majority carriers (the electron-hole pairs are undone). The recombination current is compensated by the thermal-generation current ($I_{rec} = I_{gen}$). The recombination strongly determines the performance of the photovoltaic cell, since the current that can be collected by the minority carriers within the semiconductor is reduced.

When the thermal equilibrium is broken, under illumination conditions, due to photons with enough energy (and frequency) to create an electron-hole pair, between the valence and conduction bands of the metallurgical junction, which increases the concentration of the minority carriers, is enabled the flow of holes from the n-type region to the p-type region, and vice versa in the case of the electrons. The flow of the photo-generated carriers determines a photo-generation current density (of intensity I_{ph}), which is added to the current density by thermal generation when the temperature increases within the material.

Based on the above, the behavior of the photovoltaic cell in darkness can be described as a currents balance due to recombinations and thermal generation phenomena, by 1:

$$I_{dark} = I_{01}(e^{qV/kT} - 1) + I_{02}(e^{qV/AkT} - 1) \quad (1)$$

where I_{01} and I_{02} are the saturation current in the quasi-neutral regions and the saturation current in the depletion region, respectively, q is the carrier's charge, k is the Boltzman's constant, T is the temperature, V is the built potential due internal electric field in the junction, A is the ideality factor between both regions. In this case, note that the exponential function is like the Boltzman Approximation, which has been demonstrated in the literature.

On the other hand, under illumination condition, the photogeneration current I_{ph} appears as shown in expression 2:

$$I_{total} = I_{ph} - I_{01}(e^{qV/kT} - 1) - I_{02}(e^{qV/kT} - 1) \tag{2}$$

with:

$$I_{ph} = [I_{sc} - K_i(T - Tr)](\frac{E}{1000\ W/m^2}) \tag{3}$$

where I_{sc} represents the short-circuit current, I_{total} corresponds to carriers net flow between the conducting terminals from the photovoltaic array, K_i is the short-circuit current coefficient, T_r represents the reference temperature and E is the irradiance.

The I-V and P-V curves, current-voltage and power-voltage respectively, can be obtained from equation (2), whose performance is shown in figure 3, where the I_{sc} is pointed at the left extreme when $V = 0$ and V_{oc} in the right extreme at $I_{total} = 0$.

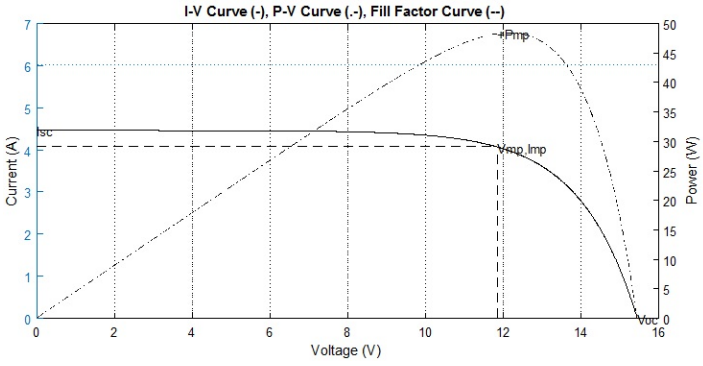


Figure 3: I-V Curve (blue), P-V Curve (red) and Fill Factor (yellow).

When the short-circuit condition is given, I_{max} not could be reached completly, in fact is a little lower due to the high resistivity presents between the ends of the junction, equivalent to a resistance in parallel to the current source and the diodes, denoted as shunt resistance (R_{sh}). In the other side, when the open-circuit condition is given the V_{max} not could be reached due to the ohmic losses in the union of the contacts with the ends of the n-p junction, denoted as resistance in series (R_s) to the current source and diodes.

$$I_{total} = I_{ph} - I_{01}(e^{\frac{q(V-I_{total}Rs)}{kT}} - 1) - I_{02}(e^{\frac{q(V-I_{total}Rs)}{AkT}} - 1) - (V + I_{total}R_s)/R_{sh} \tag{4}$$

The behavior of a single photovoltaic cell is given by the combination in parallel or series of a single photovoltaic cell, modifying expression (4) by the following:

$$I_{total} = N_pI_{ph} - N_pI_{01}(e^{q(V/N_s - I_{total}R_s/N_p)/kT} - 1) - N_pI_{02}(e^{q(V/N_s - I_{total}R_s/N_p)/AkT} - 1) - (N_pV/N_s + I_{total}R_s)/R_{sh} \tag{5}$$

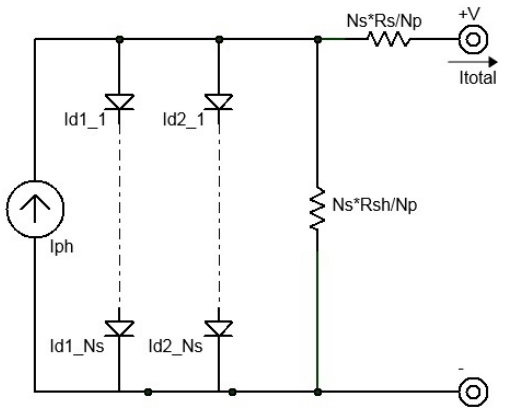


Figure 4: Equivalent circuit for a photovoltaic array.

where I_{ph} is represented by a current source and the current under darkness (I_{01} and I_{02}) by two diodes in parallel, justifying the ideal factor between both diodes denoted as A, N_p and N_s, are the number of cells (from the array) in parallel or series , respectively. The equivalent circuit to (5) is shown in Figure 4.

Using the short-circuit condition, $V = 0$ gives $I_{total} = I_{sc}$, then:

$$I_{sc} = N_pI_{ph} - N_pI_{01}(e_{qI_{sc}R_s/kT}) - N_pI_{02}(e_{qI_{sc}R_s/AkT}) - I_{total}R_s/R_{sh} \tag{6}$$

Meanwhile, using the open-circuit condition, $I_{total} = 0$ gives $V = V_{oc}$, then (5) changes to a:

$$0 = N_pI_{ph} - NpI_{01}(e^{qV_{oc}/N_skT} - 1) - NpI_{02}(e^{qV_{oc}/AN_skT} - 1) - (N_pV_{oc}/N_s)/R_{sh} \tag{7}$$

Combining (6) and (7) gives:

$$NpI_{sc} = I_{01}(e^{qV_{oc}/N_skT} - e^{qI_{sc}R_s/N_pkT}) - I_{02}(e^{qV_{oc}/AN_skT} - e^{qI_{sc}R_s/AN_pkT}) + (V_{oc}I_{sc}R_s)/R_{sh} \tag{8}$$

Differentiating (4) with respect I gives:

$$1 + qN_pI_{01}R_s/kT(e^{q(V/N_s + IR_s/N_p)/kT} + qN_pI_{02}R_s/kT(e^{q(V/N_s + IR_s/N_p)/kT} + R_s/R_{sh} \tag{9}$$

Under the open-circuit condition, $R_{s0} = -dV/dI$, then (9) changes to:

$$(R_{s0} - R_s)(\frac{qN_p}{kT}(I_{01}e^{\frac{qV_{oc}}{N_skT}} + I_{02}e^{\frac{qV_{oc}}{N_sAkT}}) + \frac{1}{R_{sh}}) - 1 = 0 \quad (10)$$

meanwhile, in the short-circuit condition, $R_{sh0} = -dV/dI$, then expression (9) changes to:

$$(R_{sh0} - R_s)(\frac{qN_p}{kT}(I_{01}e^{\frac{qI_{sc}R_s}{N_pkT}} + I_{02}e^{\frac{qI_{sc}R_s}{N_pAkT}}) + \frac{1}{R_{sh}}) = 0 \quad (11)$$

Substituting $I_{total} = I_m$ and $V = V_m$ in (5) and using Iph from (6) gives:

$$I_m(1+\frac{R_s}{R_{sh}}) = I_{01}(e^{\frac{qV_{oc}}{N_skT}} - e^{\frac{q(V_m+I_mR_s)}{N_pkT}}) + I_{02}(e^{\frac{qV_{oc}}{ANskT}} - e^{\frac{q(V_m+I_mR_s)}{ANpkT}}) + \frac{(V_{oc}-V_m)}{N_sR_{sh}} \quad (12)$$

Finally, to determine the values of I_{01}, I_{02}, R_s, R_{sh}, A, has been solved the equations system given by (8), (10), (11), (12), with the following conditions that regards the device under test:

$$R_{s0} = 0\ ;\ R_{sh0} = \frac{V_m}{I_{sc} - I_m} - \frac{V_{oc} - V_m}{I_m} \quad (13)$$

3 Experimental Setup

The values of V_{oc}, I_{sc}, V_{max} and I_{max} for the photovoltaic array setup in laboratory, were acquired from I-V curves gives via experimental, under controlled temperature and solar levels, making the necessary measurements and following what is indicated by flow diagram shown in figure 5.

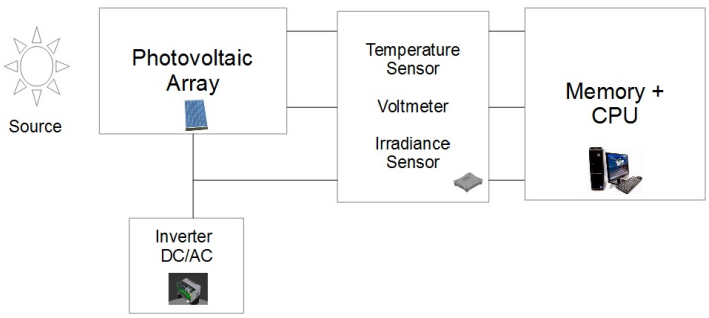


Figure 5: Flow diagram of experimental setup.

In fact, this data was acquired using a device developed in the laboratory based on PIC 18F2550, which has enough memory capacity to operate an SD memory card for data storage, as well as an RS485 interface for trasmission in real time to a computer. The device was based on the scheme proposed in [16]. The device is constituted by three sensors:
- A photodiode to measure the solar irradiance, using the BPW31.
- A temperature sensor, using the LM35.
- A voltmeter, using the ADC PIC module.

To calibrate the solar irradiance sensor, the TES 1333 was used, while for the temperature the infrared thermometer IDEAL 6165 and the multimeter Fluke 179 were used. For the acquisition of data, a script was generated in SCILAB to obtain the data via serial port, in order to plot in real time, for known conditions of irradiance (W/m2) and temperature (^{o} C), I-V curves using a variable load variable (max. 10 A). The figures 6 and 7 show some photos about the experimental setup.



Figure 6: Captures of experimental setup (UNERVEN array).



Figure 7: Captures of experimental setup (SIMENS array).

The technical specifications of photovoltaic (PV) module under test are indicated in table 1: Power (Max.), Voltage (Nominal), Series Cells (Ns) and Parallel Cells (Np).

Table 1: PV arrays under test (both polycrystalline).

Parameter	UNERVEN MSF	SIEMENS PC-2-JF
Power	195 W	35 W
Voltage	24 V	17 V
Ns	48	36
Np	1	1

The I-V curves for several solar irradiance levels during the useful solar day for each model of photovoltaic module under test are shown in figures 8 and 9.

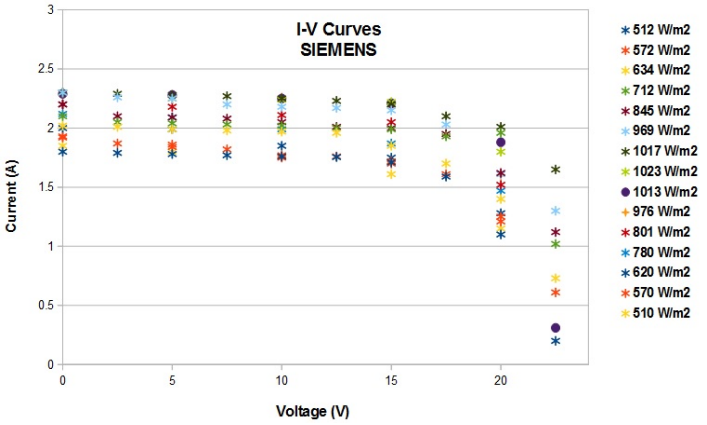


Figure 8: I-V experimental curves for UNERVEN SMF.

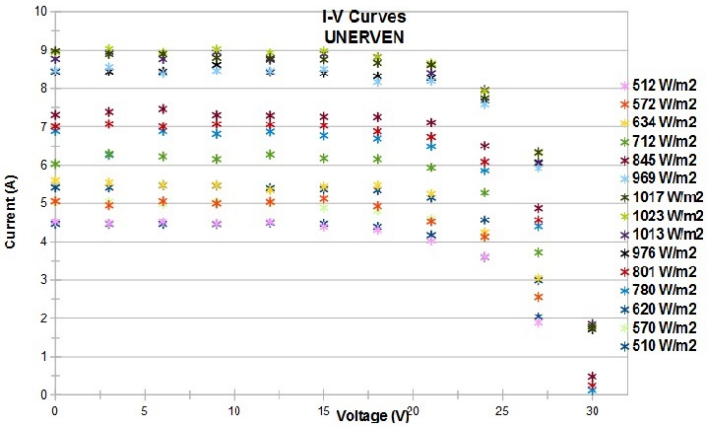


Figure 9: I-V experimental curves for SIEMENS PC-2-JF.

Table 2 shows the values obtained in laboratory for parameters: V_{oc}, I_{sc}, V_{max} and I_{max}, from photovoltaic arrays under test.

Table 2: Experimental values obtained.

Parameter	UNERVEN	SIEMENS
V_{oc}	30.25 V	23.18 V
I_{sc}	8.72 A	2.41 A
V_m	23.76 V	2.12 A
I_m	8.15 A	17.49 V

4 Software Development

A SCILAB script was developed to solve nonlinear equations system constituted by expressions (8), (10), (11), (12), using the function fsolve $[x_0, F]$, with x_0 defined as a vector with the initial values and F defined as a vector that describes equations mentioned before.

First of all, we have to define the equations to solve as a function:

$function[f] = F(x)$
$kB = value;$
$q = value;$
$T = value;$
$Voc = value;$
$Isc = value;$
$Vm = value;$
$Im = value;$
$Vt = value;$
$Rsh = value;$

$f(1) = x(1) * (exp(Voc/Vt) - exp(Isc * x(4)/Vt)) + x(2) * (exp(Voc/(x(3) * Vt)) - exp(Isc * x(4)/(x(3) * Vt))) - Isc + (Voc - Isc * x(4))/Rsh;$
$f(2) = -x(4) * ((x(1)/Vt) * exp(Voc/Vt) + x(2)/(x(3) * Vt) * exp(Voc/(x(3) * Vt)) + 1/Rsh) - 1;$
$f(3) = x(1) * (exp(Voc/Vt) - exp((Vm + Im * x(4))/Vt)) + x(2) * (exp(Voc/(x(3) * Vt)) - exp((Vm + Im * x(4))/(x(3) * Vt))) + (Voc - Vm)/Rsh - Im * (1 + x(4)/Rsh);$
$f(4) = -x(4) * ((x(1)/Vt) * exp(Isc * x(4)/Vt) + (x(2)/(x(3) * Vt)) * exp(Isc * x(4)/(x(3) * Vt)) + 1/Rsh) - 1;$
$end function$

Then, we have to define a script to solve the equations:
$x0 = [min_value; scale_factor; max_value];$
$y = fsolve(x0, F);$
$I1 = y(1);$
$I2 = y(2);$
$A = y(3);$
$Rs = y(4);$

The expression defined by (5) is a trascendental equation that was solved in this case using Xcos from Scilab. Figure 10 shows how the four terms in (5) are determined. For more details see [1].

In the last blocks a DC-AC converter has been included and using a full-bridge configuration and a LC filter in the out, with PWM control. For more details in [1].

The I-V and P-V curves generated in Scilab, under Standard Test Conditions (STC) from the mathematical two-diodes model used for each array under test: SIEMENS and UNERVEN, are shown in figures 11 and 12.

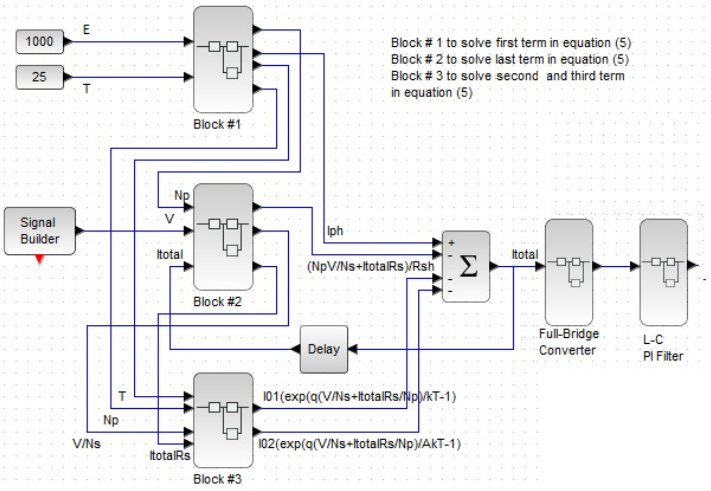


Figure 10: Xcos solution to simulate photovoltaic arrays with DC/AC inverter.

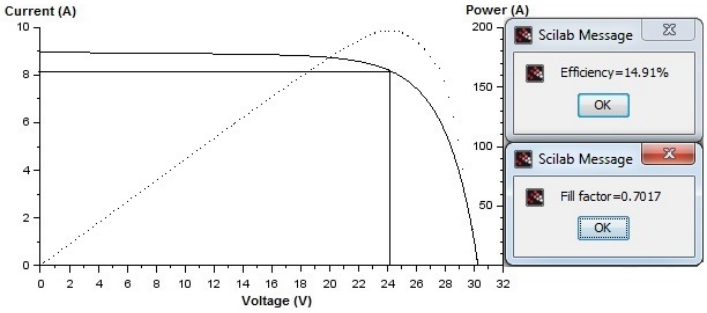


Figure 11: I-V and P-V curves through Scilab (PV array: UNERVEN).

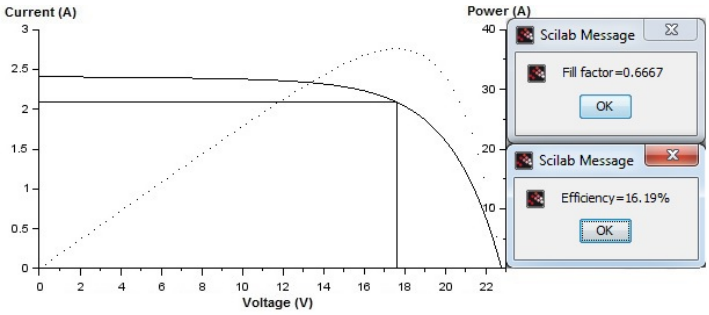


Figure 12: I-V and P-V curves through Scilab (PV array: SIEMENS).

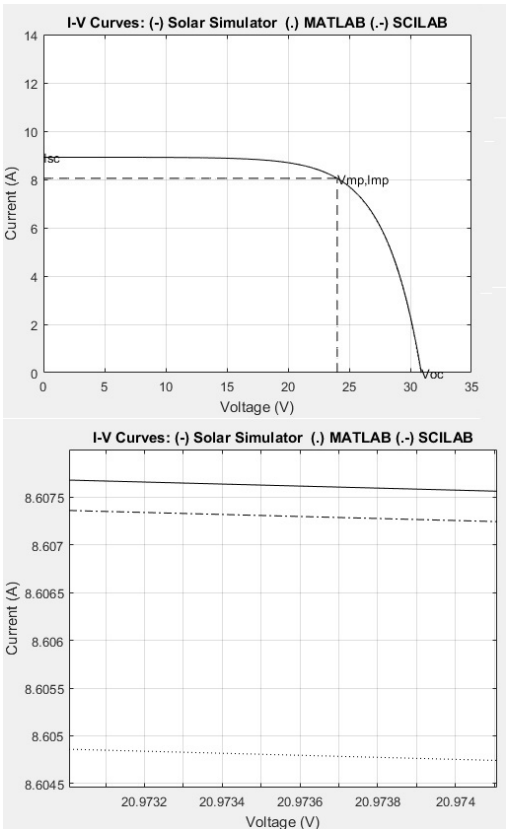


Figure 13: I-V Curves: Solar Simulator, Matlab, Scilab (PV array: UNERVEN).

5 Results

The I-V curves from different sources are shown in Figures 13 and 14: Solar Simulator (from UNERVEN, class A), Matlab Simulator (based on [1]) and Scilab Simulator, for each photovoltaic array under study on STC (1000 W/m^2 , 25 ^oC). The solid line and the dash-dotted line are very close to each other, even if some approximation algorithm was not used to solve the equations of two-diode model as in recent workpapers cited above.

The I-V curves for several solar irradiation levels using Xcos from Scilab for each photovoltaic array under study are shown in Figures 15 and 16. Again, the solid line and the dash-dotted line are very close to each other, with an accuracy superior to 99.9 percent.

6 Conclusions

A computational tool has been put to test in laboratory, based on a free software platform, which allows evaluate in real time the efficiency not only of a photovoltaic array, but also the effects generated by a DC-AC inverter coupled like a load to the system as in real enviroment, using in the develop the physic models cited. Test were coducted in two commercial photovoltaic arrays, and aditionally, the performance results were compared with other tools (licensed) reaching a high accurate.

In contrast with experimental tools used (Solar Simulator Class A from UNERVEN) and computational tool based on Matlab, using Scilab (free software) was improve the perfomance of photovoltaic arrays, even a real time estimation of fill factor and electrical efficiency as an additional feature respect to works cited above. works.

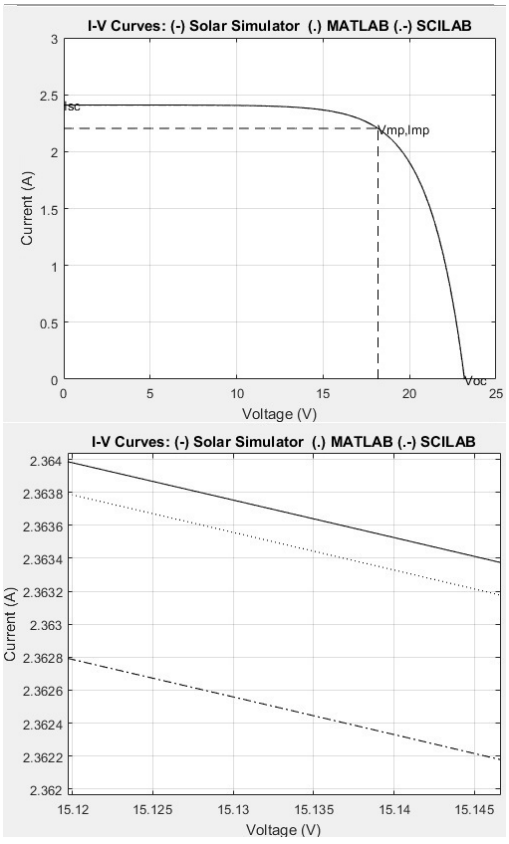


Figure 14: I-V Curves: Solar Simulator, Matlab, Scilab (PV array: SIEMENS).

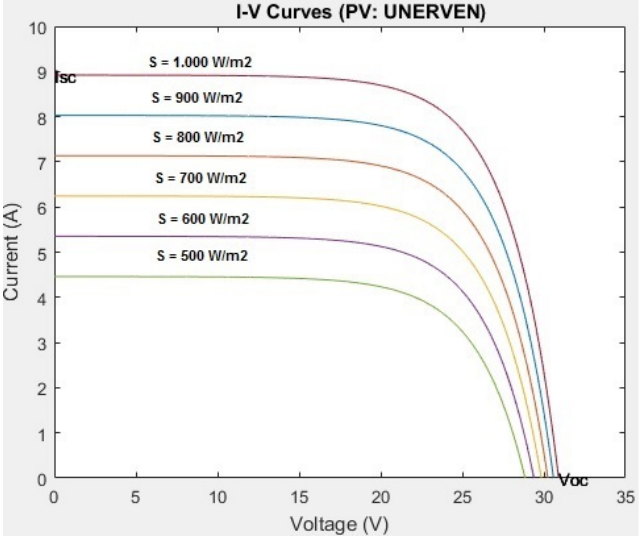


Figure 15: I-V Curves: for several irradiation levels (PV array: UNERVEN).

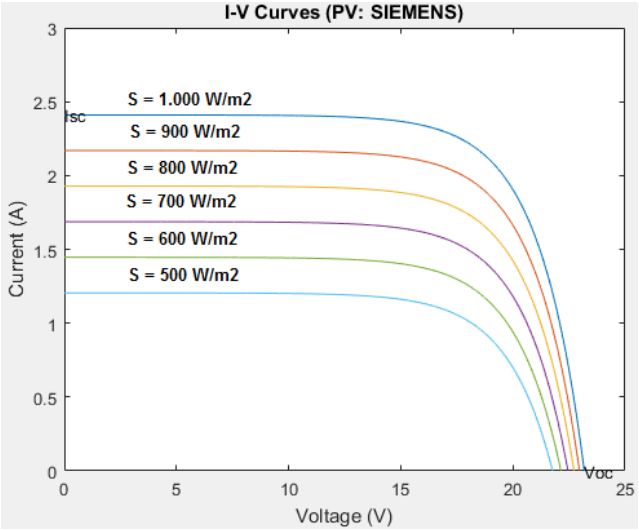


Figure 16: I-V Curves: for several irradiation levels (PV array: SIEMENS).

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank to Ministerio del Poder Popular para Ciencia y Tecnología, Ministerio del Poder Popular para la Energía Eléctrica, Unerven - PDVSA Industrial, professors Humberto Rojas and Carlos Rojas from Facultad de Ciencias de la U.C.V.

References

[1] H. L. Núñez Ramírez, G. G. Carvalho Kassar, F. C. Brito Maestre, Y. I. Gamboa Maldonado, L. A. Santos Avendaño, C. Gavidia Toro,"Highly Accurate and Inexpensive Computational Tool for Improving Photovoltaic Arrays Performance and Efficiency Measures" in Proceedings of the IEEE 2018 Open Innovations Conference (OI), Johannesburg, South Africa, 2018. https://doi.org/10.1109/OI.2018.8535899

[2] S. Chowdhury, S. P. Chowdhury, G. A. Taylor, and Y. H. Song, "Modelling, simulation, performance analysis of a PV array in an embedded environment", in Proceedings of the 42nd International Universities Power Engineering Conference (UPEC), , Brighton, UK, 2007. https://doi.org/10.1109/UPEC.2007.4469048

[3] K. Ishaque, Z. Salam, and H. Taheri, "Simple, fast and accurate two-diode model for photovoltaic modules" Solar Energy Materials and Solar Cells, vol. **95**(2), 586-594, 2011. https://doi.org/10.1016/j.solmat.2010.09.023

[4] N. Hashim and Z. Salam,"Critical evaluation of soft computing methods for maximum power point tracking algorithms of photovoltaic systems" IJPEDS,**10**(1), 548-561, 2019. https://doi.org/10.11591/ijpeds.v10.i1.pp548-561

[5] A. Hemza, H. Abdeslam, C. Rachid and N. Aouan,"Simplified methods for evaluating the degradation of photovoltaic module andmodeling considering partial shading" Measurement, **138**, 217-224, 2019. https://doi.org/10.1016/j.measurement.2019.01.098

[6] P. A. Gbadega and A. K. Saha, "Effects and Performance Indicators Evaluation of PV Array Topologies on PV Systems Operation Under Partial Shading Conditions", in Proceeding of the 2019 Southern African Universities Power Engineering Conference, Bloemfontein, South Africa, 2019. https://doi.org/10.1109/RoboMech.2019.8704823

[7] T. Ma, W. Gu, L. Shen and M. Li,"An improved and comprehensive mathematical model for solar photovoltaic modules under real operating conditions" Energy, **184**, 292-304, 2019. https://doi.org/10.1016/j.solener.2019.03.089

[8] P. Ranjan Satpathy, R. Sharma and S. Dash,"An efficient SD-PAR technique for maximum power generation frommodules of partially shaded PV arrays" Solar Energy, **175**, 182-194, 2019. https://doi.org/10.1016/j.energy.2019.03.078

[9] G. Sai Krishna and T. Moger, "Improved SuDoKu reconfiguration technique for total-cross-tied PV array to enhace maximun power under partial shading conditions" Renewable and Sustainable Energy Reviews, **109**, 333-348, 2019. https://doi.org/10.1016/j.rser.2019.04.037

[10] M. Wolf, G. J. Noel, R. J. Stirn, "Investigation of the double exponential in the current-voltage characteristics of silicon solar cells" IEEE Trans. Electron Devices, **24**(4), 419-428, 1977. https://doi.org/10.1109/T-ED.1977.18750

[11] D. S. H. Chan and J. C. H. Phang, "Analytical Methods for the Extraction of Solar Cell Single- and Double-Diode Model Parameters" IEEE Trans. Electron Devices, **34**(2), 286-293, 1987. https://doi.org/10.1109/T-ED.1987.22920

[12] N. Veissid, A. M. De Andrade, "The I-V silicon solar cell characteristic temperature dependence an experimental study using the standard deviation method", in Proceedings of the 10th European Solar Photovoltaic Energy Conference, Lisbon, Portugal, 1991. https://doi.org/10.1007/978-94-011-3622-8_11

[13] A. Gow, C. D. Manning, "Development of a model for photovoltaic arrays suitable for use in simulation studies of solar energy conversion systems", in Proceeding of the 1996 Sixth International Conference on Power Electronics and Variable Speed Drives, Nottingham, UK, 1996. https://doi.org/10.1049/cp:199608909.03.089

[14] Luque, S. Hegedus, Handbook of Photovoltaic Science and Engineering. Willey, 2nd Edition, 2011.

[15] A. Smets, K. Jäger, O. Isabella, R. van Swaaij, M. Zeman, Solar Energyhe Physics and Engineering of Photovoltaic Conversion, Technologies and Systems. UIT Cambridge, 2016.

[16] M. A. Martínez, J. M. Andújar and J.M Enrique, "A New and Inexpensive Pyranometer for the Visible Spectral Range" Sensors, **9**(6), 4615-4634, 2009. https://doi.org/10.3390/s90604615

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 46-52 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Detecting Malicious Assembly using Convolutional, Recurrent Neural Networks

Michael Santacroce[*,1], Daniel Koranek[2], Rashmi Jha[1]

[1]*University of Cincinnati, EECS, 45221, Ohio, USA*

[2]*Air Force Research Laboratory, Sensors Directorate, 45433, Ohio, USA*

ARTICLE INFO

Article history:
Received: 15 March, 2019
Accepted: 7 June, 2019
Online: 06 September, 2019

Keywords:
Malware Detection
Neural Networks
Image Classification

ABSTRACT

We present findings on classifying the class of executable code using convolutional, recurrent neural networks by creating images from only the .text section of executables and dividing them into standard-size windows, using minimal preprocessing. We achieve up to 98.24% testing accuracy on classifying 9 types of malware, and 99.50% testing accuracy on classifying malicious vs. benign code. Then, we find that a recurrent network may not entirely be necessary, opening the door for future neural network architectures.

1 Introduction

This paper is an extension of the work originally presented in the 2018 IEEE National Aerospace and Electronics Conference (NAECON) [1] .

Stakes are always escalating in the arms race that is malware detection. From personal security to breaches at large data centers, it is important to remain vigilant and always explore new methods for combating malware. Classifying executable code images may soon be one of the best methods for malware detection given all of the advances with neural networks in recent times, and we aim to further push the field by questioning the amount of information necessary and size of network necessary to make a judgment.

With our previous work, we proved in concept that using only the .text section of executable files, and not the data, is a possible avenue for malware detection. Using only the executable code allows for leaner, faster networks at the lowest possible level in the hardware stack. Unfortunately, one large issue with malware images is the extremely variable size of the input - code can be anywhere from minuscule to gigabytes in size so figuring out how to use one network to classify code of any size can be difficult. We previously used padding to standardize all files to one length, negatively impacting performance. In this work, we extend on our previous work by using a "windowing" method to allow for arbitrary-length, minimally processed input that achieves competitive performance in malware classification compared with seminal works. To do this, we use a variant of recurrent neural network called a ConvLSTM [2] that uses convolution in tandem with a recurrent architecture. With further investigation, however, we find that it may be possible to ditch the recurrent relationship altogether, opening the door for a multitude of possible architectures.

2 Related Work

Viewing executable code as an image has become a subject of interest for many studying malware detection and neural networks. We will present some of these works for comparison to our work, to both show the state of research and also how our work has some advantages and disadvantages to other methods. In particular, our method has minimal preprocessing of the images, unlike almost all other methods. To view executable code as an image, each byte of the file is taken and translated directly from hexidecimal to decimal. These values then become greyscale pixels, creating an image. Figure 1 shows an example.

Figure 1 shows the first 100 lines of a malicious image where each line of the executable code is translated to a line of greyscale pixels. We pad each line with 0's to the maximum line length. This format differs from some other formats, especially with the line padding and that we only use the .text section, or executable code, and no data. For many of the malware as image works, the results do not classify malicious vs. benign, but instead classifying which type of malware a file is. This is because obtaining and distributing a large number of malicious files is much easier than doing the same

[*]Corresponding Author: Michael Santacroce, santacml on GitHub, santacml@mail.uc.edu

for benign files. Additionally, it is in some ways a better test of the network to classify the type of malware than malicious or benign, as it can be considered a more difficult type of the same problem. If a network can classify the type of a file out of 10 or more classes of executable code, it should be easier to classify out of only 2 classes, in theory.



Figure 1: Example executable code as an image

In 2010, the authors of [3] used k-NN and feature extraction to achieve 98% accuracy on nearly 9,500 samples of 25 classes of malware. This is one of the earliest works that visualize malware as an image and shows that the concept is viable. The authors note some possible methods of avoiding detection such as inserting redundant bytes. Additionally, we note that this method includes a large amount of preprocessing for the malware images.

In [4], the authors process the images to a standard size and then use a densely-connected neural network to achieve 96.35% accuracy on 3131 files of 24 classes of malware. We note that in this work, excellent results are achieved on such a large number of classes of malware.

More than any other method, the ideas of malware image detection have been combined with the burst of developments with convolutional neural networks (CNNs). CNNs have become the go-to for image classification due to their ability to identify spatial relationships in data [5, 6, 7]. In particular, the Microsoft Malware Classification Challenge has spurred a large interest in using CNNs for malware detection [8]. For instance, the author of [9] used a CNN to acheive 95.24% accuracy on a test set of over 40,000 samples, and then used a residual network to achieve 98.21% accuracy on the same samples. The author also notes that the test set was limited to a certain size of binaries and that future work should expand to any length of binaries. In our own work, which this paper extends, we presented up to 88% accuracy in classifying malware on the same data from the Microsoft Malware Classification Challenge [1] . The largest issue from that work was the large amount of padding used. Instead of reducing or expanding images, we padded all images to be the same length. The upside of this method is that it loses no information, however, there were of course large amounts of padding for tiny images, affecting performance. Like in this work, however, we only used the .text section, proving it possible to only need limited information.

The authors of [10] use a convolutional, distributed network, topped with a recurrent neural network on over 2 million malicious files. The network uses bytes of the full executable without preprocessing and uses and embedding, pointing out the issues with preprocessing. The authors also note difficulties with RNNs due to the length of the sequence, as we also found. The network achieves a maximum of 94% accuracy and 98% AUC. This work is the most similar to ours, however, we expand in a few ways. Most importantly, we are only using the .text section of files, a much smaller subset of the total files. Next, we introduce the concept of breaking the malware image into windows, which would theoretically allow for processing only window at a time, saving large amounts of memory. Next, we apply the ConvLSTM architecture to the data which has not been done before, alleviating RNN issues. We also use global average pooling in one of the final network layers to improve generalization [11]. Finally, we are using a different dataset that is much smaller, but we do achieve better overall performance.

The best performance we could find was by the authors of [12], who combine a CNN and LSTM in ensemble, achieving 99.88% accuracy on 40,000 samples balanced equally between malicious and benign files. Then, when used on only the malicious files, they achieved 99.36% accuracy on 9 classes. Again, the input is the full disassembly of the executables and the input is again preprocessed. Additionally, the authors do not present the method of disassembly for the benign files. We will discuss in the next section some possible flaws with mixing these two types of dataset, of which there may be many.

Executable code as an image has become a hot area of research in recent years. In many of these works the images are preprocessed before classification, which might take a large amount of time or resources in practice depending on the operations, and may be bypassable by reverse-engineering the preprocessing. Additionally, all of the works we have seen use the full executable file for classification. By only using the .text section of an executable, classification can be done at a lower level in the hardware stack, with less memory, less time, and less context of what is running. We will address the issues of using arbitrary-length files by using minimal preprocessing and only using the executable code with a ConvLSTM architecture.

3 Network Input

In our previous work we generated our own data using malware obtained from a Windows installation and live malware found online [1] , number 240 total files split evenly between the two classes. There were a few issues with this dataset, namely, the limited size due to difficulty of acquiring samples and the overly-complex tokenization method. The tokenization of assembly resulted in a larger-than-needed vocabulary size and some preprocessing time which we attempting to limit as much as possible.

As a second test, we used the data from the Microsoft Malware Classification Challenge on Kaggle [8], which contains over 20,000 total malware files from 9 classes of malware. With this data, we are able to directly use the machine code and translate it into decimal per byte, which resembles other works as we have discussed. We will again evaluate performance on this dataset. Of these files, 10,869 had labels and then of those 10,375 were found to have .text sections that we could use, however, we quickly hit memory limits while testing longer files. To solve this, we had to limit to only files

that had under 50,000 lines of code, resulting in a total of 8,550 files, of which we used 6,500 training and 2,050 testing files. Because the final goal is classification of malware vs. binary code, however, we have generated an additional dataset consisting of 4,550 benign files from a default Windows installation that were disassembled using IDA pro. Again, due to the 50,000 line cutoff, we were only able to use 3,100 of these files. We then combined the benign data with randomly-chosen malicious files for a second dataset of only malicious vs. benign classification, consisting of 4,000 training and 1,000 testing files.

We will present results on both of these datasets to validate that our own benign files were not biased because of some concerns with the benign data format. While we are comfortable that the datasets are now fair to compare we will show those concerns and how we addressed them. Many benign files were initially identifiable by eye due to some certain characteristics, as in Figure 2:



Figure 2: First 100 lines of a benign code image

The first problem is that there seems to be a long portion of data bytes at the beginning of many benign files that only includes one or two executable bytes. In the Microsoft dataset, these sections are grouped to have many data bytes per line (up to 18). Next, many lines end in the same color pixel. This is because of how we extract data from the disassembled bytecode - once disassembled from IDA, we simply look at every line to see if it starts with ".text", signifying executable code. Then, we extract the hexadecimal bytes from that line until there are no bytes left. Many of these lines end in the byte "db", a signal from IDA that represents the variable size (a byte) and not code. For some reason these extra "db" tokens were never an issue in the Microsoft dataset as there is always some other non-hexadecimal token before them, causing them to be excluded in our extraction. We therefore excluded the "db" at the end of any line (which is always lowercase if it is not an actual byte in the code but added by IDA, so no innocent bytes were lost). With both of these issues it is possible we are disassembling the data differently than Microsoft did in the Kaggle challenge. Because the exact way that the Microsoft dataset was generated is unknown, we present results on both datasets with the assumption that our benign files are not, but may be, biased.

Unfortunately, malware or benignware can come in all sizes, so the question becomes how to approach classifying arbitrary-length files. Additionally, the files can reach extreme lengths, so a straight LSTM or RNN will experience difficulties in training. We therefore split the files into t windows by grouping every w lines of code. The number of windows forms the new number of timesteps, and the dataset can now be seen as an arbitrary-length sequence of images. An example is shown in Figure 3.

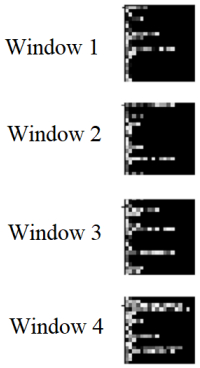


Figure 3: Malware Image Windows

Figure 3 shows a malware image split into windows (not to scale).

4 Network Architecture

Forming the malicious bytecode into an arbitrary-length, windowed format allows for input into a variety of network architectures. Before discussing the network architecture for the new input format, we first downsample and perform an embedding on the data. We previously experienced issues with the size of the input being too large to use at once, and even after purposely omitting files over a certain length, the networks would train in an unreasonably long amount of time. Therefore, we utilized pooling to downsample the input data [1] . While this is the only preprocessing we perform, which adds time to making predictions, we still perform downsampling on the data by randomly discarding 9 of every 10 lines. Along with the benefit of faster training and prediction we find that it does not negatively affect performance and actually improves testing performance in some cases, which we will discuss in the results section.

We also previously used embeddings to assist in feature detection, similar to natural language processing [10, 13], as the magnitude of the bytes have no intrinsic meaning. The embedding serves as a look up table where each singular value maps to a higher-dimensional space of values with less magnitude.

With the initial layer of an embedding, we now expand on the previously static, convolutional network. A natural pick for classification of these videos would be some kind of recurrent neural network given the time-varying nature and seeming temporal dependency of the problem. In fact, one large question we will pose throughout our work is if the windows are actually temporally dependent, in other words - does the previous window affect how we should see the next? Given the nature of the problem, it seems logical that this would be true, and is the assumption we made for beginning to explore architectures.

Convolutional LSTMs are a relatively recent form of recurrent neural network that have cropped up in a few field of study that involve both temporally- and spatially-related data [2, 14, 15]. In

Keras, this layer of a network is called ConvLSTM2d [16], meaning that it performs 2d convolution inside an LSTM. For comparison, A traditional LSTM uses the following equations:

$$\begin{aligned} i &= sigm(h_{t-1} \cdot U_i + x_t \cdot W_i) \\ f &= sigm(h_{t-1} \cdot U_f + x_t \cdot W_f) \\ o &= sigm(h_{t-1} \cdot U_o + x_t \cdot W_o) \\ g &= tanh(h_{t-1} \cdot U_g + x_t \cdot W_g) \\ c_t &= f \cdot c_{t-1} + i \cdot g \\ h_t &= o \cdot tanh(c_t) \end{aligned} \tag{1}$$

Where $W_{i,f,o,g}$ and $U_{i,f,o,g}$ are trainable weight vectors, and c_t and h_t are the hidden state and output, respectively. These equations can of course change and are only an example for comparison.

LSTMs can help alleviate "vanishing gradient" issues when training and therefore allow for training recurrent networks on much longer inputs than a vanilla RNN [17, 18]. We attempted to use a regular LSTM on input mawlware without windows and found that the model would not only take an extraordinarily long time to train, as expected due to the input length, but would not perform well. Even if a decent training accuracy was obtained the results would not translate to testing accuracy.

Recent works, including our own, have shown that classifying malware images responds well to convolution [1] . Therefore, instead of using a LSTM, it is possible to use the Convolutional LSTM (from now on, ConvLSTM). The ConvLSTM consists of nearly the same parts as a regular LSTM except that many dot products are simply switched to being convolution:

$$\begin{aligned} i &= sigm(h_{t-1} * U_i + x_t * W_i) \\ f &= sigm(h_{t-1} * U_f + x_t * W_f) \\ o &= sigm(h_{t-1} * U_o + x_t * W_o) \\ g &= tanh(h_{t-1} * U_g + x_t * W_g) \\ c_t &= f \cdot c_{t-1} + i \cdot g \\ h_t &= o \cdot tanh(c_t) \end{aligned} \tag{2}$$

Where $W_{i,f,o,g}$ and $U_{i,f,o,g}$ are trainable weight vectors, and c_t and h_t are the hidden state and output, respectively.

The ConvLSTM was introduced in 2015 by [2] as a way to capture both spatial and with temporal relations in precipitation forecasting, as opposed to an LSTM, which is only capable of capturing temporal relations. ConvLSTMs have since been used many times to capture spatio-temporal relations in data. In [15], they are used to build on feature extraction in gesture recognition, and in [19], the authors add to the ConvLSTM architecture for traffic accident prediction. More notably, in [20], the authors use a nested ConvLSTM for video classification. Our work can be seen as converting the malware to a video rather than an image, and taking inspiration from these works. In our case, the ConvLSTM captures spatial relations in blocks and functions of code and temporal relations between those blocks.

With the recurrent networks it was found best to use use two hidden recurrent layers, one that returns sequences and a second that does not. When a recurrent network in Keras has the parameter return_sequences set to true, the output of the layer is not just the final output but the output of each timestep, the result being a sequence of the same length. Using this parameter is useful as it allows the stacking of recurrent layers and therefore deeper networks.

While our models that use ConvLSTM found success, which we will discuss in the results section, we also found that they took up a large amount of time to train due to the complexity of the hidden state. Many modified RNNs were attempted by removing or adding various gates to the ConvLSTM hidden state until we tested a recurrent network that simply adds the previous hidden state to the current as so:

$$\begin{aligned} h_t &= x_t * W_t + b_t + h_{t-1} \\ y_t &= \sigma(h_t) \end{aligned} \tag{3}$$

Where σ represents the sigmoid function, W_i and b_i are trainable weight vectors, h_t is the hidden state, and y_t is the output at time t.

This network is about as minimal as a recurrent neural network can get and can be compared to a standard recurrent network where the hidden weight vector is set to all 1s. We call this the "MinConvRNN". In theory, one layer of this network that does acts the same as if each timestep was independently run through a small neural network and then summed afterwards. Because there are two hidden layers used and each timestep is returned for the first, however, the intermediate timesteps returned by the first hidden layer do not resemble a kind of distributed network.

We can perform backpropogation on the MinConvRNN for training through the widely-known procedure of BackPropogation Through Time (BPTT). We will perform BPTT on the MinConvRNN for completeness, however, gradients are calculated symbolically and automatically in our simulations, so these calculations are not stricly necessary. BPTT consists of unrolling the recurrent neuron into t neurons and repeatedly applying the chain rule. We will start by defining all parameters in Table 1, for any arbitrary objective function J.

Table 1: MinConvRNN Backpropogation Parameters

Variable Name	Purpose
x_t	Input at time t
h_t	Hidden state at time t
W_t	Weight vector at time t
b_t	Bias vector at time t
y_t	Output of neuron at time t
σ	Sigmoid Function
δ_t	$\frac{\partial J_t}{\partial y_t}$
J_t	Objective function value at time t

The goal of backpropogation is to find the partial derivative of the objective function with respect to a parameter. In this instance, $\frac{\partial J_t}{\partial W_t}$. Note that W_t for time t is a formality, and $W_i = W_j, \forall (i, j) \in t$, and the same for b_i. We begin by defining the base case of $t = 1$ such that $h_0 = 0$, making $h_1 = x_1 * W_1 + b_1$.

Then, we can define the partial derivative of J with respect to W_t using the chain rule:

$$\frac{\partial J_t}{\partial W_t} = \frac{\partial J_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_t} \tag{4}$$

We can define δ_t to be $\frac{\partial J_t}{\partial y_t}$ to simplify notation as the expression will change for a different J. We can then calculate the other terms directly as:

$$\begin{aligned} \frac{\partial y_t}{\partial h_t} &= \sigma'(h_t) \cdot 1 \\ \frac{\partial h_t}{\partial W_t} &= \frac{\partial x_t * W_t}{\partial W_t} \end{aligned} \tag{5}$$

We leave $\frac{\partial x_t * W_t}{\partial W_t}$ alone to simplify the expression, but we note it is computable. It is also known commonly known that $\sigma'(x) = \frac{\partial \sigma(x)}{\partial x} = \sigma(x) \cdot (1 - \sigma(x))$, and we will use $\sigma'(x)$ for simplification of expressions as well.

The full partial derivative at time t can then be defined as:

$$\frac{\partial J_t}{\partial W_t} = \delta_t \cdot \sigma'(x_t * W_t + b_t + h_{t-1}) \cdot \frac{\partial x_t * W_t}{\partial W_t} \tag{6}$$

Because W_t is the same $\forall t$ as stated earlier, we can compute the partial derivative for W for n timesteps as:

$$\begin{aligned} \frac{\partial J}{\partial W} &= \sum_t \delta_t \cdot \sigma'(x_t * W + b_t + h_{t-1}) \cdot \frac{\partial x_t * W}{\partial W} \\ h_0 &= 0 \\ t &\in [1, 2...n] \end{aligned} \tag{7}$$

For a recurrent neural network where only the output is considered, we could factor out δ_t from the sum. We can also perform a similar process to find the partial derivative for b for n timesteps as $\frac{\partial J}{\partial b} = \sum_t \delta_t \cdot \sigma'(x_t * W + b_t + h_{t-1})$, with the same conditions.

This partial derivative does not look too dissimilar to that of a basic recurrent neuron. The neuron is indeed recurrent in the sense that it is necessary to calculate h_{t-1} in order to calculate h_t, and that the gradient must be propogated starting with the initial output in the same manner. Regardless of these equations, however, the neuron as defined in Eq. 3 does not make intuitive sense. Taking the raw sum after convolution, in theory, might simply lead to an unbounded output in the hidden state that approaches ∞ (or $-\infty$) for all elements as the number of timesteps approaches ∞ , leading to all 1s (or 0s) in the output with the sigmoid activation.

As will be shown in the results, we found the MinConvRNN to largely match the performance of the ConvLSTM2d. The success of the MinConvRNN begs the question if a recurrent neural network is necessary, or in other words, if executable code as an image has a temporal dependency when it comes to classification. As will be presented in results, it is even possible to scramble the order of the input lines and achieve decent performance using the same net trained on ordered input lines. It seems unintuitive at first but a simple explanation is simply that a large amount of code is wrapped up in short functions. Depending on how large the window size is made, it is possible that each window encapsulates even multiple functions. Because the order in which functions are defined largely does not matter, there is not a strong temporal dependence between windows. A different explanation may be that even between windows, temporal dependency does not matter, and that the class of a file is largely determined by the types and quantities of commands issued, and not their order. Future work will need to investigate the temporal dependence of the data more closely as it will greatly affect how we classify malware with neural networks.

Regardless of the type, after the recurrent layers, the output has three dimensions. The first two dimensions can be average-pooled to reduce overfitting [11], and then connected with a final densely-connected layer for the output. The architecture of the network is shown in Figure 4.

5 Performance

We present results on both the Malware Classification Challenge dataset and the Malware vs. Benign dataset as discussed, starting with the Malware Classification Challenge dataset.

Table 2 shows the final results of the ConvLSTM2d and MinConvRNN networks. Both networks were trained until no further performance gains were seen.

Table 2: Results on Microsoft Malware Classification Challenge Dataset

Metric	ConvLSTM	MinConvRNN
Parameter Count	573,634	116,951
Train Accuracy	98.35%	97.28%
Test Accuracy	98.24%	96.39%
Time to Inference	~63.67ms	~23.73ms

In Table 2, it can be seen firstly that both networks performed competitively to other works. The MinConvRNN was purposely limited in size to have the same number of convolutional kernels as the ConvLSTM2d network, being 50 kernels of size 4 in the first hidden recurrent layer, and 25 kernels of size 3 in the second hidden recurrent layer. The network was kept the same size to demonstrate the relatively similar performance. While the ConvLSTM2d network did achieve better results, it has nearly 5 times the total number of parameters. Better results are certainly possible through tweaking the network size, activation functions, etc., but we see these networks as a proof-of-concept for future work that does not rely on recurrent networks given the performance of the MinConvRNN. Additionally, we note that the ConvLSTM2d takes nearly twice as long to make a prediction given an input due to the complexity of the hidden state. If we can cut out the recurrent nature of the network, the time to predict would drastically go down.

Table 3 shows Confusion matrix for the ConvLSTM2d network on the testing data, where the row is the actual class and the column is the predicted class. Of note is that one class in the dataset has an extremely low number of samples, class 5. Neither this network nor the MinConvRNN was able to correctly identify and class 5 samples. This might be addressed by using weighted categorical crossentropy, but again, we are presenting these results as a proof-of-concept.

Table 4 shows the confusion matrix for the MinConvRNN on the testing dataset, where the rows are the actual class and the columns are the predicted class.

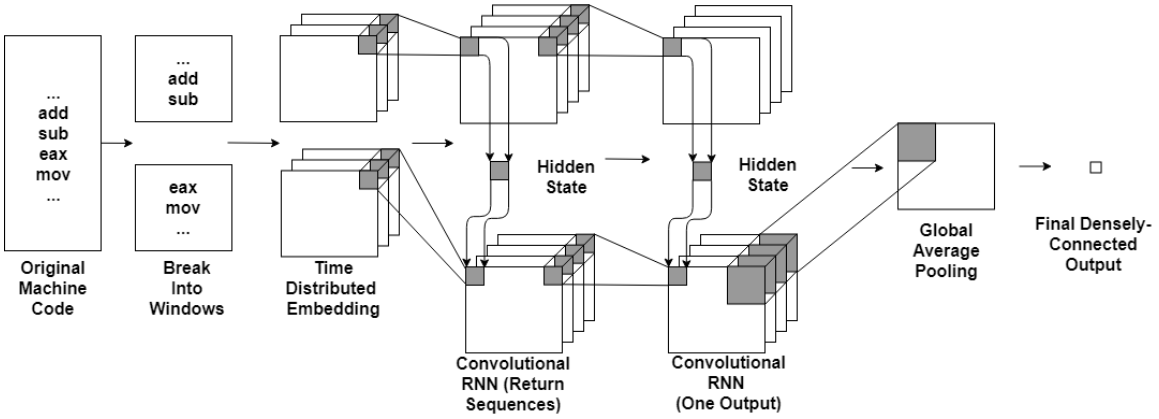


Figure 4: RNN for Malware Classification Architecture (for arbitrary hidden state)

Table 3: Confusion Matrix for ConvLSTM2d Net on Testing Data

	1	2	3	4	5	6	7	8	9
1	249	0	0	0	0	0	0	0	0
2	2	364	0	0	0	0	0	0	0
3	0	0	681	0	0	0	0	0	0
4	0	0	0	99	0	0	0	0	0
5	2	1	0	0	0	0	1	2	0
6	3	0	0	3	0	107	0	0	0
7	0	1	0	0	0	1	89	2	1
8	6	1	0	0	0	1	0	259	1
9	3	1	0	0	0	0	0	4	166

Table 4: Confusion Matrix for MinConvRNN on Testing Data

	1	2	3	4	5	6	7	8	9
1	245	0	0	0	0	0	0	2	2
2	3	356	1	0	0	2	0	3	1
3	0	0	680	0	0	0	0	1	0
4	0	0	1	98	0	0	0	0	0
5	1	0	0	1	0	1	0	2	1
6	1	0	0	4	0	105	0	2	1
7	0	1	12	0	0	1	80	0	0
8	13	2	0	1	0	0	0	252	0
9	4	1	5	1	0	1	0	2	160

Overall, we see performance for both networks, moreso the ConvLSTM2d, matching other works. The benefit to using these networks is the minimal preprocessing, the small size of the networks, and the minimal amount of input by only using the executable code.

Table 5 shows the results of the networks on the Malicious vs. Benign dataset.

Results on the Malicious vs. Benign Dataset are on par or better than the previous dataset. Despite reservations with the benign data as previously discussed we believe that the networks show competitive performance in classifying Malicious vs. Benign data to seminal works. The MinConvRNN again shows worse performance than the ConvLSTM, likely due to the smaller number of parameters. Considering that the network is smaller, the relative success of the MinConvRNN calls into question whether using a recurrent network is necessary.

Table 5: Results on Malicious vs. Benign Dataset

Metric	ConvLSTM	MinConvRNN
Parameter Count	573,634	116,951
Train Accuracy	99.78%	99.32%
Test Accuracy	99.50%	99.20%
Test Recall	99.37%	98.12%
Test Precision	99.58%	99.58%
Time to Predict	~91.16ms	~35.32ms

6 Future work

There are many topics of future work to expand on. In the big picture, the concept of a non-recurrent network would be interesting to test compared to the recurrent networks. It is possible to take more inspiration from video classification methods in this way. We have observed that a minimal recurrent network achieves decent performance, so with optimization, a purely time-distributed network that is aggregated afterwards may perform well.

With or without recurrent, it would be important to optimize the network, exploring activation function, weighted categorical crossentropy, cross-validation methods, and pruning methods. Performing an architecture optimization search would take much more work, but would improve the performance of the network.

There are also some flaws in our methodology, the largest being the limit of 50,000 line long samples. One possible method of removing this constraint is to reformat the data. The figures of malware code show that each line is padded with 0's to be the same length - this was done to preserve the ordering of the opcodes in relation to the data. This format, however, may be unnecessary and the code is often treated as one contiguous block in other works. The size of the input would be greatly reduced by removing the padding, which would not only allow for any length input, but also faster and possibly more accurate predictions. Additionally, it would also

ensure compatibility with the Microsoft and Benign datasets given the different maximum line lengths.

Finally, finding a way to extend the work to hardware would be great for evaluating the feasibility of the practical implementation of the networks. One possible method of using the malware-detecting-network in practice might be to deploy it in a custom chip or FPGA. While the network is likely too large for such acceleration now, with compression techniques, it may be possible to make the network fit. Putting the network in separate, purpose-built hardware would not only free up resources from the main system but also speed up the network considerable.

7 Conclusion

Malware detection will always be a vital part of computer systems. In recent times, classifying malware images made from the executable bytes of a file has become a possible route to take.

This extension of our previous conference paper again proves it is possible to classify malware with nothing other than the .text section of an executable, which contains the actual code that makes up the program. Additionally, we expand on the type of network for detection by creating windows out of the malware image. With this format, minimal preprocessing is required to feed the image into a convolutional, recurrent neural network, achieving competitive performance.

Additionally, we find that using a recurrent neural network may not be entirely necessary. By stripping the ConvLSTM2d of its components and making the MinConvRNN, we have shown that the temporal dependency between windows may not be necessary, opening the doors to new network architectures. Using the MinConvRNN is not the end of the story - it is highly likely that even leaner neural network architectures are possible for malicious assembly.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the University of Cincinnati, Dr. David Kapp at AFRL, WPAFB, OH, and AFRL GRANT No. 1014236 under subaward No. 1919.05.05.91 with TDKC.

References

[1] M. Santacroce, D. Koranek, D. Kapp, A. Ralescu, and R. Jha, "Detecting Malicious Assembly with Deep Learning," in *NAECON 2018 - IEEE National Aerospace and Electronics Conference*. IEEE, jul 2018, pp. 82–85. [Online]. Available: https://ieeexplore.ieee.org/document/8556657/

[2] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, and H. Kong Observatory, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1506.04214v1.pdf

[3] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware Images: Visualization and Automatic Classification," Tech. Rep., 2010. [Online]. Available: https://vision.ece.ucsb.edu/sites/vision.ece.ucsb.edu/files/publications/nataraj{_}vizsec{_}2011{_}paper.pdf

[4] A. Makandar and A. Patrot, "Malware analysis and classification using Artificial Neural Network," in *2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)*. IEEE, dec 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7492653/

[5] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, sep 2017. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/neco{_}a{_}00990

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1097–1105, 2012. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[7] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale Video Classification with Convolutional Neural Networks," Tech. Rep. [Online]. Available: http://cs.stanford.edu/people/karpathy/deepvideo

[8] R. Ronen, M. Radu, C. Feuerstein, E. Yom-Tov, M. Ahmadi, and M. . Crowdstrike, "Microsoft Malware Classification Challenge," 2018. [Online]. Available: http://doi.acm.org/10.1145/2857705.2857713,

[9] A. Singh, "Malware Classification using Image Representation," Tech. Rep., 2017. [Online]. Available: https://pdfs.semanticscholar.org/5bba/bd377616bd446a194f432a12e98326f5c637.pdf

[10] E. Raff, J. Barker, J. Sylvester, R. Brandon, B. Catanzaro, and C. Nicholas, "Malware Detection by Eating a Whole EXE," oct 2017. [Online]. Available: http://arxiv.org/abs/1710.09435

[11] M. Lin, Q. Chen, and S. Yan, "Network In Network," dec 2013. [Online]. Available: https://arxiv.org/abs/1312.4400

[12] J. Yan, Y. Qi, and Q. Rao, "Detecting Malware with an Ensemble Method Based on Deep Neural Network," *Security and Communication Networks*, vol. 2018, pp. 1–16, mar 2018. [Online]. Available: https://www.hindawi.com/journals/scn/2018/7247095/

[13] T. Schnabel, I. Labutov, D. Mimno, and T. Joachims, "Evaluation methods for unsupervised word embeddings," Tech. Rep., 2015. [Online]. Available: http://www.cs.

[14] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, apr 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7558228/

[15] G. Zhu, L. Zhang, P. Shen, and J. Song, "Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM," *IEEE Access*, vol. 5, pp. 4517–4524, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7880648/

[16] F. Chollet, "Keras Documentation." [Online]. Available: https://keras.io/

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, nov 1997. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735

[18] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, oct 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7508408/

[19] Z. Yuan, X. Zhou, and T. Yang, "Hetero-ConvLSTM," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*. New York, New York, USA: ACM Press, 2018, pp. 984–992. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3219819.3219922

[20] C. Lu, M. Hirsch, and B. Schölkopf, "Flexible Spatio-Temporal Networks for Video Prediction," Tech. Rep. [Online]. Available: http://openaccess.thecvf.com/content{_}cvpr{_}2017/papers/Lu{_}Flexible{_}Spatio-Temporal{_}Networks{_}CVPR{_}2017{_}paper.pdf

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 47-50 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

The Study of PSi & PSo Algorithm for Reducing Power of the Mobile Communication Network

Jongsin Kim[*,1], Yonggil Choi[2], Younghoon Oh[2]

[1]*Department of IT Policy and Management, Soongsil University, 06978, South Korea*

[2]*SK Telecom, Western Infra Engineering Team, 62355, South Korea*

ARTICLE INFO

Article history:
Received: 09 April 2019
Accepted: 26 June, 2019
Online: 11 July, 2019

Keywords:
Power Saving
Energy Saving
4G/5G
LTE
Algorithm

ABSTRACT

Recently, the trend of mobile communication network in Korea has been dramatically increasing the number of operating equipment for managing the explosion of wireless data traffic in 4G LTE network. This phenomenon has caused an increase in the operating cost to the mobile telecommunication service provider. Especially, the study about the reduction of the operating cost through the reduction of the power consumption in the power consumption area and the efficient operation method are under study. Through these studies, we consider the method to maintain the best service quality of mobile communication subscriber while minimizing the power consumption of LTE equipment, find out PSi & PSo algorithm that can realize this and verify it in commercial network. Applying this algorithm to a commercial network will save both power consumption and service experience of mobile subscribers to the best level, and it will be more effective in the 5G era that has already begun.

1. Introduction

This paper is an extension of work originally presented in *2018 International Conference on Information and Communication Technology Convergence (ICTC)* [1].

The world has already begun to enter the 5G world, and the first 5G signal in Korea has started to be transmitted. There are also many killer services available to satisfy customers' needs. This means that a lot of people are using smartphones and are getting into daily life. In the past, the use patterns of voice-based usage changed with data, and the network that accommodates customers changed quickly. As the LTE network matured, the frequency bandwidth and the number of equipment services are continuously increasing to accommodate the increasing subscriber traffic [2].

This provides a pleasant environment for various services and accommodates a growing number of subscribers, but the cost of operating of equipment has increased. SK Telecom's electricity costs account for about 30% of the total cost of operating a network, exceeding 30% in 2017, which is still rising. In 2019, electricity costs are expected to account for 35% of total operating costs.

This increase in network operating costs is bringing considerable stress to mobile telecom operators, and it is also evolving to a level that significantly affects corporate EBITDA [3]. As a result, each company has a great interest in reducing the cost of network operation, especially in the field of electric power. In previous mobile communication areas, there has been much interest in battery efficiency aspects such as SoC (System on Chip) upgrade [4] and C-DRX(Connected Mode Discontinuous Reception) [5]. Recently, attention has been focused on the reduction of power consumption of network equipment itself.

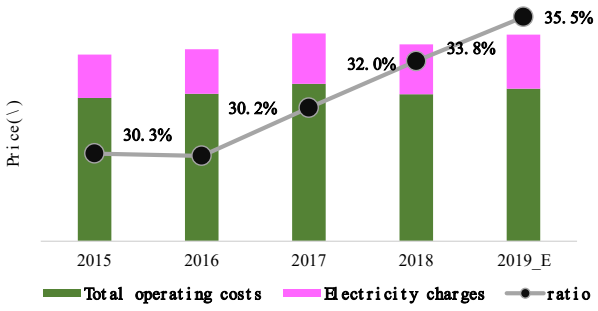


Figure 1: Electricity charges and ratio in operating costs

We used the power saving function of SK Telecom's NOKIA equipment to investigate how to reduce the power consumption of the network equipment itself. Through the derivation of optimal values of PSi and PSo, which are control variables for triggering

*Jongsin Kim, +8210-3768-6309, Email: kjs3838@sk.com

power saving to be explained later in detail, Consumption is decreasing. The optimal PSi and PSo values mean the point where the quality of the mobile communication subscriber can be maximized while minimizing the power consumption.

In this paper, we propose an operation algorithm that can reduce the power consumption of wireless network service equipment while providing optimal service without inconvenience to mobile communication subscribers. In this process, SK Telecom's commercial network is divided into several cases, and results are derived. The power saving function of the wireless network service equipment and the algorithms discussed in this study are used to reduce the actual power consumption.

2. Basic direction of power saving of LTE wireless network

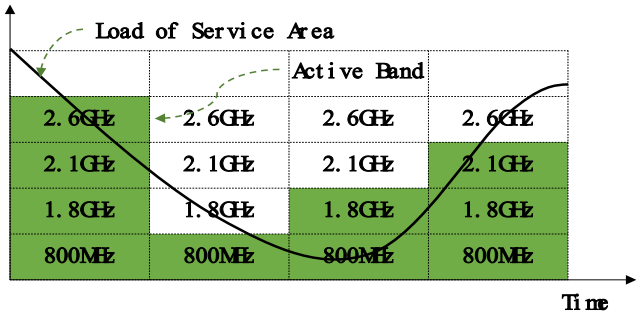
Table 1: Number of subscribers per 10Mbps

Freq. Combination	Number of eligible subscribers
800M	7.5
800M+1.8G	22.5
800M+1.8G+2.1G	30.0
800M+1.8G+2.1G+2.6G	52.5

Table 1 shows the number of subscribers per frequency configuration operated by SK Telecom. As shown in the table, the number of subscribers that can be accommodated for each frequency combination is linear. (However, in LTE network, the service required for voice, messaging, video, and FTP is different according to service type, but it is calculated based on 10Mbps speed for convenience.) In order to service many subscribers, more frequency and bandwidth are needed. On the contrary, it is not necessary to provide the frequency and bandwidth of mobile communication subscribers in every place. This is because the number of subscribers and traffic vary with time. In other words, the frequency and bandwidth need to be tailored to the region's maximum traffic, but not all-time full-band service. Taking operational efficiency in accordance with the traffic volume and characteristics of subscribers is a basic direction for power saving in LTE networks.

In order to reduce power consumption in LTE mobile communication, there is a method to utilize a SON (Self Organizing Network) [6] implemented by a mobile communication provider. As mentioned above, this study utilized the power saving function provided by NOKIA equipment.(Other vendors, however, offer similar functionality.) The basic concept of power saving is that when the same area is served by several frequencies, the subscriber traffic of the service area is judged as a combined load, and when this value is low, the power consumption is reduced by turning off the output of some frequencies.

As can be seen in Figure 2, when several frequencies are serviced in the same area, the number of subscribers and traffic varies with time, and thus the number of frequencies required varies. Typically, however, mobile operators have been providing full-band services in their area, which means that unnecessary power consumption is still happening. One thing to notice is that traffic loads do not change over time in all locations. It should be noted that the traffic varies according to the characteristics of the area and the subscriber, and thus the service quality of the mobile communication subscriber may be changed by such a change. In addition, it is necessary to avoid the factor that hinders the service quality of the mobile communication subscriber in this process.



3. PSi & PSo Algorithm

It is an ideal point of execution of power saving to minimize power consumption without affecting service quality of mobile subscriber. If you do not mind this ideal goal and run power saving, it is very likely that you will be shifting to negative side. When the service quality of the mobile communication subscriber is considered, the power saving level is lowered. On the contrary, if the power saving level is increased, the service quality of the subscriber is lowered. This means that when running power saving, you should not have to break down either your service quality or your power consumption. In this study, we have developed an algorithm that derives the optimal value between service level and power saving through three step approach.

① Power saving operation control variable regulation

② Subscriber service impact index calculation

③ PSi & PSo algorithm exploration

In order to find a clear power saving algorithm, we first define the control variables that can control the operation and clarify the input variables that the operator can set and control the algorithm. Then, in the step of measuring the change of the result, it is necessary to find an index capable of quantitatively judging the influence of the service of the mobile communication subscriber. After excavation, the amount of change must be analyzed through tracking and observation. After the analysis, we set the design direction of the algorithm by considering the determinants of the service impact. Finally, it is necessary to find a point that can maximize the power saving level while ensuring service quality for mobile phone subscribers. The PSi & PSo algorithm established and stipulated these step - by - step core requirements and repeated the process of setting and analyzing each variable and outcome indicator.

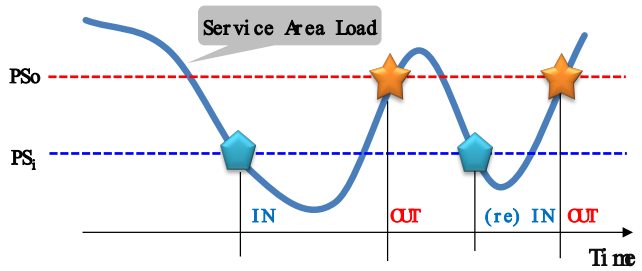


Figure 3: Power saving operation control variable

3.1. Control variables related to power saving operation

First, in order to control the power saving operation, PSi, which is the power saving operation start step and PSo, which is the step of canceling the power saving operation, are defined based on the total load value of the service area.

As shown in Figure 3, the load on the service area changes over time. There is no difference in the fact that there is a difference in the number of times the load increases and decreases in each service area, but changes with time. When the load of the service area falls below a certain level in consideration of this variability, the power saving mode is entered (IN), and the value of this moment is defined as PSi. Conversely, when the load of the service area rises to a certain level or more, power saving mode is released (OUT), and the value of this moment is defined as PSo. When the power saving mode is entered (IN), the equipment of the specific frequency is turned off, and the remaining frequencies that are not off are able to handle the traffic of the subscriber, and power consumption is reduced from this moment. When the subscriber and traffic increase and the load is increased at the moment, it is necessary to turn on the frequency that was turned off to secure the proper capacity. Is turned on. From this moment on, power consumption is not reduced. Two variables, PSi and PSo, allow specific frequencies to be turned on or off to accommodate changing traffic in the service area. However, frequent entry (IN) and release (OUT) to the power saving mode may occur depending on changing traffic and load, which negatively affects the service quality of the mobile communication subscriber. In this paper, we try to find a way to maximize the reduction of power consumption through power saving operation and to minimize the influence of service quality of subscribers in the process.

3.2. Definition of subscriber service impact indicators

Although there are various indicators for calculating the service quality in the mobile communication network, it is necessary to adopt the connection rate (%) and the truncation rate (%) that can comprehensively judge the quality of voice and data service. A new index CEI (Customer Experience Index) is calculated by multiplying two indicators and calculated by the following formula.

$$\text{Call Connection Rate}(\%) = \frac{Number\ of\ Connection\ Successful}{Number\ of\ Connection\ Attempt} \times 100 \quad (1)$$

$$\text{Call Drop Rate}(\%) = \frac{Number\ of\ Call\ Drop}{Number\ of\ Connection\ Successful} \times 100 \quad (2)$$

$$CEI(\%) = Call\ Connection\ Rate \times (100 - Call\ Drop\ Rate) \quad (3)$$

Equation 1: Estimation of the calculated index

Let's look at Equation 1. As shown (1), it is possible to determine the success rate when the mobile communication terminal attempts to connect to the wireless network for the first time, and the truncation rate can be determined based on the connection rate of the mobile communication subscriber. It is possible to judge the situation in which the connection is released irrespective of the will of the user. (1) and (2), the environment of the service to be connected to the first wireless network, and the maintenance level of the connected state after the connection, and calculates the instant when the mobile communication subscriber attempts to connect to the network through the CEI of (3). It can be used as an indicator to represent the quality from the moment to the end.

3.3. PSi & PSo algorithm

PSi and PSo, which are the control variables of power saving, are classified into three phases (high, mid, and lower) and the conditions for each stage are defined as follows.

Table 2: Understanding of PS control variable condition

Division	Step	Condition
PSi (Entry Variable)	High	Difficult to Entry
	Mid	Normal to Entry
	Low	Easy to Entry
PSo (Release Variable)	High	Easy to Release
	Mid	Normal to Release
	Low	Difficult to Release

Combining the three stages of PSi and the three stages of PSo, it can be combined the following nine cases. PSi & PSo value means integrated load of one site.

Table 3: Test Case by PS Control

Division	PSo(High): 30	PSo(Mid): 40	PSo(Low) : 50
PSi(High): 20	Case1	Case2	Case3
PSi(Mid): 30	Case4	Case5	Case6
PSi(Low): 40	Case7	Case8	Case9

Each case was tested in the commercial network of Namwon and Jangsu in Korea, which provides CA service of SK Telecom. The results of changing the PSi and PSo are summarized by the above CEI and the time when the equipment is off and not operated, and the cumulative calculated power saving value is added. Also, the values of PSi & PSo set in each case are arbitrarily set, and they are set to a wide range in order to derive the optimal value.

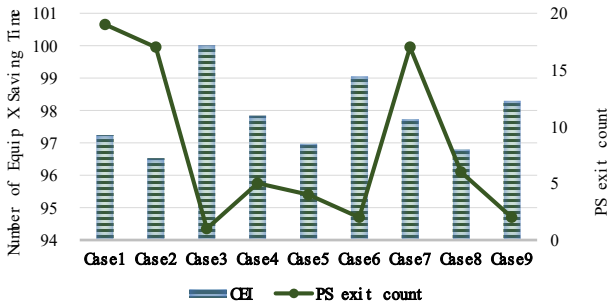


Figure 4 : Power saving exits and CEI by case

As shown in Figure 4, it can be seen that the CEI value varies depending on the case under test. And it can be clearly seen that it is inversely proportional to the number of times the power saving

mode is abandoned. In other words, the more frequent the entry and exit into the power saving mode, the more negative the service quality of the subscribers. Conversely, the lower the number of exits (the longer stay in power saving Mode), the less negative impact on the service quality of the subscriber can be interpreted. Case # 3, 6, and 9 had the fewest number of exits, and the service quality of the subscriber was maximized. As a result, it can be seen that adjusting the PSo to make it difficult to exit in power saving mode and reducing the number of exits has contributed to the optimum result preserving the CEI value.

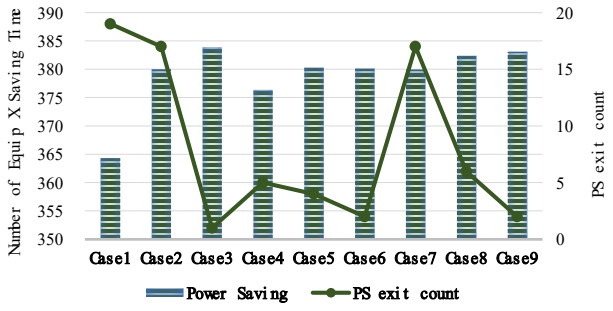


Figure 5 : Power saving level according to the number of power saving mode exits per case

The power saving level can be interpreted as PSi, which determines the entry, rather than the PSo condition, which interferes with the mobile communication subscriber experience when interpreting Figure 4 and Figure 5. In Case # 1, 2, and 3, which are difficult to enter, the power saving level is low and the condition of entry is high in Case # 7, 8, and 9.

3.4. Expanded application of PSi & PSo algorithm

The optimal values of PSi and PSo were selected as 20 and 50 respectively in SK Telecom. NOKIA developed a function to monitor the power consumed by mobile communication service equipment in real time. As a result of using this function, it is possible to save 0.12kW when 1 hour off of 2.6G equipment. We extended it to the commercial network in 2019 and saved about $ 93,000 over five months. Also, it is confirmed that CEI, which is an indicator of service quality, is maintained within the normal range.

As mentioned above, the lower the number of exits to the power saving mode, the lower the service quality of the subscriber can be prevented. As a result of applying power saving to 2.6G equipment in Jeju region of Korea, we found a device with maximum mode change count of 66 times. The site has a lot of tourists, and the traffic load is very volatile. There are two options for these areas. In order to maintain the service quality of the subscriber, it is necessary to abandon the power saving or set the PSi value low to make it difficult to enter the power saving mode (IN), thereby performing the power saving only for the necessary time. The author chose to give up power saving in the area. It is also important to reduce electricity consumption by reducing electricity consumption. However, it is because the company regards the service quality of its subscribers as more important. If one day the company's policy changes and minimizes operating costs become more important, it can turn into a situation where power savings can be applied to these areas and the service quality degradation of subscribers can be tolerated to some extent.

SK Telecom will continue the power saving in order to reduce the electricity cost to operate the network. However, it considers the quality preservation of service subscribers as the top priority and repeats the process of continuously checking the PSi & PSo algorithm and deriving the optimum value will be.

4. Conclusion

The PSi and PSo algorithms for the power saving method that maximize the reduction of power consumption while maintaining the service quality of the mobile communication subscriber are discussed. In order to derive the results, we have divided into 9 cases and the optimal value of PSi and PSo in the area managed by the author can be derived as PSi = 20 & PSo = 50. This number will vary depending on the traffic conditions in each region served. We concluded that we can adjust the power saving mode deviation to minimize PSo and access the ideal result of power saving execution if the PSi-controlled entry operation is easily set up. Although this study investigated algorithms for power saving, we did not discuss the various ISSUEs that can occur when applying this algorithm to commercial networks and the values that should be prioritized by each company's policy. This is because it depends on the value and philosophy of the mobile service provider and operator adopting this algorithm. However, the load fluctuation differs according to the service area and the load level that can be tolerated during the power saving execution may be different between the mobile communication company executing the same and the organization within it. Thus, the task of identifying the optimal value for the demand level is left as an additional research area. In addition, this approach has a weak point that it can partially obstruct the service quality of the mobile communication subscriber during the process, but it has a strong point that it is possible to draw conclusions that can be immediately expanded through direct execution in the commercial network.

Based on the results of this study, the operators who are interested in the reduction of the electricity cost among the network operation cost will be able to execute immediately because they have obtained the initial value and the method that maximizes the power consumption reduction while minimizing the quality deterioration of the subscriber.

References

[1] Kim, Jongsin, Yonggil Choi, and Younghoon Oh. "The Study of PSi & PSo Algorithm for Reducing Power of the Mobile Communication Network." 2018 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2018.

[2] 3GPP TR 36.822, LTE Radio Access Network (RAN)enhancements for diverse data applications (Release 11)

[3] Korea Communication Commission, Jung Woo-su, Kim Seung-gun. A study on "Research on the ripple effects of telecom operators' investment activities". Research on broadcasting communication policy, A Reserch report.

[4] Hyun suk-bong, Kang sung-won, Um nak-woong. Low Power SoC Technology Wireless Terminals. Electronics and telecommunications trends, Vol.23 No.6

[5] Kim tae-yong, Cho gyu-seong, Jo sung-ho. A Study on CDRX operation of eNB for battery saving in LTE smartphone. The Journal of Korean Institute of Communications and Information Sciences.

[6] Aderemi A. AtayeroOluwadamilola I. AduAdeyemi A. Alatishe, “Self Organizing Networks for 3GPP LTE”

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 47-52 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Walker Ergonomic Design for Visually Impaired Elderly

Fang-Lin Chao*

Chaoyang University of Technology, Department of Industrial Design, Taichung, 436, Taiwan R.O.C.

ARTICLE INFO

Article history:
Received: 30 March 2019
Accepted: 28 April, 2019
Online: 13 May, 2019

Keywords:
Design
Walker assistants
Pressure
Ergonomic

ABSTRACT

The study proposed a self-driven guidance walker through user-oriented integration design. The common two-wheel drive performs left and right turn. By reducing the weight of the walker body, the durability of the battery can be increased. Secondly, we use a foldable structure so that the walker can be conveniently stored in a small space. A small number of sensors are required to reduce the complexity of signal processing. The front sensors of the walker are infra-red and ultra-sonic detectors, which determine and evade obstacles. The walker handle has three to five force sensors. The combination of the pressure data from these sensors can be used to determine the intent of the user, like turning or stopping.

1. Introduction

With the significant improvements in health science; the population of the senior citizens over 65 years old will be 15% by the year 2025 in Taiwan. Visually impaired persons often encounter difficulties in living an independent life. They need assistive devices to reduce visual obstacles.

Most of the seniors prefer to stay at home. When visually impaired people are unable to complete activities in their lives, they cannot participate in social activities. The assistance is vital for them to participate in ordinary life. Postmenopausal changes decrease the ability for visually impaired women to maintain postural stability [1]. Elderly blind persons had a slower walking speed, shorter stride length and longer time of gait [2].

In [3], the author addresses the motivation of elders to initiate exercise. The elderly are less active than the general population. Fear of injury is a common concern. Therefore, an exercise environment and regimen must be safe. Gradual exercise progression, increasing duration and frequency is significant. People should have the ability to exercise independently [4]. They should be able to maintain a sense of safety throughout the exercise.

1.1. Smart environment

Wireless technology evolved in recent years, designers can bring it to people with accessibility and ergonomic thinking. Many products found are based on a technology perspective and are less conceived from a user perspective. Article [5] suggests a way of deploying ubiquitous computing to cope with these difficulties. A friendly environment reveals itself by talking to the user with a ubiquitous module prototype.

Smart cane [6] offers voice and vibration alert. There is an antenna at the front of the guide stick to sense the RFID tag in the room. The guide wand transmits the signal to the blind person. Sometimes the contents of the tag cannot be read for a fast moving stick.

1.2. Active assistants

Active assistive walker usually has navigation ability. CMU developed robotic assistant in 2002 which lead the users to the destination. An active assistant gather environmental information, achieve obstacle avoidance with sensors. More sensors could increase the processing load and system cost.

An extended Kalman filter is applied to mobile robot navigation in a public environment which relies on geometric beacon [7]. Leonard et al. [8] used the geometric beacons for mapping and localization. The depth camera's data provides indoor localization. Received signal strength (RSS) constructs the radio maps through different data sets and the environment's coordinates [9].

Another study of Tsukiyama [10] proposed a monitoring system of solitary elderly based on urination, kitchen work, and physical cleanliness. The system monitors health status by using water flow sensors attached to faucets. Semi-autonomous control provided simple access through a handheld device. In case of problems, the robot contacts informal caregivers using teleoperation [11]. Another study [12] predicts the possible

* Fang Lin CHAO, Email: flin@cyut.edu.tw

movement according to the output of the force sensor, the direction of motion conveyed by varying the center of mass elevation. A semi-passive walking robot designed using antagonistic actuation [13] with two pairs of legs, parallel links, and one torso. The robot could walk on horizontal surfaces while hung on the bar to prevent falling. The toe length is sensitive with lateral distance, and gait pattern of robot locomotion [14]. Gait function can be used as reference data in the control process. A passive exoskeleton developed with a clutch-spring mechanism and calf muscles. The engaging and disengaging timings were optimized through simulation for extending knee joint application [15].

The active assistants face the challenge of a trade-off between computing power and weight. The pursuit of the function increases the power consumption. Excessive functionality also makes the user unable to learn or unable to afford it. Heavy parts and large battery are also challenges for this to be implemented in living situations.

2. Concept design and prototyping

We proposed the concept design of an active walker with less weight [16, 17] to fulfill the need of the elderly. Smartphones have remote access and communication ability through the USB or Bluetooth connection.

2.1. Design process

Many of the previous designs focused on special features. From the user interviews, we found that such a model is large in size and complicated to construct, which makes it difficult to use. Complex functions also result in increased power consumption and battery drain. Here, conceptual design proposed from the perspective of human factors with an appropriate function, size, and weight.
Figure 1 shows the design process, and user reflection was collect with ergonomic consideration. The design process set a priority of design goal and tradeoffs of product concepts to confirm feasibly.

The digital model creates a prototype 3D printing which utilized in user evaluation. Electrical parts and interface module are placed with mechanical elements. An existing smartphone app has a built-in position sensor and voice notice. The measured data and user's feedback results are collected to adjust the pressure sensing parameters to acquire better control.

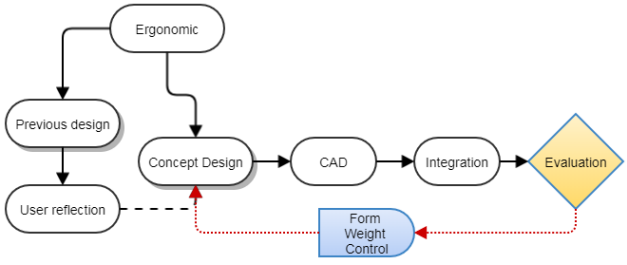


Figure 1: The design process of the people-oriented electromechanical integration design

As shown in figure 1, the ergonomic consideration ensures user's requirements and functional integration. Evaluation test confirms the product features through a series of measurements.

2.2. Electrical components and sensors

The electronic parts and mechanical components are combined to achieve appropriate functions through system integration. The electrical components primarily follow sensing and control circuits.

The walker's body has sensor units placed in front to determine and evade obstacles. The block diagram (Figure 2) shows the existing technologies such as motor controllers and sensors, which have widely been used in different fields. We proposed a link between pressure sensing and user intent.

2.3. Mechanical parts and chassis

The structural and mechanism components are configured with CAD system. The walker's handle (Figure 2) has pressure monitor sensors. The handle sensors located corresponding to palm and finger contact areas. The pressure gathered from the distribution of charges with thin-film sensors. The pressure distribution helps to verify the current status of the user and adjust the walker's speed (Figure 3).
The relationship between the visually impaired and the walker is like the way people lead them.

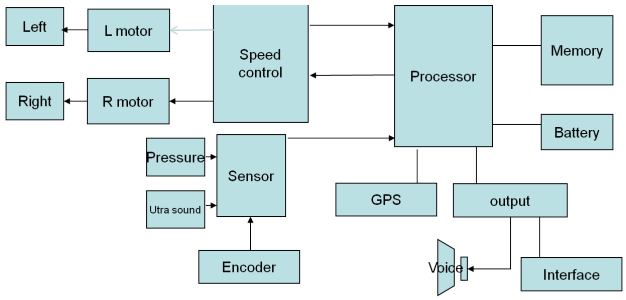


Figure 2: Functional block diagram using a smartphone to control, some software builds in the mobile app. via the Bluetooth interface.

The user configures the parameters of the walker to fit an individual's need. Typically, the handle has three to five sensors. The combination of the data from these sensors can determine the intent of the user (turning or stopping).



Figure 3: Perspective view of walker leads the visually impaired in the moving target.

3. Ergonomic design

The ergonomic design follows some rules; detail concerns: 1) the adjustable handle accommodates users of different heights within weight limit; 2) easy control and reduces the applied force; 3) fashionable style that is apart from handicap stereotypes.

3.1. Overall image

Primary purpose of the assisted walker is to take a walk with the visually impaired. Through the pre-set path and speed, the assisted walker allows the elderly to be involved in minor exercises in daily life, hence promoting a healthier lifestyle. The image of handicap is subjective to the acceptance of users. They are sensitive to negatively perceived in a social context. The form of the assisted walker is more aesthetically pleasing to customers, which may remove some of their mental pressures in purchasing and using the walker.

3.2. Weight and dimension

In [16] participants performed different lifting tasks, with three lifting heights. The working heart rate of the participants showed a significant difference during lifts. One of the design targets is to make it less heavy so that an individual can lift it comfortably. Put the user's needs to the forefront, the weight of the walker was reduced to avoid injury and easy to carry. So we need to choose to remove unnecessary elements and function.

The weight of the car is 23Kg referring to the requirements of airline baggage which is a reasonable load for user. The parts are selected and configured under the premise of this weight limit.

3.3. Way of control

Safety is an important consideration for the elderly. To avoid the sudden rotation caused by the interference of walking, we set the stop button near the thumb. When the elderly consciously can't follow or is too fast, they can manually stop the helper to prevent potential danger.
A user may choose the assistive mode and listen to the related information. Seniors face usability issues related to physical, mental, and cognitive impairments. Most of the elderly have limited experience with technology.

Observing the guidance for blind people is helpful for walker design. Usually, people follow the below steps to activate by a sighted person.

Offer an elbow and back up to the person's outstretched hand. Walk at a comfortable pace by remaining a slightly behind.

Keep arms bent at a 90-degree angle. The visually impaired person will remain a half step behind as you begin walking.

Stop when you need to turn around and tell your follower. He or she will release your arm and wait until you offer it again.

Imagine that when we assist the elderly in crossing the road, we also support him or her on one side; so this is a more natural form of assistance. We choose one-handed operation; the prerequisite is that the elderly can still walk normally. It also avoids interference with the machine while holding both hands.

4. Evaluations

The measured pressure determines the user's intention. The implement procedures are described in Figure 4.

4.1. Moldable clay

Moldable clay was used to demonstrate deformation during initial test. The handle was wrapped with clay evenly. The deformation of clay is related to the pressure distribution during usage.

User's operations would be recorded in the change of shape to determine possible positions of sensors.

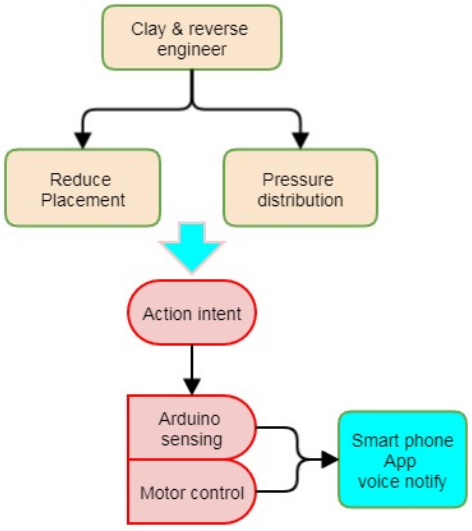


Figure 4: The ergonomic study and implement procedures

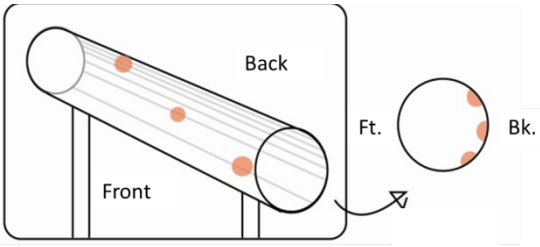


Figure 5: The side view of the pressure detector located on the handle (red dots are high-pressure area). The leading edge (Front or Ft.) is the direction of movement, and the trailing edge (Bk.) is the side of user.

There is less variation among the elderly in terms of palm-size. We choose the positions of sensors according to the magnitude of geometrical change (Figure 5).
When the old man is led forward, he may be towed and fall if the walker runs fast. If the walker turns quickly, the users also have challenges to follow. The sensors acquired the intent and conditions of the elder; so that the robot can adjust its speed.

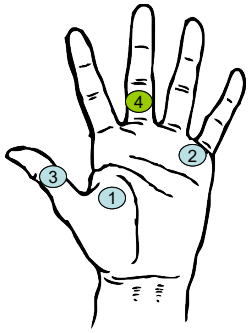


Figure 6: Sensor contact location

4.2. Pressure distribution

As shown in Figure 6, a user holds the handle during a walk. We adjusted the speed and direction of movement based on the pressure distribution. The purpose is to understand the appropriate location to detect the action intently. In this experiment, the driver

did not have a built-in processor and measurement device, so an external sensor and wire is employed to transmit directly to the host PC for subsequent design.



Figure 7: Measurement of pressure distribution.

We will start with five sensors and then re-evaluate that of fewer sensors (3 or 4) in a suitable position based on the pressure distribution observed from the clay.

Five sensors

We collect walking data of five sensors with five participants (Figure 7) to identify the pressure distribution. The measurements is based on the following situations.

- Condition 1： Right-hand control, go forward and intend to turn left,
- Condition 2： Right-hand control, go ahead and expect to turn right,
- Condition 3： Right-hand control, go forward and plan to stop.

The measured results are listed in Table 1-3.

Table 1: Pressure of left turn

User	Thumb	Index	Middle	Ring	Little
A	Low	High	Medium	Low	Low
B	Medium	High	Low	Low	Low
C	Low	High	Low	Low	Low
D	Low	Low	Low	Low	Low
E	Low	Low	Low	Low	Low

We find some variation in Table 1-3, such as:

- Table 1: the index finger of participant A.B.C is different from that of participant D and E.
- Table 2: the little finger pressure of participant E is different from that of others.
- Table 3: the index finger pressure of participant B is different from that of others.

Table 2: Pressure of right turn

User	Thumb	Index	Middle	Ring	Little
A	High	Low	Low	Medium	High
B	Medium	Medium	Low	Low	High
C	High	Low	Low	Low	High
D	High	Low	Low	Low	High
E	High	Low	Low	Low	Low

Table 3: Pressure of sudden stop

User	Thumb	Index	Middle	Ring	Little
A	Low	Medium	Medium	Medium	Medium
B	High	Low	Low	Low	Low
C	Low	High	High	High	High
D	Low	High	High	High	High
E	Low	High	High	High	High

Discussion

Depending on the size of the participant's hand, some users apply pressures by the palm, which made the finger pressure to be less apparent. This variation causes a limitation of using pressure data to determine the user's intent.

The way to overcome this difference is to reduce the unnecessary sensors and to tailor the handle's shape according to the size of the user's hand. Just as people buy a shoe with a specific size, the handle's shape needs to fit the user's hand.

The personal handle can be printed in a 3D facility at a reasonable cost — the specific handle shape made according to the clay model, and adjusted 3D model by reverse engineering. Afterwards, the sensors are then placed in the correct position of the handle.

Three / four sensors

The price of the sensor is high, so we hope to know the possible placement position with a limited number of sensors. Three and four sets of sensors placed to record the user's main motion intent sequentially. The user holds the handle and records the pressure at each position. Based on observations, we infer the position of the sensor that is suitable for the user.

The measurement of three/four sets of sensor placement summarized in (Figure 8, 9). The position for the three sets of sensors for different users is not exactly the same, which is due to the difference in the size of the palm and personal holding habits.

According to the three intentions we want to distinguish (scenario: left turn, right turn, and stop), the test results of the three sensors (Figure 8). We found that the three values at sensor 1 are identical; there is a significant difference in sensor 2; the right turn is similar to the stop at sensor 3. Although the sensor 2 is able to

resolve the three intents, in order to improve the reliability of the results, we continue to try four sensors.

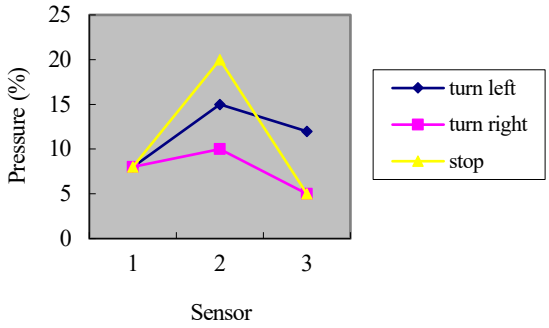


Figure 8: Pressure test results when using three sensors. The three lines represent the values of the pressures presented at each sensor when three actions are performed (left turn, right turn, and stop). The pressure sensing distribution is different when performing three actions.

The fourth pressure sensor was added to the first section of the finger joint (the leading edge of the walking handle). As shown in Figure 9, the test results of the four sensors are shown in the figure. We found that the three values of intent at sensor 1 are identical; there is a notable difference at sensor 2; the right turn is still similar to the stop at sensor 3; there is a significant difference at sensor 4. The sensor 2 and 4 are able to verify the three intents reliably.

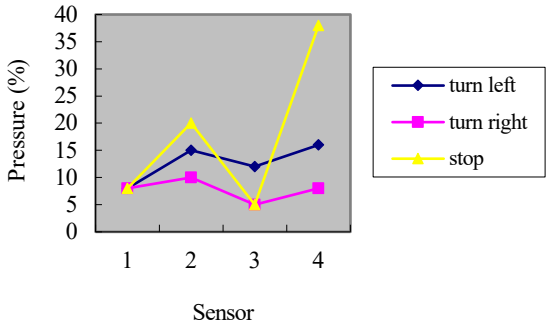


Figure 9: Pressure test results when using four sensors. The three lines represent the values of the pressures presented at the four sensors when performing three actions (left turn, right turn, and stop). The difference in pressure sensing distribution is more obvious when performing three kinds of actions.

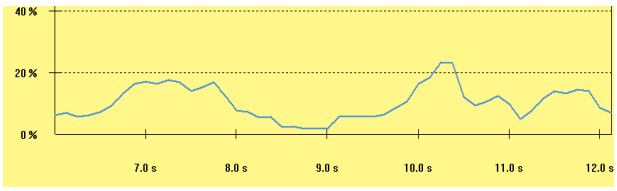
4.3. Pressure variation with time

Thin-film sensor applied to acquisition the pressure variation with time — f sensors placed on the different contact area. The sensor is 0.208mm thick and 14mm in width and sensing diameter is 9.53 mm (Figure 10). The sampling rate is 20 Hz and the data is fetched through the serial port.

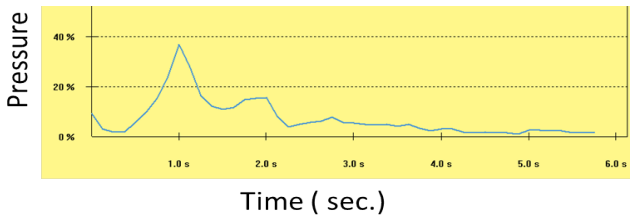
The measured pressure is plotted in Figure 11; the maximum force occurs in two situations: while turning, or when the walker hits an object. The maximum value does not last long (0.3-0.4 sec.). The result of the fourth sensor near the middle finger is shown in Figure 11(a), it can be found that in the temporary stop operation situation, the value of the fourth sensing zones becomes significantly more extensive, and scenarios can be separated. The measured data show that the 'stop' scenario, the sensor provide good variation to distinguish it at 10.6 sec. Figure 11(b) has a gently varying waveform showing a continuous force-applying state, and a glitch shows a collision.



Figure 10: Thin-film sensor and configurations



(a)



(b)

Figure 11: Sensor's responses: (a) the second sensor during turns left, (b) the fourth sensor during the collision and stop.

We divide the value of the sensor into three different levels; low-level (< 5%), high-level (>14%), the median level in between. If the signal maintained at the same level for a certain period, the phenomenon presented a user's action. The maximum pressure value maintained for 0.3 sec during turning moment (Figure 11.a).

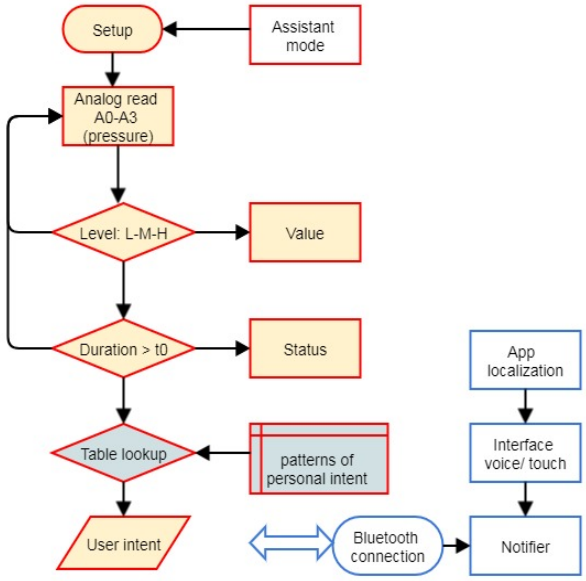


Figure 12: Flow chart of the Arduino sensing module

4.4. Implementation

The sensing module of the walker can implement with Arduino, which connects the four pressure sensors to the analog input pins of A0 to A3 (digital value 0-1023). The percentage of the Y-axis on the left side of Figure 11(b) represents the maximum measured value of the pressure sensor. We set a high pressure zone is above 14% (digital value 143), and a low pressure zone is below 5% (digital value 51), and a medium zone between the two (range 5%-14%). For example, in 2sec time (sampling rate is 20 Hz), the high pressure value is in the range of 0.75 seconds to 1.3 seconds, 1.75 seconds to 2.0 seconds; the low pressure time is between 0.1 seconds and 0.45 seconds, and the rest is medium pressure value. When maintaining the same level for a period of time, we can compare the pressure distribution of the four pressure sensors to predict the user's intention. The sensing module determines the rotation or deceleration through a logic judgment and controls the rotation speed of the motor through the voltage value of the output pin (Figure 12). There are many applications on the phone that handle positioning and navigation. The sensing module can also be connected to the mobile phone via Bluetooth, and the action to be performed is notified to the elderly by the voice of the mobile phone without being overwhelmed.

5. Conclusions

The walker is suitable for elderly in public environments, such as hospitals or retirement homes. The concept design emphasized on ergonomic features; such as weight reduction, folding, sensing the user's intent, and the like. Physical prototype and hardware circuit implementation is ready-to-expand Arduino module to perform force sensing. The smart phone app helps the elderly to move in a safe environment.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by the National Science Council, Taiwan, ROC, under grant NSC 99-2221-E-324-026-MY2. We gratefully thank Wenjie Xiao and Yusheng Huang for model making and the field test.

References

[1] Z. Maćkowiak, W. Osiński, and A. Salamon, "The effect of sensorimotor training on the postural stability of visually impaired women over 50 years of age," Journal of Women & Aging, 27(1), 68-80, 2015. https://doi: 10.1080/08952841.2014.928140

[2] T. Nakamura, "Quantitative analysis of gait in the visually impaired, disability and rehabilitation" (19)5, 194-197, 1977. https://doi: 10.3109/09638289709166526

[3] E. M. Phillips, J. C. Schneider, and G. R. Mercer, "Motivating elders to initiate and maintain ex-ercise1" Archives of Physical Medicine and Rehabilitation, (85)1, 52-57, 2004.

[4] K. Rector, "Exploring the opportunities and challenges with exercise technologies for people who are blind or low-vision" in 2015 17th International ACM SIGACCESS Conference on Computers & Accessibility, Lisbon, Portugal, 2015.

[5] C. Vlad, F. Röthenbacher, "The Chatty Environment–providing everyday independence to the visually impaired" in Workshop on Ubiquitous Computing for Pervasive Healthcare Applications, UbiComp, Seattle, Washington, USA, 2003.

[6] M. H. A. Wahab, A. A. Talib, H. A. Kadir, A. Noraziah, A. A. Mutalib, "Smart cane: Assistive cane for visually-impaired people" arXiv preprint, 1110-1115, 2011.

[7] Sorour S, Lostanlen Y, Valaee S, Majeed K., "Joint indoor localization and radio map construction with limited deployment load," IEEE Transactions on Mobile Computing, (14) 5, 1031-43, 2015.

[8] J. J. Leonard, H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons" IEEE Transactions on Robotics and Automation, (7)3, 376-382, 1991.

[9] J. Biswas, M. Veloso, "Depth camera based indoor mobile robot localization and navigation" in 2012 IEEE International Conference Robotics and Automation (ICRA), St. Paul, MN, USA, 2012.

[10] T. Tsukiyama, "In-home health monitoring system for solitary elderly" Procedia Computer Science, (63)1, 229-35, 2015.

[11] Mast, M., Burmester, M., Graf, B., Weisshardt, F., Arbeiter, G., Španěl, M., & Kronreif, G., "Design of the human-robot interaction for a semi-autonomous service robot to assist elderly people," in Ambient Assisted Living, 15-29, Springer, Cham, 2015.

[12] Granados, Diego Felipe Paez, et al., "Guiding human motions in physical human-robot interaction through COM motion control of a dance teaching robot," Humanoid Robots (Humanoids), presented at the 2016 IEEE-RAS 16th International Conference, Cancun, Mexico, 2016.

[13] T. Yano, J. H. Lee, and S. Okamoto, "Development of a biped walking robot with antagonistic actuation" International Journal of Mechanical Engineering and Robotics Research, (5)3, 196-201, 2016. https://doi: 10.18178/ijmerr.5.3.196-201

[14] V.-T. Nguyen and H. Hasegawa, "Effect of toe length on biped walking behavior," International Journal of Mechanical Engineering and Robotics Research, (7)6, 599-603, 2018. https://doi: 10.18178/ijmerr.7.6.599-603

[15] Z. Shen, S. Sam, G. Allison, and L. Cui, "A simulation-based study on a clutch-spring mechanism reducing human walking metabolic cost" International Journal of Mechanical Engineering and Robotics Research, (7) 1, 55-60, 2018. https://doi: 10.18178/ijmerr.7.1.55-60

[16] A.S. Abadi, A. Mazlomi, A.H. Jafari, "Effects of box size, frequency of lifting, and height of lift on maximum acceptable weight of lift and heart rate for male university students in Iran" Elvectronic Physician, (7)6, 1365-1370, 2015.

[17] S. Azenkot, C. Feng, and M. Cakmak, "Enabling building service robots to guide blind people: A participatory design approach," in the Eleventh ACM/IEEE International Conference on Human-Robot Interaction, Christchurch, New Zealand, 2016. https://doi: 10.1109/HRI.2016.7451727

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 2, 260-269 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Effects of Different Activation Functions for Unsupervised Convolutional LSTM Spatiotemporal Learning

Nelly Elsayed[*,1], Anthony Maida[2], Magdy Bayoumi[3]

[1,2]*School of Computing and Informatics, University of Louisiana at Lafayette, 70503, Louisiana, USA*

[3]*Department of Electrical Engineering, University of Louisiana at Lafayette, 70504, Louisiana, USA*

ARTICLEINFO

Article history:
Received:03 February, 2019
Accepted: 26 March, 2019
Online: 10 April, 2019

Keywords:
Convolution LSTM
Activation function
Video prediction
Moving MNIST
Prediction

ABSTRACT

Convolutional LSTMs are widely used for spatiotemporal prediction. We study the effect of using different activation functions for two types of units within convolutional LSTM modules, namely gate units and non-gate units. The research provides guidance for choosing the best activation function to use in convolutional LSTMs for video prediction. Moreover, this paper studies the behavior of the gate activation and unit activation functions for spatiotemporal training. Our methodology studies the different non-linear activation functions used deep learning APIs (such as Keras and Tensorflow) using the moving MNIST dataset which is a baseline for video prediction problems. Our new results indicate that: 1) the convolutional LSTM gate activations are responsible for learning the movement trajectory within a video frame sequence; and, 2) the non-gate units are responsible for learning the precise shape and characteristics of each object within the video sequence.

1 Introduction

Problems in video prediction (spatiotemporal prediction) are important in both research and industry. Examples include human action recognition [1], spread of infections, medical predictions [2], weather forecast prediction [3], and autonomous car video prediction systems [4, 5].

Spatiotemporal (e.g. video) datasets are challenging to predict due to the spatial and temporal information which the datasets carry through time. In sequence prediction problems, the prediction at the current step depends primarily on the previous history. Making a near future prediction requires an computational tool that is able to discover both the long and short-term dependencies in the data. Initially, recurrent neural networks (RNN) [6] were proposed to model data which was sequential in nature. However, straightforward RNNs have major drawbacks. RNNs suffer from the vanishing/exploding gradient problem which prevents them from learning long-term data dependencies [7–9]. There are several gate-based RNN models that attempt to solve this classic problem. These include the long short-term memory unit (LSTM) [7], peephole LSTM [10], gated recurrent unit (GRU) [10], and minimal recurrent unit (MRU) [11].

This paper studies the LSTM architecture as an attractive architecture architecture to perform spatiotemporal (video) prediction. In addition, Gref et al. [12] empirically showed that the LSTM is the most efficient recurrent architecture to solve for speech recognition, handwriting recognition, and polyphonic music modeling [12]. The LSTM has the highest number of gates compared with other gated units [9–11]. Furthermore, the choice of the activation function for gate and non-gate units within the LSTM has an essential effect on the the LSTM's function [12].

This work studies gate activations and non-gate activations within the LSTM architecture. In most APIs the gate activations and non-gate activations are known as recurrent activations and unit activations, respectively. To solve image and video related problems, different convolution-based models are used such as in Kalchbrenner et al [13], Elsayed et al. [5], Lotter et al. [4], Finn at al. [14], and Wang et al. [15]. However, the most widely used architecture for video prediction to-date is the convolution LSTM. Hence, we performed our experiments on a convolution-based LSTM (convolutional LSTM) network model [4]. We studied the effect of different gate activation and unit activation

[*]Corresponding Author: Nelly Elsayed, School of Computing and Informatics& nel.elsayed@gmail.com

https://dx.doi.org/10.25046/aj040234

functions through the learning process of the convolutional LSTM architecture [3]. Our experiments use the convolutional LSTM found in the Keras deep learning framework [16]. Training was performed on the moving MNIST dataset as the most commonly used benchmark for video prediction problems [17].

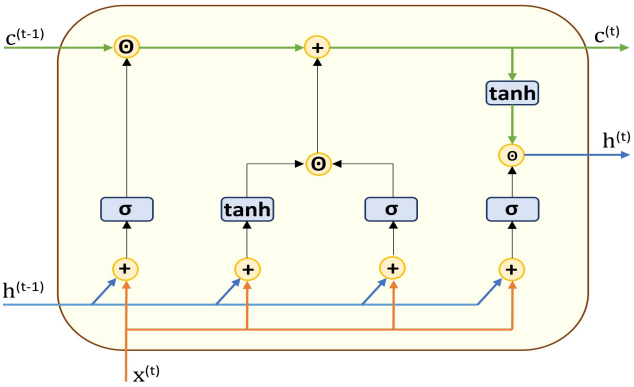


Figure 1: The unrolled LSTM architecture. The memory-state and output lines are time-evolving and labelled c and h, respectively. The CEC is indicated by the plus sign on the c line. The level of description is coarse enough so that this could either depict a standard LSTM or a convolutional LSTM.

In our empirical study, we worked with a model than what was needed for optimal prediction. Specifically, it had fewer parameters. This reduction allows us to more clearly see the effect of gate and unit activations on the convLSTM model without contamination by influences of other high-effect parameters. For example, it avoids saturation effects on the model accuracy which helps to see changes in activation function effects [18]. Our study compares different activation functions on a simple convolutional LSTM model for spatiotemporal (video) prediction task. We investigated the responsibilities of the convolutional LSTM gates and unit activations over the training process. Furthermore, we found the most appropriate function to be applied for the gates and for the unit activation for solving video prediction problem.

2 Convolutional LSTM

The LSTM [7] is the first gate-based recurrent neural network. As noted earlier, it was motivated by the need to mitigate the vanishing gradient problem and did so by using the principle underlying a constant error carousel (CEC) [7] to improve the learning of long-term dependencies. This was implemented by adding a memory-state line to the architecture (Figure 1). This memory-state line can retain (remember) the recurrent values over arbitrary time intervals. Gres et al. [8] added a forget gate to the LSTM cell to allow information to be removed from (forgotten) the memory-state line. This gate was trainable so that the module could learn when stored information was no longer needed. Gers et al. [10] improved the design by adding so-called peephole connections from the memory cell to the LSTM gates. The peephole connections enable the memory line to exert control over the activation of all three of the gates. This helps the LSTM block to avoid vanishing gradients and also helps learning of precise timing in conjunction with Constant Error Carousel (CEC) [8]. The CEC is the key ingredient mitigate the vanishing gradient problem.

LSTMs are used in several application areas including speech recognition [19,20], language modeling [21], sentiment analysis [22], music composition [23], signal analysis [24], human action recognition [1], image caption generation [25], and video prediction [3,4,17,26].

The convolutional LSTM (convLSTM) was introduced in 2015 by Shi et al [3] for weather precipitation forecast prediction task. This is an enhancement to the LSTM discussed previously. The crucial change is that the data within the module takes the form of a multichannel 2D feature map (image stack). The convLSTM improves predicted video on data sets of evolving precipitation maps compared to the non-convolutional LSTM. Apparently, it captures spatiotemporal correlations better than non-the standard variant [3,17].

An unrolled convLSTM module in Figure 1 can be either standard or convolutional because it does not make the relevant internal operations and weights explicit. Like the standard LSTM, the architecture has three gates indicated by the activation functions labeled σ. These are the forget, input, and output gates [9]. There are two non-gate units which are labeled with tanh activation functions. The convLSTM gates (input, forget, and output) have the ability to multiplicatively attenuate the outputs of tanh units and the value on the $c^{(t-1)}$ line. The forget gate attenuates information deemed unnecessary from the previous cell memory state $c^{(t-1)}$. $h^{(t-1)}$ is the convLSTM cell output at time $t-1$, and $h^{(t)}$ is the output of the convLSTM cell at time t. $x^{(t)}$ is the current video frame target at time t. $c^{(t-1)}$ and $c^{(t)}$ are the memory cell states at consecutive times.

The following equations calculate all of the gate and the unit values based on the Gers et al. [8] LSTM combined the convolutional model of Shi et al. [3]. These operations should be compared with Figure 2.

$$i^{(t)} = \sigma(W_{xi} * x^{(t)} + U_{hi} * h^{(t-1)} + b_i) \quad (1a)$$

$$g^{(t)} = tanh(W_{xg} * x^{(t)} + U_{hg} * h^{(t-1)} + b_g) \quad (1b)$$

$$f^{(t)} = \sigma(W_{xf} * x^{(t)} + U_{hf} * h^{(t-1)} + b_f) \quad (1c)$$

$$o^{(t)} = \sigma(W_{xo} * x^{(t)} + U_{ho} * h^{(t-1)} + b_o) \quad (1d)$$

$$c^{(t)} = f^{(t)} \odot c^{(t-1)} + i^{(t)} \odot g^{(t)} \quad (1e)$$

$$h^{(t)} = tanh(c^{(t)}) \odot o^{(t)} \quad (1f)$$

The the input, forget, and output gate activations s are denoted $i^{(t)}$, $f^{(t)}$, and $o^{(t)}$. σ denotes logsig. The gates have activation values falling in the interval $[0,1]$ where 1 means completely open and 0 means completely closed. $g^{(t)}$, is the current input. $c^{(t)}$ represents the current memory module state. $h^{(t)}$ represents the current output. b_i, b_g, b_f, and b_o are biases for the respective gate. W's and U's are the feedforward and

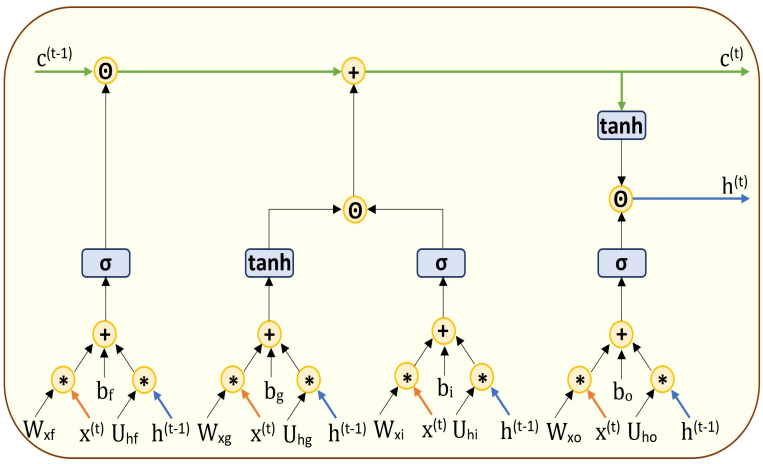


Figure 2: The unrolled convLSTM architecture at the arithmetic operation level. This rendering makes convolution operations and weights explicit and should be compared with Equation Set 1.

recurrent weights, respectively. "⊙" denotes element-wise multiplication, and "∗" denotes the convolution operator.

Figure 2 shows the arithmetic operational level of the convLSTM architecture where the convLSTM components and their corresponding weights are shown explicitly.

3 Activation Functions

As explained earlier, gate units usually have sigmoid (range $[0,1]$) activation functions and non-gated units usually have *tanh* activation functions. This paper studies the effects of using different activation functions for these units. This section reviews the most common activation functions used. All of the functions discussed are nonlinear.

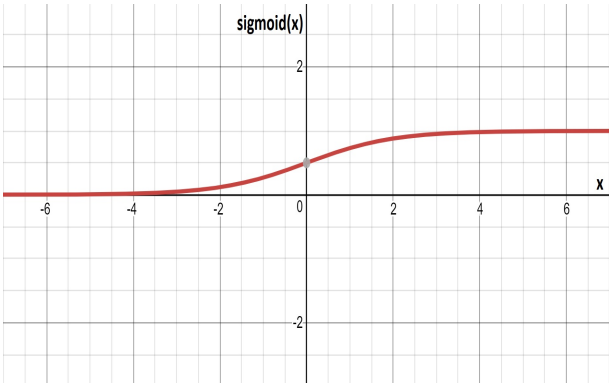


Figure 3: The curve of the logsig (σ) function.

3.1 Logsig Function

The logistic sigmoid (logsig) function [27] is shown in Figure 3. It is given by the following equation:

$$f(x) = \sigma(x) = \frac{e^x}{e^x + 1} = \frac{1}{1 + e^{-x}}, \tag{2}$$

The logsig is commonly denoted by $\sigma(x)$, where $x \in (-\infty, \infty)$ and $\sigma(x) \in (0,1)$. The logsig is historically common in network models because it intuitively represents the firing rate of a biological neuron. The logsig is still used in recurrent models for gating purposes. The input value is usually known because it appears in the feedforward sweep of the network. Because of this, its derivative can be computed efficiently as shown in (3).

$$f'(x) = \sigma'(x) = f(x)(1 - f(x)) \tag{3}$$

3.2 Hard-Sigmoid Function

The hard-sigmoid (hardSig) function is considered a noisy activation function. HardSig approximates logsig function by taking the first-order Taylor expansion around zero and clipping the result according to function's range $[0,1]$. Hardsig appears in Figure 4 and is calculated as follows:

$$hardSig(x) = \max(\min(0.25x + 0.5, 1), 0) \tag{4}$$

This function allows a crisp decision for the gradient during function saturation phase to be either fully on or fully off. In the non-saturation range near zero, hardSig is linear. This helps avoid vanishing gradients [28].

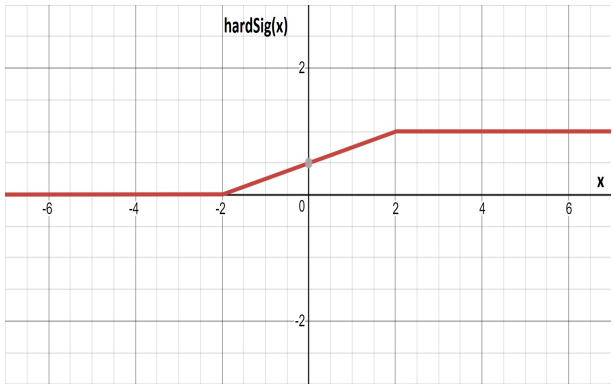


Figure 4: The curve of the hardSig function.

3.3 Hyperbolic Tangent Function

tanh is a saturating activation function. It is typically used as the recurrent activation function in RNNs [6].

In some cases, tanh is considered as cheaper for gradient evaluation in comparison to the logsig function. The function graph appears in Figure 5 and is specified by (5):

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{5}$$

$$tanh'(x) = 1 - tanh^2(x) \tag{6}$$

A convenient property of the tanh activation function is that calculating its first derivative from a known input value is easy as seen in Eqn. (6).

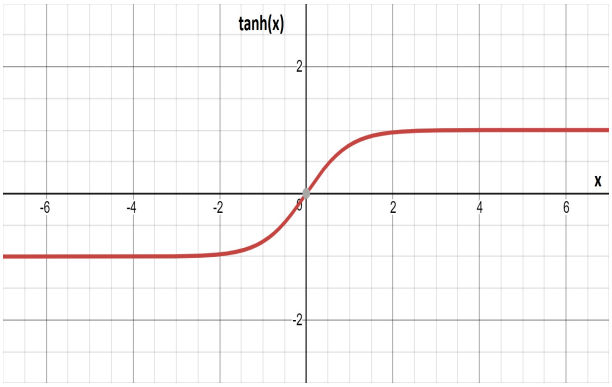


Figure 5: The curve of the Hyperbolic Tangent function.

3.4 SoftSign Function

The SoftSign function was introduced in [29]. Its grahp is given in Figure 6 and is specified by:

$$SoftSign(x) = \frac{x}{1 + |x|} \tag{7}$$

The SoftSign function range is $(-1, 1)$ and the function domain is $x \in (-\infty, \infty)$. The first derivative of SoftSign also can be evaluated simply as the following:

$$SoftSign'(x) = \frac{x}{(1 + |x|)^2} \tag{8}$$

SoftSign is an alternative function to hyperbolic *tanh*. In comparison to *tanh*, it does not saturate as quickly, improving its robustness to vanishing gradients.

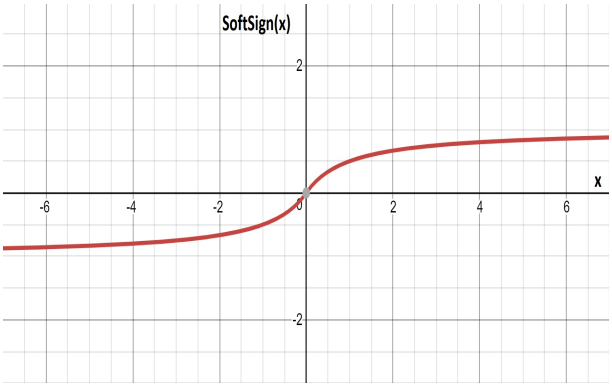


Figure 6: Curve of the SoftSign function.

3.5 ReLU Function

The rectified linear unit (ReLU) is very efficient compared to other activation functions. It was proposed in [30]. For $x \geq 0$ it is the identity function and appears in Figure 7. ReLU is defined by:

$$ReLU(x) = max(x, 0) \tag{9}$$

Its has domain $x \in (-\infty, \infty)$ and range $[0, \infty)$.

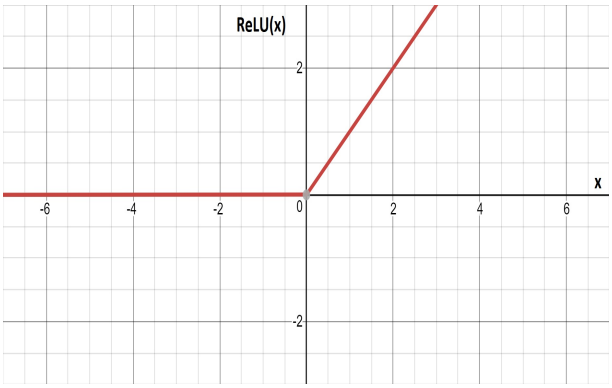


Figure 7: The curve of the ReLU function.

3.6 SoftPlus Function

The SoftPlus function (Figure 8) was proposed in [31]. It modifies the ReLU by applying a smoothing effect [29] and is defined:

$$SoftPlus(x) = \ln(1 + e^x) \tag{10}$$

with $x \in (-\infty, \infty)$ and range $(0, \infty)$. Its derivative is:

$$SoftPlus'(x) = \frac{1}{1 + e^x} \tag{11}$$

Like the SoftSign function, the SoftPlus function is used only in specific applications.

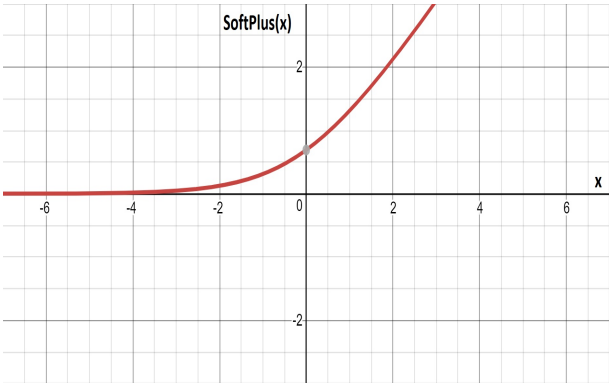


Figure 8: The curve of the SoftPlus function.

3.7 ELU Function

The exponential linear unit (ELU) was introduced in [32] to avoid a zero derivative for negative input values. The ELU achieved better classification accuracy than ReLUs [32] and is defined by:

$$ELU(x) = \begin{cases} \alpha(e^x - 1), & \text{if } x < 0 \\ x \quad , & \text{otherwise} \end{cases} \tag{12}$$

Its domain and range is $(-\infty, \infty)$. Assuming that $\alpha = 1$, its graph is shown in Figure 9 .

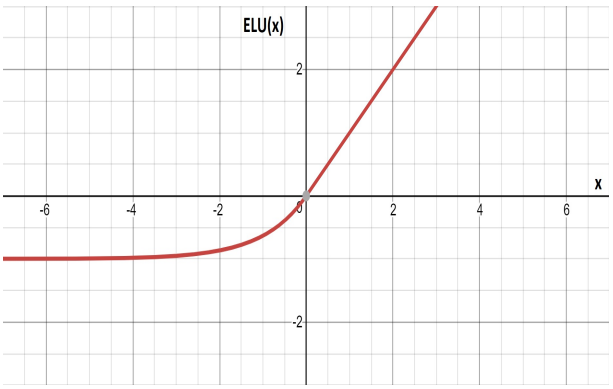


Figure 9: The curve of the ELU function.

3.8 SELU Function

The scaled exponential linear unit (SELU) an ELU function modified with two hyperparameters. Its definition follows:

$$SELU(x) = \lambda \begin{cases} \alpha(e^x - 1) & , \quad \text{if } x < 0 \\ x\,, & \text{otherwise} \end{cases} \tag{13}$$

Its domain is the same as the ELU function, but its range is limited to $-(\lambda\alpha, \infty)$. λ and α are hyperparameters that depend on the mean and variance of the training data. If the data is normalized to $\mu = 0$ and $\sigma = 1$, then $\lambda = 1.0507$ and $\alpha = 1.6732$ whose graph appears in Figure 10.

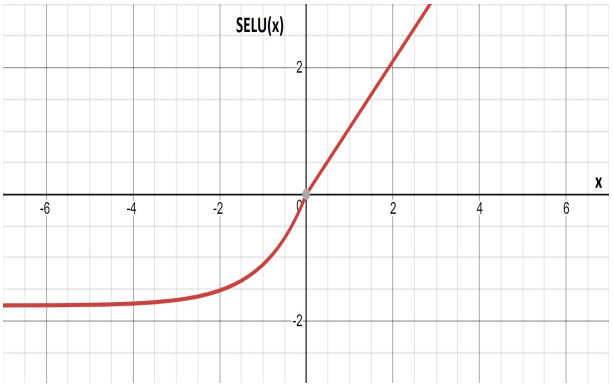


Figure 10: The curve of the SELU function.

The first derivative of SELU is calculated as shown below:

$$SELU'(x) = \lambda \begin{cases} \alpha e^x & , \quad \text{if } x < 0 \\ 1\,, & \text{otherwise} \end{cases} \tag{14}$$

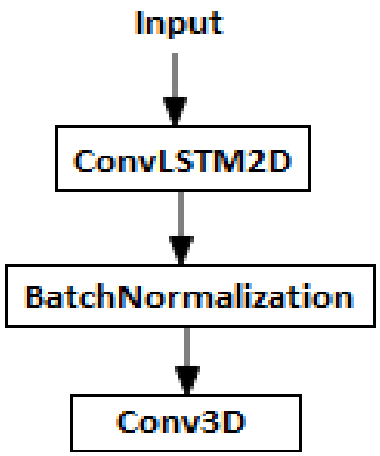


Figure 11: The model architecture used in our experiments.

4 Experiments and Results

This section the results of comparing different activation functions for both the gate units and non-gate units. We chose the functions described in the previous section. They are commonly available in deep learning frameworks. We explore the question of which activation function is best as a gate activation and which is best as a unit activation for convolutional LSTMs in an unsupervised context. Moreover, we want to understand the differential behavior of the gate and non-gate units as a result of training.

We calculated several measures: training loss, validation loss, elapsed training time, and appearance of the predicted video frames compared with ground truth. We did our experiment and analysis on the moving MNIST [17] as explained in the introduction. The moving MNIST dataset consists of 10,000 different video sequences. Each sequence has a length of 20 frames. Each frame is 64×64 pixels and contains two randomly selected handwritten digits from the standard 28×28 MNIST dataset. The two digits move randomly within the frame [17].

4.1 Framework

Our framework uses Shi's et al. [3] convLSTM. Because we had limits on our computing power, our test were serendipitously performed using a small model with insufficient learning capacity for fully accurate predictions. However, this did let us more clearly see the effects of the activation function manipulations. The reduced capacity models helped us avoid ceiling effects so that we could see differential performances in the activation functions.

The model is implemented using the Keras API [16] and appears in Figure 11. It consists of three layers. The first layer is a convLSTM with 40 size 3x3 kernels. The next layer is batch normalization. Finally, we used $3 \times 3 \times 3$ kernel to get the video output shape. Table 1 provides details of the model parameters.

The Adam optimizer [33] was used with initial learning rate $r = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$. The cost

Table 1: Description of the proposed model architecture used in this study.

Layer (type)	Output Shape	Param Number
ConvLSTM2D	(None, None, 64, 64, 10)	4000
Batch Normalization	(None, None, 64, 64, 10)	40
Conv3D	(None, None, 64, 64, 1)	271
Total params	4,311	
Trainable params	4,291	
Non-trainable params	20	

function was MSE [34].

Each input was a 20-frame video randomly sampled from moving MNIST [17]. As previously mentioned, the amount of input data is 10,000 videos each of length 20 frames. We used 9,500 for training and 500 for validating.

The task is proposed for sequence to sequence video prediction. A sequence of five future frames were predicted using the previous five frame sequence. Because of the reduced learning capacity mentioned at the beginning of this section, the trained network could predict only the two temporally closed frames clearly and the rest were unclear because of the reduced learning capacity in the model.

We used an Intel(R) Core i7 – 7820HK, 32GB RAM, and a GeForce GTX 1080 GPU. Training was for ten epochs in each experiment. In each epoch, we fit 475 batches where each batch size was set to 20 for our model experiments. All parameters were held constant over all experiments except for the recurrent activation function that was being tested.

4.2 Results

Our results pertain to the relative performance of varied activation functions for gate and non-gate units.

4.2.1 Results for Gate Activation Functions

For a fair comparison of different gate activation functions, we only changed the gate activation functions and preserved the remaining parameters with the same initializations and values during each experiment. The gate effects were studied collectively because we used the default convolutional LSTM module in found in the Keras API. The API allowed us to change all the gate functions at once (not each gate separately) for setting up an experiment.

For this experiment, the gate activation functions apply to all three gates gates described in Eqns 1a, 1c, and 1d. Accordingly, only functions with range $(0, 1)$ are eligible. Thus, this experiment only compared the logsig and hardSig functions.

Table 2 shows the performance results, including average elapsed training time (in minutes), MSE loss (Loss-MSE) for training, validation loss (Val-loss-MSE), percent accuracy (Accuracy) for $n = 3$, and the standard error (SE) for training loss ($n = 3$). The elapsed training time using the hardSig activation is higher than the logsig function by about ten percent. We are not sure why this happens. In Table 2, both logsig and hardSig have nearly exactly the same MSE and MAE.

Figure 12 shows the training and validation loss for the logsig and hardSig activation functions. From the figure, one can also see that the loss average difference between the training and validation runs using the hardSig function is smaller than the logsig function.

We also examined the differences between the accuracy of both training and validation runs using the logsig versus hardSig functions. The average ($n = 3$ for all data points) accuracy values for each function appear in Figure 13. Blue represents training and yellow represents validation accuracy. Both logsig and hardSig functions have similar validation and training accuracy It is also temporally time. The hardSig shows a more comparable accuracy among the training and validation curves. Moreover, hardSig curve does not contain frequent validation accuracy drops down as in logsig curve.

Figure 14 visually shows the prediction results for both the sigmoid and hardSig gate activations of the model. The first row shows ground truth. The two rows below show predictions using the logsig versus hardSig gate activation functions. Both logsig and hardSig are comparable. From the figure, both functions show that the model attempts to learn the digit movement trajectories as shown in each of the prediction frames. The 'double exposure' appearance indicates that the prediction retains the object's previous location within the predicted frame. Despite the double exposure look, digit shape is preserved. Moreover, the hardSig function shows better prediction accuracy than the sigmoid. This agrees with the Gulcehre et al. [28] theory stating that noisy activation functions are better than their non-noisy counterparts.

This experiment suggests the gate activation functions of the convolutional LSTM play a role learning the movement trajectories.

4.2.2 Results for Non-gate Activations

Our second experiment analyzes which activation function is best to use for non-gate units (input-update and output activation) as seen in Eqns. 1b and 1f. We applied the same setup and initialization as in the first experiment except that we retained all the convLSTM parameters. Only the set of possible unit activation functions were manipulated. Gate activations were

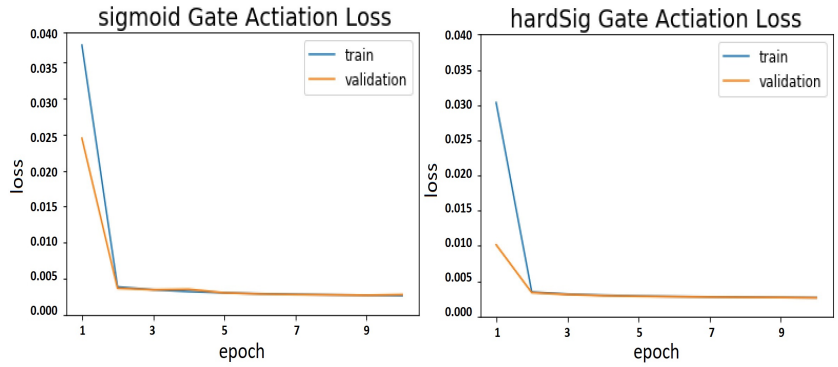


Figure 12: Training and validation loss for each gate activation function over each training epoch. In the left figure, all three gate units used the logsig activation. In the right figure, all three gate units used the hardSig activation function.

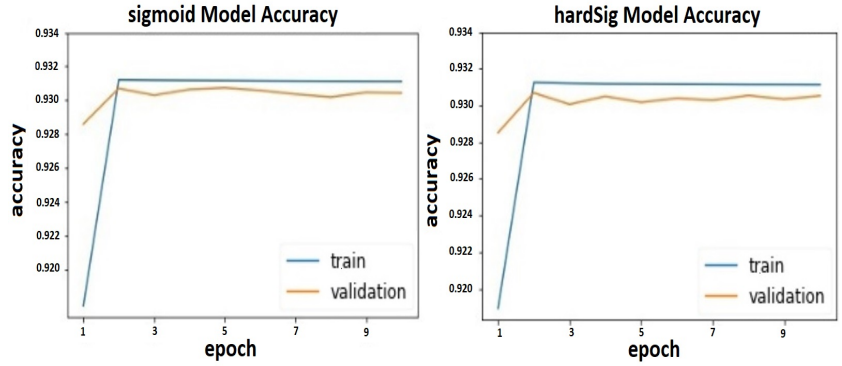


Figure 13: Training and validation accuracy for each gate activation function over each training epoch.

Table 2: Gate Activation Effect on the ConvLSTM Training Process.

Activation function	Training Results					
	Time (Minutes)	*Loss (MSE)*	*Val-loss (MSE)*	*Mean Absolute* Error (MAE)	*Accuracy (Percentage)*	*SE (Loss)*
sigmoid	20.2	0.00433	0.00510	0.01726	0.92979	0.0000498
hardSig	22.5	0.00420	0.00525	0.01699	0.92999	0.0000147

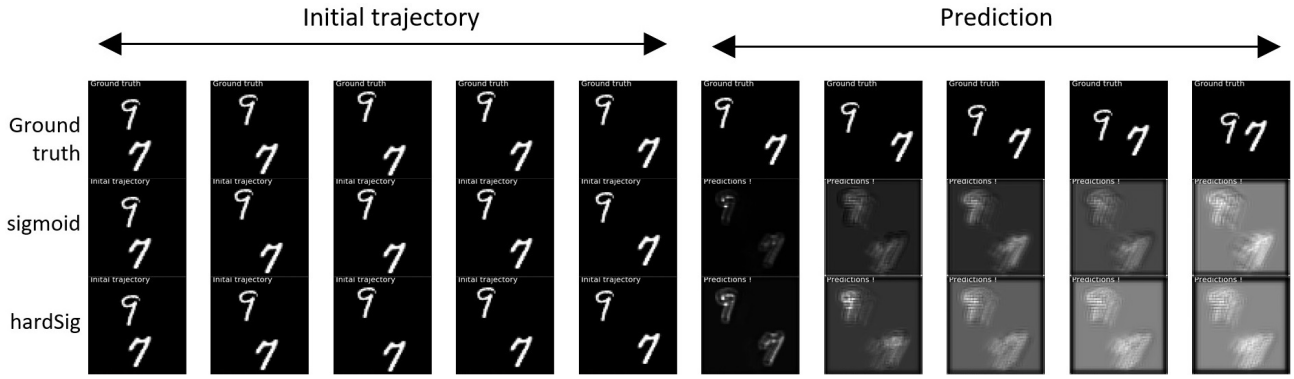


Figure 14: The history and predicted frames using logsig and hardSig as gate activations. Except for the 'double exposure effect,' the predicted shape is preserved.

Table 3: The Non-Gate Unit Effect on Training.

Activation function	Training Results					
	Time (Minutes)	*Loss (MSE)*	*Val-loss (MSE)*	*Mean Absolute* Error (MAE)	*Accuracy (Percentage)*	*SE (Loss)*
sigmoid	24.5	0.00367	0.00387	0.02597	0.93058	0.0000438
tanh	23.9	0.00423	0.00537	0.01751	0.93003	0.0000607
hardSig	25.3	0.00347	0.00386	0.02401	0.93076	0.0000300
softSign	23.9	0.00431	0.00442	0.01777	0.92972	0.0001690
softPlus	24.2	0.00342	0.00308	0.02044	0.93073	0.0000400
ReLU	23.7	0.00424	0.00451	0.01911	0.93015	0.0000723
ELU	23.6	0.00453	0.00555	0.01760	0.93001	0.0001431
SELU	26.5	0.00392	0.00407	0.01784	0.93031	0.0001984

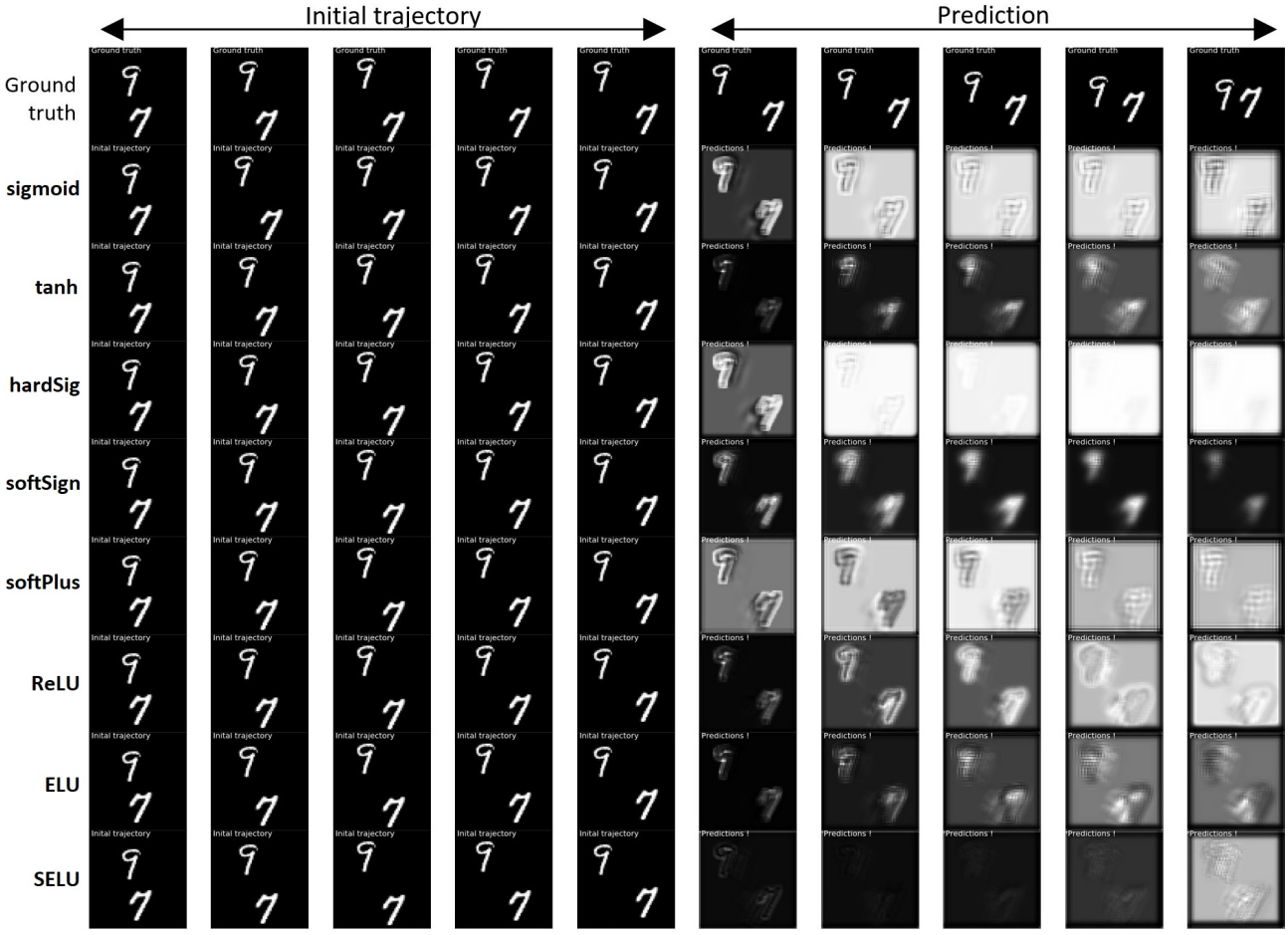


Figure 15: The initial trajectory and predicted frames for each non-gate unit activation function. Although there is shape distortion, the movement directions seem to be preserved.

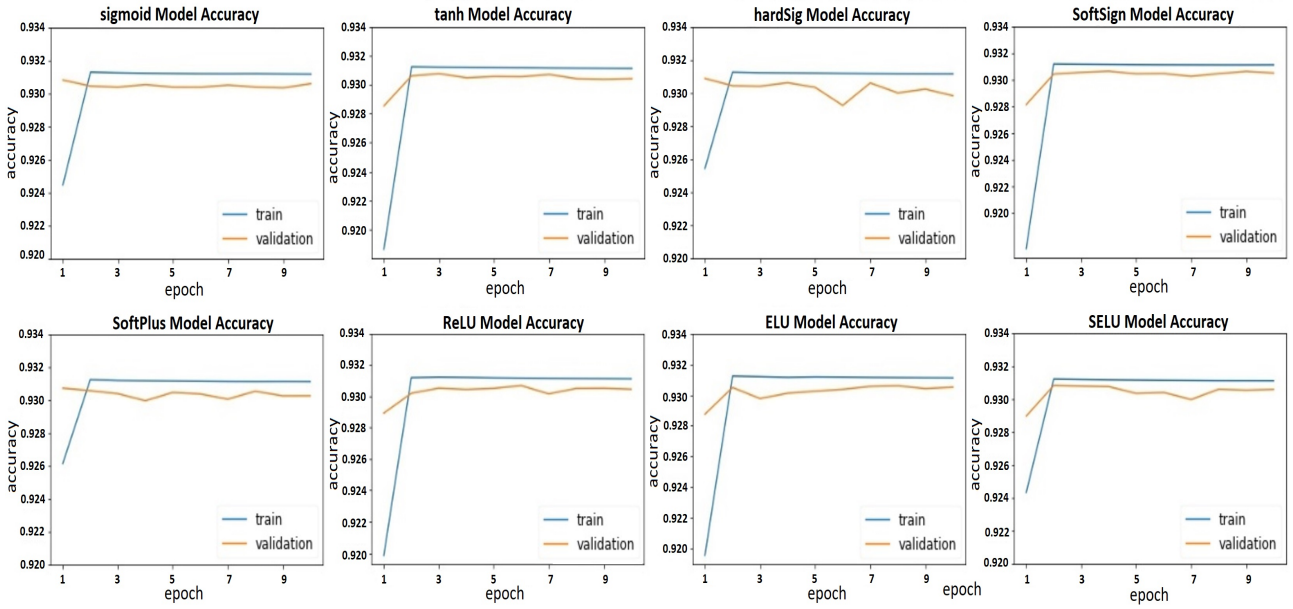


Figure 16: Training and validation accuracy non-gate unit activations. All tested functions are shown for each training epoch.

always set to hardSig. This happens to be the default recurrent activation in the Keras convLSTM API.

There is no restriction on the range of the unit activation function selection compared to the gate activation. Therefore, we tested all eight of the functions from Section 3. Table 3 shows average training time, training loss, validation loss, validation mean absolute error, training accuracy, and standard error of training loss ($n = 3$). The units of training and validation loss are MSE.

Table 3 shows that the elapsed training time of the functions is overall about the same and the highest training time (SELU) is about two minutes longer than the rest. Loss was calculated using MSE and is about the same for all tested functions.

Accuracy of the training and validation tests appears in Figure 16 where the hardSig has the lowest validation accuracy because the accuracy degraded over time. In addition, it has a significant average difference between both the training and validation accuracy curves. For the SoftPlus, SELU, ReLU, and ELU diagrams, there is a notable unstable validation accuracy compared to the training accuracy. However, the most stable validation accuracy for the other tested functions is found when the activation was sigmoid, tanh, or the softSign function. However, the average difference between training and validation accuracies for logsig is higher than the tanh and softSign.

The predicted videos as compared with ground truth for the non-gate unit functions is shown in Figure 15. The lowest visual accuracy occurs when the SELU function is used. The hardSig, logsig, ReLU, and softPlus functions show roughly the same visual prediction results. These functions were unable to maintain, the object's outline, the background, or object's color. The ELU function could not retain the digit outline through time. The softSign function was able to maintain the shape of the objects but it maintained the object's location with reducing their actual sizes. It also suppressed object movement. The tanh preserved the object shape and movement.

In important insight can be seen by looking at the differential visual prediction results in Figure 15. Although changing the non-gate activation functions changes the shape distortion, the predicted movement accuracy seems to be the same across the different activation functions. This indicates that non-gate units play more of a role in predicted shape than in predicting movement direction. In contrast, we saw the opposite pattern with the gate activation functions.

As seen in Figures 14 and 15, prediction accuracy should be improved for practical applications. The intent of our work was to study differential activation function effects. Prediction accuracy can be improved by enhancing the model architecture by increasing the number of the convolutional LSTM layers and adding regularization techniques such as dropout [35]. Further improvements can be obtained by adding more channels to each layer, increasing the number of training epochs, adjusting the input batch size, and adjusting the hyper-parameters. Moreover, using an information flow propagation transfer algorithm (such as predictive coding [4, 36, 37]) to data between the multilayered hierarchy could significantly increase the prediction performance.

5 Conclusion

We found that the choice of gate unit activation function, when used in the convLSTM video prediction problem, affects acquisition of the digit's movement trajectory throughout the video frame sequence. The role of the non-gate unit activation is to enhance the object's shape and determine its details. Only activation functions with range $[0, 1]$ are eligible as a gate activation functions. The convLSTM exhibited the best

prediction performance when the gate activation used the hardSig function and it had the best prediction visual results. In addition, the relation between training and validation accuracies was stable. The model also obtained better predictions when the tanh function was used as the non-gate activation. The tanh has similar training loss and training time to the other functions examined in our experiments.

References

[1] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *International Workshop on Human Behavior Understanding*, pp. 29–39, Springer, 2011.

[2] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Deepcare: A deep dynamic memory model for predictive medicine," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 30–41, Springer, 2016.

[3] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional *LSTM* network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, pp. 802–810, 2015.

[4] W. Lotter, G. Kreiman, and D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning," *arXiv preprint arXiv:1605.08104*, 2016.

[5] N. Elsayed, A. S. Maida, and M. Bayoumi, "Reduced-gate convolutional LSTM using predictive coding for spatiotemporal prediction," *arXiv preprint arXiv:1810.07251*, 2018.

[6] L. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, 2001.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Computation*, 2000.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[10] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *Journal of machine learning research*, vol. 3, no. Aug, pp. 115–143, 2002.

[11] G.-B. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, "Minimal gated unit for recurrent neural networks," *International Journal of Automation and Computing*, vol. 13, no. 3, pp. 226–234, 2016.

[12] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, pp. 2222–2232, Oct 2017.

[13] N. Kalchbrenner, A. v. d. Oord, K. Simonyan, I. Danihelka, O. Vinyals, A. Graves, and K. Kavukcuoglu, "Video pixel networks," *arXiv preprint arXiv:1610.00527*, 2016.

[14] C. Finn, I. Goodfellow, and S. Levine, "Unsupervised learning for physical interaction through video prediction," in *Advances in neural information processing systems*, pp. 64–72, 2016.

[15] Y. Wang, M. Long, J. Wang, Z. Gao, and S. Y. Philip, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Advances in Neural Information Processing Systems*, pp. 879–888, 2017.

[16] K. T. P. D. L. library, "Keras documentation," 2018.

[17] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using LSTMs," in *International conference on machine learning*, pp. 843–852, 2015.

[18] N. Elsayed, A. S. Maida, and M. Bayoumi, "Empirical activation function effects on unsupervised convolutional LSTM learning," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 336–343, IEEE, 2018.

[19] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, pp. 273–278, IEEE, 2013.

[20] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning*, pp. 1764–1772, 2014.

[21] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[22] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," *arXiv preprint arXiv:1609.02745*, 2016.

[23] D. Eck and J. Schmidhuber, "A first look at music composition using LSTM recurrent neural networks," *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale*, vol. 103, 2002.

[24] F. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[25] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3156–3164, 2015.

[26] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," in *Advances in Neural Information Processing Systems*, pp. 1171–1179, 2015.

[27] M. Schtickzelle, "Pierre-François Verhulst (1804-1849). La première découverte de la fonction logistique," *Population (French edition)*, pp. 541–556, 1981.

[28] C. Gulcehre, M. Moczulski, M. Denil, and Y. Bengio, "Noisy activation functions," in *International Conference on Machine Learning*, pp. 3059–3068, 2016.

[29] J. Bergstra, G. Desjardins, P. Lamblin, and Y. Bengio, "Quadratic polynomials learn better image features," *Technical report, 1337*, 2009.

[30] Y. W. Teh and G. E. Hinton, "Rate-coded restricted Boltzmann machines for face recognition," in *Advances in neural information processing systems*, pp. 908–914, 2001.

[31] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia, "Incorporating second-order functional knowledge for better option pricing," in *Advances in neural information processing systems*, pp. 472–478, 2001.

[32] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUS)," *arXiv preprint arXiv:1511.07289*, 2015.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[36] J. M. Kilner, K. J. Friston, and C. D. Frith, "Predictive coding: an account of the mirror neuron system," *Cognitive processing*, vol. 8, no. 3, pp. 159–166, 2007.

[37] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature neuroscience*, vol. 2, no. 1, p. 79, 1999.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 3, 171-182 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

An Efficient Automotive Paint Defect Detection System

Sohail Akhtar[*,1], Adarsh Tandiya[2], Medhat Moussa[1], Cole Tarry[1]

[1]*The Robotic Institute, School of Engineering, University of Guelph, Guelph, ON, N1G 2W1, Canada*

[2]*Praemo Inc., Kitchener, ON, N2G 2Z3, Canada*

ARTICLEINFO

Article history:
Received: 25 March, 2019
Accepted: 23 May, 2019
Online: 12 June, 2019

Keywords:
Defect detection
Deflectometry
Profilometry
Automatic inspection
Painted surface inspection

ABSTRACT

Vision-based defect detection techniques are widely used for quality control purposes. In this work, an efficient deflectometry based detection system is developed for semi-specular/painted surface defect detection. This system consists of a robotic arm that carries a screen/camera setup and can detect defects on large surfaces with different topologies, such as a car bumper, by traversing its profile. A hybrid pipeline is designed that utilizes multi-threading for optimal resource utilization and process speed. Specific filters are also designed to remove spurious defects introduced by acute curvature changes and part edges. The system was successful in consistently detecting various defects on small test samples as well as on large bumper parts with varying topology and color and can accommodate inherent ambient lighting and vibration issues.

1 Introduction

Quality control is a crucial factor in manufacturing industry as it affects customer satisfaction, reduces production cost and increases profitability. In the automotive industry, vehicles are usually assembled from parts shipped by various original equipment manufacturers (OEM) to the assembly plant. Example of these parts includes front and back bumper covers, side fenders, and other exterior parts which are normally manufactured and painted to specific colour before being shipped to another plant for assembly. These outer body parts are made by Thermoplastic PolyOlefin (TPO) injection moulding process. This process consists of three main steps: moulding, cleaning, and painting. Defects may induce during any of these processes, which results in part rejection or rework, causing loss of revenue. As such, it is essential to perform a full inspection of every part. This inspection process is usually carried out by human inspectors. It is a costly and a labour intensive job which requires multiple inspection lines for high volume yield. Further, the defect judgement is very subjective, which results in inconsistencies. As a result, overall productivity and quality are diminished. This paper presents a system for inspecting automotive painted semi-specular exterior body parts and is an extension of work originally presented in 15th Conference on Computer and Robot Vision (CRV 2018) [1]. This extended version includes expended testing and analysis.

There are a series of challenges that impact the development of an inspection system for this task, including:

1. Parts are in motion while they are inspected on the production line. The vibrations induced from this motion makes profiling the surface harder.
2. The inspection process must fit within the existing production cycle time.
3. The part being inspected vary in curvature, size, shape and material, with different specular characteristics.
4. The visibility characteristic of the defect depends on the external lighting conditions. Designing a proper lighting environment for such a system is a challenging task due to high reflection coefficient of the test surface [2].
5. Each type of defect varies in shape and size, and experienced inspectors even miss some defects.

In recent years, vision-based surface inspection systems have found a burst of application in areas such as defect detection in aluminium sheets [2], locomotive

[*]Corresponding Author: Sohail Akhtar, University of Guelph, 50 Stone Rd. E, Guelph, ON, soakhtar@uoguelph.ca

https://dx.doi.org/10.25046/aj040323

rail tracks [3], liquid crystal displays (LCDs) [4], plastic objects [5], sphere parts [6], agricultural food produce [7], fabric and textile industry [8] etc. However, defect detection and surface profiling of a specular surface remains a challenging problem due to the reflective nature of the surface. Profilometry is a widely used approach for characterization, reconstruction and inspection of such surfaces. Profilometry approaches can be generally divided into contact and non-contact methods. Contact-based profilometry, such as stylus profilometer, scanning tunnelling microscopy, scanning force microscopy, coordinate measuring machine (CMM), etc., is a high-resolution method. The measurement accuracy is in the order of nanometers and is considered as a gold standard for surface finish measurement. This approach is independent of surface characteristics or shape and also works well in dirty environments. As a result, it can be used for both specular and non-specular surface measurement. However, being a direct contact method, it may damage high finish inspection surfaces such as painted automotive parts and its operating speed is also sluggish.

Non-contact profilometry methods are used for high-speed surface scanning and 3D reconstruction [9]. Over the years, many different non-contact profilometry techniques are developed, such as common structured light projection, phase shifting interferometry, deflectometry, digital holographic microscopy, etc. These approaches are based on the inhomogeneous reflection of light from defects. The visibility of defects can be enhanced by the use of a structured light source. In common structured light projection method, the specular surface is coated with a thin layer of powder to make it behave like a diffused surface [10] and then fringe pattern profilometry is used for analysis. This additional coating changes the surface geometry so cannot be reliably used for surface defect detection. On the other hand, interferometry is highly accurate but requires a reference and is not suitable for complicated free-form surfaces. Further, the field of view is limited, so it cannot be used for large surfaces. Hence, deflectometry is recently used for measuring a large object with varied surface topology. A sinusoidal fringe pattern is widely used as a structured light source in deflectometry analysis [11]. In smooth defect-free regions, imaging process obeys specular reflection while the incident rays are dispersed in defected areas. As a result, an acute gradient variation is observed in defect bearing regions, which is used for defect segmentation. These approaches have a lower measurement accuracy (in micrometres range) compared to contact profilometry approaches. Since these approaches are surface dependent, so are affected by the testing environment as any dirt or external dust particles will affect the obtained results.

A new defect detection system is presented in this paper, which is based on the deflectometry principle. The developed system can overcome the practical challenges of a production line. Controlled lighting and camera configurations are designed to counterpoise vibration effects. Various filters are used to segment the region of interest and eliminate false defects due to edge effect and abrupt curvature changes. Finally, to track parts while going through the production line, the system is designed to be mounted at the end of a robotic arm. The proposed system is validated by field testing in an operational automotive painting booth.

The remainder of this paper is organized as follows: Section 2 provides a brief review of the deflectometry principle as it is the foundation of the defect detection system. Section 3 describes the proposed system and explains the working of its different components. Section 4 depicts the system setup and presents implementation details. Section 5 archives the obtained experimental results with analysis. Finally, Section 6 summarizes the findings of this research and highlights some areas of future development.

2 Deflectometry Principle

In this section, we will provide a brief overview of the deflectometry principle, which forms the basis of the proposed system's working principle. The use of deflectometry is actively explored by researchers for the measurement of objects with an abrupt slope change or large size. Many different deflectometry methods are proposed for surface profilometry and defect detection. Some of these include Moier deflectometry [12–16], Ronchi method [17, 18], laser scanning deflectometry [19–21], and phase measuring deflectometry (PMD) [11, 22–24].

PMD technique is used in this paper as it is highly accurate and provides continuous data for surface profiling. It has a large dynamic range and can give full-field measurements. This technique is first developed by Horneber et al. [25]. Over the years, researchers have modified PMD to measure different specular objects [11, 26, 27]. PMD is also successfully applied for specular and semi-specular surface defect detection [28–33]. However, various challenges arise while implementing deflectometry to inspect a larger part like a car bumper which are successfully addressed in the developed system.

2.1 Phase-shifting Deflectometry

In deflectometry, the topographical information of a specular surface is obtained by analysing the reflections of a structured light source. A sinusoidal fringe pattern, displayed on an LCD screen located at some distance from the specular test surface, acts as the light source. The attached camera setup captures the reflections of the deformed pattern from the test surface. By applying phase-shifting and phase-unwrapping algorithms, useful information is extracted from these deformed fringe patterns which are used for surface profile construction. A phase varying single frequency pattern is generally used for deflectometry analysis. A pixelated version of such a multi-phase single frequency pattern is defined as [34]:

$$f(x,y) = \frac{G}{2}\left[1 + \sin(2\pi f \frac{x}{p} + n\frac{2\pi}{7})\right] n = 1..7 \quad (1)$$

where, $f(x,y)$ is the displayed pattern intensity at (x,y) pixel location, G is the maximum of the image intensity range, f is the frequency of the displayed pattern, p is the total pixels in a time period and n is the number of phase-shifted patterns selected.

These phase-shifted patterns are one-by-one displayed on the LCD screen, and their reflections are captured from the specular surface. The intensity of the captured pattern is given by (2) [35].

$$I(x,y) = A(x,y) + B(x,y)cos(\phi(x,y)) \quad (2)$$

where $I(x,y)$ is the captured image intensity, $A(x,y)$ is the ambient light intensity, $B(x,y)$ is the amplitude of the modulated fringe pattern and $\phi(x,y)$ is the desired phase measurement.

The measured phase is directly related to the surface topology. Computing this phase at each pixel location will create a surface phase map. Captured surface reflections of the phase-shifted patterns are used to calculate this phase map by applying Windowed Discrete-Fourier Transform (WDFT). For seven phase-shifted single frequency patterns, the closed-form solution is represented by the following relation [36]:

$$\phi = \tan^{-1}\left[\frac{(I_1 - I_7) - 3(I_3 - I_5)}{4(I_4) - 2(I_2 + I_6)}\right] \quad (3)$$

where $I_{1,\ldots,7}$ correspond to the captured reflections of the seven phase-shifted patterns displayed on the LCD screen. The phase map ($\phi(x,y)$) is in $0 - 2\pi$ range, so a phase unwrapping step is needed to recover the actual phase with the inclusion of appropriate multiples of 2π [37]. Then the absolute derivative of the phase map is computed with respect to pattern variation axis (x-axis) as defined by the following equation [33]:

$$D_y(x) = \left|\frac{\mathrm{d}\left|\phi(x)\right|}{\mathrm{d}x}\right| \quad (4)$$

here, $D_y(x)$ is a continuous function which represents line topography for the corresponding y^{th}-axis.

The effect of this differential operation is the attenuation of low-frequency signal information and accentuation of high-frequency contents induced due to the presence of defects in the specular/semi-specular surface. As a result, $D_y(x)$ remains almost constant in defect-free regions, whereas an acute variation is observed in defect-bearing areas.

3 Defect Detection System

Use of deflectometry for big automobile bumper inspection brings its own set of challenges. Its impractical to examine the entire part surface in one scan. Therefore, it is divided into smaller segments based on the reflection of the displayed pattern. Since external light sources impact the detection results, controlled lighting setup is constructed to minimize the effect of these external factors. Figure 1 shows the image processing pipeline used. Captured surface reflections are pre-processed before deflectometric defect detection operation. Several false positives are observed due to the topographical variation in the test surface. A set of spurious defect removal filters are designed to eliminate these false positives. Camera shutter speed, aperture and robotic motion are adjusted to reduce vibration effects and eradicate motion blur. Since the detection operation is carried out in small segments, the results are then combined and localized on test parts. Finally, defect characteristics are gathered and archived in the designed database which can later be used for display or higher level decision making for process improvement. The following sections provide additional details on the functionality of different components.

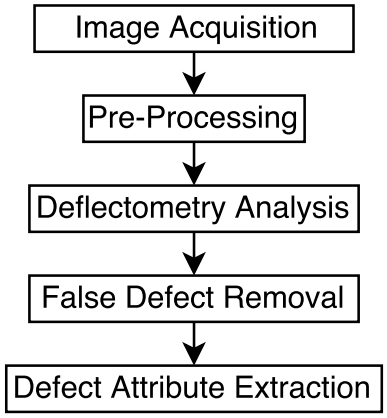


Figure 1: Defect detection scheme

3.1 Capturing and Pre-processing

Figure 2 and Figure 3 show the process involved in capturing and pre-processing of images. OpenGL library is used for the rendering of seven phase-shifted sinusoidal patterns. The patterns are generated only once during the run-time using pattern screen resolution and then stored in memory as 2D matrices. Later on, these stored patterns are displayed one-by-one for deflectometry analysis. The reflection of these patterns is captured from the part surface by an attached camera. Pattern display and capture of surface reflections are carried out sequentially. Once a set of seven surface reflections is captured, a pre-processing thread is initiated to compute the phase map and find its derivative. It is noted that a single phase-map-derivative image is created for each segment. While the pre-processing step is running, the next set of seven patterns are ready to be displayed on the screen and captured by the camera. This multi-threading approach reduces cycle time through parallel computation.

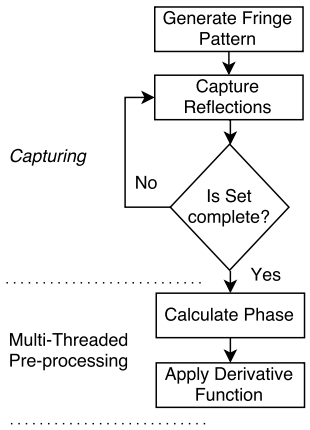


Figure 2: Sequential image capture and mutli-threaded pre-processing steps

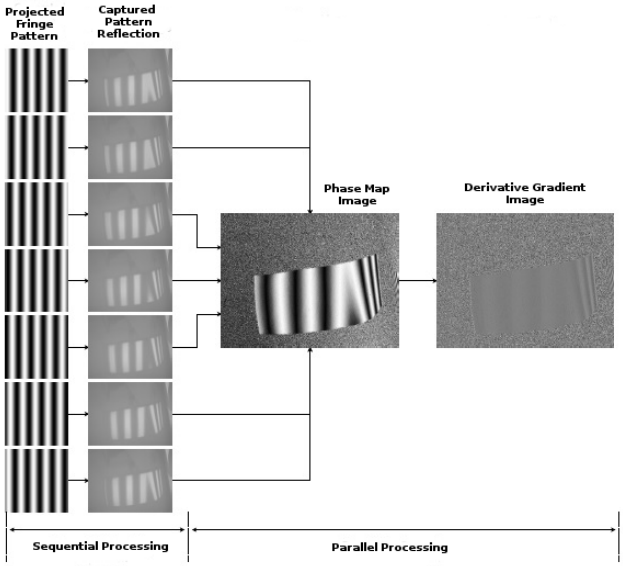


Figure 3: Surface reflection capturing and phase-map-derivative image generation

Once the Capturing and Pre-processing steps are completed, a defect detection thread is initiated, which extracts defect information from the phase-map-derivative image. Defect detection process can be divided into three main stages; Fringe region segmentation, Defect pool and spurious defect elimination and Defect attributes extraction and registration.

3.2 Fringe Region Segmentation

A fixed focus camera at a defined distance from the test surface is used in this study. As a result, the camera captures a fixed size image that may contain both the fringe projected region and non-fringe region. Further, the size of the reflected pattern varies with the surface curvature; the reflected pattern shrinks as the surface become convex. As a result, the first step in the defect detection process is to crop the fringe pattern projected region from the phase-map-derivative image.

The obtained phase-map-derivative image in the previous step is thresholded and smoothed with a Gaussian kernel. The resulted blobs are eroded to remove uneven corners and then selected based on size to create a mask. This mask is used to extract the fringe projected region as shown in Figure 4, which is later used for defect detection.

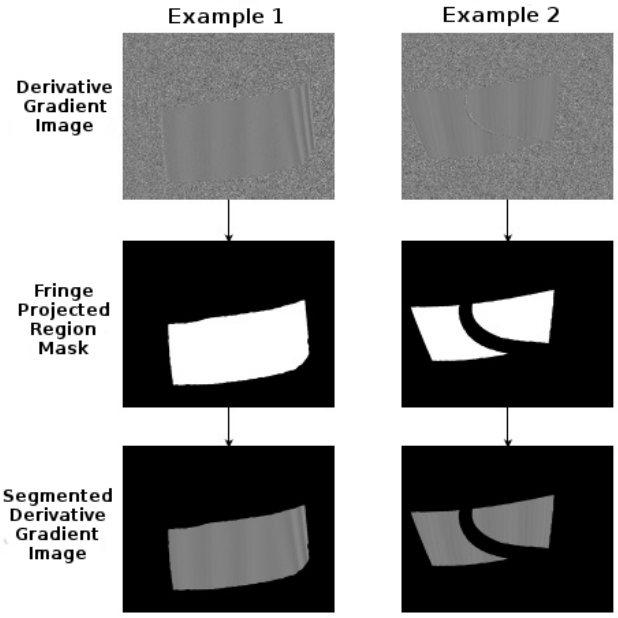


Figure 4: Fringe region extraction from the generated phase-map-derivative image

3.3 Defect Pool and Spurious Defect Elimination

Once the fringe projected region is extracted, thresholding and edge detection are used to create a defect pool. This defect pool consists of possible true and false defects. False defects are a combination of tiny defects below the accepted size criterion, edge defects and defects due to abrupt curvature changes. Three different filters; Size/Noise filter, Surface curvature filter and Edge filter are designed to remove these false defects. These filters work in parallel to remove these defects as depicted in Figure 5.

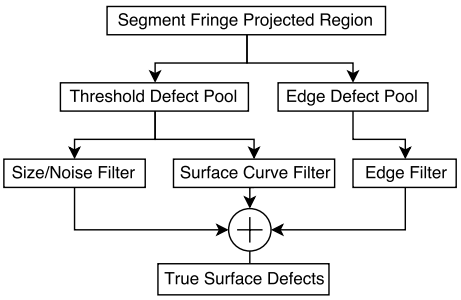


Figure 5: Processing and false defect removal

A size based filter is designed to remove tiny noise like defects. It is observed that after thresholding, several very small defects are detected in the binary image. This filter removes all those defects which are below the given size criterion. The second filter removes the false defects from the curved region. This filter uses parallel lines to determine the curved region in

the captured surface reflection. A Canny edge detection operation followed by Hough transform is used for these lines detection. The angle variation of the detected lines is exploited to determine the curved region from where the defects are removed. The last filter removes spurious defects in the edge region, which are the result of the application of WDFT. It makes use of both curved region information and defect pool generated by Canny edge detection operation to segregate defects produced due to edge effect.

The application of these filters results in three separate images that contain only legitimate defects. Later on, all these defects are aggregated together to form an image which shows all the valid defects in the inspected region as depicted in Figure 5. Application of these filters on a typical gradient image is shown in Figure 6.

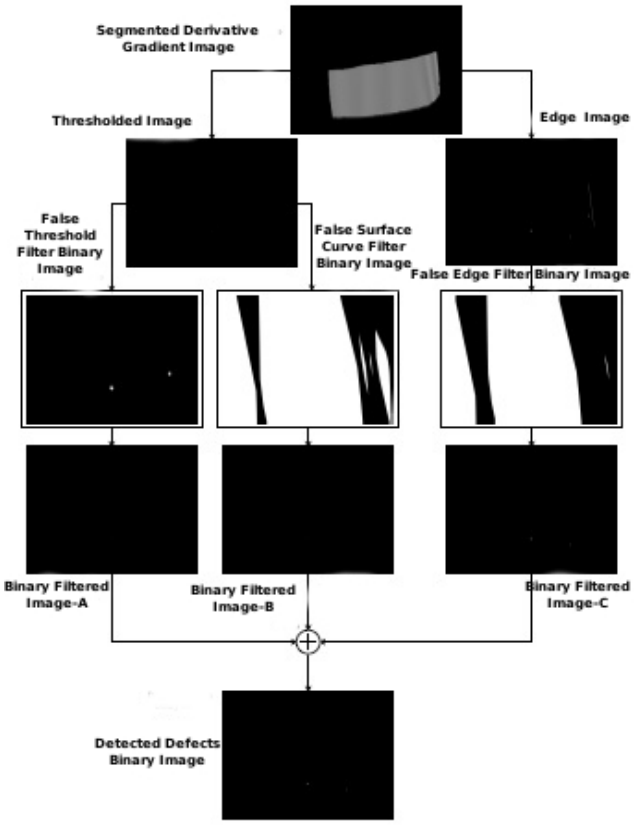


Figure 6: Application of false defect filters on defect pool

3.4 Defect Attributes Extraction and Registration

Once all the true defects are found, defect attributes such as size (in terms of pixels), bounding region and centroid location are determined. To extract these characteristics, close by defects are first fused by a morphological dilation operation. Later on, a contour-finding operation is performed to separate different defects and then above mentioned attributes are computed for each found contour.

These attributes are also recorded in the appended database, which can be used in future for tracking, decision making or other process adjustments. The results are also overlaid on a part image for physical defect localization which is saved in a directory for immediate use. All these different steps are delineated in Figure 7.

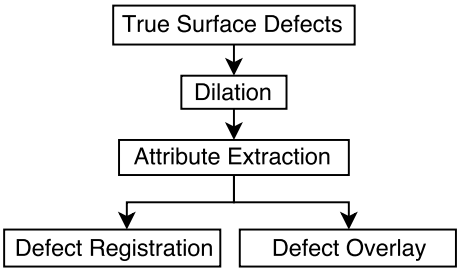


Figure 7: Defect attributes extraction and display

4 Implementation Details and Experimental Setup

4.1 System Architecture

Figure 8 shows the entire system implementation. Since the system is expected to inspect large parts, a robotic arm is used where the pattern screen and camera are mounted on the end-effector. The system includes several additional components to enable robot movement for full part inspection such as positioning of the robot end effector, generation of inspection patterns and administrative operations such as results registration and display.

The defect detection process consists of both sequential and parallel operations. As a result, a hybrid computational framework is designed. It is evident in Figure 8 that individual steps need to be completed to proceed to the next component for processing, so parallel operations (multi-threading) are only used for computationally expensive processes. First of all, robot positions the LCD pattern screen tangent to the surface to be inspected. Then seven phase-shifted sinusoidal fringe patterns are sequentially displayed on the screen, and the attached camera captures corresponding surface reflections. These surface reflections are pre-processed, and defect detection operation is applied to find the surface abnormalities. These two steps, pre-processing and defect detection, are computationally expensive, and multi-threading is used here. This multi-threading operation can be activated/deactivated by the frontal administrative control panel. All detected defect information is stored in the designed database system that can be accessed remotely. All these steps are controlled by a front end administrative control panel, which also displays a sample test part image on which detected results are overlaid. It also provides the option to configure features such as directory paths for registering captured and processed images, the number of segments to divide the test surface and displays the progress of the detection operation.

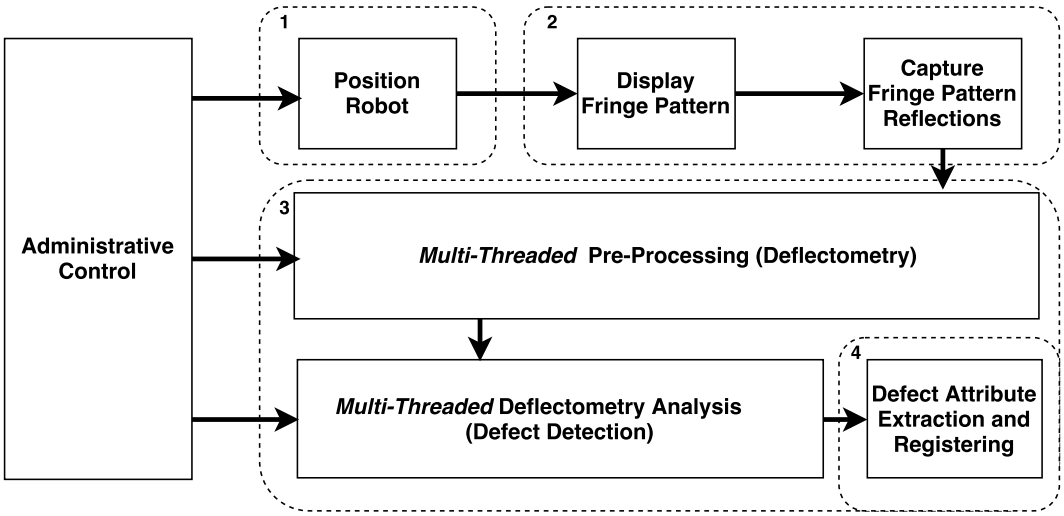


Figure 8: Schematic of implemented defect detection system

4.2 Experimental Setup

4.2.1 Overall Setup

Testing was conducted at the Robotics Institute @ Guelph lab. The experimental setup is shown in Figure 9. The test parts were placed on a carrier similar to the ones used at the part manufacturing facility. The carrier was stationary during the testing. A KUKA KR 16-2 KS industrial robotic arm is used to carry the screen/camera setup to transverse the large bumpers under test. The robotic arm was programmed to move in a predefined path to inspect each segment and hence the whole part. This operation controls the movement of the screen/camera setup. A controlled lighting environment was enforced where the pattern screen is the only source of illumination. Any direct light source incident on the part whose reflection reaches the camera will create a blind spot in the gradient image. The test surface should also be free of any dust particles as they will produce a dispersed reflection of light and will appear as a defect. A 40 inch commercial TV screen is used as the pattern screen. The computing system consists of an Intel Core i7-6700 3.40 GHz processor and has 16 GB RAM and Nvidia GTX960 graphics card running on 64-bit Windows 10 OS. The detection code is written in C++ using Qt Creator 5.6, OpenCV 3.2 and OpenGL libraries. The camera is connected via high-speed USB 3.0 connection and pattern screen is interfaced with an HDMI cable.

4.2.2 Camera Setup

A monochrome Point Grey 3.2 MP (GS3-U3-32S4M-C) camera with a 12.5 mm Fujinon lens HF12.5HSA-1 (LENS-50FS-125C) is used. The camera is attached to a specially designed fixture on the side of the pattern screen so that it looks in the middle of the projected pattern on the test sample, as seen in Figure 10. A fixed focus camera, as already mentioned, is sharply focused on the test surface, which is approximately 40 cm away from the camera. Once the robot starts traversing the surface topology, a distance of 35 cm to 45 cm is maintained so that the image remains within the camera's depth of filed and hence sharply focused.

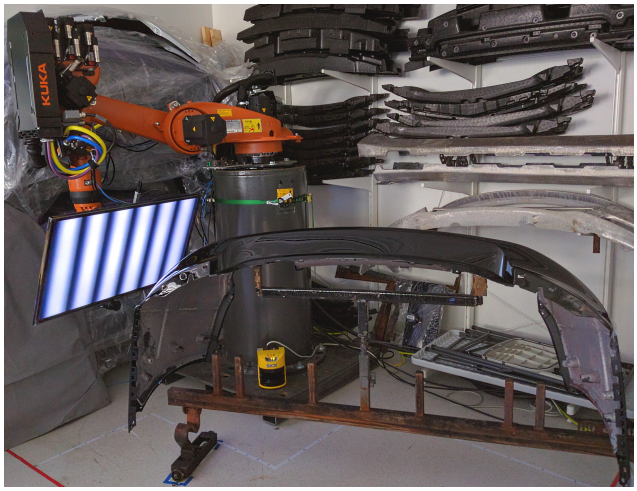


Figure 9: Experimental setup



Figure 10: Camera view

4.3 Surface Segmentation

It is virtually impossible to inspect the full test part in one capture due to its large size. There are two reasons for this. Firstly, there is a lot of variation in the surface topology of the test surface, and the pattern screen should be tangent to it to capture the displayed pattern. Secondly, only a limited payload both in terms of size and weight can be attached to the robot end effector. Hence, a huge screen which can project the pattern on the full part cannot be used. As a result, the test surface is partitioned in small regions based on the reflected pattern as shown in Figure 11.

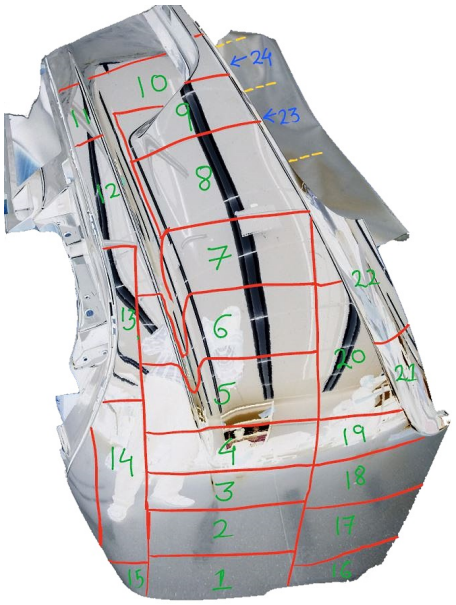


Figure 11: Bumper segmentation for inspection

Seven phase-shifted sinusoidal patterns are projected on each region, and the attached camera captures corresponding reflections. There is some overlap in the captured surface reflections among adjacent segments that accommodate for some variation in part placement. These different regions are then appended together to localize the identified defects on the entire test surface.

4.4 Test Specimen

Ten test samples are used in this study. These samples are provided by an OEM exterior body parts manufacturer. These are large SUV/Minivan bumpers of 5.5 ft long, 1.5 ft high and 2.5 ft deep, as shown in Figure 11. These bumpers are of different colours and are rejected due to the presence of various surface defects. These defects are marked on the test parts by expert human inspectors working at the manufacturing facility. In addition to these defects, some more defects are artificially created to augment the defect database used as ground truth (GT). Since the bumpers are symmetric, we only needed to inspect one half of each bumper. In real production, two robots can be used to examine the whole bumper. Each tested part is divided into 24 small segments as depicted in Figure 11.

5 Results and Analysis

Two sets of experiments are performed to evaluate the performance of the developed painted surface defect detection system. In the first study, small defect samples are used to verify the effectiveness of the deflectometry principle in detecting various defects. Once, the efficacy of the deflectometry principle is validated on these small samples; large parts are tested. Large parts have their own set of challenges in terms of size, sudden curvature change, throughput etc.

5.1 Small Test Samples Preliminary System Validation

Different type of surface defects such as dirt, cold sludge, fisheye, blister etc. occur during moulding or painting process. Small test samples bearing these defects are used to validate the effectiveness of the detection principle. Some of the sample parts are shown in Figure 12. These small parts are mounted on a tripod in front of pattern screen, as shown in Figure 13. The parts are placed 40 cm away from the camera, which captures the surface reflections. A set of seven sinusoidal fringe patterns with a phase shift of $2\pi/7$ are generated to display on the LCD screen and then their reflections are captured. These surface reflections are analysed to reconstruct the surface profile. During this testing, the fringe projected area is manually segmented for deflectometry analysis. This process is depicted in Figure 14 and the results are tabulated in Table 1. Here, detection accuracy represents whether a defect is successfully detected by the developed system.

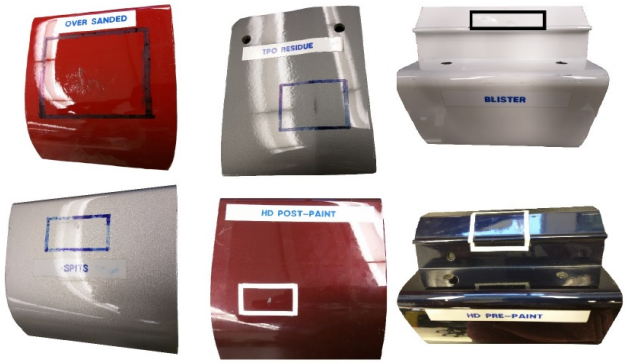


Figure 12: Some typical surface defect samples



Figure 13: Camera with fringe pattern screen and test part on tripod

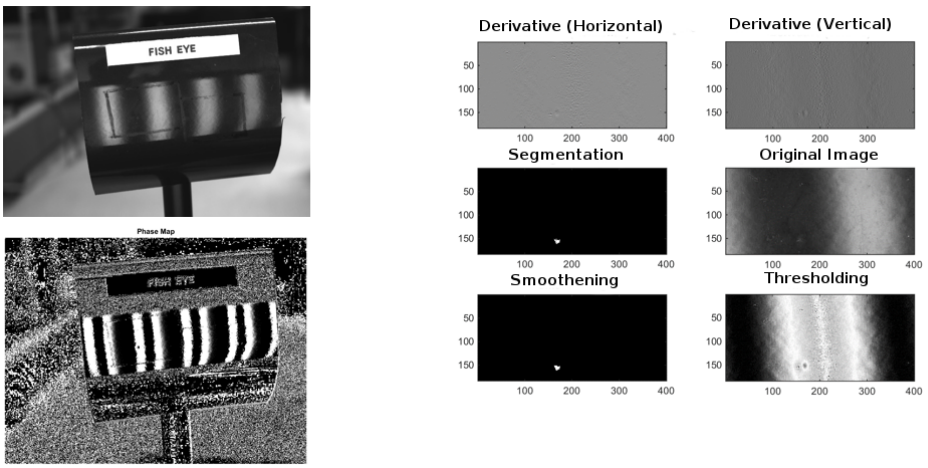


Figure 14: Fisheye defect detection

Table 1: Small sample defect detection results

Sr. No.	Defect Type	No. of Samples	Detection Accuracy
1.	Dirt	2	100%
2.	Cold Slug	4	75%
3.	Spits	1	100%
4.	Fish Eye	2	100%
5.	Flash	2	100%
6.	Blister	3	100%
7.	TPO Protusion	2	100%
8.	Over Sanded	3	100%
9.	Hot Pre-paint	1	100%
10.	Overflame Torch	2	100%
11.	Water Spots	1	100%
12.	TPO Residue	3	100%
13.	Popping	1	50%
14.	HD Post Paint	3	100%

These results indicate that the developed system mostly detects the studied defects although some low detection performance is observed for cold slug and popping defects. Popping defects are salt-and-pepper noise like small defects that occur when the solvent or air trapped in paint film escapes during the drying process. This creates a lot of tiny defects clustered in a small area. A 50% popping defect detection accuracy indicates that approximately half of these tiny defects are detected, but it is enough to declare the part defective.

5.2 Large Parts Defect Detection

Following the experimental setup outlined in section 4, various large bumpers are tested by the developed system. The developed framework was successful in detecting most of the manufacturing as well as the artificially created defects. Some of these detection results are shown in Figure 15.

Results are reported in two tables. The segment-wise results are given in Table 2 while body part color based results are reported in Table 3. These results are reported in terms of recall, precision and F-measure. Equal weighting of recall and precision is considered for computing F-measure. These metrics are defined as follows:

$$Recall = \frac{TP}{(TP+FN)} \tag{5}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{6}$$

$$F-measure = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{7}$$

where TP is true positive, FP is false positive and FN is false negative. Ground truth data to compute these measures is provided by the part manufacturer.

By reviewing the results, we can deduce that the performance of the defect detection system varies depending on part colour and segment location. The performance of the algorithm is lower in segments 8, 9 and 10 where precision drops significantly to 0.46, 0.43 and 0.30, respectively. These segments contain highly curved sections which drastically distorts the projected pattern resulting in a higher number of FP. This can be improved by tuning the system differently on a per region basis to accommodate the different variations in surface curvature. The results by part colour show that the performance of the detection system is compromised on white parts. White surfaces reflect all light incident to it, so the algorithm becomes more sensitive to surrounding light. Besides, the contrast of the projected pattern is muted by white parts affecting the visibility of defects on the part surface. The combination of these factors reduces the accuracy of the algorithm on white parts.

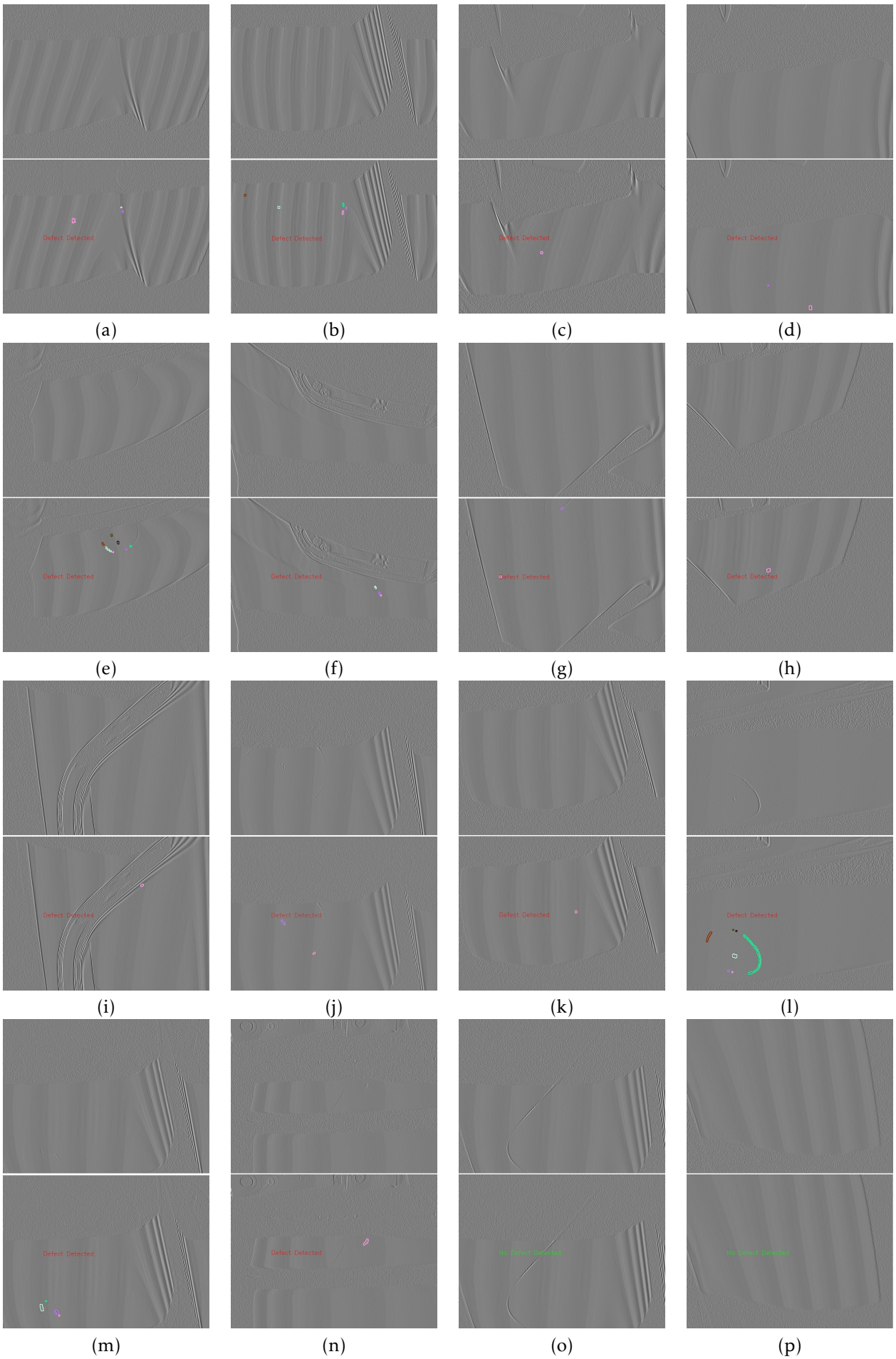


Figure 15: Some defect detection results on large bumpers First row: Gradient image, Second row: Detected defect ((o) is an example where defect is not detected as it is not enclosed by fringe region and (p) is an example of a defect free region)

Table 2: Defect detection results by segment

Seg.	Samples	*TP*	*FP*	*FN*	Recall	Precision	F-measure
1	10	6	4	0	1.00	0.60	0.75
2	10	0	0	0	NA	NA	NA
3	10	6	2	1	0.86	0.75	0.80
4	10	3	1	0	1.00	0.75	0.86
5	10	3	1	0	1.00	0.75	0.86
6	10	14	0	1	0.93	1.00	0.96
7	10	8	0	1	0.89	1.00	0.94
8	10	6	7	0	1.00	0.46	0.63
9	10	3	4	0	1.00	0.43	0.60
10	10	6	14	2	0.75	0.30	0.43
11	10	0	3	1	0.00	0.00	NA
12	10	2	0	0	1.00	1.00	1.00
13	10	2	0	0	1.00	1.00	1.00
14	10	4	0	0	1.00	1.00	1.00
15	10	3	0	0	1.00	1.00	1.00
16	10	3	0	0	1.00	1.00	1.00
17	10	0	0	2	0.00	NA	NA
18	10	1	0	0	1.00	1.00	1.00
19	10	2	1	0	1.00	0.67	0.80
20	10	3	2	1	0.75	0.60	0.67
21	10	0	0	0	NA	NA	NA
22	10	7	1	0	1.00	0.88	0.94
23	10	7	0	0	1.00	1.00	1.00
24	10	0	0	0	NA	NA	NA
Average					**0.87**	**0.76**	**0.81**

Table 3: Defect detection results by color

Part	Color	TP	FP	FN	Recall	Precision	F-measure
1	Black	14	0	0	1.00	1.00	1.00
2	Dark blue	3	1	1	0.75	0.75	0.75
3	Dark Blue	6	1	0	1.00	0.86	0.92
4	Dark silver	9	2	2	0.82	0.82	0.82
5	Dark blue	0	1	1	0.00	0.00	NA
6	Purple	16	1	2	0.89	0.94	0.91
7	Dark silver	9	2	1	0.90	0.82	0.86
8	White	5	19	2	0.71	0.21	0.32
9	White	12	9	0	1.00	0.57	0.73
10	Black	15	4	0	1.00	0.80	0.89

6 Conclusion

A deflectometry based defect detection system is presented in this research paper. The developed system has successfully detected large and subtle defects on large painted automotive parts. The system analyses gradient variation in the measured phase for defect detection and does not require an existing dataset or pre-training for its functioning. The use of robotic arm lends it the capability to inspect varying surface geometries. The system examines the whole part by operating on individual segments. These defects are then aggregated to localize them on the full part.

The system successfully detected various manufacturing defects induced during moulding and painting processes on small test samples as well as on large automotive bumpers. For large bumper testing, additional algorithms are designed and used to extract the region of interest, as well as eliminate spurious defects detected due to edge effects and abrupt curvature changes. The algorithm processing takes approximately 0.35 seconds per segment, and this doesn't include the time that it takes the robotic arm to move between segments.

It is noted that the implemented framework can only detect defects if they are enclosed by a fringe projected region from all sides. It is possible to use multiple cameras to reduce the number of segments and speed-up the inspection process. A colour fringe pattern could potentially be used to improve the system's performance on light coloured parts. Use of a phase shifted coloured pattern can be explored as it will alleviate the requirement for a stationary part-camera/screen system by eliminating the need for multiple phase-shifted patterns.

Acknowledgement This work is supported by the Natural Science and Engineering Research Council (NSERC), Canada.

References

[1] A. Tandiya, S. Akthar, M. Moussa, and C. Tarray, "Automotive semi-specular surface defect detection system," in 2018 15th Conference on Computer and Robot Vision (CRV), May 2018, 285-291. DOI: 10.1109/CRV.2018.00047

[2] C. Fernandez, C. Platero, P. Campoy, and R. Aracil, "Vision system for on-line surface inspection in aluminum casting process," in Proceedings of IECON '93 - 19th Annual Conf. of IEEE Industrial Electronics, 15-19 Nov. 1993, 1854 - 1859. DOI: 10.1109/IECON.1993.339356

[3] Z. Liu, W. Wang, X. Zhang, and W. Jia, "Inspection of rail surface defects based on image processing," CAR 2010 - 2010 2nd Int. Asia Conf. Informatics Control. Autom. Robot., **vol. 1**, pp. 472-475, 2010. DOI: 10.1109/CAR.2010.5456793

[4] Y.-G. Cen, R.-Z. Zhao, L.-H. Cen, L. Cui, Z. Miao, and Z. Wei, "Defect inspection for TFT-LCD images based on the low-rank matrix reconstruction," Neurocomputing, **vol. 149**, 1206-1215, 2015. DOI:10.1016/j.neucom.2014.09.007

[5] B. Liu, S. Wu, and S. Zou, "Automatic detection technology of surface defects on plastic products based on machine vision," 2010 International Conf. on Mechanic Automation and Control Engineering, MACE2010, 2213 - 2216, 07 2010. DOI: 10.1109/MACE.2010.5536470

[6] J. Guo and W. Shao, "Automated detection of surface defects on sphere parts using laser and CDD measurements," IECON 2011 - 37th Annu. Conf. IEEE Ind. Electron. Soc., pp. 2666-2671, 2011. DOI: 10.1109/IECON.2011.6119732

[7] S. Janardhana, J. Jaya, K. J. Sabareesaan, and J. George, "Computer aided inspection system for food products using machine vision A review," 2013 Int. Conf. Curr. Trends Eng. Technol., pp. 29 - 33, 2013. DOI: 10.1109/ICCTET.2013.6675906

[8] P. M. Cho CS, Chung BM, "Development of real-time vision based fabric inspection system," IEEE Trans. Ind. Electronics, **vol. 52**(4), 1073–9, 2005.

[9] S. V. der Jeught and J. J. Dirckx, "Real-time structured light profilometry: a review," Optics and Lasers in Engineering, **vol. 87**, 18 – 31, 2016, Digital optical & Imaging methods in structural mechanics. https://doi.org/10.1016/j.optlaseng.2016.01.011

[10] D. Palouek, M. Omasta, D. Koutny, J. Bednar, T. Kouteck, and F. Dokoupil, "Effect of matte coating on 3D optical measurement accuracy," Optical Materials, **vol. 40**, 02 2015. https://doi.org/10.1016/j.optmat.2014.11.020

[11] M. C. Knauer, J. Kaminski, and G. Hausler, "Phase measuring deflectometry: a new approach to measure specular free-form surfaces," in Optical Metrology in Production Engineering, ser. procspie, W. Osten and M. Takeda, Eds., **vol. 5457**, 366-376, Sep. 2004. https://doi.org/10.1117/12.545704

[12] I. G. O. Kafri, "Moire deflectometry: A ray deflection approach to optical testing," Optical Engineering, **vol. 24**, no. 6, 944 - 960, 1985. [Online]. Available: https://doi.org/10.1117/12.7973607

[13] M. Servn, R. Rodriguez-Vera, M. Carpio, and A. Morales, "Automatic fringe detection algorithm used for Moire deflectometry," Applied optics, **vol. 29**, 3266 - 70, 08 1990.

[14] B. Wang, X. Luo, T. Pfeifer, and H. Mischo, "Moire deflectometry based on Fourier-transform analysis," Measurement, **vol. 25**, 24 - 253, 06 1999. https://doi.org/10.1016/S0263-2241(99)00009-3

[15] A. E.-R. Ricardo Legarda-Senz, "Wavefront reconstruction using multiple directional derivatives and Fourier transform," Optical Engineering, **vol. 50**, no. 4, 1 – 4 , 2011. [Online]. Available: https://doi.org/10.1117/1.3560540

[16] D. Fontani, F. Francini, D. Jafrancesco, L. Mercatelli, and P. Sansoni, "Mirror shape detection by "Reflection Grating Moire Method" with optical design validation," Proceedings of SPIE - The International Society for Optical Engineering 5856, Jun 2005. DOI: 10.1117/12.612114

[17] S.-W. K. Ho-Jae Lee, "Precision profile measurement of aspheric surfaces by improved Ronchi test," Optical Engineering, **vol. 38**, no. 6, 1041 - 1047, 1999. [Online]. Available: https://doi.org/10.1117/1.602147

[18] G. P. Butel, G. A. Smith, and J. H. Burge, "Binary pattern deflectometry," Appl. Opt., **vol. 53**, no. 5, 923 - 930, Feb 2014. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-53-5-923

[19] J. L. H. Willem D. van Amstel, Stefan M. B. Baumer, "Optical figure testing by scanning deflectometry," Proc. SPIE 3739, Optical Fabrication and Testing, 6 September 1999. doi: 10.1117/12.360155, Available: https://doi.org/10.1117/12.369201

[20] S. Krey, W. D. van Amstel, K. Szwedowicz, J. Campos, A. Moreno, and E. J. Lous, "Fast optical scanning deflectometer for measuring the topography of large silicon wafers," Proc. SPIE 5523, Current Developments in Lens Design and Optical Engineering V, 14 October 2004. doi: 10.1117/12.559702, https://doi.org/10.1117/12.559702

[21] K. Ishikawa, T. Takamura, M. Xiao, S. Takahashi, and K. Takamasu, "Profile measurement of aspheric surfaces using scanning deflectometry and rotating autocollimator with wide measuring range," Measurement Science and Technolog, **vol. 25**, no. 6, Apr 2014. [Online]. Available: https://doi.org/10.1088%2F0957-0233%2F25%2F6%2F064008

[22] Y. Tang, X. Su, Y. Liu, and H. Jing, "3D shape measurement of the aspheric mirror by advanced phase measuring deflectometry," Opt. Express, **vol. 16**, no. 19, 15090 - 15096, Sep 2008. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-16-19-15090

[23] D. Malacara, Optical Shop Testing (Wiley Series in Pure and Applied Optics), Wiley-Interscience, 2007.

[24] A. V. Maldonado, P. Su, and J. H. Burge, "Development of a portable deflectometry system for high spatial resolution surface measurements," Appl. Opt., **vol. 53**, no. 18, 4023 - 4032, Jun 2014. https://doi.org/10.1364/AO.53.004023

[25] C. Horneber, M. Knauer, and G. Husler, "Phase measuring deflectometry - a new method to measure reflecting surfaces," Annual Report Optik, 01 2000.

[26] J. Xiao, X. Wei, Z. Lu, W. Yu, and H. Wu, "A review of available methods for surface shape measurement of solar concentrator in solar thermal power applications," Renewable and Sustainable Energy Reviews, **vol. 16**, no. 5, 2539 - 2544, 2012. https://doi.org/10.1016/j.rser.2012.01.063

[27] C. E. Andraka, S. Sadlon, B. Myer, K. Trapeznikov, and C. Liebner, "Rapid reflective facet characterization using fringe reflection techniques," Proceedings of the Energy Sustainability, 19 - 23, 2009. DOI: 10.1115/1.4024250

[28] C. Devivier, F. Pierron, and M. Wisnom, "Damage detection in composite materials using deflectometry, a full-field slope measurement technique," Composites Part A: Applied Science and Manufacturing, **vol. 43**, no. 10, 1650 - 1666, 2012. https://doi.org/10.1016/j.compositesa.2011.11.009

[29] J. Geng, "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photon., **vol. 3**, no. 2, 128–160, Jun 2011. https://doi.org/10.1364/AOP.3.000128

[30] M. Kujawinska and J. Wjciak, "High accuracy Fourier transform fringe pattern analysis," Optics and Lasers in Engineering, **vol. 14**, no. 4, 325 - 339, 1991. https://doi.org/10.1016/0143-8166(91)90056-Y

[31] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am., **vol. 72**, no. 1, 156 - 160, Jan 1982. https://doi.org/10.1364/JOSA.72.000156

[32] L. Armesto, J. Tornero, A. Herraez, and J. Asensio, "Inspection system based on artificial vision for paint defects detection on cars bodies," in Robotics and Automation (ICRA), 2011 IEEE International Conference on, 1–4, 9-13 May 2011. DOI: 10.1109/ICRA.2011.5980570

[33] C. Tarry, M. Stachowsky, and M. Moussa, "Robust detection of paint defects in moulded plastic parts," Proc. - Conf. Comput. Robot Vision, CRV 2014, 306–312, 2014. DOI: 10.1109/CRV.2014.48

[34] T. Liu, C. Zhou, Y. Liu, S. Si, and Z. Lei, "Deflectometry for phase retrieval using a composite fringe," Optica Applicata, **vol. 44**, no. 3, 451 - 461, 2014. DOI: 10.5277/oa140309

[35] Y. Surrel, "Deflectometry: A simple and efficient noninterferometric method for slope measurement," Xth SEM Int. Congr. Exp. Mech., 2004.

[36] K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase shifting for nonsinusoidal waveforms with phase-shift errors," J. Opt. Soc. Am. A, vol. 12, no. 4, p. 761, 1995. https://doi.org/10.1364/JOSAA.12.000761

[37] E. Stoykova, J. Harizanova, and V. Sainov, "Pattern Projection Profilometry for 3D Coordinates Measurement of Dynamic Scenes," Three-Dimensional Telev., pp. 85–164, 2008. DOI: 10.1007/978-3-540-72532-9_5

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 366-379 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Evaluation of the Effects of Bidding Strategy with Customized Pricing on the Individual Prosumer in a Local Energy Market

Borislava Spasova[*,1,2], Daisuke Kawamoto[1], Yoshiyasu Takefuji[2]

[1]*Sony Computer Science Laboratories Inc.,141-0022, Japan*

[2]*Keio University, Graduate School of Media and Governance, 252-0882, Japan*

ARTICLE INFO

Article history:
Received: 31 May, 2018
Accepted: 27 July, 2019
Online: 19 August, 2019

Keywords:
Energy trading
Microgrid
Heterogeneous power sources

ABSTRACT

Peer-to-peer (P2P) energy trading is a mechanism that allows people to share locally the energy they have generated from distributed renewable resources (DER), to generate profit form the unused resources and to reduce the cost for electricity for the household and in the community itself. This calls for the design of new energy markets, accompanied with the development of comprehensive exchange strategies, which reflect both the consumer?s preferences and the heterogeneity of the renewable sources. The paper presents a study of the behavior patterns of individual prosumers using bidding strategies based on the State-of-Charge of the battery and two different pricing algorithm, one with fixed prices and one based on the battery price of each standalone system as well as the effect such patterns have on a local energy market with solar panels, fuels cells and batteries. The evaluation is achieved by the use of a prototype based on the Open Energy Systems (OES), a community in Okinawa, Japan, made of 19 interconnected houses with residential storage, photovoltaic cells and AC grid connection. In order to simulate heterogeneity, a fuel cell, modeled after Ene-Farm, is added to the original configuration. Each house has a power flow management system, which uses a priority-based algorithm to maintain demand-response efficiency, capable of scheduling the operating hours of the available fuels cells. The energy market is based on Zaraba, a continuous double auction algorithm used by the Japanese Stock Exchange, in which prosumer can bid for a desired amount and price of electricity for a 30-minutes time slot in the future The preliminary numerical evaluation is based on the results from several simulations using different versions of the bidding agent.

1 Introduction

This paper is written as an extension of the work titled "Energy exchange strategy for local energy markets with heterogeneous renewable sources", presented at the 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 Industrial and Commercial Power Systems [1] in Palermo, Italy.

In the recent years a huge surge in the usage of renewables worldwide [2]. The growth of the solar photovoltaic capacity installations in particular is remarkable, nearly double those of wind power and adding more net capacity than coal, natural gas and nuclear power combined. This trend can be seen not only on global scale since the dropping prices for energy storage and Feed-In Tariffs(FIT) [3] has led to an increase in the adoption of solar energy by individual consumers. Additionally, the efforts of countries like Japan, have contributed to the improvement of the efficiency and reducing of the cost for Distributed Renewable Resources (DER), making on-site energy sources such as fuel cells more popular with consumers [4]. However, the current centralized nature of both the grid infrastructure and the electricity markets is preventing residential customers from participating and taking full advantage of cheaper renewables. The successful implementation of blockchain technology for peer-to-peer (P2P) trading in the Brooklyn Microgrid [5], has not only led to a surge in the adoption of the technology in the development of local energy markets but to new and innovative approach towards designing such markets. Still the increased heterogeneity of the microgrids as well as the needs of the individual prosumers are yet to be addressed.

The paper [1], on which the research presented here is based, studied the impact of introducing a new power source,

[*]Borislava Spasova, 3-14-13 Higashigotanda, Tokyo 141-0022, Japan, +818044094144, b.spasova@gmail.com

in this case a hydrogen fuel cells, on already existing on a local energy market with renewables and batteries. The prototype, used for the analysis was built upon and existing system, the Open Energy Systems (OES) [6], operating it real world microgrid in Okinawa, Japan. OES consists of 19 autonomous subsystems, equipped with photovoltaic cells, energy storage and loads and interconnected via DC power bus. The system, currently used by the OES microgrid, has a well-defined exchange algorithm [7], designed to enable energy exchanges between all of the connected subsystems. However, no financial rewards are attributed to the prosumers and the energy is essentially exchanged for free.

The subsequently developed prototype aimed to change that by developing a market-based trading platform which can provide financial incentives for trading and thus enabling all the prosumers to be active participants in the local energy market and to make better use of the generated renewable energy. Such approach contributes to the an increase in the self-sufficiency and the sustainability of the community itself and expands the possibility for developing decentralized energy market to reflect on the decentralized nature of the P2P exchange. Usually trading platforms require the submission of an order bid, using an internal bidding strategy, which may or may not be fully matched, or be matched only partially, with someone else?s bid using a perviously agreed matching method. The matching can be done centrally by aggregating all requests, determining a common market clearing price (MCP) in order to maximize trading volume for a given timeslot. Another approach is to use auctions, a method often used for trading goods, mainly non-fungible goods e.g. E-bay. Continuous or real-time trading, characteristic for spot markets, is used in this prototype as it provides both speed and liquidity. The bids, used here, require for the desired amount and price of the energy for exchange to be set in advance. Both parameters are determined by the individual prosumer's exchange strategy, which is based on the State of Charge (SoC) of the available battery. A version of the algorithm was already in operation in the OES system, proving to be working well for the for subsystems with solar panels and batteries. In order to further test the SoC-based exchange strategy, a hydrogen fuel cell was added to the original configuration of the subsystems in simulated environment. The fuel cell came to the attention of the authors, due to it rising popularity among Japanese households. Unlike the the fluctuating output of the solar panels, it provides stable generation level which coupled with well devised prediction of consumption patterns can contribute to the building of off-grid communities, completely self-sufficient and independent from the utility grid.

The study presented 2018 IEEE EEEIC and I&CPS showed the broader effects of the SoC-based exchange strategy have on a community with heterogeneous power sources and the strategy itself has a very simplistic pricing algorithm which didn't account for the actual price of the energy being exchanged. As a result, a new pricing method was added, based on the price of the energy, stored in the battery of each individual subsystem. This paper presents a quantitive evaluation of the effects of this change on the developed market as well as on the behavior of predetermined prosumers' groups, created based on the daily consumption average. Examining more detailed behavioral patterns allows for better understanding of each prosumer's needs which can enhanced the development of more customized bidding strategies. Moreover, it opens the possibilities to design a more diverse community, where each standalone system can use different set of power sources, to serve as testing environment in order to achieve more realistic results.

The evaluation, presented in this paper, is based on a study of the effects of SoC-based bidding strategy with simple pricing algorithm and pricing algorithm based on the current price of the energy, contained in the battery of each standalone subsystem, on the local market and on the individual prosumer. It is achiebed by comparing the results, obtained by the simulation with the OES community, with the two bidding strategies. Furthermore, the presented strategies are evaluated in the context of a single member of a prosumers' group to better understand how changes in the bidding system can affect the individual prosumer.

Section 2 of this paper introduces the basic concepts used in the development of the solution and provides a look into the current state of P2P energy trading and existing blockchain based markets. Section 3 gives a more detailed information about the observed community and the selected prosumers' groups. Section 4 provides an overview of the prototype's design and the concrete implementation which is used for the quantitative evaluation. Finally, Section 5 will focus on the analysis of the simulations' results followed by final remarks and discussion of future works.

2 Research background

This paragraph will introduce existing solutions for energy exchange between interconnected subsystems. Many studies focus on the exchange strategies between interconnected microgrids and the main grid. The research, presented by Asimakopoulou [8], suggests a Leader-Follower strategy for energy management of Multi-Microgrids, achieved with the use of bilateral contracts between microgrids and a large production unit. Wang [9] proposes a control strategy for networked microgrids and distribution systems. Other authors such as Matamoros and Gregoratti [10, 11] has focused on developing an optimization framework for a distributed energy trading algorithm between islanded microgrids. They were also among the first researchers to give a solution for direct trading between microgrids although due to the lack of clear separation between the microgrids' internal power management strategy and their exchange strategy the proposed system didn't offer many possibilities of implementing self-interest-based internal strategies.

Such opportunity was offered by Wang [12, 13, 14] who proposed a two-layered optimization method that separates internal and external strategy and helps avoiding undesired power injections and reduces peak consumption. This proposal creates new opportunities for collabotation between self-interest microgrids. Other papers [16, 17] prefer the use of Game Theory for the trading algorithm, while Nunna [18] came up with an multi-agent approach for power trading between microgrids using demand-response and distributed storage.

Dimeas' paper [19] proposed a *multiagent system* (MAS) approach, that allows independent agents,*Buyers* and *Sellers*, to submit *Bids* for *EnergyPackages* according to *GridPrices*. Other papers also discussed this kind of approaches [20, 21, 22] but mainly from a theoretical point of view without physical prototype.

Starting with Nakamoto's white paper on Bitcoin [23], Blockchain has received both academic and industry interest for being the first scalable method to successfully decentralize P2P exchanges. It can be applied not only for cryptocurrencies, but also for various other use cases as for instance in the energy sector, notably the here-addressed P2P energy trading [24]. Using Blockchain for trading was first address by Transactive Grid Project [25] which demonstrated the concept of the Brooklyn Microgrid [26], [5]. Later many projects such as Grid+ [25], NRGCoin [27], Bankymoon, Daisee, D3A and Energo Labs [28] followed but so far few pilots and prototypes are completed and the one that are focus more on developing functioning energy market from the perspective of retailers and producers and not from the customer's perspective.

As stated in previous works [1], a prototype, following the Wang's approach was developed, which focuses on developing a platform for trading between members of one microgrid. The base concept design allows in the future to expand the market by trading with neighbouring microgrids as well. For research purposes, a prototype was developed, having explicit separation between the internal strategy for energy management and the trading algorithm, defined in the external strategy built-in in its base design. Such division aims to address both the lack of incentive and the decision coupling issue which is particularly affects cases where blockchain is used without separating external and internal logic beforehand. The systems aims to create environment where different energy exchange strategies can be used to better reflect the different targets set by prosumers when using renewable energy. The internal strategy, is tasked with managing the demand-response efficiency of each subsystem and with evaluating the system needs for buying and selling energy, implemented by the priority-based internal agent and the bidding agent, respectfully. The external strategy is the software implementation of the local energy market, used in this study.

Moreover, unlike the other studies, introduced in the paragraph, this research focuses on the small prosumers and aims to create local energy market where each player can use a customized bidding strategy that is best suited to their needs. In order to achieve that, a better understanding of the behavior of different types of prosumers is needed. This paper deepens the study of the market by observing the effects a small change in the bidding strategy can have on the individual prosumer.

3 Application in real world microgrid

This section will provide more detailed information on the real world microgrid, used for the evaluation and on the prosumers' groups, designed based on the recorded consumption data in the microgrid for the year 2015.

3.1 Microgrid Architecture

The OES is a bottom-up, distributed electric power system that mainly uses renewable energy sources. It presents an alternative way of exchanging energy in-between energy subsystems in order to manage energy fluctuations within the community. Supply and demand are balanced autonomously without impacting the utility grid and thus energy autonomy is increased with minimal infrastructure costs. The concept and feasibility of OES have been demonstrated in a decentralized, peer-to-peer microgrid in Okinawa, Japan. The microgrid has been operating there since late 2014 and consists of 19 houses, each equipped with their own subsystem with photovoltaic cells, residential energy storage in a form of 4.8 kWh batteries, dc nanogrid and loads. All the subsystems are connected via dedicated, shared DC power bus as well as communication line which allows energy exchanges within the community. The subsystem use a direct current (DC) procedure for exchange, combined with pure peer-to-peer (P2P) communication infrastructure and exchange algorithm [7]. All the houses also have connection to the utility (AC) grid, which can be used as an auxiliary power supply when no other energy supply is available. Bidirectional dc-dc converters are used as an interface between subsystems: they allow for an effective control of the power flow and can be used as an abstraction of the internal subsystem design. This way, dc power can be exchanged within a community to help balancing demand-response requirements without requiring global knowledge or control.

In order to test the system capabilities in simulated environment with heterogeneous power sources, a fuel cell is added to the to the configuration of each standalone subsystem. The fuel cell, used here is largely based on Ene-Farm, a hydrogen fuel cell for home use, that has been gaining popularity in Japan. Ene-Farm is an energy system that can extract hydrogen from natural gas, usually supplied though underground pipes in most major cities in Japan, or propane gas (LPG), usually supplied to a residence via external tanks part of the building design. The hydrogen is then used in combination with ambient oxygen to generate electrical power and at the same time to capture the residual heat which in turn is used to heat up water. Ene-Farm was chosen here for both its wide spread use and its high efficiency ratio.

The extended topology of the observed microgrid, very similar to the real world one, can be seen on Figure 1.

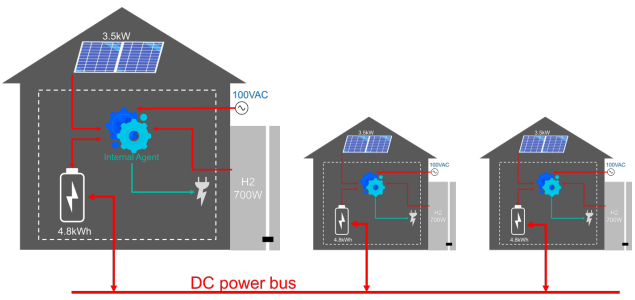


Figure 1: Sample subsystem configuration and general microgrid architecture of the chosen microgrid

3.2 Prosumers' groups in the OES community

During the process of design and development of the prototype the daily consumption of all 19 members of the OES community for the entire year of 2015 was examined in detailed. Furthermore the solar generation patterns we also mapped for the same period of time. Based on the analysis, the OES consumers were divided into three main groups, defined by their individual average daily consumption for the observed period, using the generated output from the fuel cell and the solar panels patters a baseline indicator. The groups are as follows:

1. Group 1: High consumers, with observed daily average consumption of more than 700 watts;
2. Group 2: Average consumers, with daily average consumption in the range of 350 to 700 watts;
3. Group 3: Low consumers, with daily average consumption of less than 350 watts.

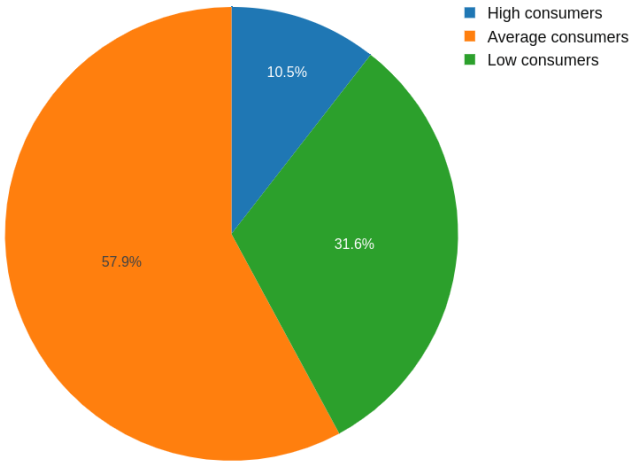


Figure 2: OES prosumers distribution, based on their consumption

Figure 2, shows that more than 50% of the consumers in the OES community are in the second category, a trend that remained unchanged for the entire year 2015. Moreover, no noticeable changes in the members of Group 2 was observed throughout the year. One notable exception was registered in April 2015 when a user moved from Group 3 to Group 1 due customers moving into previously uninhabited house.

Dividing the community into categories, based on such a crucial factor as consumption, helps further the understanding of the community and the prosumers themselves because it provides an effective way to analyze the behavior of different types of consumer since members of one category tend to exhibit similar behavior within an acceptable margin of error. That will allow for the development of different strategies aimed to satisfy the needs of variety of customers.

4 Prototype design and implementation

The prototype, presented here, is developed based on the Open Energy Systems (OES), a real world microgrid operating in Okinawa, Japan. This section will provide an overview of the basic premises and concrete algorithms used in the development of the proposed solution.

4.1 Design assumptions and limitations

During the development of the prototype, presented here, several simplification were incorporated in the design. The following paragraph focuses only on the ones that have major implication for the research, subject of this paper. For more detailed list, please refer to the cited paper [1].

1. No losses, including conversion and transmission losses, are taken into account during the powerflow management;
2. Projected transmission losses, however, are taken into consideration in the pricing algorithm;
3. Initial investment in the renewable source is disregarded when calculating the energy price;
4. Each subsystem is equipped battery, used for absorbing imbalances in the network;
5. The price of the battery is calculated for $\forall t \in \mathcal{T}$, where $\mathcal{T}$ is time series.
6. The price of the battery is calculated as a sum of the amount charged in the battery from each power source multiplied by the predefined energy price of the source and then divided by the current amount of energy in the battery. The formula can be seen bellow.

$$bp = \frac{\sum_{i \in \mathcal{N}} \mathrm{PS}_{i,t}^{\mathrm{amount}} * \mathrm{PS}_{i}^{\mathrm{price}}}{currentcapacity} \quad \forall t \in \mathcal{T} \qquad (1)$$

7. Users are charged when storing energy in the battery, not when discharging the stored energy for the consumption
8. It is presumed that the data, provided by the power meters and the controllers, can be trusted;
9. The DC line capacity is set as 2.5 kw for all subsystems;
10. Japanese yen is selected as main currency for all the financial transactions, due to the microgrid location;
11. The battery is charged solely from the renewable sources;
12. There is no Feed-in to the utility grid

4.2 Prototype design

The following paragraphs will give a short description of the internal and externals strategies, focusing on the bidding component of the former. For more information about the chosen algorithms, please refer to the cited paper [1].

4.2.1 *Internal strategy*

The internal strategy, used in this paper, is tasked with managing the demand-response efficiency of each subsystem and with evaluating the system needs for buying and selling energy. It is comprised of two main components: priority-based agent and bidding agent, each implementing different side of the internal agent. They are designed to work independently using only the same data in order to increase encapsulation. More information about the basis of the algorithms cam be found in the paragraphs bellow.

Priority-based internal agent: The internal agent is responsible for scheduling the power flows to and from the modules, configured for each individual subsystem. It is designed for standalone house equipped with residential storage, energy sources and loads. The fundamental premise built into the design of the system is inherent flexibility when it comes to changing the number and types of the installed power sources. Essentially the system allows for various sources such as photovoltaic and fuel cells, wind turbines, standard electricity grid, etc. to be added to the configuration at any moment, given that each source is a assigned a priority level, a parameter used to control when it is going to be used, and a price for electricity for kWh. The priority is normally determined by factors such as energy price and cost for curtailment. In the current system the priority is set as follows: fuel cell output; solar generation; battery; utility grid. In this case power sources with what is considered an expensive curtailment, like fuel cells, are given higher priority in order to fully utilize their generation even though the cost for electricity for other sources such as solar tends to be much lower. Once the priority is defined, the algorithms uses it to satisfy the demand-response requirements by accessing the available energy from each source until the demand is satisfied. If after using all the power from the renewables, the demand is still not satisfied, the system gets the remaining necessary power from the AC grid. If there is still available renewable energy after the demand-response requirements are satisfied, it is charged to the battery if there is enough capacity, the remaining energy is curtailed. As it can be seen, in this system, the battery plays a dual role as both a power source and a load. The role of the agent is to maintain internal power balance at any moment:

$$P^{batt}_{i,t} = \sum_{n=1}^{\infty} P^{ps}_{n,i,t} + P^{ac}_{i,t} + P^{trade}_{i,t} - P^{load}_{i,t} \quad \forall t \in \mathcal{T}, i \in \mathcal{Z} \qquad (2)$$

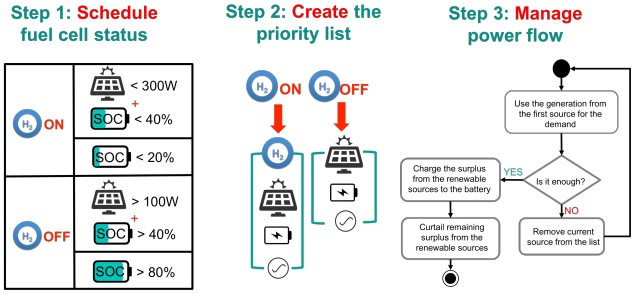


Figure 3: Scheduling conditions and simplified workflow of the internal agent.

Following the research presented in previous works [1] a scheduling component was added to the functionality of the internal agent. This software module is tasked with scheduling the operating hours of power sources with constant output like a fuel cell. The module analyzes the future energy requirements and gives command to shut down or turn on the fuel cell. Fuel cells of the Ene-Farm type require 30 minutes to stop and start operation therefore the default conditions are set based on the solar generation output and the SoC of the battery to take into account the future state of the system.

Figure 3 shows the basic workflow of the priority-based internal agent with the defined configuration. As shown the fuel cell is scheduled to shut down when the SoC of the battery is higher that 80%, which is more than enough to satisfy any upcoming changes in the consumption and to reduce to risk of curtailment from this power source. Even if the SoC is lower, if there is an observed spike in the solar generation, the fuel cell also turn off to enable the solar generation to be fully absorbed by the battery, thus utilizing better the generation and lowering the bidding price. In the opposite scenario, the fuel cell is scheduled to resume operations, when the SoC falls under 20% in order to prevent the activation of the connection to the utility grid, which has much higher price than the fuel cell generation. For the same reasons, the fuel cell turns on, when there is a decline in solar generation whether due to reduced solar radiation caused by weather conditions or during nightfall. This conditions help reduce the electricity cost for the household by decreasing the reliance on the utility grid.

Bidding agent: The bidding agent is a module that is installed together with the internal agent and is tasked with generating a sell or buy bid for the next timeslot and dispatching it for matching. Each bid consists of desired amount of electricity for a certain price for kWh. The amount is determined based on a detailed analysis of the status of each subsystem which includes the number and type of configured power sources, e.g. batteries, PV, fuel cell, etc., and current consumption and generation patterns. This is then used by a pricing function to determine the future need for electricity and is reflected in the bids that follow.

The bidding strategy, subject to evaluation of this paper, uses the State-of-charge (SoC) of the installed battery to predict the amount of electricity that will be needed for the next 30-minute timeslot but with two different pricing algorithms. The ***base*** amount, seen in Figure 4a,b, refers to the minimum amount of electricity available for exchange, calculated as a function of the DC line capacity and the trading timeslot duration to avoid line congestion.

In the case of the previously used bidding strategy, the prices are fixed based on the prices for the electricity, paid by the customer, as seen on Figure 4a. It is assumed that the lower the SoC level, the more inclined is the prosumer to buy electricity. Since all systems are connected to the AC-grid, the highest price for the electricity in the market was designed to always be less than the price of the grid. Similarly, there is a higher incentive to sell the generated renewable energy when the battery is getting full. The price in this case is set as a minimum price at which electricity can

be sold while still generating profit. This pricing, although well designed for the original OES configuration, is not as effective in the presence of a stable power source such as fuel cell, because of the difference in the price for energy, generated from the solar panels and from Ene-Farm.

Battery SoC (%)	Sell Bid		Buy Bid	
	price (yen/kWh)	amount(watt-hour)	price (yen/kWh)	amount(watt-hour)
<=20			25	(base) x4
20 - 40	30	base	20	(base) x3
40 - 60	20	(base) x2	10	(base) x2
60 - 80	15	(base) x3	5	base
>= 80	10	(base) x4		

a)

Battery SoC (%)	Sell Bid		Buy Bid	
	price (yen/kWh)	amount(watt-hour)	price (yen/kWh)	amount(watt-hour)
<=20			25	(base) x4
20 - 40	bp + 40%	base	20	(base) x3
40 - 60	bp + 35%	(base) x2	10	(base) x2
60 - 80	bp + 30%	(base) x3	5	base
>= 80	bp + 25%	(base) x4		

b)

Figure 4: Bidding prices and amounts for State of Charge (SoC)-based bidding strategy with simple pricing algorithm(**a**) and pricing algorithm based on battery price (**b**).

The new pricing algorithm, shown on Figure 4b, takes into account the price of the electricity, currently stored in each individual subsystem, noted as ***bp*** in the table. It currently affects only the sell bid with the price similarly decreasing with the increased incentive to sell. The percentage is based on a projected transmission loss of nearly 15% which in the case of the studied community is paid by the Seller.

4.2.2 *External Strategy:*

The external strategy represents the software implementation of the local energy market, used in this study. At the core of the market is the matching algorithm, a simplified version of the Zaraba method, which is a continuous double auction algorithm used by the Japanese Stock Exchange. The trading is done in real time, with multiple sell and buy bids received for the duration of each 30-minute timeslot. They are match on on first-come first-served basis with the most beneficial counter bid being chosen each time as a counterpart. At the end of the timeslot the agreed deals are sent to the internal agent to execute the physical exchange and all the unmatched bids are invalidated.

Since the DC losses are neglected, we assume that the sum of all energy trading must be equal to zero:

$$\sum_{i \in \mathcal{Z}} \mathrm{P}_{i,t}^{\mathrm{trade}} = 0 \quad \forall t \in \mathcal{T} \tag{3}$$

4.2.3 *Virtual Wallet:*

In order to keep an accurate record of all financial transaction, including detailed information about the used energy from the renewable sources and from the grid, cost of the energy charged to batter and cost/revenue from the energy trading, each subsystem has a virtual wallet. The data, kept in the wallet, is accessible via simple user interface which provides basic analysis and monthly reports.

5 Simulation Setup and Results

For numeric evaluation purposes, this study uses the results of multiple simulations, obtained using prototype, comparatively similar to the OES Simulator [7], written in Python 3.6. The prototype provides a software approximation of the internal power flow of each system and of the community. All the implemented algorithms are consistent with the descriptions given in the previous section and in other works [1]. The following subsection will display the general configuration of studied subsystems, the data, provide information of the scope of the data used as an input for the simulator and will discuss the results of the performed tests.

5.1 *Input Data and Configuration*

In order to further the research, shown in previous works [1], the simulation uses the same dataset for the consumption and solar radiation, recorded for 19 houses, part of OES in Okinawa, Japan, in the duration of year 2015 with 10 minutes granularity. The bidding is done for a 30-minutes timeslot in the future, which amounts in total to 48 trading timeslots per day.

Each subsystem in the community has the following configuration of modules:

1. ***Solar panels*** with average area of 20 m^2, with estimated panel yield of 19% and 75% performance ratio. Since the initial investment is not taken into consideration in the estimate, the price for solar energy is set to 0 yen for kWh.

2. ***Fuel cell*** with factory output of 700 watts and predefined price for electricity of14 yen for kWh, based solely on the price of the fuel used, since the initial investment is also disregarded during calculations.

3. ***Utility grid,*** used as an auxiliary power source with no power limitations. The price for electricity is considered 30 yen for kWh, which is an approximation of the real price of the electricity in Japan. This is necessary due to the varying prices for different regions as well as different usage categories.

4. ***Battery*** with capacity of 4.8kWh, charged at 50% with energy price of 10 yen for kWh, used only in the first iteration of the algorithm. For the remaining of the simulation, it is calculated in the beginning of each iteration, how much energy was charge it in the previous cycle and from which renewable source.

5. ***Wallet*** with 100,000 Japanese yen available.

5.2 Results

In order to evaluate the added functionality presented in this paper, two simulations are performed, each using different bidding strategies as listed below:

1. SoC-based bidding strategy, described in the cited paper [1].

2. SoC-based bidding strategy with pricing algorithm based on the battery price of each standalone system, presented in this paper.

All simulations are performed using a only selected potion from the available dataset with length of 1 calendar month, in this case August 2015. To simplify the visualization of the results, only one week, from August 2nd to August 9th, is taken for the analysis of the whole market. That particular week is chosen because of the variety of weather conditions that occurred during that period. The selected timeframe consisting of several sunny days and two partially cloudy day is chosen. In order to better study the behavior pattern of the individual participants, only data from 1 day, August 5th 2015, with daily generation close to the month's average is examined.

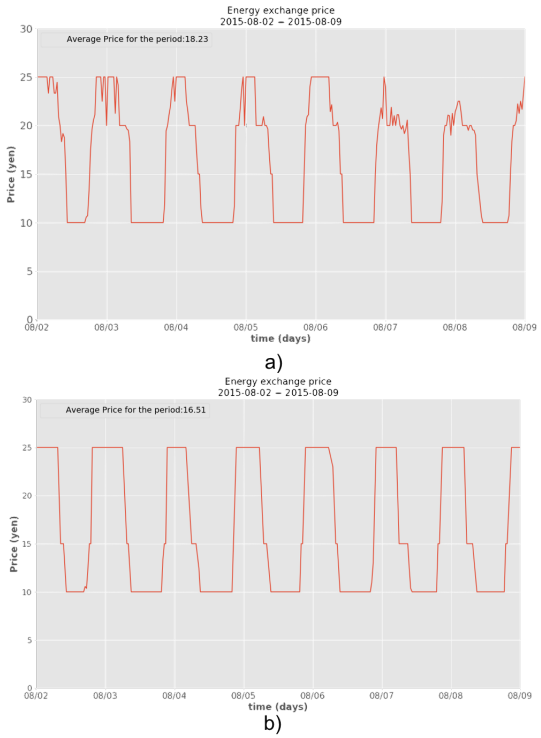


Figure 5: Market price within a community, comprised of standalone houses with solar panels and fuel cells, SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

5.2.1 Community overview

This section will provide a detailed analysis of the effects of the new bidding strategy on the entire community before focusing on selected members of the predefined prosumers' groups. The second portion of the analysis can be seen in the next section. Detailed analysis of the results of the simulations showed 55% decrease in the number of deals with the new bidding strategy. This is due to less available matches in the preferable price range of the prosumers which makes them less willing to either sell or buy energy if it is not beneficial for them. This has had an effect on the market price as well as seen on Figure 5.

Fig. 5b shows that the pricing algorithm, has led to predictable patterns for market price, with stable peaks that indicate higher prices in comparison with the case of fixed pricing, Fig. 5a. Most influenced by the new pricing algorithm are the high consumers which show 10% decrease in energy spending. Average consumers and low consumers with daily average consumption more than 100 watts spent 7% less for electricity for the same period of time while systems with daily average consumption of less than 100 watts registers a surplus increase of 10%.

Despite the lower number of deals, there has been observed in 10% increase in the utilization of solar generation due to a drop in the consumption of electricity from the AC grid. The numbers are based on the numeric evaluation of the results though a slight reduction the curtailment peaks can be seen on Figure 6. The increase is due to more customized bidding strategy and shows that better focus on customers' energy requirements and implementing them into individual bidding strategy can be used as a foundation for building self-sustainable community. There is no visible curve for the Fuel cell curtailment since it remains 0 for the entire period due to the efficiency of the scheduling component of the internal agent.

5.2.2 Analysis of individual prosumers from the predefined groups

The analysis of a selected members of the groups, described above, provides a deeper understanding of the results shown above. Consumers of all groups showed slight in the usage of the energy stored in their battery which is a contributing factor for the slight decrease in the solar curtailment.

For a Group 1 consumer the increase has led to lowering the usage of the AC grid, shown on Fig. 7, while for consumer from Groups 2 and 3, the difference is compensated by using less electricity directly from the Fuel cell, shown on Fig. 8 and Fig. 9 respectively. The solar usage remained unchanged. Since solar energy is considerable cheaper than the one from AC or Fuel cell it can be considered as a factor in the general lowering of the energy cost.

Deeper analysis of the sources of energy for the battery further shows that the new pricing algorithm has contributed to the higher utilization of the solar energy. Although the results shows very little change for customers of Group 1, Fig. 10, in the case of customers from both Group 2 and 3, Fig. 11 and Fig. 12 respectively, shows increased charging from the solar generation while the no energy has been bought from the market. That scenario shows better management of the energy and of the trading which is more consistent with the system requirement of this two types of consumers since both should have enough energy so be entirely self-sufficient and participate in the market only to sell their excess energy. The graph of the energy exchanged through the DC bus gives even better idea of what is happening. As we can see on Fig. 14 for customers of Group 2 and Fig. 15 for customers of Group 3, there is a distinctive

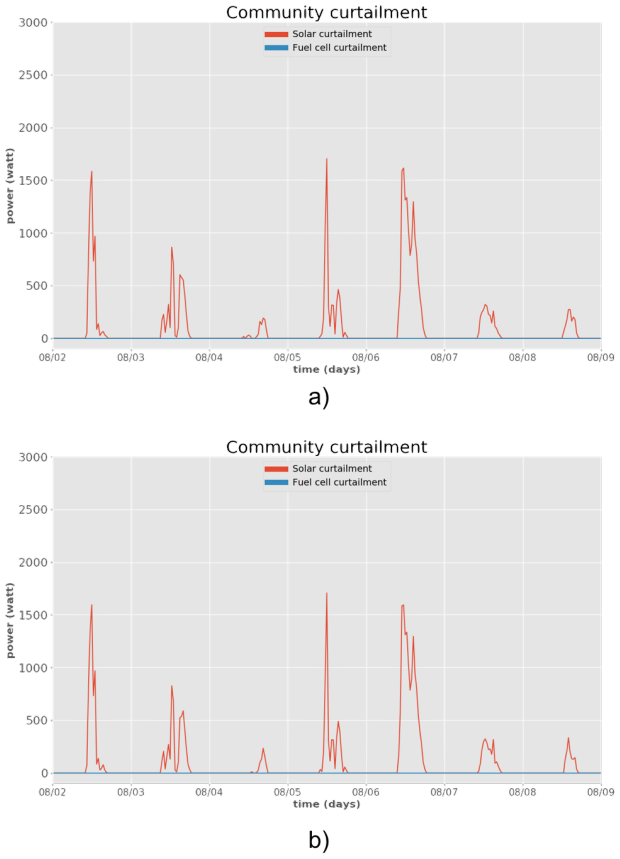


Figure 6: Average curtailment in community, comprised of standalone houses with solar panels and fuel cells, SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

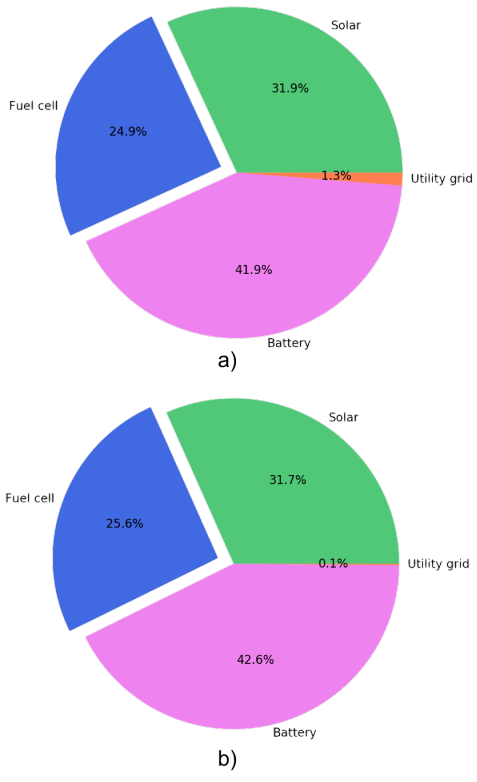


Figure 7: Energy usage by power sources for ***Group 1*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

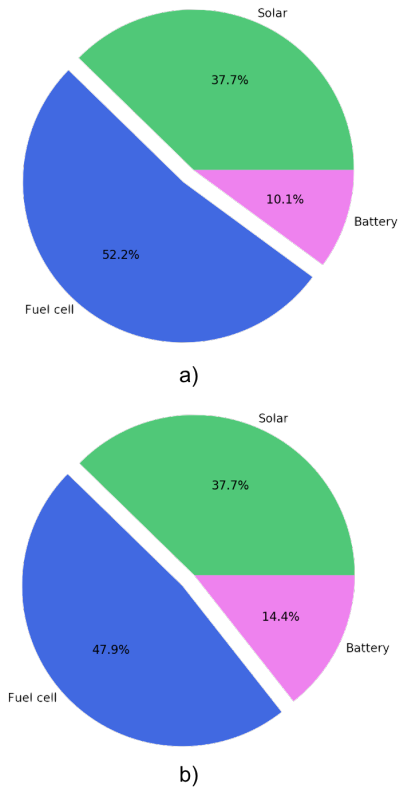


Figure 8: Energy usage by power sources for ***Group 2*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

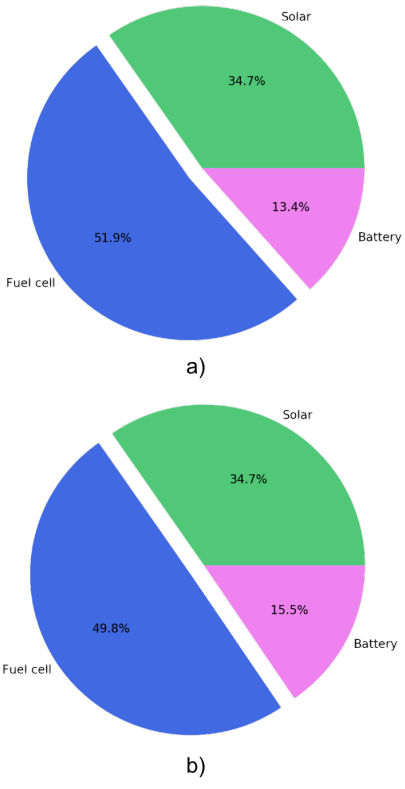


Figure 9: Energy usage by power sources for ***Group 3*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

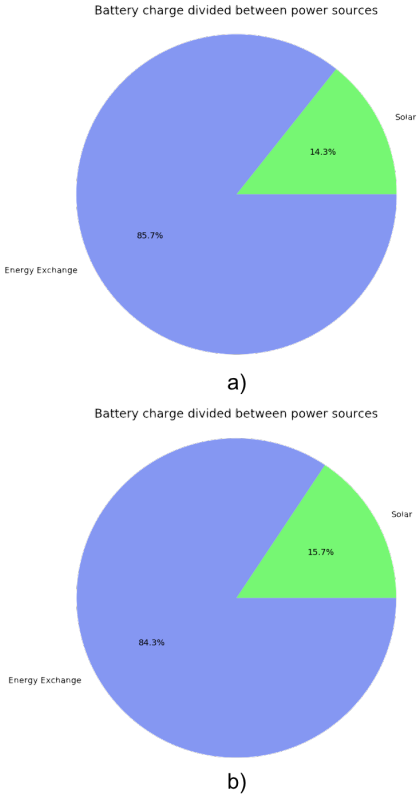


Figure 10: Energy charged to the battery by power sources for ***Group 1*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

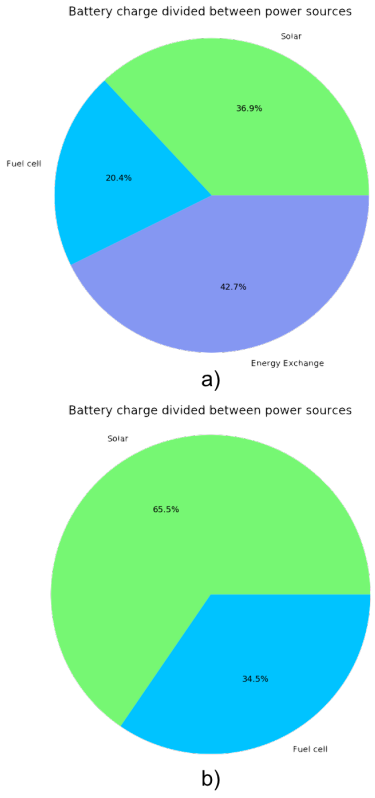


Figure 11: Energy charged to the battery by power sources for ***Group 2*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

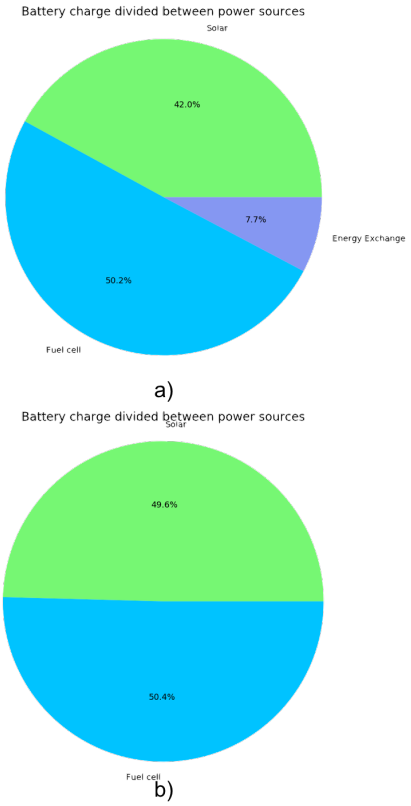


Figure 12: Energy charged to the battery by power sources for ***Group 3*** consumer that use SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

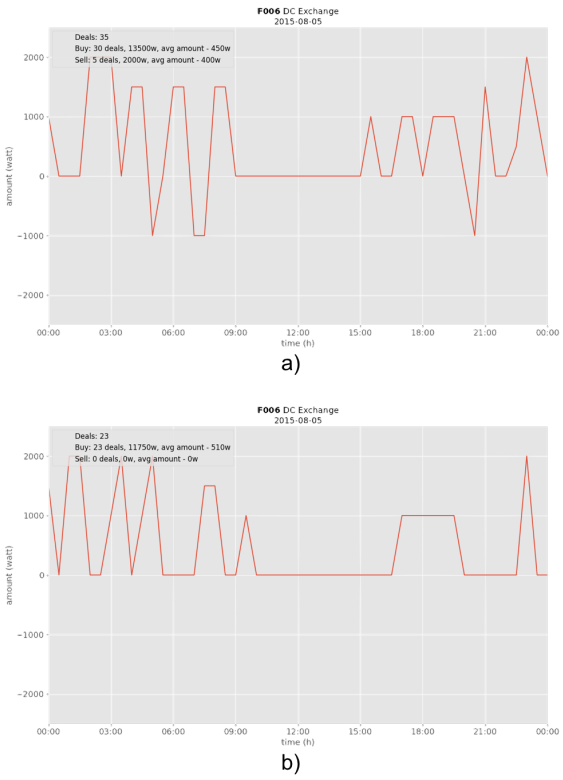
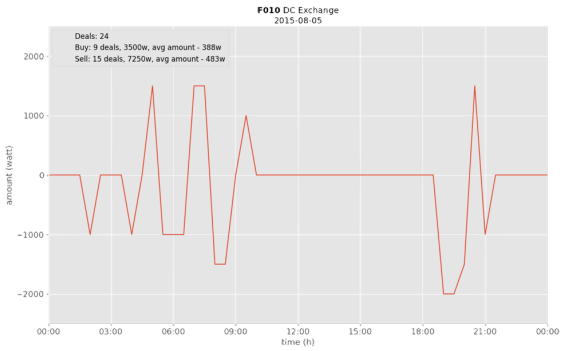
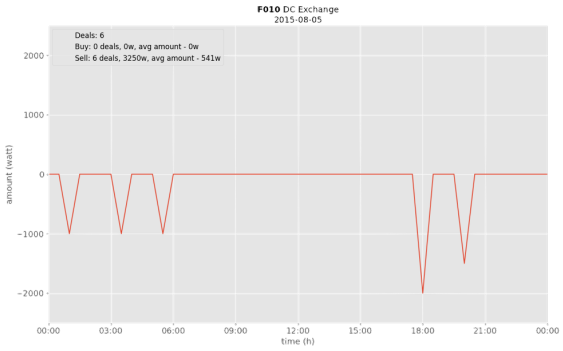


Figure 13: Deals distribution for ***Group 1*** consumer using SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

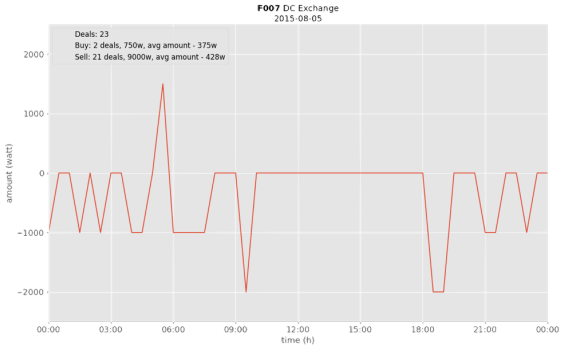


a)

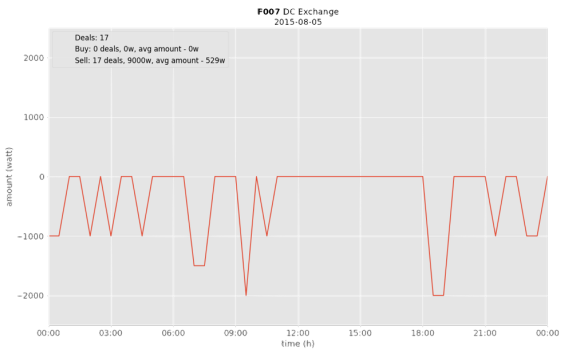


b)

Figure 14: Deals distribution for ***Group 2*** consumer using SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)



a)



b)

Figure 15: Deals distribution for ***Group 3*** consumer using SoC-based bidding strategy with fixed pricing(**a**) and SoC-based bidding strategy with pricing algorithm based on the battery price (**b**)

lack of spikes that indicate buy deals, which shows that in the case of the simple SoC strategy we have moments early morning and late evening where prosumers with enough energy are forced to buy from other prosumers electricity on higher prices because they have previously sold their for lower price. Reversed pattern can be observed for customers of Group 1, Fig. 13, because their have electricity demand that on average exceeds their generation by 50% so they should not be selling energy in those periods of time as well.

Although the effects on the community are only marginal, the new pricing method had shown significant shift in the patters of all of the observed individual customers. Further customization can provide better results and can make the prosumers more competitive in the market but in the context of their individual needs.

6 Conclusion

This study shows the impact of enhanced pricing algorithm, which accounts for the price of the electricity contained in the battery of standalone systems, part of a local energy market with renewables and batteries. The evaluation is done by analyzing the results, generated by a prototype with the use of real data from a full-scale microgrid in Okinawa, Japan. The microgrid represents a community of 19 houses each equipped with residential storage, photovoltaic cells, fuel cell and AC grid connection. All the houses are connected via dedicated, shared DC power bus which allows them to exchange energy with their neighbors. A continuous double-auction algorithm, used by the Japanese Stock exchange, is used for the implementation of the trading platform. Each house can bid on the local market, using a State-of-Charge based bidding strategy. Similar to previous works [1], some simplifications were taken during the development of the simulation platform, including neglecting all losses and ignoring the price of the energy, stored in the battery when calculating the desired exchange price. The effects of the incorporation of the losses in the algorithm are yet to be studied but it is expected that such change will cause an increase in the market price. The topic is currently being investigated and will be a subject to future works.

This paper focuses on studying how including the battery price in the pricing algorithm affects different types of consumers and the community as a whole. The results show that increasing the customization leads to higher utilization of the renewables and further reduces the cost of electricity for consumers with daily average consumption higher than 100 wats. Since this was achieved with customizing a single parameter, it is the opinion of the authors that further analysis of the behavior of the individual system is need in order to create bidding strategy, tailored to each prosumer. It is the opinion of the authors that a well design exchange strategy should allow for the customers to set individual goals such as maximizing profit, maintaining independence from the utility grid, reducing energy waste, etc. It should implement a prediction algorithm which accounts for consumption and generations patterns and include weather predictions for better evaluation of the energy needs. Furthermore, with the wide spread of renewables and distributed renewable resources (DER), more diverse communities are likely to appear so further tests in the community with standalone systems with various configuration power sources can serve as a testing environment to achieve more realistic results.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This paper would not have been possible without the continuous support of the Okinawa Institute of Science and Technology and H.Kitano. The authors would like to thank A.Werth and M.Hoshino, for their contribution in forming these ideas.

References

[1] B. Spasova, D. Kawamoto, Y. Takefuji, "Energy exchange strategy for local energy markets with heterogenous renewable sources", 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 Industrial and Commercial Power Systems, 12-15 June 2018, Palermo, Italy. https://doi.org/10.1109/eeeic.2018.8494458

[2] Renewable energy policy network for the 21st century REN21, "Renewables 2018 Global Status Report (GSR)", http://www.ren21.net/gsr-2018/

[3] A. Brautigam, T. Rothacher, H. Staubitz, and R. Trost, "The Energy Storage Market in Germany", 2017.

[4] P. De Martini, K.M. Chandy and N.A. Fromer, "Grid 2020: Towards a Policy of Renewable and Distributed Energy Resources", Technical Report, California Institute of Technology, Resnick Institute: Pasadena, CA, USA, 2012.

[5] E. Mengelkamp, J. Garttner, K. Rock, S. Kessler, L. Orsini, and C. Weinhardt, "Designing microgrid energy markets. A case study: The Brooklyn Microgrid", Applied Energy, vol. 210, pp. 870?880, 2017. https://doi.org/10.1016/j.apenergy.2017.06.054

[6] A. Werth, N. Kitamura, and K. Tanaka, "Conceptual Study for Open Energy Systems: Distributed Energy Network Using Interconnected DC Nanogrids", IEEE Transactions on Smart Grid, vol. 6, pp. 1621–1630, jul 2015. https://doi.org/10.1109/tsg.2015.2408603

[7] A. Werth, A. Andre, D. Kawamoto, T. Morita, S. Tajima, M. Tokoro, D. Yanagidaira, and K. Tanaka, "Peer-to-peer Control System for DC Microgrids", Smart Grid, IEEE Transactions on, pp. 1–8, 2016. https://doi.org/10.1109/tsg.2016.2638462

[8] G. E. Asimakopoulou, A. L. Dimeas, and N. D. Hatziargyriou, "Leader- follower strategies for energy management of multi-microgrids", IEEE Transactions on Smart Grid, vol. 4, no. 4, pp. 1909–1916, 2013. https://doi.org/10.1109/tsg.2013.2256941

[9] Z.Wang, B.Chen, J.Wang, M.M.Begovic,and C.Chen, "Coordinated energy management of networked microgrids in distribution systems", IEEE Transactions on Smart Grid, vol. 6, no. 1, pp. 45–53, 2015. https://doi.org/10.1109/tsg.2014.2329846

[10] J. Matamoros, D. Gregoratti, and M. Dohler, "Microgrids energy trading in islanding mode", 2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm 2012, pp. 49–54, 2012. https://doi.org/10.1109/smartgridcomm.2012.6485958

[11] D. Gregoratti and J. Matamoros, "Distributed Energy Trading: The Multiple-Microgrid Case", IEEE Transactions on Industrial Electronics, vol. 62, no. 4, pp. 2551–2559, 2015. https://doi.org/10.1109/tie.2014.2352592

[12] H. Wang, "Economic Mechanisms for Integrating Smart Grid Technologies into Power System", PhD thesis, The Chinese University of Hong Kong, 2016.

[13] H. Wang and J. Huang, "Bargaining-based energy trading market for interconnected microgrids", IEEE International Conference on Communications, vol. 2015–Septe, pp. 776–781, 2015. https://doi.org/10.1109/icc.2015.7248416

[14] H. Wang and J. Huang, "Cooperative Planning of Renewable Generations for Interconnected Microgrids", IEEE Transactions on Smart Grid, vol. 7, no. 5, pp. 2486–2496, 2016. https://doi.org/10.1109/tsg.2016.2552642

[15] A. Werth, N. Kitamura, I. Matsumoto, and K. Tanaka, "Evaluation of centralized and distributed microgrid topologies and comparison to Open Energy Systems (OES)", 15th International Conference on Environment and Electrical Engineering (EEEIC), pp. 492–497, IEEE, June 2015. https://doi.org/10.1109/eeeic.2015.7165211

[16] L. Zhang, Z. Li, C. Wu, "Randomized Auction Design for Electricity Markets between Grids and Microgrids", SIGMETRICS '14 The 2014 ACM International Conference on Measurement and Modeling of Computer Systems, pp. 99–110. https://doi.org/10.1145/2591971.2591999

[17] W. Saad, Z. Han, and H. V. Poor, "Coalitional game theory for cooperative micro-grid distribution networks", IEEE International Conference on Communications, pp. 6–10, 2011. https://doi.org/10.1109/iccw.2011.5963577

[18] H. S. V. S. Kumar Nunna and S. Doolla, "Energy management in microgrids using demand response and distributed storage - A multiagent approach", IEEE Transactions on Power Delivery, vol. 28, no. 2, pp. 939–947, 2013. https://doi.org/10.1109/tpwrd.2013.2239665

[19] A. Dimeas and N. Hatziargyriou, "A multiagent system for microgrids", in IEEE Power Engineering Society General Meeting, 2004., vol. 2, pp. 55–58, IEEE, 2004. https://doi.org/10.1109/pes.2004.1372752

[20] A. Amato, B. D. Martino, M. Scialdone, and S. Venticinque, "A Negotiation Solution for Smart Grid Using a Fully Decentralized, P2P Approach", 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 527–534, 2015. https://doi.org/10.1109/cisis.2015.78

[21] J. Lagorse, D. Paire, and A. Miraoui, "A multi-agent system for energy management of distributed power sources", Renewable Energy, vol. 35, pp. 174–182, jan 2010. https://doi.org/10.1016/j.renene.2009.02.029

[22] H. Algarvio and F. Lopes ,"Multi-agent Retail Energy Markets : Bilateral Contracting and Coalitions of End-use Customers", 12th International Conference on the European Energy Market (EEM), p. 5, 2015. https://doi.org/10.1109/eem.2015.7216750

[23] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", Consulted, pp. 1–9, 2008. https://bitcoin.org/bitcoin.pdf

[24] M. Swan, "Blockchain: Blueprint for a New Economy", O'Reilly, 2015.

[25] Consensys, "GRID+ Welcome to the Future of Energy", 2017.

[26] J. Lilic, "Building the Decentralized Utility on the Ethereum Blockchain", in EcoSummit Berlin, 2016.

[27] M. Mihaylov, S. Jurado, N. Avellana, K. Van Moffaert, I. M. De Abril, and A. Nowe, "NRGcoin: Virtual currency for trading of renewable energy in smart grids", International Conference on the European Energy Market, EEM, 2014. https://doi.org/10.1109/eem.2014.6861213

[28] E. Labs, "White Paper: Decentralized Autonomous Energy System", 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 51-57 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Frequency-Tunable Narrow-Band Ladder-Shape Microstrip Patch Antenna for TV Applications

Maria Moussa*, Mervat Madi, Karim Kabalan

Electrical and Computer Engineering Department, American University of Beirut, 1107 2020, Lebanon

ARTICLE INFO

Article history:
Received:11 April 2019
Accepted:26 June, 2019
Online: 11 July, 2019:

Keywords:
Digital TV
UHF
Ladder shape
Frequency tuning
Varactors

ABSTRACT

In this paper, a frequency-tunable antenna is proposed based on a previous antenna which has a ladder shape and a size of 120 x 50 mm^2. It is fed using a coaxial cable and it is operational at 700 MHz which is in the UHF band with a bandwidth of 30 MHz. Varactors, having a very low resistance value and a wide range of capacitance values, are implemented on the surface of the antenna which achieves tunability in the DTV applications band used for mobile handheld devices. By using such varactors, the electromagnetic characteristics of this antenna could be changed by changing its electrical length. Multiple cases are studied; either adding a varactor on each step of the ladder antenna or adding only one varactor in a certain well justified position. The operating frequency is swept in different ranges where all of the obtained frequencies could be used for digital TV broadcasting on mobile devices.

1. Introduction

Digital TV broadcasting is now the median of exchanging information and that is thanks to the progress in digital TV technology in the form of wireless devices as well as the ways of sending and receiving signals. The new mobile devices have an attractive feature which is to allow users to watch TV on their own handheld devices; however, this feature presents some challenges in the field of designing antennas for this purpose. The antenna has to have a small size to fit in the devices and a high performance to satisfy interested users.

This paper is an extension of the work originally presented in the conference paper entitled "Ladder Shape Microstrip Patch Antenna", [1]. It is known that the operational frequency and the size of the antenna are inversely proportional. Since our targeted frequency is in the UHF band, it is then obvious that the size of the antenna operating in the high MHz range will increase. One advantage in the design process is that the designer is capable of slightly sacrificing the efficiency of the antenna because digital TV is a receive-only system in which lower total efficiency is acceptable as stated in [2].

Other than the fact that the antenna has to have a small size, it has to have good impedance matching conditions also. This is where the designer is required to add a matching circuit in the purpose of achieving the best possible power transfer between the amplifier and the antenna, which leads to the optimal antenna efficiency. Also, the designer has to take into account the fact that all the used components in the matching circuit add some losses to the design, which affects the whole performance of the system.

In the literature, there exists many papers that present designs of digital TV broadcasting antennas for mobile handheld devices. Some of these antennas are the grating monopole antenna, [3], having dimensions of 240 x 35 mm^2 and operating between 458 MHz and 960 MHz, the asymmetric fork shaped monopole antenna, [4], having dimensions of 245 x 35 mm^2 and operating between 451 MHz and 912 MHz, the coplanar waveguide-fed dipole antenna, [5], having dimensions of 241 x 58.5 mm^2 and operating between 430 MHz and 1180 MHz and the monopole slot antenna, [6], of dimensions 128 x 51 mm^2 and operating at 600 MHz. In [7], two small microstrip antennas and one planar inverted F antenna are proposed. These antennas are used in the TV band for transmission of cognitive radios. The three of them operate between 700 MHz and 900 MHz. Some new references could be found in the literature such as [8] in which the authors designed and fabricated a flexible printed active antenna having a meander line shape for the DTV reception. It has a size of 180 x 50 mm^2 and a wide operational band ranging from 510 MHz to 790 MHz due to the usage of coplanar waveguide feeding technique. This antenna has an omnidirectional radiation pattern and a maximum gain of about 18 dB. A compact 46 x 46 mm^2 antenna is designed in [9] covering a wide band from 490 MHz to 720 MHz. The bandwidth enhancement is achieved by changing the size of the ground plane of the antenna and the size reduction

* Maria Moussa, Email: mnm33@mail.aub.edu

is achieved by incorporating the meander line shape in the radiating patch.

In [1], the goal was to design a small antenna for digital TV after modifying the antenna proposed in [10], which is also a ladder shaped antenna but operational at 90 GHz. Considerable modifications were done on the latter one to function in the MHz frequencies despite the fact that wireless devices impose some size constraints on the antenna. The resulting patch antenna has the shape of a ladder and a size of 120 x 50 mm^2 which overrides some of the antennas presented in the literature. Many of the ladder dimensions were studied in [1] to try to improve the matching conditions. The software used for the design and simulation of the antenna is HFSS. It is then fabricated and tested. The simulated and measured results showed good agreement. It has a narrow operational bandwidth of 30 MHz only ranging from 685 MHz to 715 MHz, which do not cover the whole UHF band channels. The solution proposed for that issue is to add surface mounted varactors to the ladder patch antenna to be able to tune the operational bandwidth. A DC bias voltage should be applied to the varactors to change their capacitance values and hence achieve frequency tunability to different UHF channels.

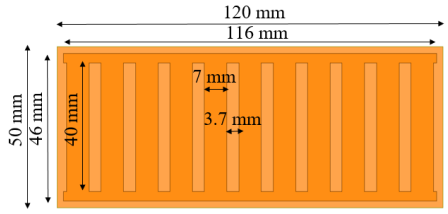
In the upcoming wireless communication systems, reconfigurable antennas are used widely for many reasons. A re-configurable antenna could have the ability of changing many of its characteristics. Frequency, radiation pattern, bandwidth and polarization could be altered to adapt to the environment, [11], using a reconfigurable antenna. Besides their re-configurable capability, re-configurable antennas contain many other features such as reduced cost, size miniaturization and multipurpose functions and they use microstrip antennas as a platform. In addition, recent systems must be able to receive signals over a wide range of frequencies so they require, as stated in [12], either wide-band or tunable narrow-band antennas. The requirement of receiver filters is relaxed because tunable narrow-band antennas, in contrary to broadband antennas, provide frequency selectivity. Tunable antennas, [13], are of interest for wireless communication systems because they could substitute multiple antennas operating at different frequencies, which will reduce the implementation size and its cost as well as the complexity of the system. Another advantage presented by such antennas is the ability to reject the interference from services coexisting in the spectrum. Slot antennas are one of the common types of antennas used in frequency tuning because varactors or switches could be used easily to change their resonant frequency. The frequency tuning is achieved by varying the effective length of the slots using varactor diodes embedded across the slots. Therefore, a tunable antenna could be obtained by loading the patch antennas with varactor diodes, [14], thus having a change in the resonant frequency of the patch.

The authors of [15] present a slot antenna that uses switches to change its electrical length and obtain an effective wide bandwidth. A small patch antenna that could be tuned to different frequencies, ranging from 800MHz to 900 MHz, is designed in [16], this antenna uses variable capacitors and transistors. A radio baseband processor sends commands to a digital control unit implemented in the antenna to control it. In [17], two lumped variable capacitors, also known as varactors, are placed in properly chosen positions on a slot antenna. The antenna is changed to a dual-band antenna having two operational frequencies that could be controlled individually.

In this paper, the varactors are placed according to three different scenarios. The first thing done is to plot the current magnitude on the surface of the antenna in [1] and indicate the points where it is the highest and the lowest. The first scenario is designed by placing a varactor component where the current gets cancelled out, which is in the middle of each ladder step. The two other scenarios are designed by choosing two different points where the current is circulating with high magnitude. These two points appeared to be on the side strip connecting the ladder elements at the same side of the feeding point, one between the third and the fourth element and the other one is between the fourth and the fifth ladder element. Two designs were tested for the second and third scenarios; first, the copper strip facing the varactor had to be removed to avoid short-circuiting the inductors and drive the current to pass through the varactor. The second design was done by removing all strips similar to the one facing the varactor from the ladder patch. All the proposed designs were fabricated and tested and good agreement was obtained between the simulated and the measured reflection coefficients.

2. Original Ladder Antenna: Design and Fabrication

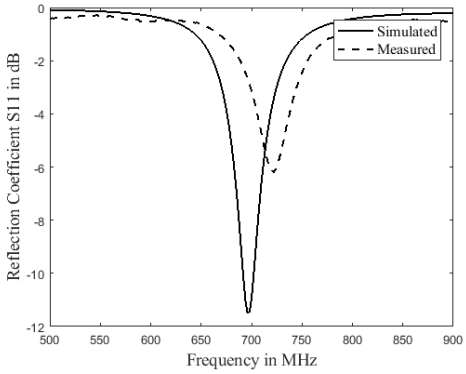
The antenna in [1] is designed with a size of 120 x 50 mm^2 on an FR4 epoxy substrate with a dielectric constant of 4.4 and a thickness of 1.6 mm. All the dimensions of the antenna are shown in Figure 1(a).



(a) Dimensions.



(b) Prototype.



(c) Simulated and measured S_{11}.

Figure 1: Original ladder shape antenna.

The substrate has a rectangular shape of length 120 mm and a width of 50 mm. The patch has also a rectangular shape of length 116 mm and a width of 46 mm which means that it is smaller than the substrate by 2 mm from each side. The ladder is composed of eleven copper rectangular steps of length 40 mm and of width 7 mm with a spacing of 3.7 mm between each two steps. Finally, the width of each of the two side strips connecting the eleven steps is 3 mm. As already said, the antenna is fed using a coaxial cable and the position of the feeding point was chosen to be in the middle of one of the side strips connecting the ladder elements. A parametric study was done before making the best choice of the feeding position to get the best possible performance of the antenna. The fact that this antenna occupy the entire space of the PCB does not affect the addition of other antennas necessary for the functionality of the device because most of the antennas designed now are very small strip antennas and could use different PCBs.

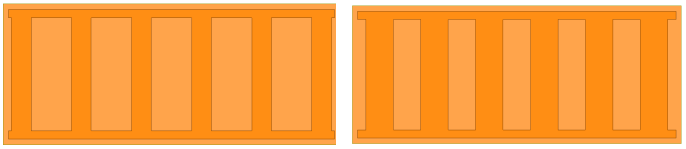
After the fabrication of the antenna as shown in Figure 1(b), measurements were done on its reflection coefficient. Both the simulated and the measured S_{11} are shown in Figure 1(c). There is a good agreement between the simulated and the measured results. The value of -6 dB is the threshold of acceptable S_{11} values for TV antennas therefore the antenna is found to be operational at 700 MHz with a bandwidth of 30 MHz.

The reflection coefficient of the antenna shows that it has a narrow band covering only one TV channel. However, to be able to cover a wider bandwidth in the UHF band, many solutions could be proposed one of which is the use of surface mounted varactors and changing their capacitance values which allows the tuning of the operational frequency of the antenna to different UHF channels.

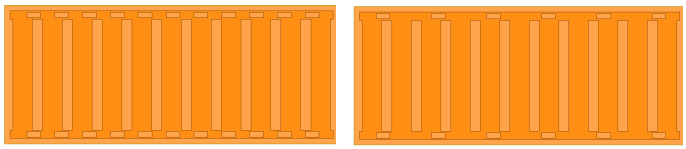
3. Optimization Approaches

To improve matching conditions at 700 MHz, four approaches were tested in [1]. The first approach, Figure 2(a), was to decrease the number of ladder steps to six. The second, Figure 2(b), was to use the design obtained after applying the first approach and increase the width of the steps to 10 mm which is an increase of 43% of their original width. The third approach, Figure 2(c), was to reuse the original design and subtract eleven slots of size 5 x 2 mm^2 from both, upper and lower, side strips connecting the steps making them look like film strips. The final approach tested, Figure 2(d), was to decrease the number of slots subtracted to six and check their influence on the reflection coefficient. The four approaches mentioned are combined with Figure 2(e) that shows the reflection coefficients of the antennas obtained after applying these approaches as well as that of the original antenna simulated using Ansoft HFSS.

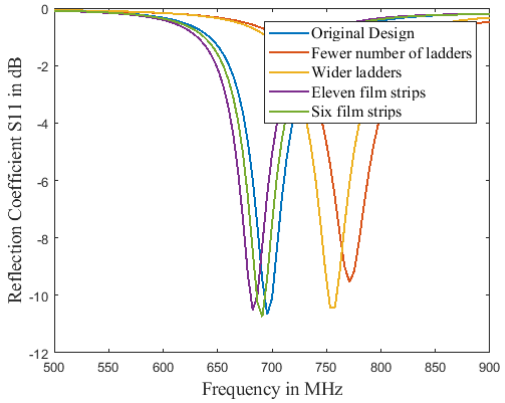
No major improvement is shown in the reflection coefficient of the antenna. Instead, resonant frequency is shifted away from 700 MHz in the four cases. Therefore, since the chosen frequency is 700 MHz and also thanks to the simplicity of the design and its easy fabrication, the antenna of the original design is the one chosen to be fabricated and measured. The network analyzer used to do the measurements is the "FieldFox RF Analyzer N9914A" from the Keysight Company. A small acceptable difference between the simulated and the measured S_{11} could be noticed and it is caused by some fabrication losses therefore, the ladder antenna fabricated is a good candidate to be used for TV applications.



(a) Approach 1. (b) Approach 2.



(c) Approach 3. (d) Approach 4.



(e) Simulated S_{11} of the above approaches.

Figure 2: Investigated optimization approaches.

4. Frequency Tunable Ladder Antenna Scenarios: Designs and Fabrication

The addition of varactor diodes along the current path leads to changing the electrical length of the patch and hence the operating frequency. It all depends on the capacitance value of the varactor; for some capacitances the electrical length increase and the resonant frequency decrease and for some other cases, the inverse occurs.

To plainly explain the effect of the capacitance on the transmission line length, it is necessary to introduce the Smith Chart. The Smith Chart is a way of graphically determining transmission line parameters rather than mathematically. Consider a transmission line, it has an impedance which is usually a complex number expressed by (1) where R is the resistance and X is the reactance and can be plotted on the smith chart using the resistance and the reactance circles. A complete circle on the Smith Chart represents a half wavelength transmission line. One of the important applications of a Smith Chart is the length calculation of a short-circuited piece of transmission line to provide a required capacitive or inductive reactance. The addition of a series RC circuit will change the impedance of the transmission line by adding R to its real part and X_c, calculated using (2), to its imaginary part.

$$Z = R + jX \quad (1)$$

$$X_c = \frac{1}{2\pi f C} \quad (2)$$

f is the frequency in Hertz and C is the capacitance in Farads. Any change in the position of the transmission line impedance point on the Smith Chart, caused by adding matching circuits such as a series RC circuit, will lead to a change in the electrical length of the transmission line as well as its impedance which modifies its matching conditions. The electrical length will either increase or decrease, therefore the operational frequency which is directly related to it, will also change. The distance traveled by a current signal along a transmission line is expressed in relation to its source frequency. The wavelength calculated by (3), where c is the speed of light, is this distance considered for a given frequency.

$$\lambda = \frac{c}{f} \quad (3)$$

In the two suggested scenarios, the varactors are connected serially across a slot subtracted from the radiating patch, which means between two halves of the patch as in [18] and [19]. According to [20], the tunability of the antenna is therefore achieved by cutting small slots from the copper lines and inserting varactors in their place.

The diode model used here is a serial connection between a capacitor and a resistor as shown in Figure 3. The resistor has a very small value of 0.01 Ω. The capacitance of the varactor that is used, 1SV325, is varied between 7 pF and 80 pF when a reverse voltage between 0V and 6V is applied.

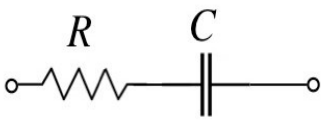
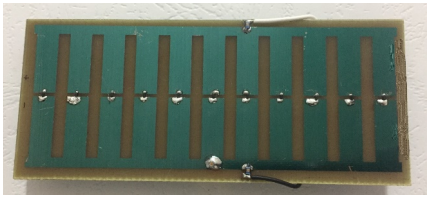


Figure 3: Used configuration of the varactor model depending on resistance R and capacitance C implemented in series.

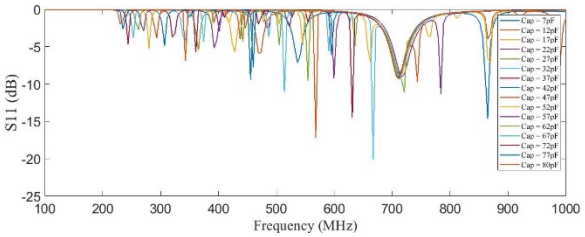
In what follows, the simulated reflection coefficients are presented for some capacitance values ranging between 7 pF and 80 pF. The measured reflection coefficients are captured while varying the DC bias voltage. It is varied between 0V and 6V for the case of only one varactor added to the surface of the ladder patch antenna represented by scenarios 2 and 3 in the following figures. However, the voltage could be varied in a greater range for the antenna of scenario 1 that has a varactor added to each of its ladder steps because the varactors are in a parallel connection to each other so the voltage would be divided across all the varactors.

Figure 4(a) represents the first antenna scenario where eleven varactors are added each in the middle of one of the steps, with the simulated S_{11} in Figure 4(b) and the measured S_{11} in Figure 4(c).

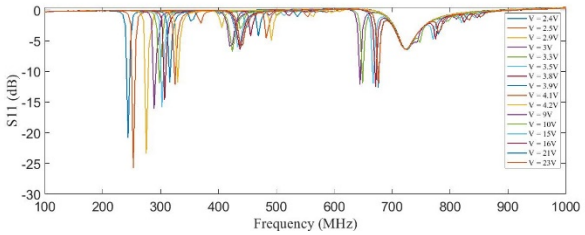
In the second antenna scenario, the varactor is added between the third and the fourth ladder step. Figure 5(a) represents the first design of the second scenario with only one copper strip removed. The simulated S_{11} is represented in Figure 5(b) and the measured S_{11} in Figure 5(c).



(a) One Varactor in the middle of each ladder step.

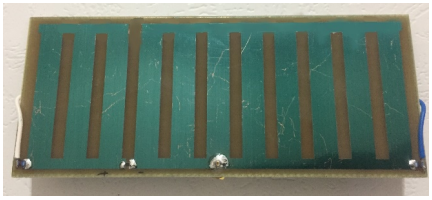


(b) Simulated Reflection Coefficient.

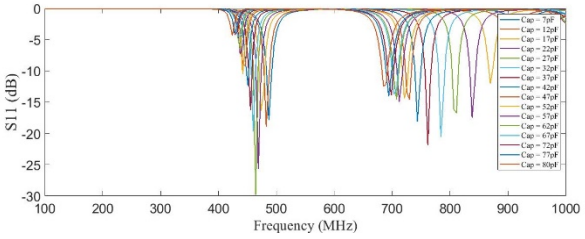


(c) Measured Reflection Coefficient.

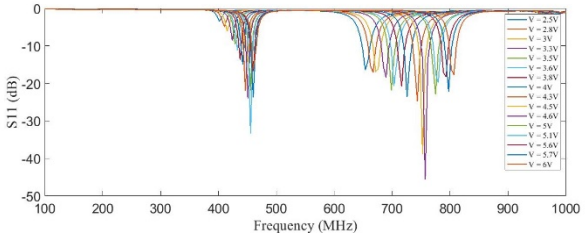
Figure 4: Scenario 1.



(a) One Varactor between the third and the fourth step with one facing copper slot removed.



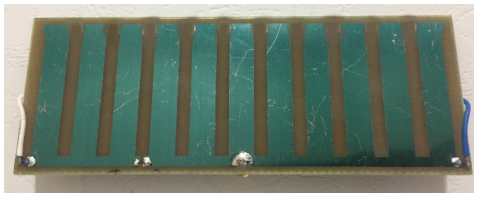
(b) Simulated Reflection Coefficient.



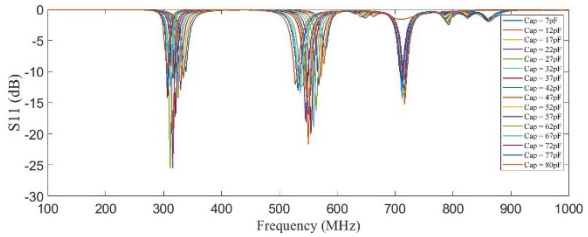
(c) Measured Reflection Coefficient.

Figure 5: Scenario 2 - Design 1.

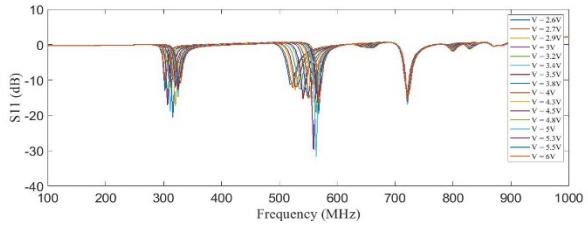
Figure 6(a) represents the second design of the second scenario with all copper strips removed. Its simulated S_{11} is represented in Figure 6(b) and the measured S_{11} in Figure 6(c).



(a) One Varactor between the third and the fourth step with all copper slots removed.

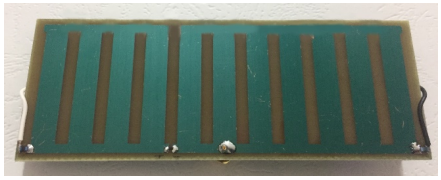


(b) Simulated Reflection Coefficient.

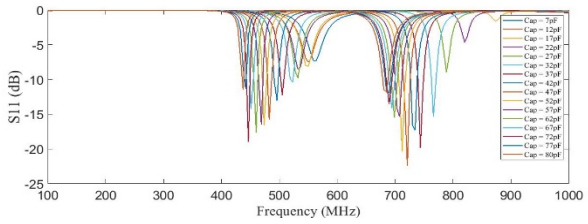


(c) Measured Reflection Coefficient.

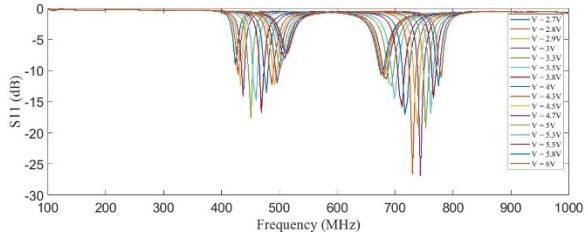
Figure 6: Scenario 2 - Design 2.



(a) One Varactor between the fourth and the fifth step with one facing copper slot removed.



(b) Simulated Reflection Coefficient.



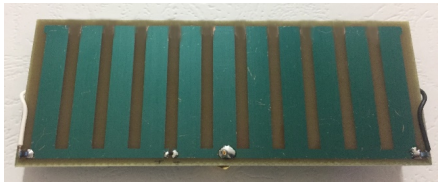
(c) Measured Reflection Coefficient.

Figure 7: Scenario 3 - Design 1.

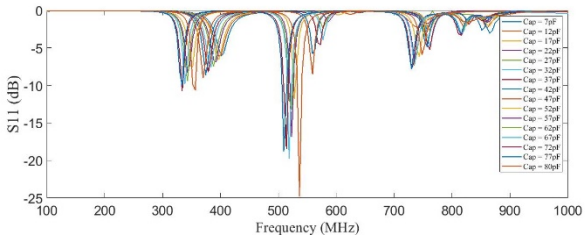
In the third antenna scenario, the varactor is added between the fourth and the fifth ladder step. Figure 7(a) represents the first design of the last scenario where only one copper strip is removed. The simulated S_{11} is represented in Figure 7(b) and the measured S_{11} in Figure 7(c).

Figure 8(a) represents the second design of the last scenario with all copper strips removed. Its simulated S_{11} is represented in Figure 8(b) and the measured S_{11} in Figure 8(c).

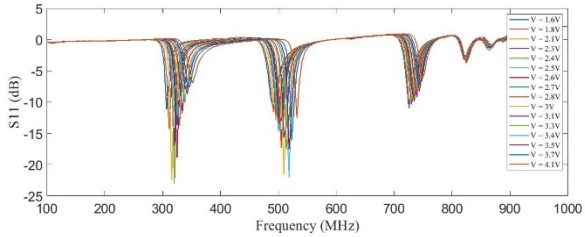
In the figures showing the simulated and the measured reflection coefficients for all the antennas deigned, a slight difference could be noticed. This difference is due to the fabrications inaccuracies as well as the losses that could be encountered in the measurement process. However, it is well shown that all the tunable frequencies fall in the same frequency ranges when comparing simulated and measured results for each design.



(a) One Varactor between the fourth and the fifth step with all copper slots removed.



(b) Simulated Reflection Coefficient.



(c) Measured Reflection Coefficient.

Figure 8: Scenario 3 - Design 2.

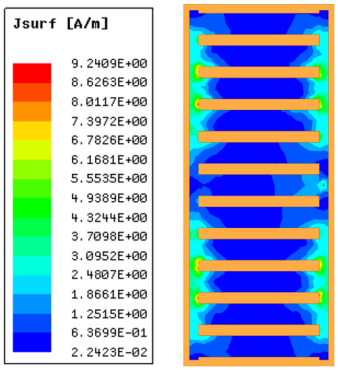
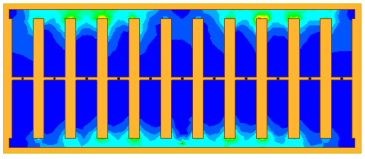


Figure 9: Current distribution for antenna [1].

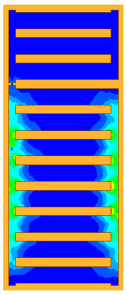
5. Current Distribution

Figure 9 shows the current distribution on the surface of the ladder patch antenna in [1] at the resonant frequency 700 MHz. It is shown that the current magnitude is not equal in all of the points on the patch; the current on the ladder steps is lower than the one circulating on the side strips connecting them. This is obvious since the current is being canceled out in the middle of the steps because of the fact that it is traveling in opposite directions.

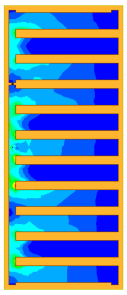
Figure 10 represents the current distribution on the surface of the ladder patch for the five tunable designs discussed above at 700 MHz frequency.



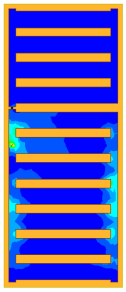
(a) Scenario 1.



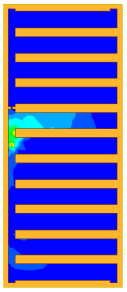
(b) Scenario 2 - Design 1.



(c) Scenario 2 - Design 2.



(d) Scenario 3 - Design 1



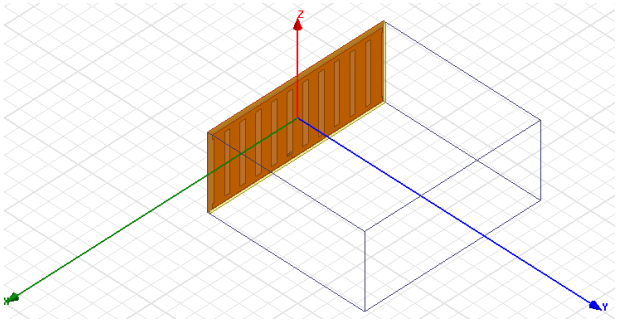
(e) Scenario 3 - Design 2

Figure 10: Current distribution for the tunable antennas.

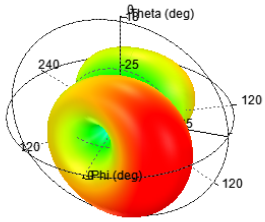
6. Radiation Characteristics

Figure 11(a) shows antenna [1] positioned in the xyz-plane. The three-dimensional radiation pattern of the antenna is shown in Figure 11(b). The main lobe of the antenna is directed in the plane formed by theta equals 90° and phi equals 45°.

Figure 12(a) represents the first antenna design of the second scenario which is having the varactor placed after the third step and removing the copper strip facing it. Two plots of the radiation pattern are presented in Figures 12(b) and 12(c) for resonating frequencies 840 MHz and 870 MHz respectively. Both plots show that the main lobe changed its direction when compared to that of antenna [1].

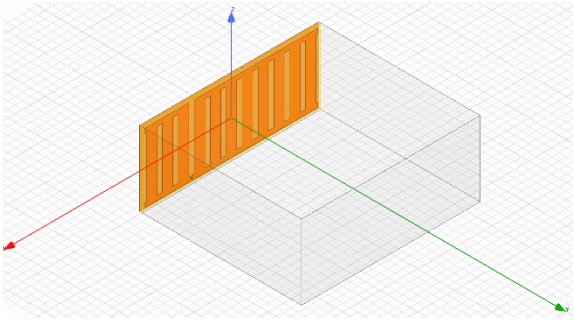


(a) Antenna [1] in the xyz plane.

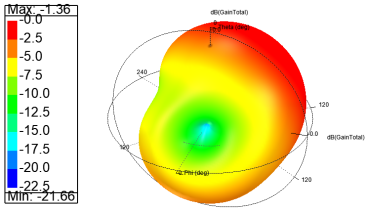


(b) 3D Radiation pattern.

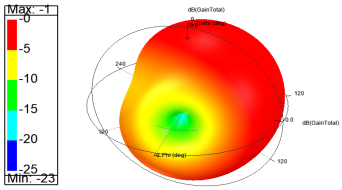
Figure 11: Antenna [1] - Radiation Characteristics.



(a) Antenna of scenario 2 - Design 1.



(b) 3D Radiation pattern at 840 MHz.



(c) 3D Radiation pattern at 870 MHz.

Figure 12: Antenna of scenario 2 - Design 1 - Radiation Characteristics.

It is shown that the gain of this antenna is negative; that is due to its small size compared to what it should have been for operational frequencies in the UHF band. For 700 MHz frequency, the wavelength is around 400 mm however the length of this antenna is 120 mm, which means that the current is circulating on a path that is much smaller than what it should be. However, as

already stated in the introduction, this antenna is considered as a receive-only system in which efficiency could be sacrificed more than in a transmit-system.

7. Conclusion

The antennas presented in this paper are an extension of the ladder-shape microstrip patch antenna design presented in the conference paper [1]. When comparing the antenna presented in [1] to other antennas in the literature designed for the same purpose, it shows good improvement in terms of the compactness of its size. It has a length of 120 mm whereas other antennas are of length 240 mm, [3], and 245 mm, [4]. Even though its size suits most of the mobile phones used in the market, this antenna has one drawback which is its narrow operational bandwidth of 30 MHz.

The proposed solution in this paper is to transform this antenna into a frequency tunable antenna by adding varactors onto its surface. A slight change in the geometrical shape of the antenna will create new current paths and new radiation edges which will introduce multiple shifts in the resonant frequency. According to [21], the position of the varactors obviously affect the current flow path, so to tune the obtained resonant frequency, changing these positions is a must.

The varactors are placed in certain positions that achieve frequency tuning while keeping sufficient matching conditions with minimal perturbation. The antennas presented in this paper resonated on new sets of frequencies falling between 200 MHz and 1 GHz, which allows them to be used for Digital TV broadcasting applications on mobile handheld devices.

These antennas have a negative gain because of their small size according to their frequency of operation however, many techniques could be used to improve the antenna gain such as changing the substrate and using a different material having lower losses such as air, adding parasitic patches or meandering edges, cutting parts of the ground plane, etc…

Acknowledgment

The authors would like to acknowledge the support of Mr. Issam Darwish and Mr. Mohammad Darwish for supporting this work.

References

[1] M. Moussa, M. Madi and K. Kabalan, "Ladder Shape Microstrip Patch Antenna," in *International Conference on High Peformance Computing and Simulation*, Orleans, France, July 2018.

[2] P. Vainikainen, J. Holopainen and M. Kyro, "Antennas for Digital Television Receivers in Mobile Terminals," *Proceedings of the IEEE,* vol. 100, no. 7, pp. 2341-2348, July 2012.

[3] C.-Y. Huang, B.-M. Jeng and J.-S. Kuo, "Grating Monopole Antenna for DVB-T Applications," *IEEE Transactions on Antennas and Propagation,* vol. 56, no. 6, pp. 1775-1776, June 2008.

[4] C.-Y. Huang, B.-M. Jeng and C.-F. Yang, "Wideband monopole antenna for DVB-T applications," *Electronics Letters,* vol. 44, no. 25, pp. 1448-1450, December 2008.

[5] O. T.-C. Chen and C.-Y. Tsai, "CPW-Fed Wideband Printed Dipole Antenna for Digital TV Applications," *IEEE Transactions on Antennas and Propagation,* vol. 59, no. 12, pp. 4826-4830, December 2011.

[6] C.-Y. Tsai and O. T.-C. Chen, "Compact Broadband Monopole Slot Antenna for Digital TV Applications," in *IEEE Asia-Pacific Conference on Antennas and Propagation*, Singapore, October 2012.

[7] M. Y. Abou Shahine, M. Husseini, A. H. Ramadan, K. Kabalan and Y. Nasser, "Antenna Designs for Cognitive Radio Applications in the TV Band," *International Journal of Scientific & Engineering Research,* vol. 6, no. 2, pp. 975-980, February 2015.

[8] T. Pratumsiri and P. Janpugdee, "Flexible Printed Active Antenna for Digital Television Reception," in *Progress in Electromagnetics Research Symposium (PIERS-Toyama)*, Toyama, Japan, August 2018.

[9] S. Bhole, A. Rathod, S. Doddipalli, S. Kannaiyan and A. Kothari, "A Compact Planar Antenna with Meander Lines for TV White Space Applications," in *IEEE International Students Conference on Electrical, Electronics and Computer Science*, Bhopal, India, November 2018.

[10] R. K. Venkata, D. Mrthy and I. a. N. M. N. Govardhani, "Multi-Band Ladder-Shape Micro-Strip Patch Antenna," *International Journal of Scientific & Engineering Research,* vol. 3, no. 3, pp. 1-6, March 2012.

[11] G. Singh and m. Kumar, "Design of Frequency Reconfigurable Microstrip Patch Antenna," in *6th International Conference on Industrial and Information Systems*, Kandy, Sri Lanka, 2011.

[12] M. S. Nishamol, C. K. Aanandan, P. Mohanan and K. Vasudevan, "Dual Frequency Reconfigurable Microstrip Antenna Using Varactor Diodes," in *XXXth URSI General Assembly and Scientific Symposium*, Istanbul, Turkey, August 2011.

[13] A. Khidre, F. Yang and A. Elsherbeni, "A Patch Antenna with a Varactor-Loaded Slot for Reconfigurable Dual-Band Operation," *IEEE Transactions on Antennas and Propagation,* vol. 63, no. 2, pp. 755-760, February 2015.

[14] B. Babakhani and K. S. Sharma, "Investigations on Frequency Agile Microstrip Patch Antenna Loaded with Varactor Diode," in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Orlando, FL, USA, July 2013.

[15] D. Peroulis, K. Sarabandi and L. Katehi, "Design of Reconfigurable Slot Antennas," *IEEE Transactions on Antennas and Propagation,* vol. 53, no. 2, pp. 645-654, February 2005.

[16] J. T. Aberle, S.-H. Oh, D. T. Auckland and S. D. Rogers, "Reconfigurable Antennas for Wireless Devices," *IEEE Antennas and Propagation Magazine,* vol. 45, no. 6, pp. 148-154, December 2003.

[17] N. Behdad and K. Sarabandi, "Dual-Band Reconfigurable Antenna with a Very Wide Tunability Range," *IEEE Transactions on Antennas and Propagation,* vol. 2006, no. 2, pp. 409-416, February 2006.

[18] N. Fayyaz, S. Safavi-Naeini, E. Shin and N. Hodjat, "A Novel electronically Tunable Rectangular Patch Antenna with One Octave Bandwidth," in *IEEE Canadian Conference on Electrical and Computer Engineering*, Waterloo, Ontario, Canada, May 1998.

[19] S. V. Hum and H. Y. Xiong, "Analysis and Design of a Differentially-Fed Frequency Agile Microstrip Patch Antenna," *IEEE Transactions on Antennas and Propagation,* vol. 58, no. 10, pp. 3122-3130, October 2010.

[20] J. Costantine, C. Christodoulo and S. E. Barbin, "A New Reconfigurable Multi Band Patch Antenna," in *SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference*, Brazil, October 2007.

[21] M. A. Madi, M. Husseini, A. Ramadan, K. Y. Kabalan and A. El-Hajj, "A Reconfigurable Cedar-Shaped Microstrip Antenna for Wireless Applications," *Progress in Electromagnetics Research,* vol. 25, pp. 209-221, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 6, 53-60 (2019)

www.astesj.com

Special Issue on Recent Advances in Engineering Systems

ASTESJ

ISSN: 2415-6698

Creating a Digital Twin: Simulation of a Business Model Design Tool

Kira Rambow-Hoeschele*, Nick Giani Rambow, Matthias Michael Hampel, David Keith Harrison, Bruce MacLeod Wood

School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow, UK

ARTICLE INFO

Article history:
Received: 03 December, 2018
Accepted: 02 November, 2019
Online: 22 November, 2019

Keywords:
Business model development
Digital twin
Simulation tool

ABSTRACT

Digitization forces industry players to adapt to transforming market situations and buyer behavior. Technological advances, buyer power, and sharpened competitive intensity imply that businesses are confronted with the menace of commoditization. For companies to perform successfully in the market, outdated business models ought to be rethought and new business models should be created. Unique selling propositions and differentiation through research, innovation, and holistic stakeholder involvement help industry players to master the change. A tool was built to support businesses facing the consequences of digital transformation: the Business Model Builder. This research paper explores the steps of creating a software version of the analog Business Model Builder. The digital twin enables firms to simulate the iterative adaptation of business models to permanently changing market circumstances and customer demands on an ongoing basis. The user of the tool can edit single variables, understand interrelations, and see the effect on the outcome of the business case, e.g., earnings before interest and taxes or economic value added. Accordingly, the simulation offers the opportunity to have a dynamic view of the business model where any variances of input parameters are reflected in the business case. Thus, profitability, feasibility, and functionality of a business model can be validated, tested, and reviewed in the digital simulation tool.

1. Introduction

Digitization has an impact on most industries, value-creating architectures, and business models based upon them. The impact depends on the market and can – in several cases will – make current business models outdated.

Digitally connecting devices – TVs, household appliances, watches, etc. –, the Internet of Things creates chances for new business models in B2C commerce. Nevertheless, they need to be created and implemented as a first step. Furthermore, the Internet of Things substitutes business models in the B2B segment [1].

To adjust current business models to the transforming market situations, or to come up with new business models, the Business Model Builder was developed within the cooperative research project "low-carbon city" [2] – a project, state subsidized by the German Federal Ministry of Education and Research from August 2016 to December 2018 (references 02K12A150 and 02K12A151). The Business Model Builder represents a digital tool that depicts the modules, drivers, and interrelations vital for the development or modification of business models.

2. Methodology

The creation of the digital twin simulating a business model design tool follows three major steps: conceptualization and design of the analog Business Model Builder, identification of interdependencies within the tool, and coding of the software-supported version to create the digital twin.

3. Digitization Process of the Tool

To customize single elements of a business model during its development, and to reveal consequences of changes on the other elements, the Business Model Builder is designed using application software. This software builds a digital twin of the business model. A digital twin stands for to a digital replica of a business model [3, 4].

This digital duplicate of the business model simulates adaptations or improvements of e.g., a product involved in the

* Kira Rambow-Hoeschele, Email: kira.rambow-hoeschele@t-online.de

business model. This includes testing the effects on function, handling, visual appearance, etc.

Internal links and databases in the background of the application enable to digitally show how adjustments of individual elements impact other ones. Additionally, the business case is calculated in the tool.

Not merely does the digital version of the Business Model Builder provide an iterative and efficient creation of business models, it also provides real-time information. Through the real-time depiction, transformations of business models can be reviewed in a convenient way.

4. Reference to Existing Work and Theories

Over time, the term "business model" has changed and has been defined in various ways. In order to fully comprehend the concept, it is therefore necessary to look at the past definitions as well as future requirements that business models need to fulfill. Business models illustrate a business idea as well as the ways and means by which the idea can be successfully implemented [5]. A business is defined as a product or service that owns a market presence and creates market- and resource-based synergies through collaboration with other entities [6].

Other researchers define the term "business model" as the synthesis of how a company selects its customers, how it defines and differentiates its products, how it structures its internal and external activities, how it uses its resources, how it enters the market, how it creates customer value, and how it earns profits [7]. A business model represents the complete system of how customer value is created and how it is monetized [8]. For several researchers, the elements of value propositions as well as customer segments are of high importance. Moreover, key resources and key activities that are necessary to create customer value as well as revenue streams are considered vital.

According to Timmers, a business model is "an architecture for the product, service and information flows, including a description of the various business actors and their roles and a description of potential benefits for the various business actors and a description of the sources of revenues" [9].

Business models can also be seen as planning and analysis tools. Companies dominating their markets therefore constantly adapt, refine, and develop current or new business models in order to remain competitive [10].

With the new economy boom at the beginning of the current century, the term "business model" was transferred to the business world, as competition was no longer defined as produces and services competing against each other but rather as a clash among various business models [11].

This led to changes in thinking about how to conduct business and where to best allocate resources in order to maximize competitiveness and value creation, since companies no longer possessed competitive advantages due to a superior product or service but rather through the design of complex and advanced business models. Instead of producing a good and simply selling it to the market, the business model needs to tell a reasonable story that is applicable to today's accelerated and more volatile business environment [12].

Moreover, a business model can be defined as a business concept that is applied in the real world. A business concept identifies what value the business is providing to its stakeholders. Also, a business concept is an architecture that should describe how value is generated for customers and business partners. This architecture should include the different steps of value creation. Moreover, it should show how the business model will be able to generate revenues and profits to enable decision makers to identify the value of business ideas and the sustainability of their long-term success [13].

Most researchers see the term "business model" as a simplified depiction of business activities. However, the understanding of the importance of certain aspects of the business model varies [14–17].

Osterwalder and Pigneur define a business model as "nothing else than the value a company offers to one or several segments of customers and the architecture of the firm and its network of partners for creating, marketing, and delivering this value and relationship capital, in order to generate profitable and sustainable revenue streams" [18].

To visualize and to test the realization of business models, they developed the so-called "Business Model Canvas." This is a tool for business model development that can also be used as a checklist to provide an overview of the tasks and requirements that need to be accomplished in order to turn the idea into a real-life business. It is especially helpful to create a common basis for all stakeholders that are involved in the business development process [19].

The concept consists of nine different components that should depict the model in regard to all important parts: customer segments, value proposition, channels, customer relationship, revenue streams, key resources, key activities, key partnerships, and cost structure.

The customer segments component includes consumers as well as businesses. This building block enables a clearer understanding of who the different parties are that are being served by the business. It provides clarity about whom value should be created for and how important those parties are for the business. The different customer segments can be structured according to their needs, their behavior, their specific characteristics, their financial buying power, and their preferred channels.

The value proposition consists of products or services that create value for specific customer segments. This value provided can differ with regard to its nature. It might include performance optimization, branding, work simplification, or entirely new products or services.

The customer relationships component describes how the business interacts with its customers to showcase how it can gain and retain customers. It shows how the relationship between the customer and the company is formed and what kind of relationship they have when conducting business.

The channels component includes all sorts of customer touchpoints as well as the different channels that are used to communicate with the customer and that are needed to distribute and sell the product or service.

Key activities include all important activities that are necessary for the entity in order to conduct business. This includes the processes that are needed to create and provide the value proposition to the customer, that are necessary to build and retain customer relationships, and that are used to reach all markets that are relevant for the business idea.

Key resources include all kinds of resources of physical, financial, human, or intellectual natures. This component gives an overview of all resources that are necessary to conduct the business.

In the key partnerships component, the network of all partners and suppliers that are needed for the business idea is described. This might include strategic alliances, joint ventures, and buyer-seller relationships.

Moreover, in the cost section of the business model canvas, all kinds of key fixed and variable costs are displayed in order to identify the biggest cost drivers that will affect the profitability of the business.

One of the most pivotal components of this model, however, is the revenue part. This section, in which the actual revenue streams are explained, will ultimately provide guidance on whether the business will be successful or not. This might include revenues from sales, rental services, leasing contracts, advertisements, commissions, or subscription fees.

In the following section, several innovative business model ideas are described. Some have existed for a long time in the business world, while others are comparably new, being enabled through technological developments such as e-commerce.

5. Innovative Business Model Approaches

One business model example is the sale of products or services with optional add-on features. This model enables the customer to buy a basic product at a lower price and then be charged a higher amount for additional options. Examples of such business practices can be seen in a variety of businesses such as the airline industry, the automotive industry, or the software industry [20].

A relatively new business model is revenue generation through affiliate links. Through the usage of such links, companies receive a commission based on the generation of sales of products or services through those links. Such business practices are used in e-commerce businesses, e.g., through review or price comparison websites generating additional sales channels for producers [21].

Another prominent business model, especially in the e-commerce sector, is based on auctions. This model enables customers to set the price, leading to higher ambiguity for the seller. Whereas the seller might be able to achieve higher returns in comparison to a traditional sales approach, the business risk increases as the predictability of revenues decreases [22].

Moreover, some companies use a barter approach to acquire new customers by providing free samples of their product. This is often found in the area of consumable products, where potential customers receive a small sample to convince them to convert to paying customers who buy the real product [22].

Bundling and cross-selling approaches capture customers through ecosystems of products and services that, once a customer is acquired, make it rather difficult for them to leave the ecosystem again. Often, technology firms provide a software and hardware ecosystem that enables the customer to synchronize devices and use services on a variety of devices [23, 24].

Another innovative business approach is crowdfunding. This enables companies to make investments they would not have been able to make without the financial support of a broad variety of sponsors [25].

Crowdsourcing business models can be found in a variety of industries. Some companies only try to focus on their specific area of expertise and follow a crowdsourcing approach instead of a vertical integration approach, leading to potential cost savings and more flexibility [26].

Moreover, some companies base their business model on customer loyalty. Many of those companies earn revenues through the collection and analysis of data that is gained from customers who profit from customer loyalty programs, paying with their data [27].

An additional business model that has been developing over the past decades with the rise of digital technology is digitalization. Many businesses have been successful in providing digitized content to customers instead of more resource-intensive and less flexible physical content [28].

Furthermore, the development of the internet created higher market transparency for customers as well as companies and established new ways for producers to earn revenues through direct selling. By leaving out intermediaries, businesses can turn profitable and establish new market opportunities that had not existed before [29].

E-commerce business models are currently transforming the retail business and a variety of industries in ways that have not been seen before. Through lower costs, targeted advertising, broader product offerings, and further advantages, these businesses threaten many well-established companies [30].

Other businesses use flat-rate business models that require users to pay a specific amount in order to use a product or service to an unlimited extent. This has been visible in the online media world, e.g., for music or movie streaming [31].

Additionally, companies use franchising models that enable large scaling effects. After a concept has been developed and a brand established, e.g., in the food industry, local business owners pay for the usage of and supply from the brand [32].

A business model that is frequently visible in the web services industry is the freemium model. This allows customers to use a product in a basic form or for a limited time and offers a premium version that must be paid for [33].

An alternative business model that is often used in the web services industry is hidden revenues. Through the display of advertisements or promotions of complementary products, companies are able to generate revenues [31].

Also, many companies track, analyze, and sell customer data collected through a variety of channels in order to earn revenue. This model is especially used by social media platforms [34].

Intellectual property that might have not been commercialized in the past can be monetized through licensing. Many software firms base their business models on this architecture [33].

Furthermore, through easier accessibility and lower costs due to digitization, long-tail business models have been created that offer products and services that include not only mass products but also niche products that have now become profitable due to economies of scale and lower barriers of entry into a market [35].

Another business model, which became attractive to companies over the past decade, is mass customization, allowing customers to buy products that are mass-produced but customized for them. This has been enabled through technological developments in production, e.g., 3D printing, as well as automation of processes and further technologies. Exemplary industries are the automotive sector as well as the technology hardware sector [36].

Moreover, through lean startup approaches and the reduction of certain product features, no-frills business models have enabled market entry for lower-cost products. A variety of consumable product companies have been established that focus on specific, highly demanded features of a product while leaving out less important features in order to reduce price and increase product attractiveness as well as market attractiveness [37].

Other companies might use open source business models that allow external parties to improve the product. This is a commonly used approach in the software industry, where developers use certain tools that are more adaptable than standard software [38].

An approach that creates flexibility in regard to pricing is the "pay what you want" business model. This architecture can be found in, for example, the cultural sector or the web application sector [39].

With an increased usage of data analytics, pay-per-use business models have become more viable over the past years. Mobility services providers have been using this business model extensively [31].

The razor and razor blade business model architecture is also seen quite frequently in today's consumables industry. Oftentimes, a basic product is provided at a low cost or even free. However, when a replacement of particular parts is needed, the purchase of those parts is more cost-intensive for the customer [33].

A very prominent business model, especially in the web-based industry, is a subscription architecture. This enables companies to earn recurring revenues that might lead to less price-sensitive customers who accept their current costs due to automatic payments. They thereby receive products or services periodically, such as consumables, media, or clothing [40].

Finally, everything-as-a-service (XaaS) business models have been seeing an increasing presence over the last years. In particular, in the software industry, product purchases have been transformed into consuming a service [41].

Those briefly introduced business models are only examples of currently existing business models and should not be seen as an exhaustive list. However, they give a good indication of the variety and complexity of business models that exist nowadays. The aforementioned business models are not only used exclusively but are often implemented as a combination of different models. Due to the potential complexity of business models, it is of high importance to use a development approach that is comprehensive, easy to understand, and flexible to changes.

6. Comparative Analysis of Business Model Definitions and Elements

The creation of a business model consists of multiple connected subcomponents, which build several components, which again join up to form modules. The business model building tool presents the individual elements of a business model and their underlying logic. The Business Model Builder is shown in low granularity in Figure 1.

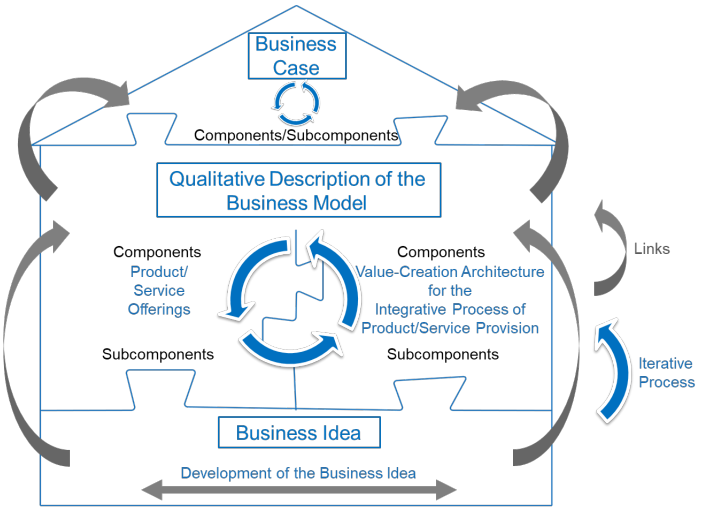


Figure 1: Business Model Builder Depicted in Low Granularity

The application of the Business Model Builder provides the user with business-specific yield numbers indicating the anticipated commercial performance of a business model. The application and field test of the analog Business Model Builder were a success [1].

There have been some former approaches for the creation of business models, e.g., from Gassmann et al. (2013) [42], Schallmo (2013) [43], Osterwalder and Pigneur (2011) [19], or Timmers (1998) [9]. These approaches often consist of diverse checklists, which are mainly grounded on theoretical concepts.

Including all important business model elements, a comparison and evaluation of established business model definitions took place (Figure 2) [7, 9, 10, 13, 15–18, 29, 40, 42–47].

Business models are built of several single elements with differing effect on the outcome of the business case. Various approaches for the creation of a business model can be chosen. For the digitization process of a business model creation tool, nonetheless, it is relevant to individually list all parts. This breakdown eases converting the impact and interrelationships of the single parts into a digital tool since precise relations can be assigned.

To comprehensively include all characteristics required for the development of the digital twin, the Business Model Builder was developed, originating from and enhancing former concepts.

As mentioned earlier, a checklist lacks some details for the complete development of a business model. Every business model requires a well-considered idea as its basis. Several checklists and related concepts still need to examine the soundness of a business idea and review the concept for its resistance. More than fifty percent of new business formations fail [48]. A key reason are unsustainable business ideas [49].

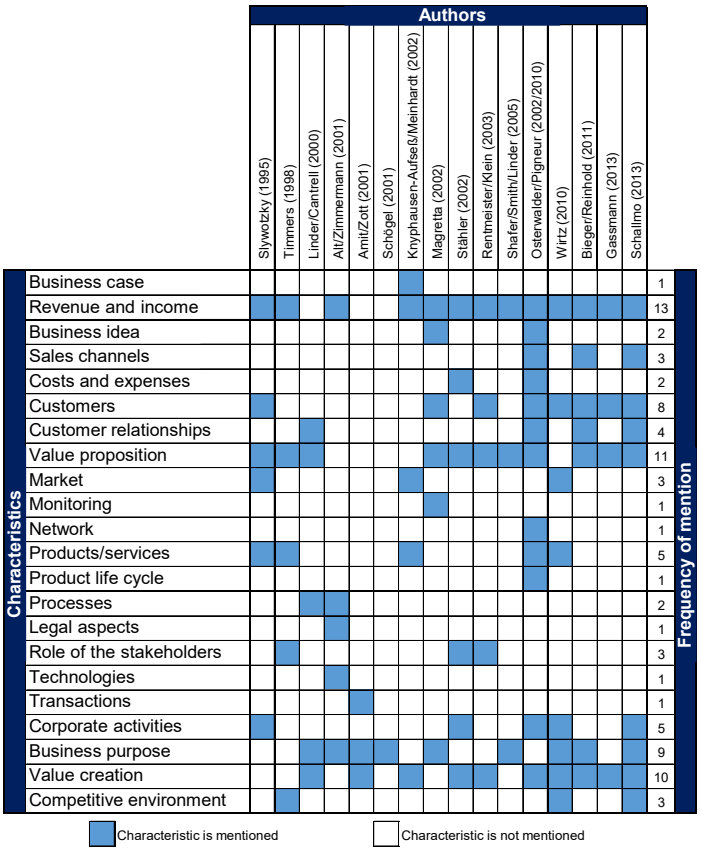


Figure 2: Comparative Analysis of Existing Business Model Theories

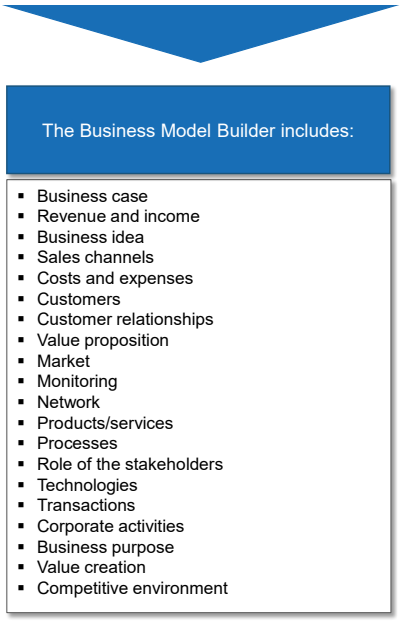


Figure 3: The Business Model Building Tool Extends the Existing State of Knowledge

Figure 3 shows that the Business Model Builder involves all characteristics that are vital for the success of a business model. The tool structures the elements to give a thought-out overview pursuing a distinct logical pattern. The goal of this tool is to enable effective creation of business models.

Compared to the concepts reviewed in Figure 2, the business model building tool follows an explicit architecture to allow for an adjustment of customer demands during and after the process of developing the business model. Pivotal advantages are that the application of this approach is both dynamic and iterative.

Some existing software-supported tools for the creation of business models offer the visualization of the business model elements. However, interrelated effects among the parts are not yet disclosed and precise figures, e.g., of the business case, are not yet calculated.

The software-based version of the business model building tool bridges this gap by paying attention on how individual parts impact each other. Also, business-specific earning numbers are calculated as a result. The digital tool of the Business Model Builder represents a self-contained, cohesive system.

7. Research and Innovation Approach

Within the state-subsidized project, various business ideas were created with the aid of the Business Model Builder for the loading of electric vehicles with self-produced electricity, generated by photovoltaic plants.

The single elements of the Business Model Builder are shown in Figure 4. The subcomponents of the tool form components, which sum up to condensed modules. These modules build on one another. Derived from the current state of research, the order of the subcomponents represents a recommendation on how to develop the business model step by step.

7.1. Module 1: Business Idea

The first module works as an idea filter for business ideas identified as fruitful. The value proposition is investigated from the customer's point of view and reviewed with regard to an implementation in the market.

7.2. Module 2: Qualitative Description of the Business Model

This module comprises of the components product/service offerings as well as creation of offerings / value creating architecture, characterizing the business model under development or revision.

At the end of this process, a definite product/service offering is determined, including a rise in quantity and price on one side and a defined value creating architecture with respective quantities and specific costs on the other side

7.3. Module 3: Business Case

The core of this module is the measurement of how performant a business model will be. The success is quantified by key performance indicators, e.g., earnings before interest and taxes or economic value added. The basis, therefore, includes selling volume and price as well as production volume and costs from the second module. The business case – as a quantified business model – can also serve as the foundation for negotiations with investors, for instance.

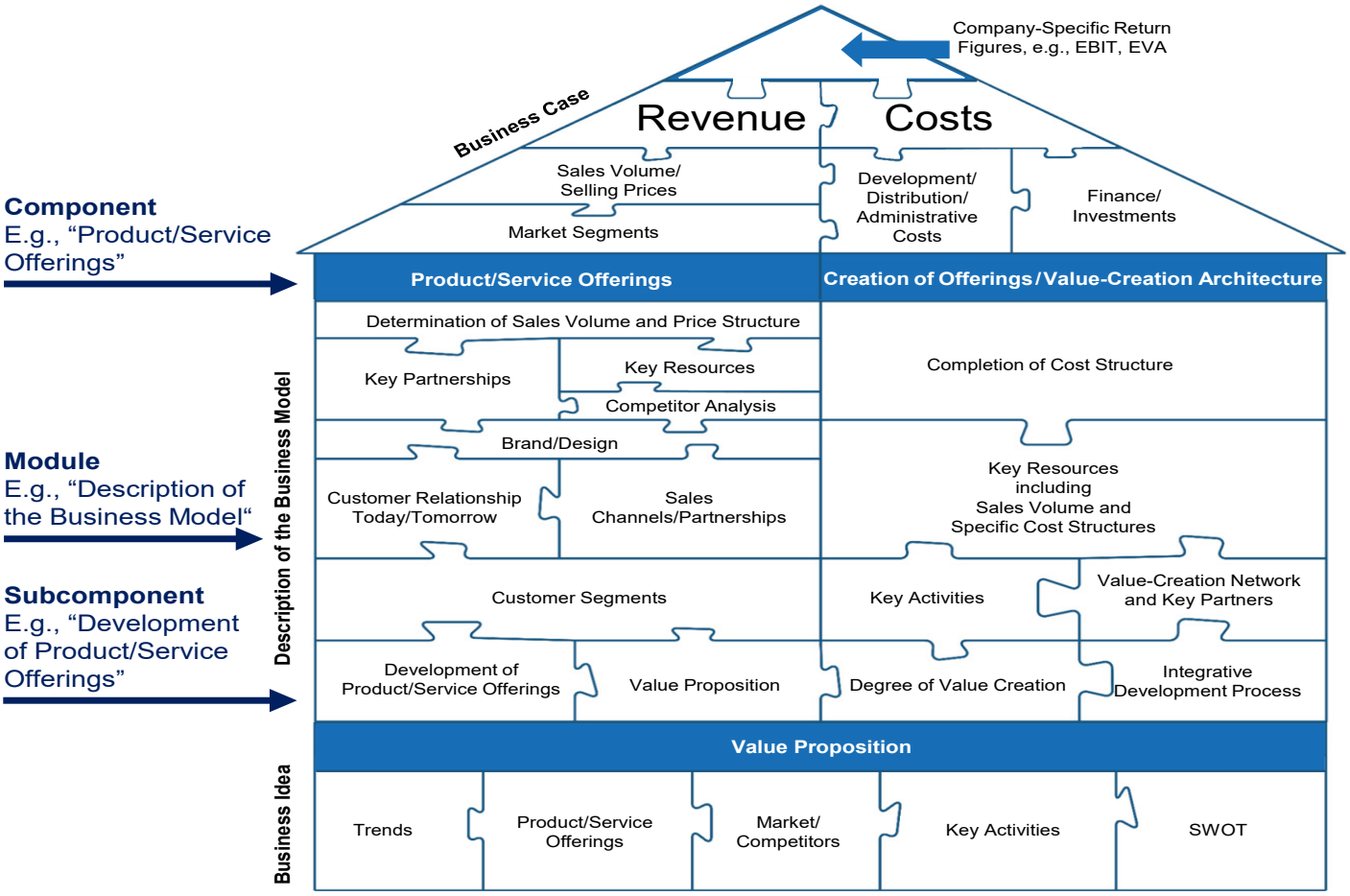


Figure 4: Business Model Builder Depicted in Fine Granularity

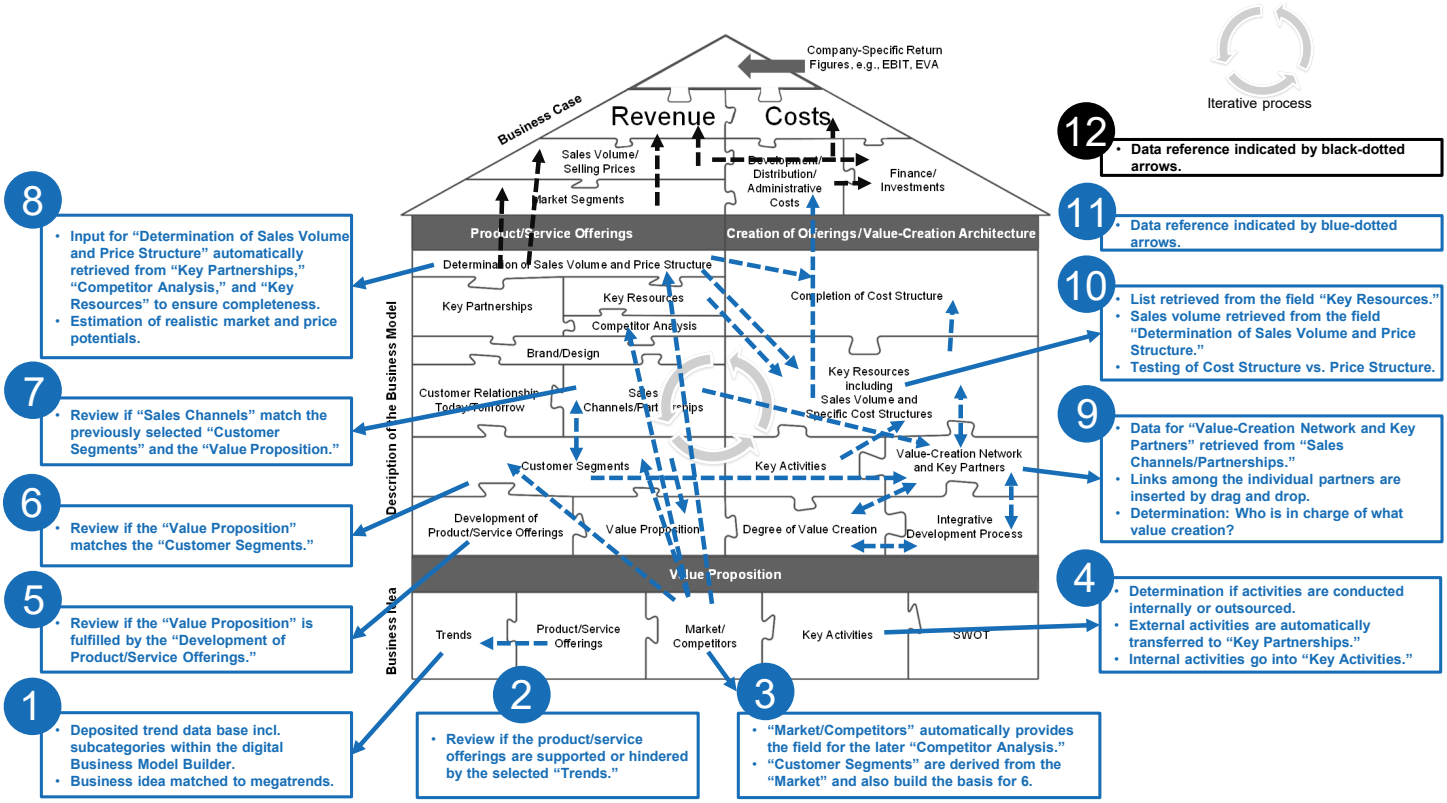


Figure 6: Interdependencies Within the Software-Supported Version of the Business Model Builder

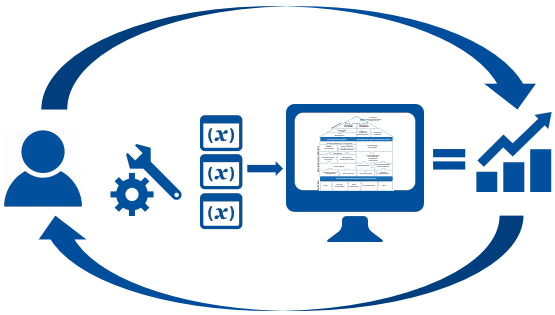
8. Conceptualization of the Tool and Its Digital Twin

The Business Model Builder is designed such that interrelated elements are linked and that effects from changes in one part on another part are revealed. Depending on the input variable, a change can also affect the business case.

The digitization of the Business Model Builder facilitates complex analyses. As multiple elements of the tool depend on other elements and are automatically updated through the links, impact on business results is easily uncovered and visualized.

Figure 5 depicts the schematic setup of the software-based tool. The user is responsible for the input of the elements. The elements automatically match as well as offset against each other. Hence, the operational results are automatically calculated.

Besides, the tool checks the user's input for plausibility and impact on other business model elements. In case of a mismatch, the user is given an error warning, and they are asked to reassess and correct the data entry.



9. Digitization Process of the Tool and Elaboration of Interrelations Among Elements

To digitize the business model building tool, it is pivotal to define which elements are freely definable, which import data from deposited databases, and which are the result of interrelated elements. The details of the interdependencies are shown in Figure 6. For the creation of this tool, all elements are studied for their impact on other elements involved. After the definition of these links, the attendant effects are programmed within the application software. This automizes the steps for the iterative business model development.

Providing an example, the first subcomponent in the tool is the trends. The trends are drawn from an external database. Thus, the user assigns their business idea to a megatrend at first. Subsequently, the idea is assigned to the macrotrends related to the megatrend. Later, the trends chosen are collated with the subcomponent "product/service offerings" to check the content of these subcomponents is consistent.

In addition to the yield numbers, the software-based version of the tool generates a digital twin of the business model. This offers convenient further editing of the business model. For instance, if the figures of the business case calculation lead to a "no go" decision, the subcomponents of the business model can be edited so that yield numbers finally result in a "go" decision. The digital Business Model Builder offers users to gradually adapt and improve the digital twin over the total lifetime.

10. Findings

Digitizing the Business Model Builder, an instrument is developed enabling users to create new, or innovate current business models in a convenient and structured manner with the desired granularity. In the software-based version of the business model building tool, the elements are automatically interconnected. Hence, business models can be modified, and results can be simulated. The tool can holistically depict the complete lifecycle of business models. Thus, various scenarios can be simulated with little expenditure of human labor and time.

11. Limitations

At this phase of the research project, the coding of the digital twin is taking place. As a next step, pilot testing will be implemented to test the application under real-world conditions, collect feedback, and potentially revise and adjust the tool. Only after the rollout will it be possible to see the impact of how the digitized Business Model Builder enhances effectiveness and efficiency of businesses.

12. Conclusion

Digitization pushes businesses aiming for sustainable commercial success to adapt business models to new market situations or create new business models when the former ones become obsolete through technological change. The Business Model Builder was developed to serve this demand. As digital version, the tool enables the permanent iterative adaptation of business models to transforming market circumstances.

The software-based version of the tool allows firms to set up their current business models in a digital twin. In this way, firms have the possibility to conveniently simulate amendments to business models and review their profitability, functionality, and feasibility. Furthermore, new business models can be created, validated, and tested. The digital tool offers established businesses as well as start-ups the option to iteratively adapt their business models to continuously transforming market situations and customer demands over a lifecycle. The tool's mission is to support firms in growing sustainably, in generating and keeping unique selling propositions to stand their ground in the market, and in achieving durable success in the world's markets.

References

[1] A. Nagl and K. Bozem, *Geschäftsmodelle 4.0 – Business Model Building mit Checklisten und Fallbeispielen.* Wiesbaden: Springer Gabler, 2018. ISBN 978-3-658-18841-2

[2] Hochschule Aalen für Technik und Wirtschaft, "BMBF Cooperative Research Project CO_2-arme Stadt / Low-Carbon City." Accessed: February 1, 2018. [Online]. Available: www.co2-arme-stadt.de

[3] S. Boschert and R. Rosen, "Digital twin – the simulation aspect," in *Mechatronic Futures,* P. Hehenberger and D. Bradley, Eds. Cham: Springer, 2016. DOI 10.1007/978-3-319-32156-1_5

[4] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems,* J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham: Springer, 2017. DOI 10.1007/978-3-319-38756-7_4

[5] A. Nagl, *Der Businessplan: Geschäftspläne erstellen – Mit Checklisten und Fallbeispielen.* Wiesbaden: Springer, 2015.

[6] R. Grünig and R. Kühn, *Methodik der strategischen Planung: Ein prozessorientierter Ansatz für Strategieplanungsprojekte.* Bern: Haupt, 2000.

[7] A. J. Slywotzky, *Value Migration: How to Think Several Moves Ahead of the Competition.* Boston: Harvard Business Press, 1995.

[8] A. Zolnowski and T. Böhmann, "Grundlagen service-orientierter Geschäftsmodelle," in Service-orientierte Geschäftsmodelle: Erfolgreich umsetzen T. Böhmann *et al.,* Eds. Berlin: Springer, 2013, pp. 1–29.

[9] P. Timmers, "Business models for electronic markets," *Electronic Markets,* vol. 8, no. 2, pp. 3–8, 1998.

[10] J. Linder and S. Cantrell, *Changing Business Models: Surveying the Landscape.* Chicago: Institute for Strategic Change, Accenture, 2000. Accessed: April 12, 2016. [Online]. Available: businessmodels.eu/images/banners/Articles/Linder_Cantrell.pdf

[11] G. Hamel, *Leading the Revolution.* Boston: Harvard Business School, 2000.

[12] B. W. Wirtz, *Business Model Management: Design – Instrumente – Erfolgsfaktoren von Geschäftsmodellen.* Wiesbaden: Springer Gabler, 2010.

[13] P. Stähler, *Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen.* Lohmar: Eul, 2002.

[14] P. Stähler, *Merkmale von Geschäftsmodellen in der digitalen Ökonomie.* Köln: Eul, 2001.

[15] D. Knyphausen-Aufseß and Y. Meinhardt, "Revisiting strategy: Ein ansatz zur systematisierung von geschäftsmodellen," in *Zukünftige Geschäftsmodelle,* T. Bieger *et al.,* Eds. Berlin Heidelberg: Springer, 2002, pp. 63–89.

[16] J. Rentmeister and S. Klein, "Geschäftsmodelle ein modebegriff auf der waagschale," *Zeitschrift für Betriebswirtschaft,* Ergänzungsheft 1, pp. 17-20, 2003.

[17] S. M. Shafer, H. J. Smith, and J. C. Linder, "The power of business models," *Business Horizons,* vol. 48, no. 3, pp. 199–207, 2005.

[18] A. Osterwalder and Y. Pigneur, "An e-business model ontology for modeling e-business," Electronic Commerce Conf., Bled, 2002.

[19] A. Osterwalder and Y. Pigneur, *Business Model Generation: Ein Handbuch für Visionäre, Spielveränderer und Herausforderer.* Frankfurt a. M.: Campus, 2011.

[20] O. Gassmann, K. Frankenberger, and M. Csik, *The Business Model Navigator: 55 Models That Will Revolutionise Your Business.* Harlow: Pearson. 2014.

[21] M. A. Rappa, "The utility business model and the future of computing services," *IBM Systems Journal,* vol. 43, no. 1, pp. 32–42, 2004.

[22] M. Dubosson-Torbay, A. Osterwalder, and Y. Pigneur, "E-business model design, classification, and measurements," *Thunderbird International Business Review,* vol. 44, no. 1, pp. 5–23, 2002.

[23] C. Zott and R. Amit, "Business model design: An activity system perspective," *Long Range Planning,* vol. 43, no. 2–3, pp. 216–226, 2010.

[24] D. J. Teece, "Strategies for managing knowledge assets: the role of firm structure and industrial context," *Long Range Planning,* vol. 33, no. 1, pp. 35–54, 2000.

[25] P. Belleflamme, T. Lambert, and A. Schwienbacher, "Crowdfunding: An industrial organization perspective," prepared for the workshop Digital Business Models: Understanding Strategies, Paris, June, 2010, pp. 25–26.

[26] D. C. Brabham, "Crowdsourcing as a model for problem solving: An introduction and cases," *Convergence,* vol. 14, no. 1, pp. 75–90, 2008.

[27] N. A. Morgan and L. L. Rego, "The value of different customer satisfaction and loyalty metrics in predicting business performance," *Marketing Science,* vol. 25, no. 5, pp. 426–439, 2006.

[28] J. Gray and B. Rumpe, "Models for digitalization," *Software & Systems Modeling,* vol. 14, no. 4, pp. 1319–1320, 2015. DOI 10.1007/s10270-015-0494-9

[29] J. Magretta, "Why business models matter," *Harvard Business Review,* vol. 80, no. 5, pp. 86–92, 2002.

[30] K. C. Laudon and C. G. Traver, *E-commerce: Business, Technology, Society.* Harlow: Pearson, 2016.

[31] E. Fleisch, M. Weinberger, and F. Wortmann, *Business Models and the Internet of Things: Interoperability and Open-Source Solutions for the Internet of Things.* Cham: Springer, pp. 6–10, 2015. DOI 10.1007/978-3-319-16546-2_2

[32] W. Gillis and G. J. Castrogiovanni, "The franchising business model: An entrepreneurial growth alternative," *International Entrepreneurship and Management Journal,* vol. 8, no. 1, pp. 75–98, 2012.

[33] D. J. Teece, "Business models, business strategy and innovation," *Long Range Planning,* vol. 43, no. 2–3, pp. 172–194, 2010.

[34] R. Amit and C. Zott, "Creating value through business model innovation," *MIT Sloan Management Review,* vol. 53, no. 3, pp. 41–49, 2012.

[35] C. Anderson, *The Long Tail: Why the Future of Business is Selling Less of More.* New York: Hachette Books, 2006.

[36] F. Salvador, P. M. De Holan, and F. Piller, "Cracking the code of mass customization," *MIT Sloan Management Review,* vol. 50, no. 3, pp. 71–78, 2009.

[37] M. W. Johnson, C. M. Christensen, and H. Kagermann, "Reinventing your business model," *Harvard Business Review,* vol. 86, no. 12, pp. 57–68, 2008.

[38] H. W. Chesbrough, "Why companies should have open business models," *MIT Sloan Management Review,* vol. 48, no. 2, p. 22, 2007.

[39] J. Y. Kim, M. Natter, and M. Spann, "Pay what you want: A new participative pricing mechanism," *Journal of Marketing,* vol. 73, no. 1, pp. 44–58, 2009.

[40] A. Osterwalder and Y. Pigneur, *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers.* New Jersey: Wiley, 2010.

[41] P. Banerjee *et al.,* "Everything as a service: Powering the new information economy," *Computer,* vol. 44, no. 3, pp. 36–43, 2011. DOI 10.1109/mc.2011.67

[42] O. Gassmann, K. Frankenberger, and M. Csik, *Geschäftsmodelle entwickeln: 55 innovative Konzepte mit dem St. Galler Business Model Navigator.* München: Hanser, 2013.

[43] D. Schallmo, *Geschäftsmodell-Innovation: Grundlagen, bestehende Ansätze, methodisches Vorgehen und B2B-Geschäftsmodelle.* Wiesbaden: Gabler, 2013.

[44] R. Alt and H.-D. Zimmermann, "Introduction to special section: Business models," *Electronic Markets,* vol. 11, no. 1, pp. 3–9, 2001.

[45] R. Amit and C. Zott, "Value creation in e-business," *Strategic Management Journal,* vol. 22., pp. 493–520, 2001.

[46] K. Schoegel, *Geschäftsmodelle: Konstrukt-Bezugsrahmen-Management.* München: FGM, 2001.

[47] T. Bieger and S. Reinhold, "Das wertbasierte Geschäftsmodell: Ein aktualisierter Strukturansatz," in *Innovative Geschäftsmodelle,* T. Bieger *et al.,* Eds. Berlin Heidelberg: Springer, 2011, pp. 11–70.

[48] C. Triebel and C. Schikora, "Scheitern bei Unternehmensgründungen," in *Failure Management,* S. Kunert, Ed. Berlin Heidelberg: Springer, 2016, pp. 235–248.

[49] J. Egeln, U. Falk, D. Heger, D. Höwer, and G. Metzger, *Ursachen für das Scheitern junger Unternehmen in den ersten fünf Jahren ihres Bestehens.* Mannheim Neuss: Zentrum für Europäische Wirtschaftsforschung, 2010.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 53-57 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Flame Distribution and Attenuation in Narrow Channels Using Mathematical Software

Bayan Sapargaliyeva[*,1], Aigul Naukenova[1], Bakhyt Alipova[2], Javier Rodrigo Ilarri[3]

[1]*Life Safety and Environment Protection Department, M. Auezov South-Kazakhstan State University, 160012, Kazakhstan*

[2]*Department of Mathematical and Computer Modeling, International IT University, 050040, Kazakhstan*

[3]*Hydraulic Engineering and Environment Department, Polytechnic University of Valencia, 46022, Spain*

ARTICLE INFO

Article history:
Received: 01 March 2019
Accepted: 03 May, 2019
Online: 13 May, 2019

Keywords:
Explosion suppression composition
Fire extinguishing powders
Burning reagents
Efficiency of powder
Heat and mass properties
Mathematical and computer modeling
Flame attenuation, narrow channels

ABSTRACT

The article of given research presents the mathematical and computer modeling of flame attenuation in narrow channels for different fire extinguishing and explosion suppression compositions. Such kind of flame distribution and attenuation could be considered for different media and channels but in this case it was considered for narrow channels. The effect on the combustion reaction is possible with gas mixture components concentration reducing, cooling the combustion zone and slowing down of chain reactions with the help of a phlegmatizing or inhibiting substances, of which the most universal and perspective are powder materials. Therefore, a necessary condition for solving the problems of developing effective flame arresters is to find common indicators and properties of substances that can become criteria for their phlegmatizing ability. The computer model in MatLab was created for methane allowing the choice of diameter and length of channel and velocity of explosion suppression composition, showing the dependence of temperature on length of channel.

1. Introduction

This paper is an extension of work originally presented in 2018 IEEE International Conference" Management of Municipal Waste as an Important Factor of Sustainable Urban Development" (WASTE). The influence of flame propagation as a result of heat losses from the reaction zone is the main factor of ensuring explosion safety for combustible gas systems. Thermal losses make it impossible to spread the flame beyond certain boundaries of composition and instrumental conditions. The mechanism of thermal losses plays a leading role in solving the emerging problems of explosion safety. The objective is to provide a general introduction to the Mechanics of heat transfer, together with its elementary applications to Solid and Fluid Mechanics. As result it would assimilated the principal concepts and laws governing the kinematics and dynamics of deformable media. In addition, this research should understand the application of this theory to the cases (i) of infinitesimal Thermo-Elasticity, and (ii) of Newtonian and perfect Fluid Mechanics [1-4].

[*]Bayan Sapargaliyeva, Kazakhstan, +77017326344, bonya_sh@mail.ru

2. Description of Problem

Mathematical and computer modeling provided a medium of exchange for the diverse disciplines (Mechanics and IT) utilizing mathematical or computer modeling as either a theoretical or working tool. Equal attention was given to the mechanics, methodology and theory of modeling with an attempt to advocate either mathematical or computer modeling, or a combination of the two, in an integrative form [5-7]. When burning in a narrow channel, the key to the mechanism of thermal losses from the reaction zone is the conductive heat conductivity to the walls of the channel. In this case, the reaction zone of obviously explosive gas systems is cooled. In sufficiently narrow channels, heat loss is possible, leading to quenching even of the most rapid-blowing known explosive systems. This feature combustion in narrow channels used in flame arrestors with narrow channels separating apparatus in which possible initiation foci combustion fire-retardant protected by the surrounding space, filled with a potentially explosive environment. The gas and flame have no other way out of the dangerous apparatus.

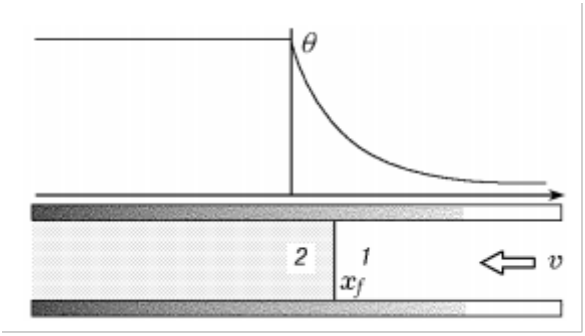


Figure 1: Flame attenuation in narrow channels

The Figure 1 shows a diagram of a flame in a flat channel. As in the classical formulation, approximations are used about the constant density, heat capacity, and thermal conductivity of the gas.

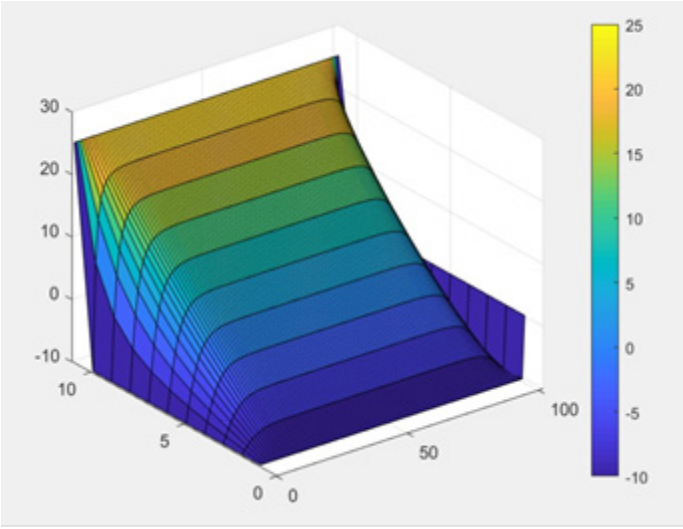


Figure. 2: Temperature distribution from wall to center of channel

This Figure 2 represents the temperature distribution from wall to center of channel. Identifying the relevant heat transfer processes on a heating when the foamed insulation sheet of channel is placed on inner surface and on the outer surface of channel. To avoid singular solution, the round configuration of channel is preferred.

3. Flame attenuation, mathematical modeling

Processes with heat regeneration include filtration combustion of gases, where the flame can stabilize in microchannels between the solid phases, even if the dimensions of the microchannels are substantially smaller than the critical diameter determined from the initial temperature. Combustion in this case is possible provided that the gas phase is surrounded by walls with a temperature exceeding the ambient temperature. This work can be considered as a continuation of the classical works on flame propagation taking into account the transfer of heat from the combustion products to the fresh mixture by an inert medium due to its high thermal conductivity.

Combustion of gas in a channel with heat exchange is the simplest example of a system with a convective heat recovery method. A characteristic feature of combustion in such a system is the thermal interaction between the combustion products and the fresh mixture through the heat-conducting wall. In this linear configuration, unlike devices with special combustion chambers, the flame stabilizes inside the inlet tube. This mechanism of flame stabilization is typical for small-sized systems with recirculation of heat. Studies of the dynamic behavior of a flame in a system with counter-current heat transfer make it possible to study the characteristics of combustion in systems with heat recovery, and also covers a whole series of new phenomena.

It is necessary to solve the Cauchy problem for the one-dimensional heat equation (a second-order partial differential equation that describes the temperature distribution in a given region of space and its variation over time) using parallel programming tools. The problem is solved for a homogeneous heat equation (the system is insulated) on the interval [0, L]: One-dimensional dimensionless equations for temperature have the form

$$\frac{\partial T_{1,2}}{\partial t} = \frac{\partial^2 T_{1,2}}{\partial x^2} + V\frac{\partial T_{1,2}}{\partial x} - \Omega(T_{1,2} - \theta(x))$$

Here the dimensionless time t is measured in units of correlation between thermal properties of gas in channel and adiabatic velocity of gas, i.e. æ/Vb², the dimensionless spatial coordinate x is in units of æ/Vb, where æ is the thermal diffusivity of the gas, and Vb is the adiabatic velocity of the flat flame. The channel wall temperature θ and the gas temperature T are measured in terms of the adiabatic flame temperature Tb [8-10].

The subscripts 1 and 2 refer to the fresh mixture and combustion products, respectively. The gas velocity V is measured in units of Vb. The dimensionless heat transfer parameter $\Omega = 4Nu/Pe^2$ is related to the Nusselt number $Nu=\alpha d_o/\lambda g$ and the Peclet number $Pe=Vbd_o/æ$, which are determined for the channel of height d_o (here α is the heat transfer coefficient; λg is the thermal conductivity of the gas). The boundary conditions for equations (1) far from the flame have the following form:

$$x \to +\infty:\ T_1 \to \sigma,\ C_1 \to 1,\ \theta \to \sigma$$
$$x \to -\infty:\ T_2 \to \Theta,\ \theta \to \Theta > \sigma$$
$$\sigma = \frac{T_0}{T_b},\quad \Theta = \frac{T_w}{T_b}$$

where T0 is the initial temperature of the fresh mixture, Tw is the wall temperature at the hot end of the channel.

Thus, Equation (1) together with boundary conditions (2) at the ends of a narrow channel determines the spread of the flame in the channel. In this case, the temperature distribution in the channel is calculated (the flame temperature of the combustible gas and its mixture before and after mixing, the temperature of the channel walls and their combination are taken into account) as a function of time. Also, the differential equation of the second order includes the velocities and accelerations of the temperature change. The gas characteristics were calculated from some physical parameters: thermal diffusivity of the gas, dimensionless heat transfer parameter (as correlation of numbers of Nusselt and Peclet), heat transfer coefficient; the thermal conductivity of the gas. Thus, taking into account the boundary conditions (2), the boundary value problem was modeled and coded in Matlab.

4. Numerical calculations in MatLab

The numerical realization of heat transfer process was implemented in MatLab, and Adomian Decomposition Method was used for calculation of components of displacement of parts

of fire extinguishing powders and burning reagents. This calculation needs to use some special functions for discretization of components of temperature.

```
% variable definition
Nx = 11;
Ny = 101;
X = 1.0;
Y = 10.0;
dx = X / (Nx - 1);
dy = Y / (Ny - 1);
dt = 1e-4;

% matrix definition
% initializing temperature in our  fire extinguishing and explosion
suppression compositions
U1 = ones(Nx, Ny) * (-10);
U2 = ones(Nx, Ny) * (-10);
U3 = ones(Nx, Ny) * (-10);

% initializing temperature on right border in our fire extinguishing and
explosion suppression compositions
for j=1:Ny + 1
  U1(Nx, j) = 25;
  U2(Nx, j) = 25;
  U3(Nx, j) = 25;
end

% loop till it converges
while 1
%    Using Adomian Decomposition Method to solve this eq
   A = 1 / (dx * dx);
   B = - (1 / dt + 2 / (dx * dx));
   C = 1 / (dx * dx);
   gamma = zeros(Nx);
   beta = ones(Nx) * (-10);

%    First loop using the definition temperature by X axis
   for j=2:Ny - 1
      for i=2:Nx
         F = -(discretize(U1, "y", i, j) + U1(i, j) / dt);
         temp_div = B + C * gamma(i);
         beta(i + 1) = (F - C * beta(i)) / temp_div;
         gamma(i + 1) = -A / temp_div;
      end
%       U2(Nx, j) = - beta(j) / (1 - gamma(j));
      for i=Nx:-1:2
         U2(i - 1, j) = gamma(i) * U2(i, j) + beta(i);
      end
   end

   gamma = zeros(Nx);
   beta = ones(Nx) * (-10);

   A = 1 / (dy * dy);
   B = - (1 / dt + 2 / (dy * dy));
   C = 1 / (dy * dy);
%    Second loop using of temperature by Y axis
   for i=2:Nx - 1
      for j=2:Ny
         F = -(discretize(U2, "x", i, j) + U2(i, j) / dt);
         temp_div = B + C * gamma(j);
         beta(j + 1) = (F - C * beta(j)) / temp_div;
         gamma(j + 1) = -A / temp_div;
      end
      for j=Ny:-1:2
         U3(i, j - 1) = gamma(j) * U3(i, j) + beta(j);
      end
   end
%    for j=Ny:-1:2
%       U3(Nx - 1, j) = U3(Nx - 2, j);
%    end
% And after every iteration check for convergence
   if converge(U1, U3)
      break;
   end
% In the end of every iteration - definition of new layer as old, and
continuation till difference between them become less than epsilon
   U1 = U3;
   surf(U3)
end
% Plot surface to image easier the result
surf(U3)
colorbar
U3((Nx - 1) / 2, (Ny - 1) / 2)

% This function checks for convergence
function result = converge (U1, U2)
% Initial max difference is -1, because of it less than 0, and as we know
% absolute value of difference never be less than 0
max_diff = -1;
% My error rate (Epsilon)
eps = 1e-4;
Nx = 11;
Ny = 101;
% This loop gets max difference Between old and new layers
for i=1:Nx
   for j=1:Ny + 1
      if abs(U1(i, j) - U2(i, j)) > max_diff
         max_diff = abs(U1(i, j) - U2(i, j));
      end
   end
end
% As the result we get Boolean value
result = eps > max_diff;
end

% This function helps me make code easier to understand
function result = discretize (U, axis, i, j)
Nx = 11;
Ny = 101;
X = 1.0;
Y = 10.0;
dx = X / (Nx - 1);
dy = Y / (Ny - 1);
if axis == "x"
   result = (U(i + 1, j) - 2 * U(i, j) + U(i - 1, j)) / (dx * dx);
else
   result = (U(i, j + 1) - 2 * U(i, j) + U(i, j - 1)) / (dy * dy);
end
end
```

The basic mechanisms of combustion of gases in systems with heat recirculation can be clarified from the study of combustion of gases in a narrow channel with a controlled wall temperature. Therefore,

studies of gas combustion in micro-channels with controlled wall temperature are necessary for understanding processes in both microsystems with heat regeneration, where the heating of the combustible mixture of gases occurs due to the heat of the exhaust combustion products, and for the development of new technologies that use combustion of gases in various media. The small dimensions of the channel in which combustion occurs allow the creation of an almost uniform temperature distribution in the cross section. Experiments show that the temperature gradient in the walls of the channel, which was created by an external heat source, remains practically constant, both with the combustion of gas inside the channel and without burning. These observations make it promising to use a microchannel with a temperature gradient to study the structure of the flame and to test existing models with detailed kinetics of chemical reactions.

Burning studies usually consider gas flows in which transport phenomena and chemical reactions play an important role. In order to analyze the combustion process, in addition to the chemical transformations of the substance, it is necessary to take into account the phenomena caused by the release of heat. If we take into account the interconnectedness of the above processes, then the complexity of the theoretical analysis of combustion becomes obvious.

Advances in the theory of combustion of gases were associated, first of all, with the use of simple and at the same time realistic mathematical models.

Software Package of given formulas are special calculations that can contain class folders, function, and class definition files, and other packages. The names of classes and functions in Matlab are scoped to the package folder. A package with name" TED in channels" (Thermoelastodynamics in channels) is unique. Function and class names must be unique only within the package. Using a package provides a means to organize classes and functions. Packages also enable you to reuse the names of classes and functions in different packages.

The proposed code of the software product allows you to transparently change the approaches to calculation, change the parameters to be set. Explicit schemes for solving the differential equation are proposed. Also in Matlab, visualization of the computational process was carried out.

The Graphical User Interface GNUI requires the following parameters:

- Diameter of channel (in m),
- Velocity of motion gas or fluid (in m/sec),
- Length of channel (in m).

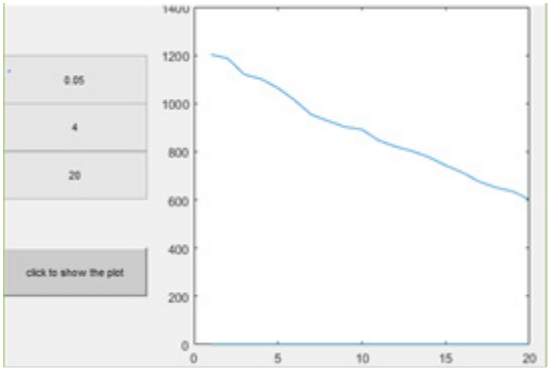


Figure 3: Visualization of flame attenuation in narrow channels in TED in channels

Figure 3 represents the Interface of package and input/output parameters. The graph of the dependence of the change in the combustion temperature of methane on the length of a narrow channel clearly indicates a decrease in temperature during motion away from the flame front, which confirms the physical nature of the problem and the correctness of the chosen mathematical model.

Thus, a mathematical model of flame burning in a narrow channel has been developed, the properties of burning and quenching of gas or liquid in a narrow channel have been studied, depending on the combustion temperature and the flame propagation velocity in the channel. In this case, the inhibitory properties of the substance, flame propagation in thin channels are investigated [11-13]. Based on the 2-D model, a computer model was created in MatLab14. This model requires the specification of some input parameters and provides a graphical representation of the temperature dependence of the length of the channel. The computer model allows you to transform and supplement the subsequent calculations.

As a result of computer modeling of combustion processes in a system with convective heat recirculation, confirmed by some experiments, it was shown that

- in this system

a) it is possible to sustain combustion in a channel with a diameter less than critical,

b) combustion with velocities greater than the normal velocity of the free flame and gas temperature may exceed the adiabatic temperature of the free flame,

c) the flame is resistant to external influences caused by a change in the ambient temperature, changes in the gas flow rate and other factors.

The possibility of flame stabilization in a straight channel with a temperature gradient in the walls, which is created by the combustion products, is shown. These results allow us to conclude that this system is promising for creating, for example, micro-burners.

- A new type of instability has been discovered in the combustion of gas in micro-channels with a longitudinal temperature gradient in the walls of the channel. A feature of this type of flame instability is the presence of an inhomogeneous velocity profile of the gas in the channel and heat exchange with hot walls. The range of values of the gas flow rate at which this type of instability manifests itself is found. It is shown that in a radial channel this type of instability leads to the formation of rotating flame structures, in particular, a spiral flame detected in experiments.

5. Conclusion

The mathematical model of flame attenuation in narrow channels for different fire extinguishing and explosion suppression compositions is given. Such kind of flame distribution and attenuation could be considered for different media and channels but in given research it was considered for narrow channels. The effect on the combustion reaction is possible with gas mixture components concentration reducing, cooling the combustion zone and slowing down of chain reactions with the help of a phlegmatizing or inhibiting substances, of which the most universal and perspective are powder materials. As in the classical formulation, approximations are used about the constant density, heat capacity, and thermal conductivity of the gas. Caushy task

for a second-order partial differential equation that describes the temperature distribution in a given region of space and its variation over time was solved. The numerical realization of heat transfer process was implemented in MatLab, and Adomian Decomposition Method was used for calculation of components of displacement of parts of fire extinguishing powders and burning reagents. The interface with input/output parameters was launched and tested. A package with name" TED in channels" (Thermoelastodynamics in channels) enables to reuse the names of classes and functions in different packages.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Authors are grateful for scientific and methodological support to M. Auezov South-Kazakhstan State University, Kazakhstan.

References

[1] Alipova B.N., Alexeyeva L.A., Dadaeva A.N. Shock waves as generalized solutions of thermoelastodynamics equations. On the uniqueness of boundary value problems solutions // AIP Conference proceedings, 2017, V. 1798, I. 020003, pp. 1-8 DOI: 10.1063/1.4972595

[2] Fennandez-Pello A. C. Micro-power generation using combustion: issues and approaches. Twenty-ninth intern. symp. on combustion, Pittsburg, 2002, V. 40, no. 5, pp. 883– 899.

[3] Sitzki, L., Borer, K., Wussow, S. Schuster, E., Maruta, K., Ronney, P. and Cohen, A., "Combustion in Microscale Heat-Recirculating Burners", 38th AIAA Space Sciences & Exhibit, Reno, NV, 2001, V. 22, no. 3, p. 1087.

[4] Vican, J. Gajdeczko, B.F., Dryer, F.L., Milius, D.L., Aksay, I.A. and Yetter, R.A."Development of a Microreactor as a Thermal Source for MEMS Power Generation" Proceedings of the Twenty-Ninth International Symposium on Combustion, Sapporo, Japan, 2002, V. 40, no. 5, pp. 906-916.

[5] Lloyd S. A. and Weinberg F. J. A burner for mixtures of very low heat content NATURE, 1974, V. 251, pp. 47–49.

[6] Jones A. R., Lloyed S. A., and Weinberg F. J. Combustion in heat exchangers // Proc. Roy. Soc. Lond., 1978, V. A369, pp. 97–115.

[7] Buckmaster J. D. The effects of radiation on streched flames // Combust. Theory Modelling, 1997, no. 1 pp. 1–11.

[8] Zamaschikov V.V., Minaev S.S. Limits of flame propagation in a narrow channel during gas filtration // Physics of Combustion and Explosion, 2001, V. 37, no. 1, pp. 25–31.

[9] Zeldovich Ya. B. The theory of propagation of a quiet flame // Journal of Energy and Solids, 1941, V. 11, I. 1, pp. 159–168.

[10] Minaev S.S., Babkin V.S. Propagation of a flame in a channel of variable cross section during gas filtration // Physics of Combustion and Explosion, 2001, V. 37, no. 1, pp. 16–24.

[11] Sapargalieva B., Naukenova A., Alipova B., Javier Rodrigo Illari, Shapalov Sh. The analysis of heat and mass properties of the fire extinguishing powder in effectiveness criteria // News of the National Academy of Science of the Republic of Kazakhstan. Series of Geology and Technical sciences, 2018,. ISSN 2518-170X (Online), ISSN 2224-5278 (Print), V. 3, no. 4 (430), pp. 51-61.

[12] M. Raoofian Naeeni, M. Eskandari-Ghadi, A. A. Ardalan, S. Sture, and M. Rahimian, "Transient response of a thermoelastic half-space to mechanical and thermal buried sources," ZAMM **95** (4), 354–376 (2015).

[13] H. Mahmoodi Kordkhieli, G. Ghodrati Amiri, and M. Hosseini, "Axisymmetric analysis of a thermoelastic isotropic half-space under buried sources in displacement and temperature potentials," J. Thermal Stresses **40** (2), 237–254 (2017).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 58-62 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

An Expanded Derivation of Suppressing Anisotropic Birefringence in a Rectangular Waveguide

Gregory Mitchell*

U.S. Army Research Laboratory, 20783, USA

ARTICLE INFO

Article history:
Received: 06 February 2019
Accepted: 03 May, 2019
Online: 13 May, 2019

Keywords:
Anisotropic media
Birefringence
Anisotropic dispersion
Rectangular waveguide
Biaxial anisotropy

ABSTRACT

Birefringence is a known property of anisotropic media where incidence of a single propagating wave on a biaxial boundary gives rise to what scientists refer to as an ordinary wave and extraordinary wave. This article explores the anisotropic electromagnetic theory that gives rise to the phenomenon of birefringence focusing on both ordinary and extraordinary plane wave propagation. The dispersion equation in an unbounded anisotropic medium leads to a fourth order polynomial solution for the wave number which is shown to be the root cause of birefringence. Finally, a derivation of the unique properties for a partially filled biaxial rectangular waveguide that serves to suppress wave propagation in the vertical direction will negate the effect of birefringence in the case of the first propagating mode.

1. Introduction

At the invitation of the associate editors of the Advances in Science, Technology and Engineering Systems Journal (ASTESJ), the author presents the following manuscript on an in-depth theoretical expansion of a conference paper previously submitted in the Applied Computational Electromagnetics Society (ACES) Symposium proceedings in 2018 [1]. Originally, the author only discussed the explanation of the suppression effect of anisotropic birefringence inside a rectangular waveguide in brief, and the author feels that scientific interest in the subject warrants a more in depth theoretical examination of the phenomenon. The author has also presented other portions of the theory relating to anisotropic wave propagation in a second publication [2]. Therefore, the author is happy to contribute an expansion on this aspect of the anisotropic birefringence theory to a special issue of ASTESJ at the invite of the associate editors. The article will focus on how the derivation of the anisotropic wave equation in a biaxial anisotropic medium leads to birefringence through a fourth order dispersion equation. The final section will explain how constraints on the boundary conditions of modes in rectangular waveguides can negate the effect of birefringence.

Scientists have understood electromagnetic wave propagation in homogeneous anisotropic optical media since the 1950s. At optical frequencies, one has to rely on naturally occurring crystalline media with anisotropic properties. However, as early as 1958 Collin showed that at microwave frequencies, where the wavelength is much greater, it is possible to fabricate artificial dielectric media having anisotropic properties [3].

Recently engineered materials have come to play a dominant role in the design and implementation of electromagnetic devices and especially antennas. Metamaterials, MetaFerrites, and magneto-dielectrics have all come to play a crucial role in advances made in both the functionality and characterization of such devices. In fact, a movement towards utilizing customized material properties to replace the functionality of traditional radio frequency (RF) components such as broadband matching circuitry, ground planes, and directive elements is apparent in the literature and not just replacement of traditional substrates and superstrates with engineered structures. A firm theoretical understanding of the electromagnetic properties of these materials is necessary for both design and simulation of new and improved RF devices.

Inherently, some of these engineered materials have anisotropic properties. Previously, the study of anisotropy had been limited mostly to the realm of optical frequencies where the phenomenon occurs naturally in substances such as liquid crystals and plasmas. However, the recent development of a separate class engineered materials at microwave frequencies has encouraged the study of electromagnetic anisotropy for RF applications. Low loss anisotropic magneto-dielectrics greatly expands the current antenna design space. Anisotropic media provide unique properties not available using isotropic media. These include a high effective refractive index reducing the electrical size of wavelengths in a medium, relatively low magnetic loss at MHz

*Gregory Mitchell, U.S. Army Research Laboratory,
Email: gregory.a.mitchell1.civ@mail.mil

frequencies, and lower density and weight than traditional isotropic ferrite materials. Anisotropic media can be up to five times less dense than traditional ferrites while exhibiting magnetic properties with lower loss tangents. Relevant work in this area includes plane wave solutions to propagation in anisotropic media [4-6]. Researchers have also performed modal decomposition of rectangular waveguides filled with anisotropic magneto-dielectrics along with in depth analysis of cut-off wave numbers and field distributions [7, 8].

An anisotropic electromagnetic medium defines permittivity ($\underline{\underline{\varepsilon_r}}$) and permeability ($\underline{\underline{\mu_r}}$) as separate tensors where the values differ in all three Cartesian directions ($\varepsilon_x \neq \varepsilon_y \neq \varepsilon_z$ and $\mu_x \neq \mu_y \neq \mu_z$). This is known as the biaxial definition of anisotropic material which is more encompassing than the uniaxial definition which makes the simplifying assumption that $\varepsilon_x=\varepsilon_y=\varepsilon_t$ and $\mu_x=\mu_y=\mu_t$. The anisotropic definition also differs from the traditional isotropic definition where ε_r and μ_r are the same in all three Cartesian directions defined by a single value. Anisotropic media yield characteristics such as conformal surfaces, the focusing and refraction of electromagnetic waves as they propagate through a medium, high impedance surfaces for artificial magnetic conductors as well as high index, low loss, and lightweight ferrite materials. The following sections aim to discuss in more detail some RF applications directly affected by the incorporation of anisotropic media and will also present a rigorous derivation of the wave equation and dispersion relationships for anisotropic magneto-dielectric media. All results agree with those presented by Meng, et. al [7, 8]. Furthermore, setting $\underline{\underline{\mu_r}} = \underline{\underline{I}}$, where $\underline{\underline{I}}$ is the identity matrix, yields results that agree with those presented by Pozar and Graham for anisotropic dielectric media [6, 9].

2. Applications of Anisotropy in Radio Frequency (RF) Devices

Traditionally, the study of anisotropic properties for RF was limited to a narrow application space where traditional ferrites exhibiting natural anisotropy were the enabling technology. These types of applications included isolators, absorbers, circulators and phase shifters [10, 11]. Traditional ferrites are generally very heavy and very lossy at microwave frequencies and those are the two main limiting factors narrowing their use in RF devices; however, propagation loss is an important asset to devices such as absorbers. Anisotropy itself leads to propagation of an electromagnetic wave in different directions. This phenomenon is already important in devices such as circulators and isolators [10]. For phase shifters and other control devices, changing the bias field across the ferrite controls the microwave signal [11]. However, newer versions of some of these devices, utilizing FETs and diodes in the case of phase shifters, rely on isotropic media to enable higher efficiency devices.

As early as 1958, Collin showed that at microwave frequencies, where the wavelength is larger, it is possible to fabricate artificial dielectric media having anisotropic properties [3]. This has led some to investigate known theoretical solutions to typical RF problems, such as a microstrip patch antenna, and extend them utilizing anisotropic wave propagation in dielectric media [5, 6]. The anisotropic dielectric antenna shows interesting features of basic antenna applications featuring anisotropic substrates. While these solutions establish a framework for electromagnetic wave propagation in anisotropic media, they simplify the problem by necessarily setting $\mu_r=1$ and only focusing on dielectric phenomena of anisotropy.

The proliferation of metamaterials research over the last few decades exemplifies the growing interest in the area of artificial media [12-22]. Metamaterials incorporate the use of artificial microstructures made of subwavelength inclusions implemented with periodic and/or multilayered structures known as unit cells [13]. These devices operate where the wavelength is much larger than the characteristic dimensions of the unit cell elements. One characteristic feature of some types of metamaterials is wave propagation anisotropy [14]. Applications such as directive lensing [15], cloaking [16], electronic beam steering [17], metasurfaces [18], anisotropic oscillators [19], and optics [20] among others all utilize anisotropic media as the enabling technology. More recent research focuses on the proper modeling schemes for computational electromagnetics [21] or examining phenomenon such as scattering off of anisotropic objects [22].

A class of engineered materials exists, known as MetaFerrites, that exhibits positive refractive index, anisotropy, and magneto-dielectric properties with reduced propagation loss at microwave frequencies compared to traditional ferrites. These materials show the unique ability to provide broadband impedance matches for very low profile antennas by exploiting the inherent anisotropy to redirect surface waves thus improving the impedance match of the antenna when very close to a ground plane. Designers have demonstrated antenna profiles at 200 MHz to 500 MHz on the orders of a fortieth of a wavelength using these anisotropic media with over an octave of bandwidth and positive realized gain [2, 23].

3. Anisotropic Wave Equation

The recent development of low loss anisotropic magneto-dielectrics greatly expands the current antenna design space. Here we present a rigorous derivation of the wave equation and dispersion relationships for anisotropic magneto-dielectric media. All results agree with those presented by Meng, et. al [7, 8]. Furthermore, setting $\underline{\underline{\mu_r}} = \underline{\underline{I}}$, where $\underline{\underline{I}}$ is the identity matrix, yields results that agree with those presented by Pozar and Graham for anisotropic dielectric media [6, 9]. Incorporating a fully developed derivation of anisotropic properties of both $\underline{\underline{\varepsilon_r}}$ and $\underline{\underline{\mu_r}}$ expands upon the simplification imposed by both Pozar and Graham that uses an isotropic value of $\mu_r=1$. An expansion on the results of Meng, et. al given in section 4 develops the waveguide theory including a full modal decomposition utilizing the biaxial definition of anisotropy versus their simplified uniaxial definition. The derivation of anisotropic rectangular waveguide resonance in section 6 differs from that of Meng, et. al by addressing the direct relationship between boundary conditions and the existence of birefringence. Furthermore, the analysis of anisotropic properties is not restricted to double negative materials, which is the case for the Meng, et. al studies.

In order to solve for the propagation constants we will need to formulate the dispersion relationship from the anisotropic wave equation. This allows us to solve for the propagation constant in the normal direction of the anisotropic medium. We start with the anisotropic, time harmonic form of Maxwell's source free equations for the electric and magnetic fields

$$\nabla x\underline{\boldsymbol{E}} = -j\omega\mu_o\underline{\underline{\boldsymbol{\mu}}}_r\cdot\underline{\boldsymbol{H}}, \tag{1}$$

$$\nabla x\underline{\boldsymbol{H}} = j\omega\varepsilon_o\underline{\underline{\boldsymbol{\varepsilon}}}_r\cdot\underline{\boldsymbol{E}}, \tag{2}$$

where ω is the frequency in radians, ε_o is the permittivity of free space, μ_o is the permeability of free space, $\underline{\boldsymbol{E}}=\underline{\boldsymbol{x}}_oE_x+\underline{\boldsymbol{y}}_oE_y+\underline{\boldsymbol{z}}_oE_z$, and $\underline{\boldsymbol{H}}=\underline{\boldsymbol{x}}_oH_x+\underline{\boldsymbol{y}}_oH_y+\underline{\boldsymbol{z}}_oH_z$. We define $\underline{\underline{\boldsymbol{\mu}}}_r$ and $\underline{\underline{\boldsymbol{\varepsilon}}}_r$ as

$$\underline{\underline{\varepsilon}}_r = \begin{bmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{bmatrix}, \tag{3}$$

$$\underline{\underline{\mu}}_r = \begin{bmatrix} \mu_x & 0 & 0 \\ 0 & \mu_y & 0 \\ 0 & 0 & \mu_z \end{bmatrix}. \tag{4}$$

Combining (1-4) yields

$$\begin{aligned} &\underline{x}_o\left(\frac{dE_Z}{dy}-\frac{dE_Y}{dz}\right)+\underline{y}_o\left(\frac{dE_X}{dz}-\frac{dE_Z}{dx}\right)+\underline{z}_o\left(\frac{dE_Y}{dx}-\frac{dE_X}{dy}\right), \\ &\quad = -j\omega\mu_o\left(\mu_xH_X\underline{x}_o+\mu_yH_Y\underline{y}_o+\mu_zH_Z\underline{z}_o\right) \end{aligned} \tag{5}$$

$$\begin{aligned} &\underline{x}_o\left(\frac{dH_Z}{dy}-\frac{dH_Y}{dz}\right)+\underline{y}_o\left(\frac{dH_X}{dz}-\frac{dH_Z}{dx}\right)+\underline{z}_o\left(\frac{dH_Y}{dx}-\frac{dH_X}{dy}\right). \\ &\quad = j\omega\varepsilon_o\left(\varepsilon_xE_X\underline{x}_o+\varepsilon_yE_Y\underline{y}_o+\varepsilon_zE_Z\underline{z}_o\right) \end{aligned} \tag{6}$$

Using the radiation condition, we assume a solution of $\underline{\boldsymbol{E}}(\underline{\boldsymbol{r}})=\underline{\boldsymbol{E}}(x,y)\exp(-jk_zz)$. Now isolate the individual components of (5) by taking the dot product with $\underline{\boldsymbol{x}}_o$, $\underline{\boldsymbol{y}}_o$, and $\underline{\boldsymbol{z}}_o$ respectively. This operation yields the following equations

$$dE_z/dy - jk_zE_y = -j\omega\mu_o\mu_xH_x, \tag{7}$$

$$jk_zE_x - dE_z/dx = -j\omega\mu_o\mu_yH_y, \tag{8}$$

$$dE_y/dx - dE_x/dy = -j\omega\mu_o\mu_zH_z. \tag{9}$$

The same procedure assuming a solution of $\underline{\boldsymbol{H}}(\underline{\boldsymbol{r}})=\underline{\boldsymbol{H}}(x,y)\exp(-jk_zz)$ for (6) yields

$$dH_z/dy - jk_zH_y = j\omega\varepsilon_o\varepsilon_xE_x, \tag{10}$$

$$jk_zH_x - dH_z/dx = j\omega\varepsilon_o\varepsilon_yE_y, \tag{11}$$

$$dH_y/dx - dH_x/dy = j\omega\varepsilon_o\varepsilon_zE_z. \tag{12}$$

Equations (7-12) allow us to solve for the transverse field components of the electric and magnetic fields in terms of the derivatives of H_z and E_z as

$$E_x = -\frac{j}{k_o^2\mu_y\varepsilon_x-k_z^2}\left(\omega\mu_o\mu_y\frac{dH_z}{dy}+k_z\frac{dE_z}{dx}\right), \tag{13}$$

$$E_y = \frac{j}{k_o^2\mu_x\varepsilon_y-k_z^2}\left(\omega\mu_o\mu_x\frac{dH_z}{dx}-k_z\frac{dE_z}{dy}\right), \tag{14}$$

$$H_x = \frac{j}{k_o^2\mu_x\varepsilon_y-k_z^2}\left(\omega\varepsilon_o\varepsilon_y\frac{dE_z}{dy}-k_z\frac{dH_z}{dx}\right), \tag{15}$$

$$H_y = -\frac{j}{k_o^2\mu_y\varepsilon_x-k_z^2}\left(\omega\varepsilon_o\varepsilon_x\frac{dE_z}{dx}+k_z\frac{dH_z}{dy}\right). \tag{16}$$

Now that we have relationships for the transverse field components, we can solve (1) and (2) for $\underline{\boldsymbol{H}}$ and $\underline{\boldsymbol{E}}$ respectively

$$\underline{H} = -\underline{\mu}_r^{-1}\cdot\left(\nabla x\underline{E}\right)/\left(j\omega\mu_o\right), \tag{17}$$

$$\underline{E} = \underline{\varepsilon}_r^{-1}\cdot\left(\nabla x\underline{H}\right)/\left(j\omega\varepsilon_o\right). \tag{18}$$

Taking the cross product of both sides and substituting (1) for $\nabla x\underline{\boldsymbol{E}}$ and (2) for $\nabla x\underline{\boldsymbol{H}}$ yields [8]

$$\nabla\times\underline{E} = \nabla\times\underline{\varepsilon}_r^{-1}\cdot\left(\nabla x\underline{H}\right)/\left(j\omega\varepsilon_o\right), \tag{19}$$

$$\nabla\times\underline{H} = -\nabla\times\underline{\mu}_r^{-1}\cdot\left(\nabla x\underline{E}\right)/\left(j\omega\mu_o\right), \tag{20}$$

$$\nabla x\underline{\mu}_r^{-1}\cdot\left(\nabla x\underline{E}\right) = -k_o^2\underline{\varepsilon}_r\cdot\underline{E}, \tag{21}$$

$$\nabla x\underline{\mu}_r^{-1}\cdot\left(\nabla x\underline{E}\right) = -k_o^2\underline{\varepsilon}_r\cdot\underline{E}. \tag{22}$$

Equations (21) and (22) represent the vector wave equations in an anisotropic medium.

4. Dispersion Equation for H_z

We expand the $\nabla x\underline{\boldsymbol{H}}$ term of (22) in terms of (13-16), and take the dot product with $\underline{\underline{\boldsymbol{\varepsilon}}}_r^{-1}$

$$\begin{aligned} \nabla x\Bigg[&\frac{\underline{x}_o}{\varepsilon_x}\left(\frac{dH_Z}{dy}-\frac{dH_Y}{dz}\right)+\frac{\underline{y}_o}{\varepsilon_y}\left(\frac{dH_X}{dz}-\frac{dH_Z}{dx}\right) \\ &+\frac{\underline{z}_o}{\varepsilon_z}\left(\frac{dH_Y}{dx}-\frac{dH_X}{dy}\right)\Bigg] = k_o^2\underline{\mu}_r\cdot\underline{H} \end{aligned}, \tag{23}$$

Evaluating (23) and taking the dot product with the $\underline{\boldsymbol{z}}_o$ direction allows us to isolate the H_z component of the magnetic field on the right hand side of the equation

$$-\frac{1}{\varepsilon_y}\frac{d^2H_z}{dx^2}-\frac{1}{\varepsilon_x}\frac{d^2H_z}{dy^2}+\frac{1}{\varepsilon_y}\frac{d^2H_x}{dxdz}+\frac{1}{\varepsilon_x}\frac{d^2H_y}{dydz} = k_o^2\mu_zH_z, \tag{24}$$

Setting E_z=0, if we differentiate (13) by $d^2/dxdz$ and (14) by $d^2/dydz$, keeping in mind that $d/dz = -jk_z$, we arrive at the following result

$$\begin{aligned} &\frac{k_z^2}{\varepsilon_y\left(k_z^2-k_o^2\varepsilon_y\mu_x\right)}\frac{d^2H_z}{dx^2}+\frac{k_z^2}{\varepsilon_x\left(k_z^2-k_o^2\varepsilon_x\mu_y\right)}\frac{d^2H_z}{dy^2} \\ &\quad -\frac{1}{\varepsilon_y}\frac{d^2H_z}{dx^2}-\frac{1}{\varepsilon_x}\frac{d^2H_z}{dy^2} = k_o^2\mu_zH_z \end{aligned}, \tag{25}$$

Combining the d^2H_z/dx^2 and d^2H_z/dy^2 terms in (25) gives the following second order differential dispersion equation for H_z

$$\frac{k_o^2\mu_x}{k_o^2\mu_x\varepsilon_y-k_z^2}\frac{d^2H_z}{dx^2}+\frac{k_o^2\mu_y}{k_o^2\mu_y\varepsilon_x-k_z^2}\frac{d^2H_z}{dy^2}+k_o^2\mu_zH_z = 0. \tag{26}$$

5. Dispersion Equation for E_z

We expand the $\nabla x\underline{\boldsymbol{E}}$ term of (21) in terms of (13-16), and take the dot product with $\underline{\underline{\boldsymbol{\mu}}}_r^{-1}$

$$\begin{aligned} \nabla x\Bigg[&\frac{\underline{x}_o}{\mu_x}\left(\frac{dE_z}{dy}-\frac{dE_y}{dz}\right)+\frac{\underline{y}_o}{\mu_y}\left(\frac{dE_x}{dz}-\frac{dE_z}{dx}\right) \\ &+\frac{\underline{z}_o}{\mu_z}\left(\frac{dE_y}{dx}-\frac{dE_x}{dy}\right)\Bigg] = k_o^2\underline{\varepsilon}_r\cdot\underline{E} \end{aligned}. \tag{27}$$

Evaluating (27) and taking the dot product with the $\underline{z}_o$ direction allows us to isolate the E_z component of the magnetic field on the right hand side of the equation

$$-\frac{1}{\mu_y}\frac{d^2E_z}{dx^2}-\frac{1}{\mu_x}\frac{d^2E_z}{dy^2}+\frac{1}{\mu_y}\frac{d^2E_x}{dxdz}+\frac{1}{\mu_x}\frac{d^2E_y}{dydz}=k_o^2\varepsilon_z E_z . \quad (28)$$

Setting H_z=0, if we differentiate (15) by $d^2/dxdz$ and (16) by $d^2/dydz$, keeping in mind that $d/dz = -jk_z$, and plug the results into (28), then we arrive at the following result

$$\begin{gathered}\frac{k_z^2}{\mu_y\left(k_z^2-k_o^2\mu_y\varepsilon_x\right)}\frac{d^2E_z}{dx^2}+\frac{k_z^2}{\mu_x\left(k_z^2-k_o^2\mu_x\varepsilon_y\right)}\frac{d^2E_z}{dy^2}\\ -\frac{1}{\mu_y}\frac{d^2E_z}{dx^2}-\frac{1}{\mu_x}\frac{d^2E_z}{dy^2}=k_o^2\varepsilon_z E_z\end{gathered} . \quad (29)$$

Combining the d^2E_z/dx^2 and d^2E_z/dy^2 terms in (38) gives the following second order differential dispersion equation for E_z

$$\frac{k_o^2\varepsilon_x}{k_o^2\mu_y\varepsilon_x-k_z^2}\frac{d^2E_z}{dx^2}+\frac{k_o^2\varepsilon_y}{k_o^2\mu_x\varepsilon_y-k_z^2}\frac{d^2E_z}{dy^2}+k_o^2\varepsilon_z E_z=0 . \quad (30)$$

6. Suppression of Birefringence in a Rectangular Waveguide

Birefringence is a characteristic of anisotropic media where a single incident wave entering the boundary of an anisotropic medium gives rise to two refracted waves as shown in Figure 1 or a single incident wave leaving gives rise to two reflected waves as shown in Figure 2. We call these two waves the ordinary wave and the extraordinary wave. For low order modes, and especially the first mode, a rectangular waveguide suppresses the birefringence inherent to anisotropic media by suppressing propagation in the vertical direction. This suppression assumes that the dimensions of the waveguide are such that the horizontal dimension is at least twice the size of the vertical dimension [10]. To see how the geometry of the waveguide acts to cancel out the extraordinary wave, we need to solve for k_z from our dispersion equations.

Equations (26) and (30) yield the following solutions in unbounded anisotropic media restricted by the radiation condition (previously stated in section 3)

$$E_z(x,y,z)=E_o e^{-j\left(k_x x+k_y y+k_z z\right)}, \quad (31)$$

$$H_z(x,y,z)=H_o e^{-j\left(k_x x+k_y y+k_z z\right)} . \quad (32)$$

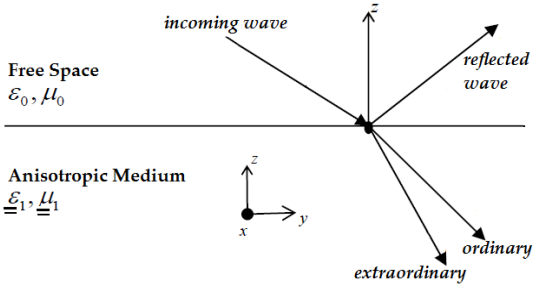


Figure 1: Free space plane wave incident on an anisotropic boundary [9].

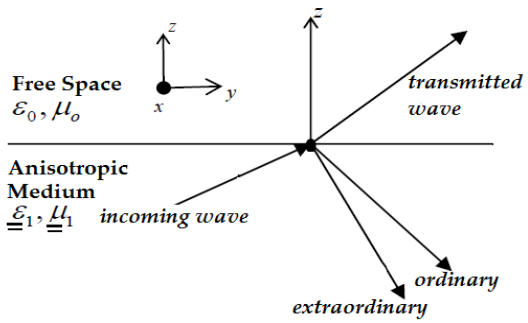


Figure 2: Anisotropic plane wave incident on a free space boundary [9].

Plugging (31) into (30) (equivalently we could substitute (32) into (26)) yields a polynomial equation whose solutions give the values of k_z in the anisotropic medium. Noting that $d^2/dx^2 = -k_x^2$ and $d^2/dy^2 = -k_y^2$, (30) simplifies to

$$\frac{k_o^2k_x^2\varepsilon_x E_z}{\left(k_o^2\mu_y\varepsilon_x-k_z^2\right)}+\frac{k_o^2\varepsilon_y k_y^2E_z}{\left(k_o^2\mu_x\varepsilon_y-k_z^2\right)}-k_o^2\varepsilon_z E_z=0 . \quad (33)$$

Dividing out the $k_o^2E_z$ term and multiplying through by both denominators gives us the following factored polynomial

$$\begin{gathered}\left(k_o^2\mu_y\varepsilon_x-k_z^2\right)\left(k_o^2\mu_x\varepsilon_y-k_z^2\right)\varepsilon_z-k_x^2\varepsilon_x\left(k_o^2\mu_x\varepsilon_y-k_z^2\right)\\ -\varepsilon_y k_y^2\left(k_o^2\mu_y\varepsilon_x-k_z^2\right)=0\end{gathered} . \quad (34)$$

Finally, multiplying out (34) yields a fourth order polynomial whose roots yield the four values of k_z describing the ordinary wave and extraordinary wave in the positive and negative propagation directions

$$\begin{gathered}k_z^4\mu_z+\left[k_x^2\mu_x+k_y^2\mu_y-(\varepsilon_x\mu_y+\varepsilon_y\mu_x)k_o^2\mu_z\right]k_z^2\\ +\left[k_o^4\varepsilon_x\varepsilon_y\mu_x\mu_y\mu_z-k_o^2k_x\varepsilon_x\mu_y\mu_x-k_o^2\varepsilon_y k_y\mu_x\mu_y\right]=0\end{gathered} . \quad (35)$$

Equation (35) is directly responsible for the existence of the two ordinary and extraordinary waves that are characteristic of the birefringence phenomenon. In an isotropic medium, the resulting polynomial for k_z is a second order polynomial, which yields only the values for the positive and negative propagation of the single ordinary wave. A fourth order polynomial allows for positive and negative propagation of both the ordinary and extraordinary wave shown in Figures 1 and 2.

The suppression of birefringence requires (35) to reduce to a second order polynomial. This is clearly not possible in an unbounded anisotropic medium as currently formulated. However, the boundary conditions in a rectangular waveguide assume that the first resonance suppresses the propagation constant in the vertical direction [10]. In other words, k_y=0 and d^2/dy^2=0, and this completely eliminates the second E_z term in (33). This simplifies (34) to

$$-\frac{k_o^2k_x^2\varepsilon_x}{k_o^2\mu_y\varepsilon_x-k_z^2}E_z+k_o^2\varepsilon_z E_z=0 , \quad (36)$$

which resembles the form of the traditional second order wave equation. Solving for k_z thus leads to the following second order polynomial

$$k_z^2 = \varepsilon_x \left(k_o^2 \mu_y - k_x^2 / \varepsilon_z \right). \quad (37)$$

Since (37) yields the positive and negative square root of the right-hand-side as the two solutions for k_z, this indicates the existence of only a single propagating wave. The suppression of the k_y term in the first resonant mode of the rectangular waveguide yields a second order differential equation for the wave number in the propagation direction, thereby eliminating the property of birefringence for this case. For a practical validation of this theory, the author would point the reader to reference [2]. This paper shows an antenna application that exploits the suppression of anisotropic birefringence to facilitate wideband impedance matching of a cavity backed antenna.

The reader should remember that (37) applies only to the first order resonance mode of the rectangular waveguide. As higher order modes are introduced propagation birefringence will exist for some modes depending on the boundary conditions. A similar analysis should be carried out for each propagating mode.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] G. Mitchell, "Suppression of anisotropic birefringence in a rectangular waveguide" in Proceedings of the Applied Computational Electromagnetics Society (ACES) Symposium, 2018. https://doi.org/10.23919/ROPACES.2018.8364258

[2] G. Mitchell and W. Wasylkiwskyj, "Theoretical anisotropic transverse resonance technique for the design of low profile wideband antennas" IET Microwaves, Antennas and Propagation, 10(5), 487-493, 2016. https://doi.org/10.1049/iet-map.2015.0470

[3] R. Collin, "A Simple Artificial Anisotropic Medium" IRE Transactions on Microwave Theory and Techniques, 6(4), 206-209, 1958. https://doi.org/10.1109/TMTT.1958.1125216

[4] M. Pardavi-Horvath, "Microwave Applications of Soft Ferrites" Journal of Magnetism and Magnetic Materials, 215-216(1), 171-183, 2000. https://doi.org/10.1016/S0304-8853(00)00106-2

[5] K. Buell, H. Mosallaei, K. Sarabandi, "A Substrate for Small Patch Antennas Providing Tunable Miniaturization Factors" IEEE Transactions on Microwave Theory and Techniques, 54(1), 135-146, 2006. https://doi.org/10.1109/TMTT.2005.860329

[6] D. Pozar, "Radiation and Scattering from a Microstrip Patch on a Uniaxial Substrate" IEEE Transactions on Antennas and Propagation, 35(6), 613-621, 1987. https://doi.org/10.1109/TAP.1987.1144161

[7] F.Y. Meng, Q. Wu, L. W. Li, "Transmission Characteristics of Wave Modes in a Rectangular Waveguide filled with Anisotropic Metamaterial" Applied Physics A: Materials Science and Processing, 94(4), 747-753, 2009. https://doi.org/10.1007/s00339-008-5057-2

[8] F.Y. Meng, Q. Wu, J.H. Fu, "Miniaturized Rectangular Cavity Resonator based on Anisotropic Metamaterials Bilayer" Microwave and Optical Technology Letters, 50(8), 2016-2020, 2008. https://doi.org/10.1002/mop.23556

[9] J. Graham, "Arbitrarily Oriented Biaxially Anisotropic Media: Wave Behavior and Microstrip Antennas," Ph.D Thesis, Syracuse University, 2012.

[10] D. Pozar, Microwave Engineering: 3rd edition, John Wiley and Sons, 2005.

[11] W.J. Ince, E. Stern, "Non-Reciprocal Remanence Phase Shifters in Rectangular Waveguide" IEEE Transactions on Microwave Theory and Techniques. 15(2), 87-95, 1967. https://doi.org/10.1109/TMTT.1967.1126381

[12] D. Torrent, J. Sanchez-Dehesa, "Radial Wave Crystals: Radially Periodic Structures from Anisotropic Metamaterials for Engineering Acoustic or Electromagnetic Waves" Physics Review Letters, 103(6), 2009. https://doi.org/10.1103/PhysRevLett.103.064301

[13] J. Sanchez-Dehesa, D. Torrent, J. Carbonell, "Anisotropic metamaterials as sensing devices in acoustics and electromagnetism" in Proceedings of the International Society for Optics and Photonics (SPIE), 2012. https://doi.org/10.1117/12.916043

[14] Y.G Ma, P. Wang, X. Chen, C.K. Ong, "Near-field plane-wave-like beam emitting antenna fabricated by anisotropic metamaterial" Applied Physics Letters, 94(4), 2009. https://doi.org/10.1063/1.3077128

[15] Q. Cheng, "Directive Radiation of Electromagnetic Waves Based on Anisotropic Metamaterials" in Proceedings of IEEE Asia-Pacific Conference on Antennas and Propagation, Singapore, Singapore, 2012.

[16] D. Schurig, J. Mock, B. Justice, S. Cummer, J. Pendry, A. Starr, D. Smith, "Metamaterial electromagnetic cloak at microwave frequencies" Science, 314(5801), 977-980, 2006. https://doi.org/10.1126/science.1133628

[17] J. Wong, K. Balmain, "A beam-steerable antenna based on the spatial filtering property of hyperbolically anisotropic metamaterials" in Proceedings of IEEE International Symposium of the Antennas and Propagation Society, Honolulu, HI, USA, 2007. https://doi.org/10.1109/APS.2007.4396438

[18] T. Cai, G.M. Wang, "Polarization-controlled Bifunctional Antenna based on 2-D Anisotropic Gradient Metasurface" in Proceedings of IEEE Conference on Microwave and Millimeter Wave Technology, Beijing, China, 2016. https://doi.org/10.1109/ICMMT.2016.7761687

[19] W. Zhong, X. Meng, Q. Hong, "Radiation power of half wave oscillator antenna in magnetic anisotropic medium" in 2018 International Conference on Electronics Technology (ICET), Chengdu, China, 2018. https://doi.org/10.1109/ELTECH.2018.8401446

[20] M. Born, E. Wolf, Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light: 7th edition, Cambridge University Press, 1999.

[21] O. Ozgun, M. Kuzuoglu, "Finite element modeling of anisotropic half-space problems by a simple mesh truncation scheme" in 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, San Diego, CA, USA 2017. https://doi.org/10.1109/APUSNCURSINRSM.2017.807283

[22] J. Zhuo, F. Han, L. Ye, Z. Yu, Q.H. Liu, "Simulation of Electromagnetic Scattering of 3-D Inhomogeneous Biaxial Anisotropic Magnetodielectric Objects Embedded in Uniaxial Anisotropic Media by the Mixed-Order BCGS-FFT Method" IEEE Transactions on Microwave Theory and Techniques, 66(8), 3745-3755, 2018. https://doi.org/10.1109/TMTT.2018.2840984

[23] G. Mitchell, S. Weiss, "An Overview of ARL's Low Profile Antenna Work utilizing Anisotropic MetaFerrites", in Proceedings of the IEEE International Symposium on Phased Array Systems and Technology, Waltham, MA, USA, 2016. https://doi.org/10.1109/ARRAY.2016.7832581

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 58-64 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Performance Analysis of Thyristors Switched Capacitors used for Reactive Power Compensation of Induction Motor

Dmitry Ivanovich Panfilov[*,1], Ahmed Elsayed ELGebaly[2], Michael Georgievich Astashev [3], Alexander Nikolaevich Rozhkov [1]

[1]*Department of industrial electronics, Moscow power engineering institute, Moscow, 111250, Russia*

[2]*Electrical power and machines department Faculty of engineering, Tanta university, Tanta, 31527, Egypt*

[3]*A G. M. Krzhizhanovsky power engineering institute (JSC ENIN), Moscow, 119991, Russia*

ARTICLE INFO

Article history:
Received:29 April 2019
Accepted:26 June, 2019
Online: 11 July, 2019

Keywords:
Reactive power compensation
Thyristors switched capacitors
Induction motor starting

ABSTRACT

This paper analyzes the process of reactive power compensation using thyristors switched capacitors TSC topology based on multiterminal switch during the starting of induction motor. The paper explains the control principle of this TSC topology which depends on the multiterminal switch. To actively track the required reactive power for the dynamic load and to perfectly transit from it from one level to other, proper control system is established. MATLAB/SIMULINK is used to develop a simulation model to prove the dynamic performance of the proposed TSC during the compensation of reactive power required for induction motor during starting.

1. Introduction

This paper is an extension of work originally presented in 2018 X International Conference on Electrical Power Drive Systems (ICEPDS) under title (Performance Assessment of Thyristors Switched Capacitors during Reactive Power Compensation of Dynamic Load) [1]. It is well known that; the induction motors are the most broadly utilized motors in the world due to their high rigidness [2]. The powers of the used induction motors vary from the fraction of kilowatt up to several megawatt [3]. The induction motors require reactive power during their operation either they are supplied from conventional AC source or through drive system [4]. But, the connection of reactive power compensators with the induction motors improves the power system quality where they are supplied [5]. The compensation of the reactive power may be achieved by the connection of fixed capacitor on the terminals of the induction motor. But, this method doesn't adapt the requirements of the motor for the various operation modes. The application of dynamically changed reactive power supply improves the power factor for wider operation range particularly during the starting of the motors [6]. The induction motors speed may be changed using dynamically changed reactive power source [7].

Several references [8,9] study the usage of conventional static VAR compensators needed by the induction. The production of harmonics is an essential demerit of the conventional SVCs. Therefore, they require the application of filters [10,11]. The switched-based SVC present a vital solution to remove the harmonics during the control of the reactive power. But, it should be considered that, the switched SVCs have discrete performance during the control of reactive power; consequently, they require different control and design methodology to obtain the smoothest characteristic [12].

This research analyzes the process of reactive power compensation for the starting of induction motor depending on switched SVC. The thyristor switched capacitors TSC which has 25 discrete steps of operation is the type of SVC applied in this study. The design of such TSC topology requires special design to dynamically change the developed capacitive reactive power. The control methodology applied for the TSC topology is established to regulate the proper amount of reactive power during the starting of the induction motor. The detailed induction motor model is illustrated to study the effect of the SVC on its operation. By using MATLAB/SIMULINK package, Simulation model of the whole system is established to demonstrate the TSC performance during the compensation of reactive power.

[*]Department of industrial electronics, Moscow power engineering institute Moscow, Russia, +79859699550, Dmitry.panfilov@inbox.ru

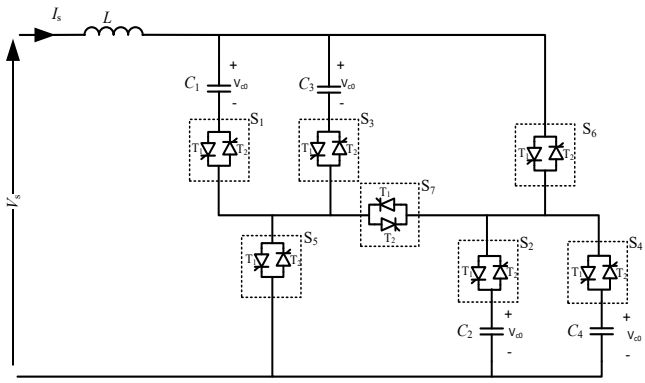


Figure 1: TSC produces 25 discrete levels of capacitive reactive power

2. Principle of operation of multiterminal switch TSC

Figure 1 illustrates the proposed TSC scheme which can produce 25 level of operation by providing 25 different equivalent capacitances [13]. In compared with binary TSC and for the same number of capacitors, the developed scheme produces more steps of operation where the binary TSC only produces 16 steps [14]. Therefore, the developed TSC provides more flexible operation to produce capacitive reactive power while there is no need for application of thyristor-controlled reactor TCR for reactive power regulation. The proposed TSC topology produces zero harmonics during its operation.

Table 1 demonstrates the several equivalent capacitances produced by the proposed TSC. The connected reactor L in the topology is important related to the reduction of current rate of change [15] and related to the restoring of the stored energy on capacitors back to the source [13].

The proposed TSC parameters is obtained depending on an optimization technique which assures the most smooth variation of the distinct characteristic [16,17]. The developed capacitive reactive power has discrete characteristic as shown in Figure 2. The function of control system of this topology is the determination the required reactive power in another words the required susceptance. Therefore, the control system should regulate the operation state of each switch of the seven switches to produce specific capacitance as in Table 1. The TSC reactive power control system needed by induction motor at starting will be explained in section 4.

As shown in Table 1, some of the equivalent capacitance include series connection of certain capacitors. The series connection of capacitors during generation of certain levels of reactive power leads in appearance of permanent different DC voltages on capacitors after termination of the reactive power level generation. To change the equivalent capacitance to another level, the capacitors which were connected in series on previous stage should be discharged. The discharging process should be completed in very short time to make the system response faster. The discharging of capacitors achieved by regenerative method by recovery the capacitors stored energy to the power supply. The

recovery process can be accomplished by connection of reactor L shown in the scheme Figure 1.

Table 1: The equivalent capacitance obtained by the proposed TSC with four capacitors and seven switches

No.	Equivalent capacitance	Switches state						
		S_1	S_2	S_3	S_4	S_5	S_6	S_7
1	0	OFF	OFF	OFF	OFF	OFF	OFF	OFF
2	C_1+C_2	ON	ON	OFF	OFF	OFF	OFF	ON
3	C_2+C_3	OFF	ON	ON	OFF	OFF	OFF	ON
4	$(C_1//C_3) + C_2$	ON	ON	ON	OFF	OFF	OFF	ON
5	C_2	OFF	ON	OFF	OFF	OFF	ON	OFF
6	C_1+C_4	ON	OFF	OFF	ON	OFF	OFF	ON
7	$(C_2//C_4) + C_1$	ON	ON	OFF	ON	OFF	OFF	ON
8	C_3+C_4	OFF	OFF	ON	ON	OFF	OFF	ON
9	$(C_2//C_4) + C_3$	OFF	ON	ON	ON	OFF	OFF	ON
10	C_1	ON	OFF	OFF	OFF	ON	OFF	OFF
11	$(C_1//C_3) + C_4$	ON	OFF	ON	ON	OFF	OFF	ON
12	$(C_1//C_3)+ (C_2//C_4)$	ON	ON	ON	ON	OFF	OFF	ON
13	$C_1//C_2$	ON	ON	OFF	OFF	ON	ON	OFF
14	C_3	OFF	OFF	ON	OFF	ON	OFF	OFF
15	C_4	OFF	OFF	OFF	ON	OFF	ON	OFF
16	$C_2//C_3$	OFF	ON	ON	OFF	ON	ON	OFF
17	$C_2//C_4$	OFF	ON	OFF	ON	OFF	ON	OFF
18	$C_1//C_3$	ON	OFF	ON	OFF	ON	OFF	OFF
19	$C_1//C_4$	ON	OFF	OFF	ON	ON	ON	OFF
20	$C_1//C_2//C_3$	ON	ON	ON	OFF	ON	ON	OFF
21	$C_1//C_4//C_2$	ON	ON	OFF	ON	ON	ON	OFF
22	$C_3//C_4$	OFF	OFF	ON	ON	ON	ON	OFF
23	$C_3//C_4//C_2$	OFF	ON	ON	ON	ON	ON	OFF
24	$C_1//C_4//C_3$	ON	OFF	ON	ON	ON	OFF	OFF
25	$C_1//C_2//C_3//C_4$	ON	ON	ON	ON	ON	ON	OFF

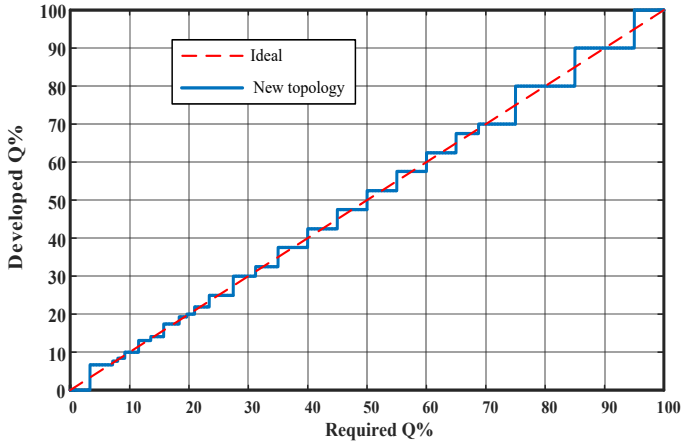


Figure 2: The characteristic of the proposed TSC

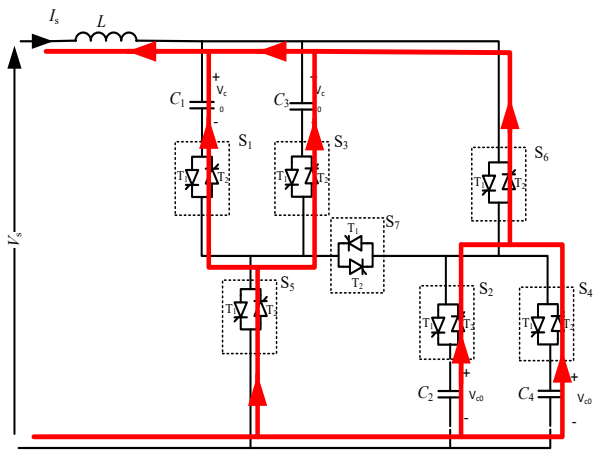
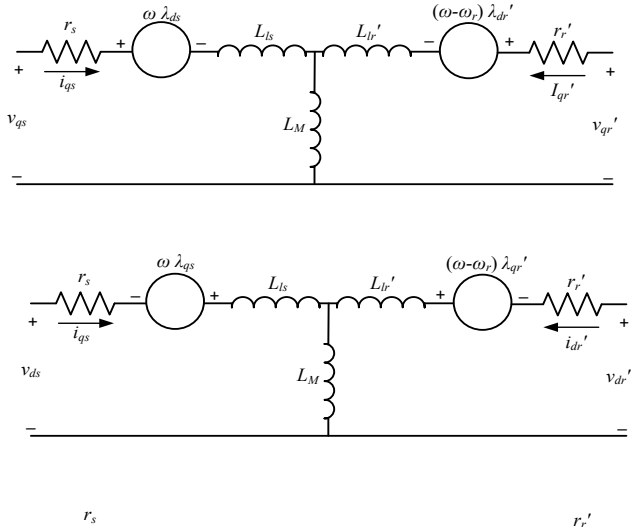


Figure 3: Discharging path for positively charged capacitors of TSC



To discharge all capacitors, the time presents a critical factor, So, in this study, the TSC system is designed to discharge the capacitors in very short less than quarter of the voltage period. The added reactor and any combination of capacitors is considered resonance circuit. By the help of this resonance circuit, the stored energy on capacitors can be recovered to the power source. Figure 3 illustrates the discharging path for the charged capacitors. The regeneration can be completed within time less than the one quarter of the entire period [14]. The semiconductor switches in this scheme has a vital rating related to the maximum amplitude of current during discharging. The inductance value and firing angle to start the process affect the performance of discharging process. The resonance frequencies ω_k relates between the inductance L and the total capacitance C_{eq} by the following relation:

$$\omega_k = \frac{1}{\sqrt{L\,C_{eq}}} \tag{6}$$

The frequency ω_k determines the time of the discharging time which should be fewer than one-quarter of the fundamental period. So, the relation between ω_b and ω_k is as the following, while $k > 4$:

$$k = \frac{\omega_k}{\omega_b} \tag{7}$$

Therefore, the inductance is obtained consistent with the subsequent relation:

$$L = \frac{1}{\left(k\,\omega_0\right)^2 C_{eq}} \tag{8}$$

In this study, k is set at 24. The total required capacitance is determined consistent with the reactive power requirement of the induction motor as will be explained in section 5.

3. The Detailed Dynamic Model of the Induction Motor

The three-phase induction motor consists of six windings; three windings are in the stator and three are in the rotor. The equations of the stator voltage are as the following:

$$\begin{bmatrix} v_{as} \\ v_{bs} \\ v_{cs} \end{bmatrix} = \begin{bmatrix} R_s & 0 & 0 \\ 0 & R_s & 0 \\ 0 & 0 & R_s \end{bmatrix} \begin{bmatrix} i_{as} \\ i_{bs} \\ i_{cs} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_{as} \\ \lambda_{bs} \\ \lambda_{cs} \end{bmatrix} \tag{1}$$

where, v_{as} , v_{bs} and v_{cs} are the voltages on the stator windings, R_s is stator resistance and λ_{as}, λ_{bs} and λ_{cs} are the linkage fluxes of the three windings. The flux linkage can be described as the following

$$\begin{bmatrix} \lambda_s^{abc} \\ \lambda_r^{abc} \end{bmatrix} = \begin{bmatrix} L_{ss}^{abc} & L_{sr}^{abc} \\ L_{rs}^{abc} & L_{rr}^{abc} \end{bmatrix} \begin{bmatrix} i_s^{abc} \\ i_r^{abc} \end{bmatrix} \tag{2}$$

where subscript 'r' specifies rotor and 's' specifies stator. Moreover, L_{ss}^{abc} and L_{rr}^{abc} are stator and rotor inductance matrixes. As well known, both of L_{sr}^{abc} and L_{rs}^{abc} are stator and rotor mutual inductance matrixes which have time variant parameters. To solve the dynamic equation (1), some difficulties should be faced. So, the using of transformation theory through Klark matrix (equation (3)) presents a solution to solve the dynamic model.

$$\left[T_{qd0}(\theta)\right] = \frac{2}{3} \begin{bmatrix} \cos\theta & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \sin\theta & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ 0.5 & 0.5 & 0.5 \end{bmatrix} \tag{3}$$

The voltage equations after transformation are presented as in the following equation.

$$\left[V_{sqd0}\right] = R_s\left[i_{qd0}\right] + p\left[\lambda_s^{qd0}\right] + \omega\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\left[\lambda_s^{qd0}\right] \quad (4)$$

where λ_s^{qd0} is the stator windings flux linkage, i^{qd0} the arbitrary reference frame currents and ω is the rotational speed reference frame. The arbitrary reference frame equivalent circuits for symmetrical three-phase induction motor are illustrates in Figure 4.

It is common to present the parameters of machines and power system in ohms, per-unit or percent of base impedance, so, it will be more suitable to express both of flux linkage and voltage equations in terms of reactances instead of inductances as in the following equations [1, 2, 3].

$$v_{qs} = r_s i_{qs} + \frac{\omega}{\omega_b}\psi_{ds} + \frac{p}{\omega_b}\psi_{qs} \quad (5)$$

$$v_{ds} = r_s i_{ds} - \frac{\omega}{\omega_b}\psi_{qs} + \frac{p}{\omega_b}\psi_{ds} \quad (6)$$

$$v_{0s} = r_s i_{0s} + \frac{p}{\omega_b}\psi_{0s} \quad (7)$$

$$v'_{qr} = r'_r i'_{qr} + \left(\frac{\omega - \omega_r}{\omega_b}\right)\psi'_{dr} + \frac{p}{\omega_b}\psi'_{qr} \quad (8)$$

$$v'_{dr} = r'_r i'_{dr} - \left(\frac{\omega - \omega_r}{\omega_b}\right)\psi'_{qr} + \frac{p}{\omega_b}\psi'_{dr} \quad (9)$$

$$v'_{0r} = r'_r i'_{0r} + \frac{p}{\omega_b}\psi'_{0r} \quad (10)$$

$$\psi_{qs} = X_{ls} i_{qs} + X_M\left(i_{qs} + i'_{qr}\right) \quad (11)$$

$$\psi_{qs} = X_{ls} i_{ds} + X_M\left(i_{ds} + i'_{dr}\right) \quad (12)$$

$$\psi_{0s} = X_{ls} i_{0s} \quad (13)$$

$$\psi'_{qr} = X'_{lr} i'_{qr} + X_M\left(i_{qs} + i'_{qr}\right) \quad (14)$$

$$\psi'_{dr} = X'_{lr} i'_{dr} + X_M\left(i_{ds} + i'_{dr}\right) \quad (15)$$

$$\psi'_{0r} = X'_{lr} i'_{0r} \quad (16)$$

$$T_e = \left(\frac{3}{2}\right)\left(\frac{P}{2}\right)\left(\frac{1}{\omega_b}\right)\left(\psi'_{qr} i'_{dr} - \psi'_{dr} i'_{qr}\right) = T_l + Jp\omega_r \quad (17)$$

where ω_b is the base electrical angular velocity used to calculate the reactances. The flux linkages λ in equation 4 are converted to ψ flux linkages per second. ω_r is the rotational speed of rotor. X_l is leakage reactance and X_M is the mutual reactance. T_e and T_l are electromagnetic and load torque respectively. J is the inertia for motor and load. By using the previous dynamic equations, the motor can be analyzed using any simulation package such as MATLAB/SIMULINK used in this paper. All variable states of the induction motor can be determined by this model specially the variables estimate the required reactive power.

4. Control System of Reactive Power for the Operation of Induction Motor

Figure 5 shows the block diagram of the control system for the proposed TSC to regulate the reactive power of the motor. The voltages and currents of the motor terminal are measured to get the feedback signals. Therefore, the amount of capacitive reactive power is determined. The corresponding amount of susceptance B can be estimated from both of terminal voltage and reactive power. Consequently, the required susceptance presents the input of TSC lookup table characteristic in Table 1 and Figure 2. One step of the 25 steps is the lookup table output. The changing of the developed

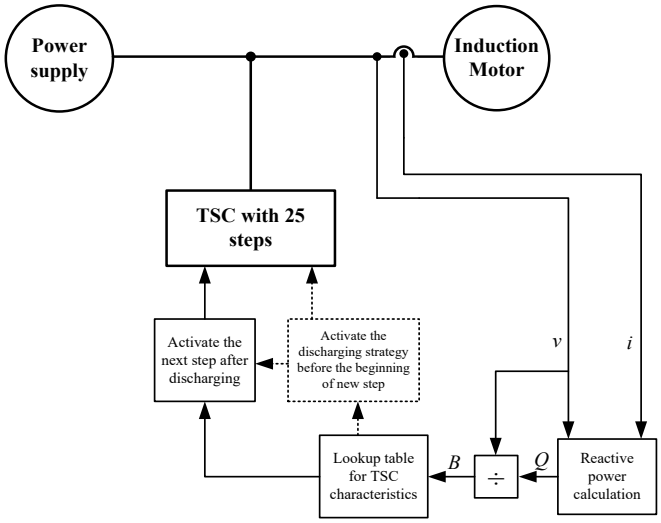


Figure 5: The block diagram for the TSC control system for reactive power compensation

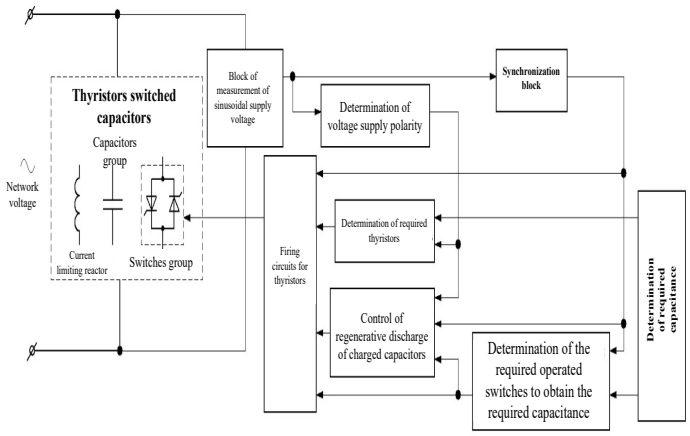


Figure 6: Block diagram of new topology of TSC control system

steps from look up table activates the capacitors discharging process once before the implementation of the following step [13]. Figure 6 illustrates the proposed TSC control system for both implementation of any susceptance and the activation of the discharging process [14].

When the discharging process is activated, the produced capacitive reactive power equals zero. If the next step is activated after the discharging process, the SVC continues the production of reactive power. The control proposed algorithm depends on the tracking of the required reactive power for the motor to produce it by the TSC. The transfer function describes the performance of the TSC can be described as:

$$\frac{B_{TSC}}{B_{ref}} = \frac{e^{-sT_d}}{1 + s\,T_b} \tag{18}$$

where B_{TSC} is the produced susceptance, B_{ref} is the reference susceptance, T_d is the dead-time and T_b is the delay-time [18].

5. Simulation Results of the compensated motor by TSC

In this study, the reactive power compensated induction motor by the proposed TSC is simulated depending on MATLAB/SIMULINK package. The reactive power compensation is imitated throughout the induction motor starting to illustrate the performance of effective control system proposed to regulate the reactive power. Figure illustrates the parallel connection of the TSC in parallel with the motor. The line-to-line voltage of power supply equals 400 V. The induction motor has the following ratings: nominal power of 37.3 kVA, 400 V line-to-line voltage, 50 Hz, 4-poles. The motor has the following parameters: r_s = 0.087 Ω, X_{ls} = 0.302 Ω, r_r' = 0.228 Ω, X_{lr}' = 0.302 Ω, X_m = 13.08 Ω and inertia constant J = 8 kg.m2.

The ratings of the connected TSC are as the following: rated line voltage of 400 V and rated power of 243 kVAR capacitive. Although the power rating of the TSC is much higher than the rating of the motor itself, the TSC reactive power is enough to compensate the reactive power required by the motor during starting. At free running, the TSC can compensate the reactive power of several connected motors because the required reactive power by the motor is reduced by the speed increment.

If the induction motor starts running, the TSC starts to compensate the reactive power compensation with the full rating then the compensated reactive power is gradually decreased. The TSC control system adjusts the reactive power by producing the maximum reactive power presented by level number 25 (Figure 2 and Table 1) at zero speed. At no-load, the required reactive power by the motor is reduced to lower level presented by level number 2. The motor with and without TSC is simulated to illustrate the TSC influence on the motor performance. The motor needs around 12 seconds to reach its full speed; consequently, the emulation period is 15 seconds which are sufficient to study the influence of the reactive power compensation on the motor performance. Figure 7 shows the supply reactive power in two cases; compensated and uncompensated motor. The using of TSC enables to reduce the total apparent power produced by the power source. Figure 8 shows the produced capacitive reactive power if the motor is compensated. Figure 9 illustrates the step of operation required to produce the corresponding reactive power. Several benefits are achieved by the reactive power compensation. The first advantage is shown in Figure 10; the compensation causes the reduction of the line current of the power source. The second benefit is the enhancement of the motor terminal voltage as illustrated in Figure 11. Therefore, the enhancement of voltage causes the increasing of motor acceleration as illustrated in Figure 12. Due to the reduction of the supply current, the internal losses of the motor are reduced where the internal winging resistance equals r = 0.5 Ω as in Figure 13. The changing of the compensation level leads to the appearance of spikes in the system response. The transition needs one full period to discharge the installed capacitors as explained in section 2.

The power factor of the system is improved due to compensation. This is due to the reduction of the phase shift between the voltage and current as illustrated in Figure 14

As explained earlier, the transition between two steps of reactive power requires the discharging process to be activated. To discuss this interval, the instantaneous waveforms of currents and voltages are illustrated. Throughout the changing of the developed steps from 25 to 24, Figure 15(a) shows the waveform of supply voltage and its produced current. As shown in the figure, the phase shift between current and voltage is around zero. In steady state, the waveform of the supply current is pure sin wave except some distortion during the transition period. The time span of the distorted current equals about one period. As shown in Figure 15(b), the phase shift between the TSC current and the supply terminal voltage is 90° lead. During the transient period, the TSC supplies zero current except spike current owing to the process of capacitors discharging. Furthermore, Figure15(c) illustrates that the voltage and the motor current are not in phase; therefore, the power factor below unity. Nevertheless, the usage of TSC enhances the power factor to unity by reducing the phase-shift between the voltage and current.

6. Conclusion

TSC has been developed as a reactive power supply to produce the required reactive power for the induction motor during starting period. The circuit topology of the proposed TSC is introduced to control the reactive power with 25 discrete steps. The principle of operation and characteristic of the proposed TSC scheme is analyzed to produce compensation system suitable for dynamically changed load. The induction motor and TSC dynamics have been introduced to analyze the dynamic performance of the whole system. The control system block diagram is constructed to track the needed reactive power for the motor. The concept of regeneration of capacitive unit energy has been explained to assure the dynamic reactive power production for the TSC. The design parameters of the TSC have been determined to produce the required reactive power and to carry out the discharging process effectively. using MATLAB/SIMULINK package, simulation and model have been developed for the whole system contains induction motor, the TSC and its control system. The simulation results demonstrate the efficiency of the proposed TSC topology during the reactive power compensation needed by the induction motor through the dynamic performance. The control system efficiently controls the TSC response to actively produce the proper reactive power.

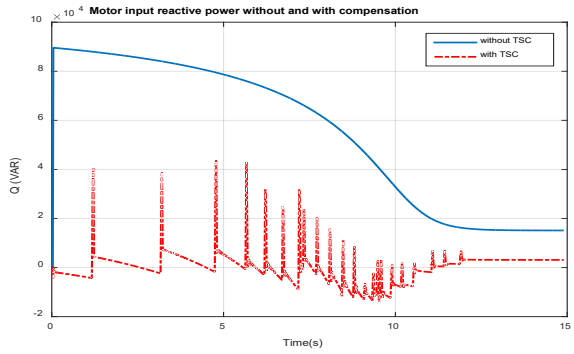


Figure 7: The supply reactive power produced for the motor in the compensated and uncompensated cases

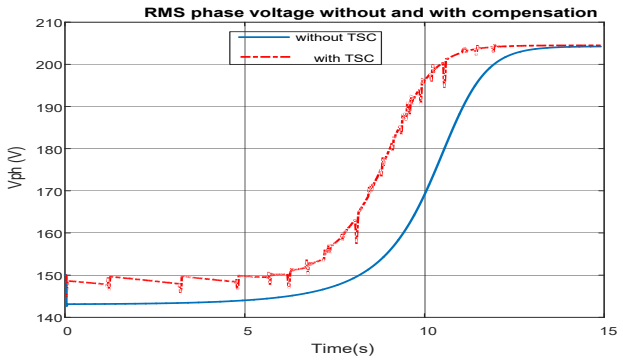


Figure 11: The IM motor terminal phase voltage throughout starting period

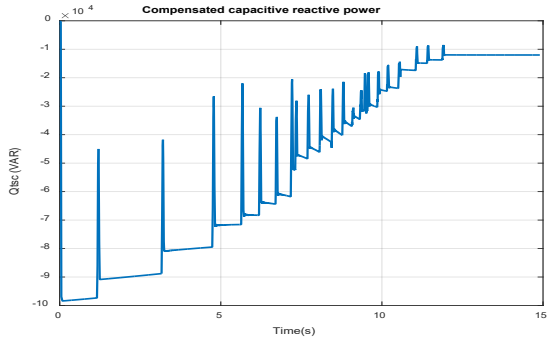


Figure 8: TSC produced capacitive reactive power

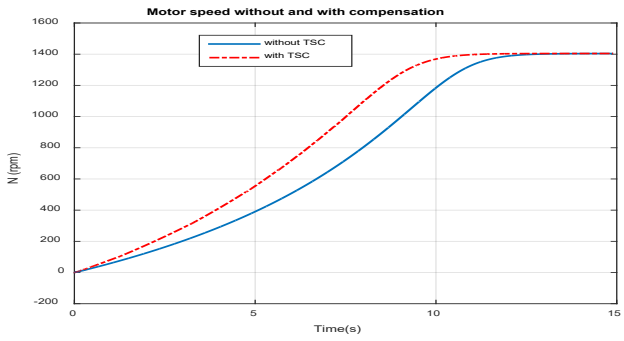


Figure 12: The speed of the IM during starting period

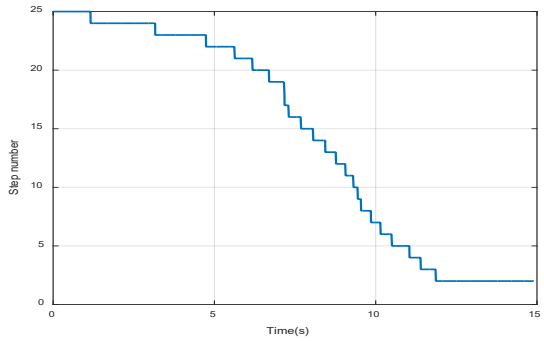


Figure 9: TSC step number to regulate the produced reactive power.

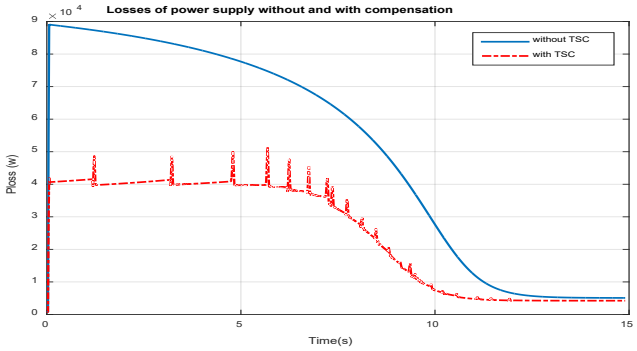


Figure 13: The internal supply losses during starting

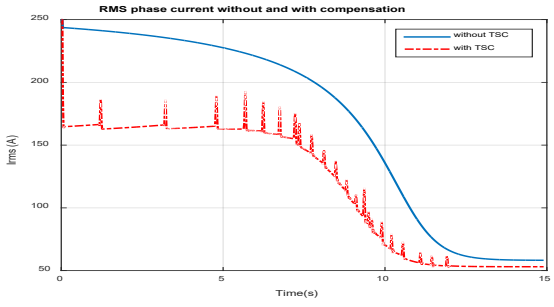


Figure 10: The power supply phase current throughout the motor starting

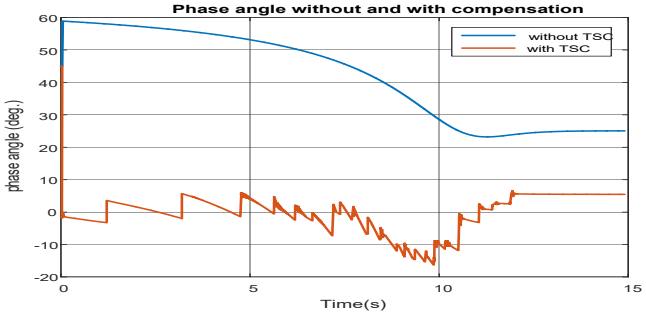


Figure 14. Phase shift between phase voltage and current

Acknowledgment

The study is achieved at Stock Company G. M. Krzhizhanovsky Power Engineering Institute in the framework of the project " Development of a controlled source of reactive power with the absence of the higher-order harmonics currents during the regulation of the electrical energy and with the

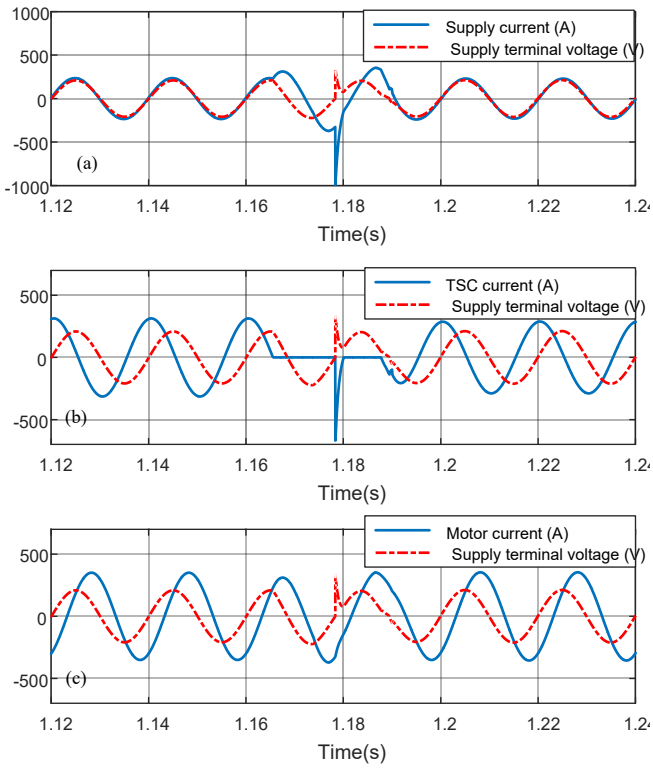


Figure 15. Waveforms of the source terminal phase voltage and current, motor current and TSC current throughout the transition between two levels between 25 and 24 steps

improved technical and economic indicators on the basis of the domestic component base of power electronics for automatic control of the voltage and the power flows in the electric power distribution networks of 6-110 kV (RFMEFI57917X0140)" with the financial support of the Ministry of Education and Science of the Russian Federation.

References

[1] D. I. Panfilov, A. E. ElGebaly, M. G. Astashev and Alexander N. Rozhkov. "Performance Assessment of Thyristors Switched Capacitors during Reactive Power Compensation of Dynamic Load." 2018 10th International Conference on Electrical Power Drive Systems, ICEPDS, Novocherkassk, Russia, 2018. https://doi.org/10.1109/ICEPDS.2018.8571426

[2] Krause, Paul C., Oleg Wasynczuk, and Scott D. Sudhoff. Analysis of Electric Machinery and Drive Systems. IEEE, 2002.

[3] Bird, J. in Electrical and Electronic Principles and Technology 467–485 (Routledge, 2018). doi:10.4324/9781315561875-26.

[4] Nigim, K. in Handbook of Automotive Power Electronics and Motor Drives 387–404 (CRC Press, 2017). doi:10.1201/9781420028157

[5] El-Saady, Gaber, Ali M. Yousef, and Farag K. Abo-Elyousr. "Fuzzy FACTS Voltage Regulator for Isolated Wind Energy Conversion Systems under Different Wind Speed and Loading Conditions." 2016 Saudi Arabia Smart Grid Conference, SASG 2016. doi: 10.1109/SASG.2016.7849677

[6] Omran, A. S., Abbasy, N. H. & Hamdy, R. A. Enhanced performance of substation dynamics during large induction motor starting using SVC. Alexandria Engineering Journal 57, 4059–4070 (2018)

[7] K. Smitanjali et.al "Design of Digital-Static VAR Compensator for Induction Motor" Imperial Journal of Interdisciplinary Research (IJIR), Vol-3, Issue-5, 2017

[8] A. Abd Elraouf et.al, "Ant lion-Based Optimization of Facts Devices' Controllers to Enhance Three Phase Induction Motor Dynamic Performance" European Journal of Research № 4 (4), 2017

[9] A. E. ElGebaly, A. El-Wahab Hassan, M. K. El-Nemr, "Reactive Power Compensation by Multilevel Inverter STATCOM for Railways Power Grid." in Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus, Moscow, Russia, 28-31 January 2019. doi: 10.1109/EIConRus.2019.8657058

[10] D. I. Panfilov, A. E. ElGebaly, "Modified Thyristor Controlled Reactors for Static VAR Compensators" 2016 IEEE 6th International Conference on Power and Energy (PECON 2016), Melaka, Malaysia, November 2016.

[11] D. I. Panfilov, A. E. Elgebaly and M. G. Astashev, "Design and Assessment of Static VAR Compensator on Railways Power Grid Operation under Normal and Contingencies Conditions", 16th EEEIC conference, Florence, Italy, 7-10 June 2016

[12] D. I. Panfilov, A. E. ElGebaly and M. G. Astashev, "Design and evaluation of control system for static VAR compensators with thyristors switched reactors" IEEE 58th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 12-13 October 2017

[13] D. I. Panfilov, A. E. ElGebaly, M. G. Astashev and A. Rozhkov "New Approach for Thyristors Switched Capacitors Design for Static VAR Compensator Systems" 19th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices June 29 - July 3, 2018

[14] Panfilov, D. I., ElGebaly, A., Rozhkov, A. N. & Astashev, M. "Control Strategy of Thyristors Switched SVCs with High Power Quality". Transactions on Environment and Electrical Engineering 3, 15 (2018)

[15] Panfilov, D. I., Elgebaly, A. E., Astashev, M. G. & Rozhkov, A. N. "Control System Operation in Thyristors Switched SVCs with Improved Quality of Reactive Power". in Proceedings - 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I and CPS Europe 2018 (Institute of Electrical and Electronics Engineers Inc., 2018). doi:10.1109/EEEIC.2018.8493791

[16] D. I. Panfilov, A. E. ElGebaly and M. G. Astashev, "Design and Optimization of New Thyristors Controlled Reactors with Zero Harmonic Content", 18th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices June 29 - July 3, 2017

[17] A. E. Elgebaly, D. I. Panfilov and M. G. Astashev "Comparative Evaluation of Binary and Conventional Static VAR Compensators" Mechanics, Materials Science & Engineering journal (mmse), vol. 17, 2018

[18] Panfilov, D. I., Elgebaly, A. E. & Astashev, M. G. Implementation of thyristors controlled reactors for reactive power control with zero harmonics content. in 17th IEEE International Conference on Smart Technologies, EUROCON 2017 - Conference Proceedings 353–358. doi:10.1109/EUROCON.2017.8011134

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 6, 61-66 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Fast Determination of Tsunami Source Parameters

Mikhail Lavrentiev[*,1,2], Dmitry Kuzakov[1,2], Andrey Marchuk[1,3]

[1]*Institute of Automation and Electrometry SB RAS, 630090, Russia*

[2]*Novosibirsk State University, 630090, Russia*

[3]*Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 630090, Russia*

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 23 September, 2019
Online: 22 November, 2019

Keywords:
Tsunami wave source
Sea surface displacement
Orthogonal decomposition

ABSTRACT

Source parameters of tsunami waves are an essential part of any modern tsunami warning system. Recalculation of a measured time series (wave profile obtained by a seabed-based pressure sensor) in terms of initial sea surface displacement at tsunami source is among the most (or) one of the promising approaches to be applied in a warning center. The "orthogonal decomposition", that was proposed and studied earlier by the authors, is numerically studied here. Realistic shape of sea surface displacement and digital bathymetry of the southern part of Japan are used. To study functionality of the proposed approach, wave profiles are obtained by the sensors of DONET – Dense Oceanfloor Network system for Earthquakes and Tsunamis – pressure gauge network of Japan. We stress on the quality of tsunami source parameters reconstruction as well as on time required. As observed, just a part of the first wave period is enough for robust determination of such source parameters as amplitude and total volume of water. Results of numerical tests are summarized in tables and then discussed.

1. Introduction

This paper is an extension of work originally presented in OCEANS'18 MTS/IEEE Kobe/Techno-Oceans 2018, May 28-31, Kobe Japan [1].

Typical tsunami warning system includes the following parts: wave generation, propagation over the water area under study, and inundation of a dry land. Aiming to achieve real time tsunami danger prediction, we address the question of fast determination of tsunami source parameters. Having such information, there are several software packages for numerical simulation of wave propagation and inundation mapping. Here we refer the reader to MOST (Method of Splitting Tsunami – official tool of the NOAA UnS tsunami warning centers), see [2-4], and TUNAMI, see [5,6].

In any case numerical simulation of wave propagation requires the initial data, say initial sea bed displacement at tsunami source. In this paper we address the question of rather fast reconstruction of the sea surface displacement at tsunami source. As an input data we use direct measurements of the wave profile (time series), obtained at any sensor system. As is well known, a number of deep-water pressure recorders are located all around Pacific Ocean, they measure and send such data via satellite channels. Among these systems we would like to mention the USA DART (Deep-ocean Assessment and Reporting of Tsunami) buoys (https://nctr.pmel.noaa.gov/Dart/), and DONET (Dense Oceanfloor Network system for Earthquakes and Tsunamis) pressure gauge network system, which is under construction in Japan (https://www.jamstec.go.jp/donet/e/).

Our goal is to recalculate these direct measurements of the water column over a single sensor (which is nothing but a curve of the wave profile) in terms of 2D function of initial sea surface displacement at tsunami source. In general mathematical statement it is not possible to solve this problem. So, we use the so-called "Calculation in Advance" inversion strategy [7]. It has been proposed to cover the subduction zone (area of possible tsunami sources) with the system of *100x50* km rectangles (typical size of the seabed displacement zone for tsunamigenic 7.5 M earthquake). In case of Alaska-Aleutian subduction zone there are 50 of such Unit Sources (UnSs), composed in two alongshore lines. Then a typical (for this region) shape of the initial sea-bed displacement, normalized with an amplitude of *0.57* m, is used at each UnSs as tsunami wave source. Propagation of such wave from all the UnSs were numerically calculated (in advance) over the entire Pacific

[*]Mikhail Lavrentiev & mmlavrentiev@gmail.com

Ocean according to the approximation of shallow water system. As a result, a database of the calculated tsunami wave profiles, generated by the source of the same initial shape at all UnSs, is available for all grid points of the Pacific Ocean.

Now, having the measured (real) wave profile at certain sensor location, one can approximate this curve with the help of linear combination of calculated wave profiles from a number of UnSs, picking up at the same point of sensor location. In other words, the problem is to determine the amplification coefficients for the UnSs such that the corresponding linear combination of UnSs (say, a Composed Source – CS) provides a good approximation of the initial sea-bed displacement.

The similar approach named SIFT – Short-term Inundation Forecasting for Tsunamis (https://nctr.pmel.noaa.gov/Pdf/SIFT_3_2_2_Overview.pdf) is used by NOAA (USA) for wave height assessment in grid points of computational domain. As well as in the method presented here, on the basis of "computations in advance" of the tsunami height distribution from the set of Unit Sources the authors [7] reconstruct the CS. Unlike the method presented here, by search of set of coefficients in linear combination of UnSs, the SIFT system uses the combinatory search of values of coefficients giving a discrepancy minimum between real and calculated tsunami wave series in several tsunami detectors. Such approach requires a lot of processor time if CS consists from more, than 5 UnSs. There is no such a shortcoming of the approach here proposed, where CS restoration time practically does not depend on quantity of UnSs involved. By the way, authors of this article also took part in development of an algorithm for source inversion in the SIFT forecasting system.

This approach has been many times tested against the field observations after near field and trans-oceanic events. Amazing agreement has been observed by comparison of computed wave, generated by a calculated CS, and measurements taken at various gauge stations. Nowadays, all subduction zones around Pacific, Atlantic, and Indian Oceans were covered with such systems of UnSs unit sources (https://nctr.pmel.noaa.gov/propagation-database.html); synthetic marigrams at every grid point from every source have been saved in a specialized database. One of the problems is to choose the aforementioned amplification coefficients in an automated mode. The orthogonal decomposition approach to handle this was proposed in [8] and then tested in [1,9]. It happens that even a part of the first wave period is enough to determine the CS – the set of amplification coefficients. In this paper we continue to study the problem of robust determination of sea surface displacement at tsunami source by a part of the wave profile.

The paper is arranged as follows. We first describe the proposed orthogonal decomposition algorithm and provide the mathematical ground behind it. In the numerical section we briefly describe digital bathymetry in use and the parameters of realistic CS, which has a feature of the reconstructed source of historical event of 1707 [10]. Then we provide the obtained numerical results in the form of a table. Finally, numerical results are discussed.

2. Orthogonal decomposition algorithm

2.1. Mathematical ground of algorithm

Here we provide the brief explanation of the algorithm for reconstruction of the CS. More detailed description one can find in [8].

Let *f(t)* be the marigram – time series, obtained at any particular sensor (DART buoy, bottom cable sensor of DONET, or other). We will use notations $f_k(t), (k = 1, \dots, N)$ for simulated wave profiles, calculated (according to the linear or nonlinear shallow water approximation) at the same point of sensor location, by direct numerical modeling of the tsunami wave propagation. It is understood, that the subindex *k* refers to a tsunami wave initiated by the normalized sea surface initial profile of a given shape from the *k*-th UnSs. We assume that all these functions, $f_k(t)$, are linearly independent. We are looking for coefficients, b_k, for the linear combination of these functions, which provide the best approximation of the function *f(t)* in L_2 norm:

$$\int_0^T \left(\sum_{k=1}^N b_k f_k(t) - f(t)\right)^2 dt \ \rightarrow min \qquad (1)$$

So, we consider the problem of optimal approximation of a given function, *f(t),* by the linear combination of the finite subset of functions from the system $\{f_k(t)\}$. Suppose that the system is orthogonal and normalized according to:

$$\left(f_i(t), f_j(t)\right) = \int_{t_0}^T f_i(t) f_j(t) dt = 0, \text{ if } i \neq j \qquad (2)$$

$$\left(f_i(t), f_i(t)\right) = 1. \qquad (3)$$

Here t_0 stands for the time moment when the wave approaches the closest receiver, *T* is the time of the latest (of all receivers) full period on marigram which is concerned as the end of the second phase of tsunami. For sensors which are more distant off the coast than a source, the second phase of tsunami is positive, and for those which are closer to the shore, it is negative.

As is well known in the Fourier series theory, the coefficients of such optimal approximation are nothing but the Fourier coefficients of expansion of *f(t)* in a series with respect to $\{f_k(t)\}$ (see, for example, [11]).

2.2 Algorithm description

Based on the statement above, the following algorithm was proposed in [8] and tested in [1,9] for searching the coefficients of the linear combination allowing optimal approximation of the given function. It includes three steps.

(a) The calculated set of "marigrams", $\{f_k(t)\}$, from the unit sources should be orthogonalized and normalized, according to equations (2)-(3).
(b) The measured marigram, *f(t),* should be expanded to the Fourier series with respect to obtained orthonormal system of functions.
(c) The so-determined Fourier coefficients should be recalculated in terms of initial functions $\{f_k(t)\}$.

Let us transform the given system $\{\mathrm{f_k}\}$ to the corresponding ortho-normal system, $\{e_k\}$, with respect to the scalar product (2). Let

$$\mathrm{e_1} = \mathrm{f_1} \qquad (4)$$

Suppose that pairwise orthogonal nontrivial functions $\mathrm{e_1}, \mathrm{e_2}, .., \mathrm{e_{k-1}}$ are already constructed. We will look for the function $\mathrm{e_k}$ in the form:

$$\mathrm{e_k} = \mathrm{f_k} + \mathrm{l_{k,1}e_1} + \dots + \mathrm{l_{k,k-1}e_{k-1}} \qquad (5)$$

i.e. we obtain the next function, e_k, as the linear combination of the new function, f_k, and the already constructed ortho-normal ones, $e_1, e_2, .., e_{k-1}$. The coefficients, $l_{k,1}, l_{k,2}, .., l_{k,k-1}$, in formula (5) should be found from the conditions that the new function, $e_k = f_k + l_{k,1}e_1 + .. + l_{k,k-1}e_{k-1}$, is orthogonal with all the previous ones, $e_1, e_2, .., e_{k-1}$, and normalized. Thus, the coefficients in (5) are calculated according to:

$$l_{k,i} = -\frac{(f_k, e_i)}{(e_i, e_i)}, \ (i = 1, \cdots, k-1) \qquad (6)$$

As is well known in calculus, the optimal L_2 approximation, according to criteria (1), of a given function f by the linear combination of orthogonal and normalized functions, e_k

$$f_appr = \sum_{i=1}^{k} a_i e_i \qquad (7)$$

is achieved at the Fourier coefficients, a_i, of the function *f(t)*, decomposition with respect to the system e_k:

$$a_i = \int_{t_0}^{T} f(t)e_i(t)dt \Big/ \int_{t_o}^{T} e_i^2(t)dt \qquad (8)$$

So, to solve the original problem (1), we only need to provide the coefficients (8) in terms of the original basis functions f_k.

It is just an algebraic exercise to conclude that the desired coefficients, b_k, in expression (1), are calculated through the given coefficients, a_k, from the formula (8), as follows:

$$\begin{cases} b_k = a_k, \\ b_{k-1} = a_{k-1} + a_k F(L', k, k-1) \\ b_{k-2} = a_{k-2} + a_{k-1}F(L', k-1, k-2) + a_k F(L', k, k-2) \\ b_i = a_i + \sum_{j=i+1}^{k} a_j F(L', j, i) \end{cases}$$

$$F(L', j, i) = \begin{cases} 0, \ \text{if } j = i, \\ l_{j,i}, \ \text{if } j = i+1, \\ l_{j,i} + \sum_{k=i+1}^{j-1} l_{j,k} F(L', k, i), \ \text{if } j \neq i, \ j \neq i+1. \end{cases}$$

So, the proposed algorithm to approximate the given marigram, *f(t),* by the linear combination of synthetic marigrams, $\{f_k(t)\}$, *k=1,...,N*, in order to achieve minimal L_2 difference (1), is as follows.

We first transform the system $\{f_k(t)\}$ to the orthogonal system, $e_k(t)$, and normalize it according to formulae $(4) - (8)$.

Then the recorded marigram, f(t), is decomposed to the part of Fourier series with respect to obtained orthogonal and normalized (see (1), (2)) basis, $e_k(t)$, by the formula (8).

Finally, the obtained Fourier coefficients, a_i, are recalculated in terms of coefficients, b_i, of the original linear combination (1), by means of (9).

3. Numerical experiments

Numerical experiments were arranged in the area around the south-east coast of Japan (eastern Kyushu, Shikoku islands and Kii peninsula. Geography, bathymetry and location of UnSs and CSs in this region are shown in Figure 1.

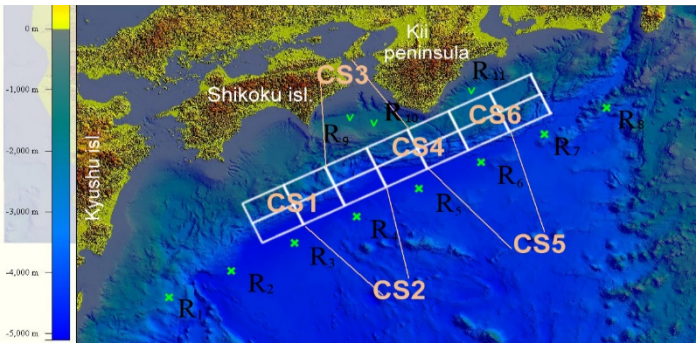


Fig. 1. Digital bathymetry of the water area under study. White rectangles show positions of model tsunami sources. Green crosses (R1-R8) indicate locations of artificial sensors. Green check marks (R9-R11) show the selected sensors of DONET.

3.1. Digital bathymetry

Numerical experiments were arranged at the gridded bathymetry around Eastern Kyushu, Shikoku Island and Kii Peninsula (southern part of Japan), built on the basis of Japanese bathymetric data produced by the Japan Oceanographic Data Center (JODC), presented in Figure 1. This digital gridded bathymetry has the following characteristics:

Size of array is 3000 x 2496 nodes;

Grid steps are 0.003 and 0.002 degrees (which means 280.6 and 223 meters, respectively);

Domain covers the area approximately between 131° E and 140° E, 30° N and 35° N.

3.2. Composition of tsunami source

In the course of numerical calculations model tsunami sources were formed of UnSs of *100x50* km in size located in two rows along the coast (Figure 1). Each of the six composed sources (CS), having the sizes of 200x100 km, is presented as the linear combination of four UnSs with given amplification coefficients *(1.4; 1; 1.5; 1.1).* For this set of coefficients, the maximum surface elevation in the source area is equal approximately to 3 m, and minimum water level there is estimated as -1.5 m (Figure 2). Bigger coefficients correspond to UnSs more distantly located off the coast. Numbering of CSs begins with UnSs which are at the western edge the source area (Figure 1). Two of them (the first and the last) are indicated in Figure 1. The ocean bottom displacement in the CS used for computations is shown in Figure 2.

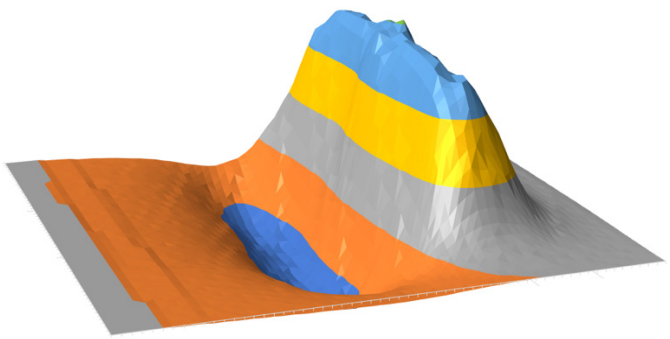


Fig. 2. Shape of the initial surface displacement at the Composite Source used for numetical experiments.

We assume precision of 10% as the value of reasonable accuracy of the tsunami height estimation. This value allows to determine parameters of the wave near the shore with the same precision, which is enough for the rapid tsunami hazard evaluation. It means that maximum difference between the initial and calculated surface displacement is less than 10 cm for a typical initial displacement height of 1 m. Accordingly, the deviation of 1 m is considered acceptable in case the wave height is 10 m.

As a rule, a part of the wave profile from the beginning to the zero level after the highest amplitude point (see Figure 2) is sufficient to reconstruct the amplification coefficients with a particular precision. So, we tried to determine how smaller could be a part of wave profile to keep the error in reconstructed ISSD amplitude within 0.1 discretization. The location of the model tsunami source centers for numerical calculations are chosen according to estimation of the source of the major tsunami, expected, according to [10], in the near future in the Nankai subduction zone (Figure 3 is taken from [10]). The size of the model sources corresponds to a tsunamigenic earthquake of magnitude 8.0.

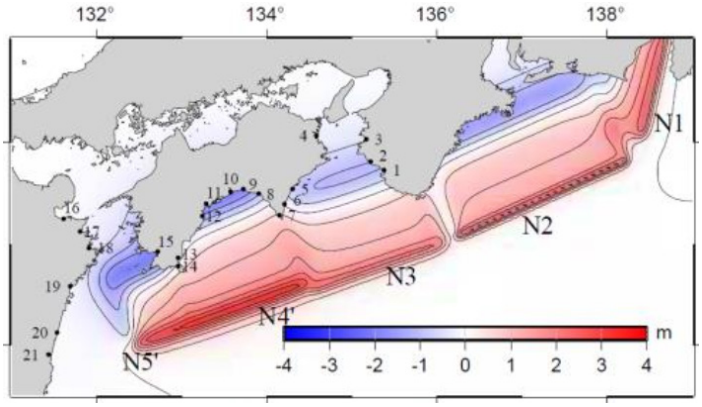


Fig. 3. The expected distribution of the initial water surface displacement in the tsunami source at the Nankai subduction zone, [10].

3.3. Positions of sources and sensor network

The virtual sensors network consists of one row of virtual deep-water wave detectors, installed along the Nankai Trough (R_1-R_8, indicated by green crosses in Figure 1). In addition, 3 sensors of DONET were also included to observation network system for source restoration in our numerical experiments. These sensors are indicated in Figure 4 as black squares, R_9, R_{10} and R_{11}. Reconstruction of a source means determination of coefficients in the linear sum of UnSs composing CS.

3.4. Numerical experiments

As a result of numerical experiments, the tsunami wave signal from each of the six locations of CS (see Figure 1) was recorded by 11 detectors, R1-R11, given in Figures 1 and 4. By means of the described method, the minimum length of a wave signal, sufficient for determination of coefficients in the linear sum of UnSs (Necessary Part of Wave Profile – NPWP) have been determined with the required accuracy. These lengths as a percentage to length of the full wave period which means complete first two phases (positive and negative) of a wave are summarized in Table 1. The first column shows receiver's number (1-11). In the second column one can find the Composed Source number (1-6). In the third and the fourth columns the length in seconds of NPWP and the full wave period are presented, respectively. Column number 5 gives the length of NPWP in percentage of the full tsunami period. The last two columns present the NPWP ending time instance and the wave arrival time to a shore.

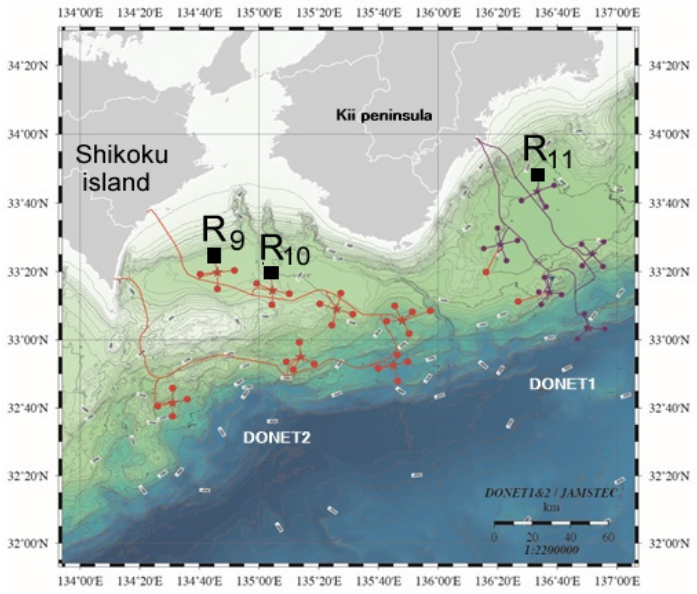


Fig. 4. Configuration of DONET-1 and DONET-2 system at the south-east of Japan (https://www.jamstec.go.jp/donet/e/). The existing sensors are in violet and sensors under construction are indicated in red. Black squares show three sensors, which positions were used for artificial sensors.

The duration of NPWP counted from the moment of wave arrival at receiver. More important for us is the end of NPWP on the record (column 6 in the Table 1). This record starts simultaneously with the initial ocean surface displacement. It is the same moment as regarded as zero in tsunami travel time. So, if NPWP (on record) is less than the tsunami travel time to the coast, then we have some time for reconstruction of CS and for estimation of the expected wave height near a shore.

Table 1: Numerical tests Parts of the entire wave profile (in percents), enough to determine the source parameters

Receiver	CS	NPWP, sec	Period, sec	NPWP, % of full wave period	Time required for accuracy to be achieved (from the generation moment), sec	Wave arrival time to a shore, sec
1	1	1511	1736	87.0	1511	837
2	1	224	1561	14.3	224	837
3	1	730	2553	28.6	998	837
4	1	442	1422	31.1	1731	837
5	1	419	1262	33.2	2313	837
6	1	1365	1608	84.9	3575	837
7	1	121	3867	3.1	2565	837
8	1	141	2949	4.8	3178	837
9	1	530	1257	42.2	698	837
10	1	496	1260	39.4	519	837
11	1	223	981	22.7	223	837

1	2	409	1311	31.2	409	876
2	2	223	1070	20.8	223	876
3	2	1724	2614	66.0	1724	876
4	2	474	2640	18.0	1327	876
5	2	679	1056	64.3	2057	876
6	2	542	2509	21.6	2469	876
7	2	247	3837	6.4	2369	876
8	2	328	2875	11.4	3040	876
9	2	219	1116	19.6	223	876
10	2	261	1031	25.3	261	876
11	2	293	1080	27.1	293	876
1	3	480	1359	35.3	480	525
2	3	223	1149	19.4	223	525
3	3	2043	2489	82.1	2043	525
4	3	728	1803	40.4	1012	525
5	3	922	1859	49.6	1786	525
6	3	1226	1594	76.9	2617	525
7	3	632	3844	16.4	2450	525
8	3	428	2769	15.5	2827	525
9	3	221	1003	22.0	223	525
10	3	311	904	34.4	311	525
11	3	380	1282	29.6	475	525
1	4	811	1322	61.3	1075	465
2	4	223	1134	19.7	223	465
3	4	419	2479	16,9	419	465
4	4	698	3203	21.8	698	465
5	4	868	1735	50.0	1183	465
6	4	1595	2767	57.6	2489	465
7	4	340	1503	22.6	1776	465
8	4	377	2808	13.4	2483	465
9	4	221	1359	16.3	223	465
10	4	223	1274	17.5	223	465
11	4	503	1482	33.9	928	465
1	5	596	1572	37.9	1405	510
2	5	988	2057	48.0	1090	510
3	5	1442	2585	55.8	1442	510
4	5	1450	1470	98.6	1450	510
5	5	1065	1882	56.6	1065	510
6	5	925	2183	42.4	1215	510
7	5	406	1895	21.4	1411	510
8	5	447	3062	14.6	2080	510
9	5	294	1484	19.8	301	510
10	5	350	1437	24.4	403	510
11	5	598	1746	34.2	1280	510
1	6	515	1138	45.3	1690	630
2	6	735	1201	61.2	1355	630
3	6	825	2457	33.6	1243	630
4	6	223	1497	14.9	223	630
5	6	568	1552	36.6	568	630
6	6	654	2124	30.8	654	630
7	6	630	2226	28.3	1093	630
8	6	460	3260	14.1	1625	630
9	6	389	1418	27.4	662	630
10	6	476	1361	35.0	851	630
11	6	539	1424	37.9	1523	630

Numerical experiments showed that in most cases use of even a wave record only from one sensor allows to restore the source (coefficients in the linear sum of UnSs) with the required accuracy, using only a part of a wave signal (less than a half of its full period). From the results presented in Table 1 it is visible that for each CS there are several sensors, processing of a wave signal from which allows to restore the source with the required accuracy even before arrival of a wave to the nearest coast. In particular, for CS1 those are detectors R1, R2, R9, R10 and R11. For CS2 those will be sensors R1, R2, R3, R9, R10 and R11. And so on.

4. Discussion

The fact that among such sensors there is always a tsunami detector which is included to observation system DONET (Figures 1 and 4) is important. In these sensors (R9-R11) the first phase of tsunami on the wave record is negative (surface inundation), and then the surface elevation is coming. Finally, we can conclude that with the help of observation network described the tsunami warning system is able to estimate expected wave heights along the coast before its arrival there.

By results of the numerical experiments given in table 1 one can notice that the length of NPWP for restoration of some CSs by records of some sensors exceed 80% of the full period of a wave. Though, in most cases duration NPWP does not exceed 50 percent of the tsunami period. In particular, more than 80 percent of the full wave period is required for sufficient-quality restoration of CS having number 1 by means of sensors R1 and R6. The CS3 by means of sensors R3 and R6. And, at last, CS5 on the basis of measurements by sensor R4. For the last case the length of NWDP reaches 98.6 percent of wavelength. From Figure 1 it is visible that in all these cases the sensor is located in the direction of a long axis of CS (a rectangle 200 x 100 km) or near it. It means that such relative positioning of the CS and the tsunami detector negatively affects speed and quality of restoration of coefficients in a linear combination of UnSs composing the source. CSs are most successfully restored using wave records taken from the detectors which are located in the direction of a short axis of the source.

5. Conclusion

The numerical method which was used for calculating the direct initial value problem on propagation of tsunami generated by a UnSs have been many times tested on exact solutions and on real tsunami data. So, correlation between initial surface displacement at the UnSs and synthetic marigrams in receivers seems to be reliable. According to the mathematical model the increase in height of Unit Source by N times causes the proportional growth of wave height in all grid-points of computational domain.

As a result of this research it is shown that at the reasonable relative location of the area of the possible tsunami sources and deep-water tsunami detectors for correct restoration of the composed source it is enough only a part of the recorded wave signal. At the same time, a part of a wave signal, necessary for restoration, can be used by the restoration algorithm even before arrival of a wave to the coast. It gives the chance for quick solution of a direct problem of the tsunami propagation from the initial water surface displacement to estimate the expected wave height along the entire coast. Besides the carried-out numerical

experiments will allow find optimal positions for installation of the new deep-ocean tsunami detectors capable to restore the composite tsunami sources more quickly.

Acknowledgment

The research was supported by the Russian Federal Targeted Program Grant 14.574.21.0145. Unique project ID - RFMEFI57417X0145.

References

[1] D. Kuzakov, M. Lavrentiev, An. Marchuk, “Determination of Initial Sea Surface Displacement at Tsunami Source by a Part of Wave Profile” in Proc. OCEANS'18 MTS/IEEE Kobe, Japan, May 28-31, 2018.

[2] V. Titov, “Numerical modeling of tsunami propagation by using variable grid” in Proc. IUGG/IOC International Tsunami Symposium, 46-51, 1989.

[3] V. Titov, C. Synolakis, “Numerical modeling of tidal wave runup” J. Waterway, Port, Coastal and Ocean Engineering, **124**(4), 157-171, 1998.

[4] J. Borrero, S. Cho, J. Moore, H. Richardson, C. Synolakis, “Could it happen here” Civil Engineering, **75**(4), 55-67, 2005.

[5] N. Shuto, C. Goto, F. Imamura, “Numerical simulation as a means of warning for near field tsunamis” Coastal Engineering in Japan, **33**(2), 173-193, 1990.

[6] A. Yalciner, B. Alpar, Y. Altinok, I. Ozbay, F. Imamura, "Tsunamis in the Sea of Marmara: Historical Documents for the Past, Models for Future" Marine Geology, **190**, 445-463, 2002.

[7] E. Gica, M. Spillane, V. Titov, C. Chamberlin, J. Newman, “Development of the Forecast Propagation Database for NOAA's Short-Term Inundation Forecast for Tsunamis (SIFT)”, NOAA Technical Memorandum, 2008. http://www.ndbc.noaa.gov/dart.shtml

[8] A. Romanenko, P. Tatarintsev, “Algorithm for reconstruction of the initial surface disturbance at the tsunami epicenter”, NSU J. of Information Technologies, **11**(1), 113-123, 2013.

[9] M. Lavrentiev, D. Kuzakov, A. Romanenko, A. Vazhenin, “Determination of Initial Tsunami Wave Shape at Sea Surface”, OCEANS 2017-Aberdeen, IEEE, 1-7, 2017.

[10] M. Hyodo, T. Hori, K. Ando, T. Baba, “The possibility of deeper or shallower extent of the source area of Nankai Trough earthquakes based on the 1707 Hoei tsunami heights along the Pacific and Seto Inland Sea coasts, southwest Japan”, Earth, Planets and Space, **66**(123), 2014. http://www.earth-planets-space.com/content/66/1/123.

[11] D.R. Kincaid, E.W. Cheney, Numerical analysis: mathematics of scientific computing, **2**, American Mathematical Soc., 2002.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 61-68 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Case Study

Renato Mio[1,*], Midori Sanchez[2], Quino Valverde[3], José Lara[4], Francisco Rumiche[4]

[1] *Laboratory of Biomechanics and Applied Robotics, Pontifical Catholic University of Peru, 15088, Peru*

[2] *Digital Manufacturing Laboratory - VEO 3D, Pontifical Catholic University of Peru, 15088, Peru*

[3] *INACOM Group / Aula PUCP-CIMNE, Pontifical Catholic University of Peru, 15088, Peru*

[4] *Mechanical Engineering Section, Engineering Department, Pontifical Catholic University of Peru, 15088, Peru*

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 18 August, 2019
Online: 04 September, 2019

Keywords:
Upper-limb prosthetics
Material testing
3d-printing
Benchmark testing
Body-powered prostheses

ABSTRACT

3D-printing technologies have greatly influenced the field of fabrication of medical devices. In particular, Fused Deposition Modeling 3D printing has emerged as one the most popular and most promising technologies for fabricating upper-limb prostheses. Over the last years, a variety of types and designs of 3D-printed hand prostheses have been created and are commercially available. However, there are no standards or established procedures for testing these devices. Available information regarding their long-term performance and functionality is very limited. This paper presents a case study of mechanical testing methods applied to a specific design of an upper-limb prosthesis. The device and its subassemblies were subjected to flexion test in hyperextension and abduction conditions, fatigue/wear test, and tensile test. The experimental results are presented and examined. Testing procedures, adaptations and recommendations are described and discussed to demonstrate ways of generating reliable data that serve for comparison among different hand prostheses designs.

1. Introduction

The fabrication of medical devices is one of the most influenced fields by 3D printing technologies advances. Particularly, in the last years, 3D printed hand prostheses have emerged as a promising low-cost alternative, compared to conventional hand prostheses, to fulfill the need for these devices in developing countries [1]. Currently, Fused Deposition Modeling (FDM) is one of the most common and inexpensive technologies for fabricating functional end-use parts by 3D printing, including hand prostheses. However, FDM has a disadvantage related to the anisotropy generated by the process and its corresponding effect on mechanical properties. A material fabricated by 3D printing shows a ductile behavior in the orientation of the bead deposition, but a brittle one in the perpendicular orientation. Therefore, FDM printed components are susceptible to fracture in the building orientation; a failure mechanism called delamination. While the effect on mechanical properties of many fabrication parameters such as layer height, infill pattern, density, nozzle diameter, build plate temperature, printing orientation and velocity [2, 3, 4, 5, 6, 7] has been evaluated using 3D-printed test specimens, little advances have been made in testing and characterizing entire components fabricated using FDM 3D-printing.

Although the popularity and the use of 3D printed hands keep increasing, there is a lack of standards for fabrication and testing of these devices and very limited information available in terms of their long-term performance and functionality. Existing hand prostheses standards and testing procedures are not directly applicable, even for currently commercial multi-fingered and dexterous hand prostheses, because they were conceived mainly for conventional hook prostheses. Moreover, hand prostheses manufacturers use their design-specific tests to provide full specifications of their commercial products.

In this context, different researchers have explored the possibility of adapting existing methods and implementing variations, according to their investigation's objectives and available resources, to evaluate the performance of open-source and specific models of 3D printed upper-limb prostheses. Tests have been conducted to evaluate mechanical properties such as tensile strength, impact resistance, flexural strength, and fatigue

* Renato Mio & Email: r.mioz@pucp.edu.pe

resistance, both in the whole prostheses and sub-assemblies, with parameters (test velocity, applied load, number of cycles, etc.) that resemble conditions of activities of daily life (ADL). A review of these tests has been presented by the authors in the conference BIBE 2018 [8] and is summarized as follows.

Tenim [9] and Koprnicky [10] analyzed the mechanical behavior of phalanges of prosthetic hands employing hyperextension (see Figure 1a) and lateral bending test and cyclic flexion-extension test (see Figure 1b). The bending tests were intended to evaluate the maximum load that a phalange can withstand and extrapolate the results to compute the maximum load that can be carried by all of the fingers of a prosthetic hand. These bending tests have been also applied to individual fingers from robotic grippers [11]. Furthermore, the purpose of the cyclic flexion-extension test was to measure the endurance of a finger under the application of several fatigue cycles and to determine the mode of failure under this condition (e.g. damage of one of the finger components or wear of the flexion cables channels). In addition, Crisco and Wolfe [12] analyzed the finger biomechanics and proposed testing procedures to mimic the conditions to which a prosthesis is subjected to when conducting activities of daily living. While the ASTM F1781 standard [13] provides some guidance for finger fatigue test, Joyce and Unsworth [14] have proposed a set of parameters to be applied, such as the flexion-extension load magnitude: 10-15 N, pinch load: 100 N, the angular amplitude of the motion: 0-90 degrees, and the frequency: 1.5 Hz.

Among the tests to which prosthetic hand subassemblies could be subjected to, the ISO 22523 standard [15], annex A, proposes the distal tensile test (see Figure 1c). In this test, a hand subassembly is induced to perform a power grip on a 19 mm diameter cylindrical object. As schematically observed in Figure 1c, the object is pulled away, forcing the fingers to extend or to break. In this way, it is possible to establish the maximum load that the prosthesis can withstand without failure. Additional tests involving grasping movements with a cylindrical object have been proposed in [16, 17, 18]. Cylindrical grip test is of key importance because it mimics the movement conditions of most of the body-powered hand prostheses, i.e. opening and closing of all of the fingers at the same time. Furthermore, the ISO 22523 standard describes the procedures to conduct hand or wrist flexion tests in the case of transradial prostheses.

This paper presents a case study of mechanical testing methods applied to an upper-limb prosthesis. The study is focused on the 3D-printed D1M Hand (acronym in English for Dando Una Mano, *Giving a Hand*) [19] (see Figure 2). The D1M hand was designed for transmetacarpal or partial hand amputations and was fabricated by the FDM 3D printing process. Due to its design and functionality the D1M hand was subjected to finger flexion test (in the hyperextension and abduction/adduction directions) to measure individual finger maximum load, finger cyclic flexion-extension to measure individual finger durability, and hand tensile test to measure the maximum possible load that the hand can withstand while performing a cylindrical grasp. Schematic drawings and a description (references, conditions, outcomes, and objectives) of these tests are shown in Figure 1 and Table 1 respectively.

This work complements the one presented by the authors in BIBE 2018 [8]. This manuscript includes a summary of the design and fabrication process of the D1M hand, the experimental details and testing procedures description; with particular emphasis placed on the machine adaptations for positioning and supporting the tested hand and finger samples, and the discussion of obtained results.

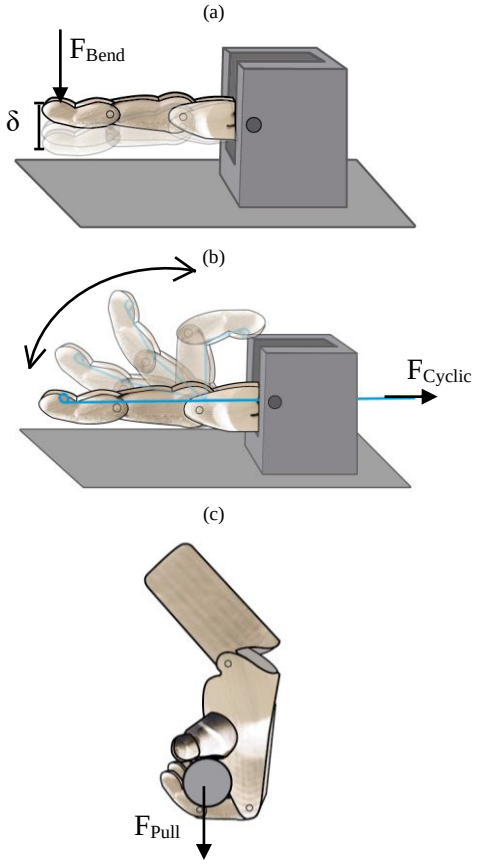


Figure 1. Schematic drawings of the different mechanical tests found in the literature. (a) Finger flexion test, (b) endurance/fatigue test and (c) hand tensile test. Adapted from [8].

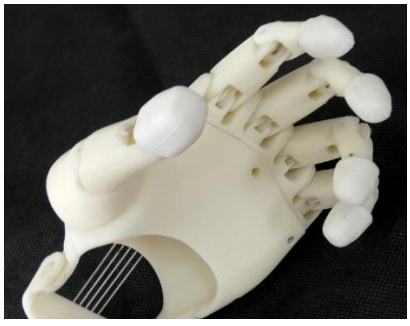


Figure 2: D1M hand prosthesis prototype. Image from [19].

2. Experimental Procedure

The mechanical tests were performed using universal testing machines. The D1M hand was fabricated using a commercially available desktop 3D printer. The use of commercially accessible equipment will allow the experimental procedures to be replicated and the results to be compared with those obtained by other researchers. The assemblies and accessories presented in this work can be modeled in CAD software and then minor modifications can be made to accommodate different designs. Validation of the printing parameters and methodology was performed with detailed testing applied to the D1M hand.

Table 1: Description of the mechanical resistance tests applied to the D1M Hand (adapted from [8]).

Type of test	*Test description*			
	Based on (guide or standard)	***Test conditions***	***Recorded outcomes***	***Objective***
Finger flexion	ISO 178, ASTM 790 and the works of Tenim [9] and Koprnicky et al. [10]	Velocity of the test: 1% the gage length per minute. Finalize when: Maximum deflection is 5% or failure. Different tests: Hyper-extensive and lateral loading.	Displacement and force against time.	Estimate the mechanical strength of the fingers to withstand forces that over-extend (hyper-extensive loading) or bend the fingers laterally (lateral loading). The target load value is 30 N according to [20]
Finger cyclic flexion-extension	ISO 14243-1 and the works of Tenim [9] and Joyce et al. [14].	Cyclic loading frequency: 1.5-2 Hz. Finalize when: achieve 1,200,000 cycles or failure of whichever component. Load at the MCP joint: 10 - 15 N. Pinch force: 100 N. Finger base joint range of motion: 0 - 90 degrees.	Cycles can achieve the prosthesis until the failure. Wear of the components measured by weight difference before and after the test.	Calculate the number of cycles the prosthetic finger can withstand. The target value is 1,200,000 cycles (equivalent to 4 years of use) according to [21]
Hand tensile test	ISO 22523-2006, annex A.8.2 [15]	The prosthesis grasps a 100mm x diameter 19 mm cylinder. An external force pulls the cylinder out of the hand. Preload: 10 N Load: increases at a rate of 1 – 10 N/s until failure.	Displacement and force against time.	Estimate the maximum payload of the whole prosthesis when performing a cylindrical grasp.

2.1. Fabrication of the D1M hand prosthesis

The D1M hand was fabricated by FDM 3D printing with an Ultimaker 3+ Extended machine; the material employed was acrylonitrile butadiene styrene (ABS). Table 2 shows the general 3D printing process fabrication parameters that were applied to all of the main components (fingers, palm and forearm).

The prosthesis was designed based on the anthropomorphically structure of the user and the functionality of the 4 linkages mechanism [19]. ABS is an optimal polymer for this use because of its strength and low density. As mentioned previously, the mechanical resistance of the 3D printed hand depends on the manufacturing orientation, caused by the intrinsic anisotropy of the manufacturing and its deposition of consecutive layers. For this reason, optimal printing orientation for each component was chosen based on factors such as maximizing the area of the critical cross-section under stress (to prevent delamination) and attaining the best surface quality with no support material for important geometrical features (like shaft supports). Table 3 and Figure 3 shows the orientation of each component of the hand and its printing parameters. In addition, it was observed that for the palm less infill density is placed than in the fingers, because in the continuous use of the prosthesis it was identified that the fingers were critical pieces, due to the small thickness in the lace. Many other existing hand prostheses designs also come with suggested infill percentages, which adds another variable influencing the resulting mechanical properties of the resulting device. With the purpose of finding the exact technical specifications of the prosthesis, mechanical tests were performed based on the experimental parameters found in the literature.

2.2. Experimental Procedure on D1M hand

Three different tests were conducted in the D1M hand: first the hand tensile test, chosen because the main actions performed by upper limb prostheses require a cylindrical grip; then, finger flexion tests were performed in order to know the maximum weight that an individual finger can resist; and finally, the finger fatigue / wear test to estimate the prosthesis' durability and the wear due to its use. These tests required the design of custom fixation components to be placed between the hand components and the testing machine.

Table 2: 3D Printing parameters for the components of the D1M Hand.

Parameter	**Value**
Nozzle Temperature	230 °C
Build plate Temperature	100 °C
Printing speed	65 mm/s
Layer height (Z-axis resolution)	0.2 mm
Wall thickness	1.2 mm
Infill pattern	Linear

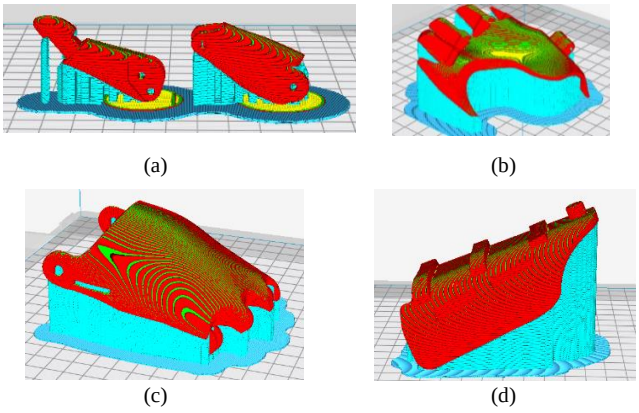
Table 3: 3D Printing Parameters specific to each component.

Piece	**Layer thickness [mm]**	**Infill [%]**	**Angle orientation**	**Figure**
Finger	0.1	50	17° and 25°	Figure 3a
Palm (front side)	0.2	30	28°	Figure 3b
Palm (backside)	0.2	30	15°	Figure 3c
Forearm	0.2	50	32°	Figure 3d

2.2.1. Hand tensile test

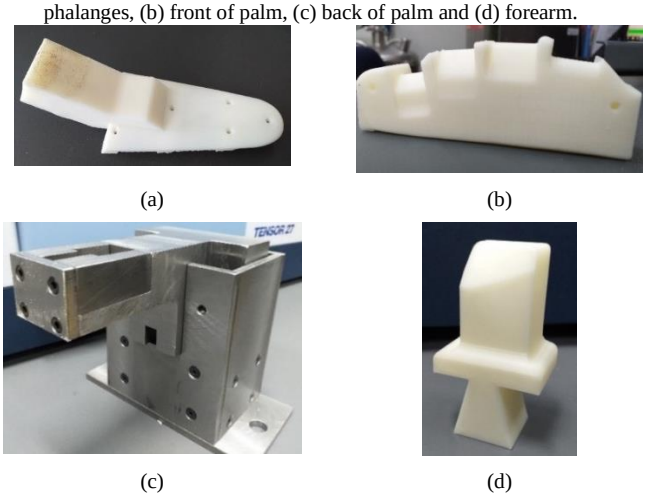
The tensile test was performed in the whole hand assembly to evaluate the maximum load that the prosthetic hand can withstand when it is performing a cylindrical grasp movement. The test was conducted using a Zwick Roell Z050 tensile test machine with the test parameters defined by ISO 22523:2006 [15]. Three whole hand assemblies were tested. As shown in Figure 4a, a flat area accessory to connect the machine jaws with the hand was fabricated, it also served to keep the assembly aligned with the axis on which the machine exerts force. This accessory was printed in ABS and 100% filling density was used. Additionally, a grasping device (see Figure 4b) was fabricated to allow the prosthesis to make a cylindrical grip. On the other hand, a support device was added to maintain the wrist flexion needed to close the fingers around the grasping device. To ensure that the test can reproduce a real-use condition, the prosthesis must be tested while

activated with its fingers closed. For this reason, a pair of accessories were designed to keep the prosthesis locked as if it were activated by a user exerting force with the stump. A steel accessory (see Figure 4c) was manufactured to support the hand and withstand relatively high loads without deforming. In addition, an accessory acting as a support point of the inner part of the palm was printed (see Figure 4d). Figure 5 shows the complete assembly to perform the testing.



(a) (b) (c) (d)

Figure 3. Orientation of the hand prosthesis' main components: (a) finger phalanges, (b) front of palm, (c) back of palm and (d) forearm.



(a) (b) (c) (d)

Figure 4. Accessories for the optimal fix between the prosthesis and the testing machine.

The test is performed until the failure of any component (or until the prosthesis stops holding the grip due to the deformations that occur in the threads that cross the phalanges).

2.2.2. *Finger Flexion test*

Flexion tests were conducted in the fingers to evaluate their maximum load-carrying capacity. The tests were carried out simulating two conditions: a finger extended in a horizontal plane (hyperextension test), and a finger extended in a vertical plane (abduction test). The tests were performed using a Zwick Roell Z0.5 multi-test machine. Three samples of middle and little fingers (longest and shortest finger of a hand, respectively) were subjected to each condition of the test. The test speed was proportional to the length of the finger, which in turn depended on the test condition. For the hyperextension condition, the test speed was 6.6 mm/min and 4.6 mm/min for the middle and little finger, correspondingly. For the abduction condition, the speed was 7.3 mm/min and 5 mm/min for the middle and little finger respectively. Three accessories were needed for this test: the general coupling, the secondary coupling and the bending punch. Figure 8a shows the assembly designed in the CAD software. The general coupling (see Figure 6) was fixed on the machine table and housed the secondary coupling and the finger. The secondary coupling (see Figure 7) varied depending on the size of the finger and the type of test (hyperflexion or abduction). The bending punch was fabricated from a sheet of steel and had a U shape in order to avoid the slip with the 3D printed finger, as shown in Figure 8b.

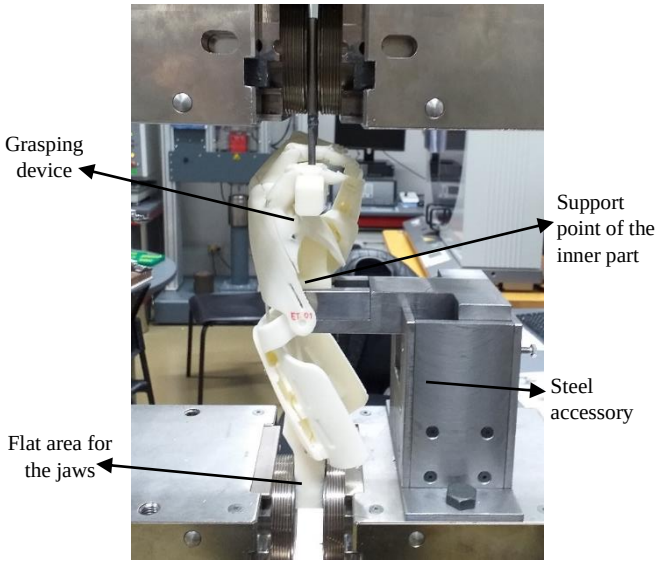


Figure 5. The complete assembly of the hand tensile test. Arrow indications correspond to the accessories shown in Figure 4.

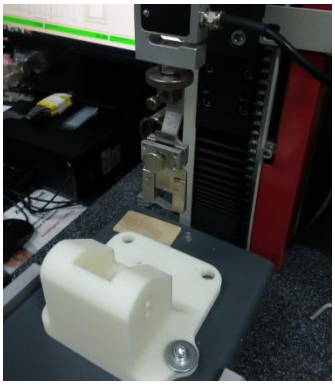


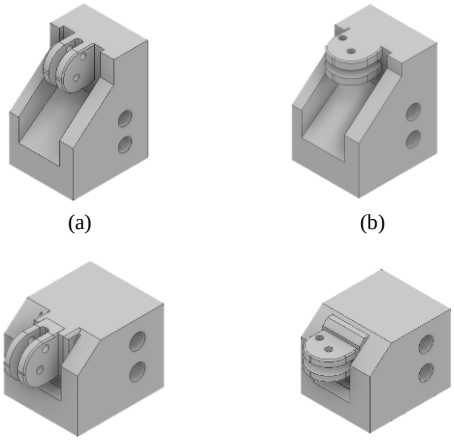
Figure 6. 3D printed coupling made of white ABS and punch that exerts the bending force.

2.2.3. *Finger fatigue/wear test*

The fatigue/wear test was performed to evaluate the capacity of the index finger assembly (3D printed phalanx, wire and pins) to perform a cyclic pincer grasp movement and measure the possible wear between the threads and the internal channels of the prosthesis, and the metal pins with the phalange holes. A Zwick Roell Z0.5 machine was used for the test, subjecting samples of the index finger to 24000 cycles in flexion mode. The test load variation was from 2 N to 80 N to ensure the force in the fingertip of 15N (as suggested in [21]), with a displacement speed of 1000 mm/min. The wire that pulls the mechanism to operate the finger was placed in the machine's jaw (see Figure 9) and was connected with the finger through a bronze pulley (see Figure 10b) that also that also serves as an alignment tool. Then the finger was set up in

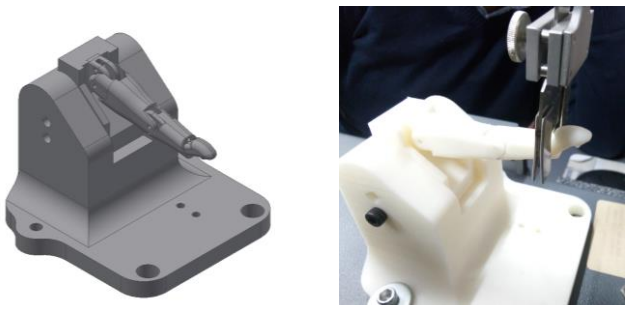
a 3D printed holder to facilitate the test and a sensor (a Force Sensitive Resistance sensor with a magnitude range: 100 g – 10 kg) was positioned in the area of contact with the fingertip in order to evaluate the pincer grasp load (see Figure 10a).

Three samples were tested and, besides analyzing their capacity to withstand cyclic loads, the wear resistance of the assembly was evaluated by weighing the samples before and after the test.



(a) (b) (c) (d)

Figure 7. Accessories designed for the (a) middle finger hyperextension, (b) middle finger abduction, (c) little finger hyperextension and (d) little finger abduction tests.



(a) (b)

Figure 8. (a) CAD design and (b) fabricated components of the finger flexion test assembly.

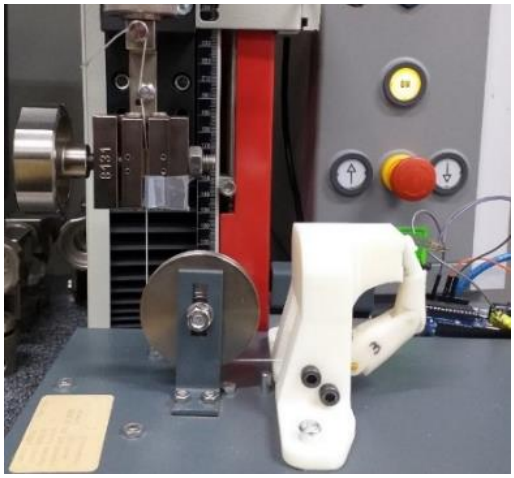
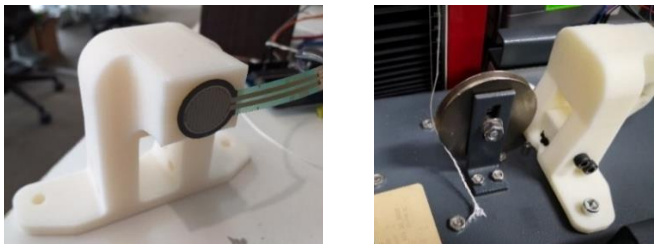


Figure 9. The complete assembly for the fatigue/wear test.



(a) (b)

Figure 10. Accessories for the finger fatigue/wear test: (a) sensor on its 3D printed holder and (b) pulley.

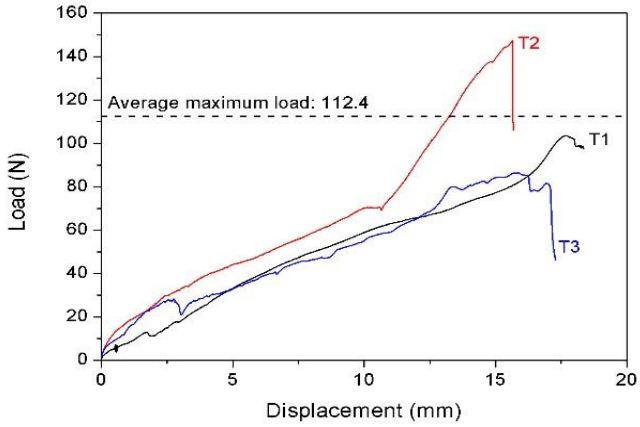


Figure 11: Tensile test curves for the D1M prosthetic hand samples

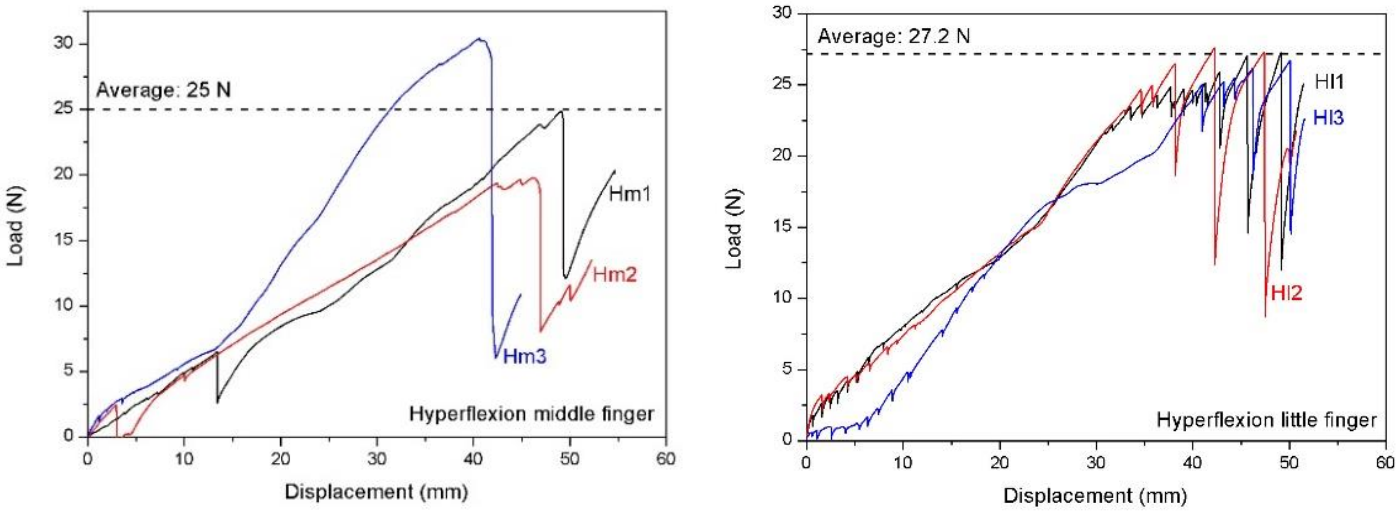
3. Results and Discussion

3.1.1. Hand tensile test

Figure 11 shows the load-displacement curves for the tensile test in the three assembled prosthetic hands (samples T1, T2 and T3). The average maximum tensile load that the hand can withstand before failure is 112.4 N. Discontinuities in the curves and differences in behavior for each sample could be attributed to factors inherent to the 3D printing process, and to re-accommodation events of the hand assembly through the test due to the elastic and plastic deformations experienced by the wire.

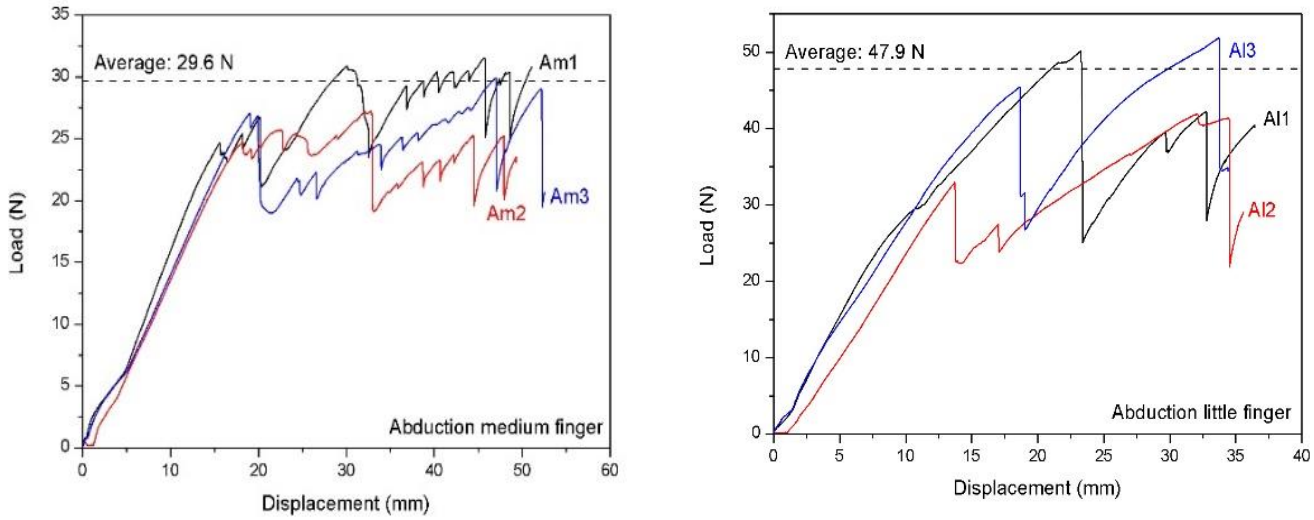
3.1.1. Finger Flexion test

Figure 12a and Figure 12b show the load-displacement curves for the middle and little finger samples obtained from the flexion tests in hyperextension condition (hyperflexion). The average maximum load for the middle and little finger in hyperflexion mode is 25 N and 27.2 N, correspondingly. The average maximum displacement is 50.6 mm for the middle finger and 51.3 mm for the little one. Although the average maximum load and the average maximum displacement values are similar for both fingers, the dispersion among the obtained values is bigger for the case of the middle finger. Figure 13a and Figure 13b present the load-displacement plots for the middle and little fingers obtained from the flexion tests in abduction condition. The average maximum load for the middle finger in hyperextension and abduction condition is 25.0 N and 29.6 N, correspondingly. For the little finger, the average maximum force is 27.2 N in hyperextension mode, and 47.9 N in abduction condition. The little finger can withstand bigger loads in both test conditions due to the fact that the length in cantilever mode is shorter, compared to that for the middle finger. The test curves exhibit discontinuities due to re-accommodation of the finger assembly. Table 4 presents the set of tested samples and the corresponding designation.



(a) (b)

Figure 12. Finger hyperextension test curves for the middle (a) and little (b) fingers of the D1M hand. The horizontal dashed lines represent the average of the maximum load values.



(a) (b)

Figure 13. Finger abduction test curves for the middle (a) and little (b) fingers of the D1M hand. The horizontal dashed lines represent the average of the maximum load values.

As shown in Figure 14a, failure of the fingers in hyperextension occurs at the link of the proximal phalanx. On the other hand, failure in the abduction condition takes place at the MCP joints, as observed in Figure 14b; cracks nucleate in this area and propagate producing misalignment of the pins. This mechanism might be the reason why significant variations in the load-displacement curves beyond the linear region are evidenced (see Figure 15). Despite being thinner and thus having a smaller cross-sectional area for load distribution, the little finger can withstand a greater load than the middle finger due to the shorter moment arm from its base to the fingertip, which should be taken into consideration for the design.

In the case of the fatigue test, after 24000 cycles of loading, no component from the finger assembly (3D printed pieces, metallic pins, and wires) show any type of failure. Table 5 shows the weight of the fatigue/wear test samples before and after the test. A relatively small difference in weight was recorded as a result of the cyclic load and wear on the contacting parts, the average percentage of variation of weight was 0.05%. The 24000 cycles represent a period of 27 days of use of the prosthetic hand, considering that an average person performs 3680 grasp movements per day [22] and that a prosthetic hand user employs it 4 hours per day [23]. Based on the obtained results, the expected annual weight loss of the D1M hand would be 0.67%, which is a low value when compared to the wear rate for common materials. Figure 16 shows one of the fatigue test samples after the test; the encircled area highlights the ABS material particles resultant from the wear on the contacting parts.

Table 4: Designation of test specimens.

Test type	Sample	Number of samples	Samples designation
Hyperflexion	Middle finger	3	Hm1, Hm2 and Hm3
Hyperflexion	Little finger	3	Hl1, Hl2 and Hl3
Abduction	Middle finger	3	Am1, Am2 and Am3
Abduction	Little finger	3	Al1, Al2 and Al3

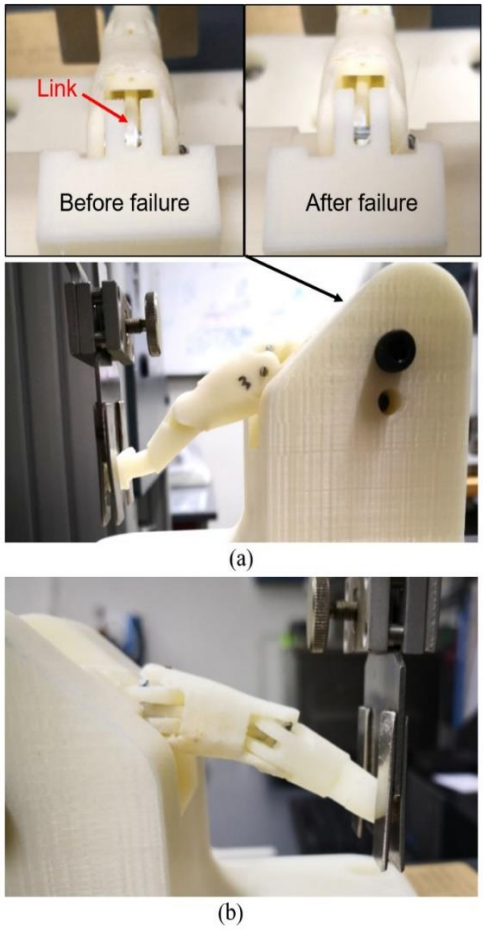


Figure 14: D1M prosthetic hand finger failure location under hyperextension (a) and abduction (b) conditions.

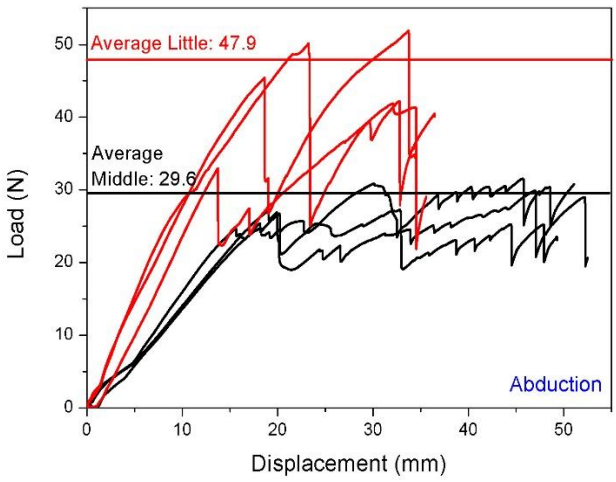


Figure 15: Finger abduction test curves for the middle (black) and little (red) fingers of the D1M hand. The horizontal lines represent the average of the maximum values per finger.

Table 5. Weight of fatigue/wear test samples before and after the test.

Sample	Weight (g)		Weight loss (%)
	Before test	After test	
1	6.7117	6.7109	0.012
2	6.7090	6.7020	0.104
3	6.8066	6.8040	0.038



Figure 16: Fatigue test sample

4. Conclusions

In this work, mechanical testing methods for upper limb prostheses were adapted and applied to a specific design, the D1M hand. Details of the experimental procedure have been described and test results have been discussed. The maximum tensile load that the D1M prosthetic hand can withstand is 112.4 N. Furthermore, maximum flexion strength of fingers in hyperextension and abduction conditions is 27.2 N and 48 n, respectively. It was also determined that after 24000 cycles of flexion no component of the fingers assembly of the D1M prosthetic hand exhibits any type of failure.

Obtained mechanical resistance tests results can be utilized to compare future experimental and simulation results. Moreover, our findings and the case described contribute in generating data for the 3D-printed prostheses community since standardization is needed in order to evaluate properties and performance of existing and future designs, which will allow for objective comparison between them and the continuous improvement of their mechanical specifications.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was funded by the National Council of Science, Technology and Innovation of Peru - Cienciactiva from Concytec (Grant Number 163-2017). The authors would like to thank Carlos Romero and Jorge Pajuelo for their support in the testing machines' operation.

References

[1] B. Phillips, G. Zingalis, S. Ritter and K. Mehta, "A review of current upper-limb prostheses for resource-constrained settings," in IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 2015.

[2] A. E. Tontowi, L. Ramdani, R. V. Erdizon and D. K. Baroroh, "Optimization of 3D-Printer Process Parameters for Improving Quality of Polylactic Acid Printed Part," International Journal of Engineering and Technology (IJET), vol. 9, no. 2, pp. 589-600, 2017.

[3] J. Torres, M. Cole, A. Owji, Z. DeMastry and A. P. Gordon, "An Approach for mechanical property optimization of fused deposition modeling with polylactic acid via design of experiments," Rapid Prototyping Journal, vol. 22, no. 2, pp. 387-404, 2016.

[4] B. Tymrak, M. Kreiger and J. Pearce, "Mechanical properties of components fabricated with open-source 3-D printers under realistic environmental conditions," Materials and Design, vol. 58, pp. 242-246, 2014.

[5] M. Fernandez-Vicente, W. Calle, S. Ferrandiz and A. Conejero, "Effect of Infill Parameters on Tensile Mechanical Behavior in Desktop 3D Printing," 3D Printing and Additive Manufacturing, vol. 3, no. 3, pp. 183-192, 2016.

[6] A. R. Torrado and D. A. Roberson, "Failure Analysis and Anisotropy Evaluation of 3D-Printed Tensile Test Specimens of Different Geometries and Print Raster Patterns," Journal of Failure Analysis and Prevention, vol. 16, no. 1, pp. 154-164, 2016.

[7] J. Cantrell, S. Rohde, D. Damiani, R. Gurnani, L. DiSandro, J. Anton, A. Young, A. Jerez, D. Steinbach, C. Kroese and P. Ifju, "Experimental Characterization of the Mechanical Properties of 3D Printed ABS and Polycarbonate Parts," Rapid Prototyping Journal, vol. 23, no. 4, pp. 89-105, 2017.

[8] R. Mio, M. Sánchez and Q. Valverde, "Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Review," in IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, 2018.

[9] S. Tenim, Design of an Affordable Anthropomorphic Mechanical Prosthetic Hand, M. A. Thesis, University of Cape Town, South Africa, 2014.

[10] J. Koprnicky, J. Safka and M. Ackermann, "Using of 3D Printing Technology in Low Cost Prosthetics," Materials Science Forum, vol. 919, pp. 199-206, 2018.

[11] Robotiq, "How we tested the Robotiq Adaptive Gripper," [Online]. Available: https://blog.robotiq.com/bid/31435/How-we-Tested-the-Robotiq-Adaptive-Gripper. [Accessed 7 Junio 2018].

[12] J. J. Crisco and S. W. Wolfe, "Mechanical testing of orthopedic implants: Hand and wrist," Mechanical Testing of Orthopaedic Implants, Vols. Duxford, United Kingdom, pp. 63-98, Woodhead Publishing, 2017.

[13] American Society for Testing and Materials, ASTM F1781 - 15, Elastomeric Flexible Hinge Finger Total Joint Implants, West Conshohocken: PA: ASTM International, 2015.

[14] T. J. Joyce and A. Unsworth, "A test procedure for artificial finger joints," Proceedings of the Institution of Mechanical Engineers,Part H: Journal of Engineering in Medicine, vol. 216, no. 2, pp. 105-110, 2002.

[15] International Organization for Standardization, ISO 22523 External limb prostheses and external orthoses, Switzerland: Requirements and test methods, 2006.

[16] National Institute of Standards and Technology, "Performance Metrics and Benchmarks to Advance the State of Robotic Grasping," [Online]. Available: https://www.nist.gov/programs-projects/performance-metricsmetricsand-. [Accessed 4 June 2018].

[17] J. T. Belter, M. T. Leddy, K. D. Gemmell Jr. and A. M. Dollar, "Comparative Clinical Evaluation of the Yale Multigrasp Hand," in IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), UTown, Singapore, 2016.

[18] R. U. N. &. Media, "A hand for the hand-makers," 3 April 2016. [Online]. Available: http://news.rice.edu/2016/04/13/a-hand-for-the-hand-makers-3/. [Accessed April26 2018].

[19] M. Bustamante, R. Vega-Centeno, M. Sánchez and R. Mio, "A Parametric 3D-Printed Body-Powered Hand Prosthesis Based on the Four-Bar Linkage Mechanism," in IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, 2018.

[20] G. Smit, D. H. Plettenburg and F. C. T. van der Helm, "The Lightweight Delft Cylinder Hand: First Multi-Articulating Hand That Meets the Basic User Requirements," in 431-440, 23, 2015.

[21] J. T. Belter, J. L. Segil, A. M. Dollar and R. F. Weir, "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review," Journal of Rehabilitation Research & Development, vol. 50, no. 5, pp. 599-618, 2013.

[22] V. Nagaraja, J. Bergmann, D. Sen and M. Thompson, "Examining the needs of affordable upper limb prosthetic users in India: A questionnaire-based survey," Technology and Disability, vol. 1, pp. 1-10, 2016.

[23] M. Vergara, J. Sancho, V. Gracia and A. Pérez, "An introductory study of common grasps used by adults during performance of activities of daily living," Journal of Hand Therapy, vol. 27, pp. 225-234, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 65-72 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

All-Pass RC-Filters Architecture with Independent Adjustment of the Main Parameters Based on Differential Difference Amplifiers

Darya Denisenko[1,2,*], Nikolay Prokopenko[1,3], Nikolay Butyrlagin[1]

[1]*Information systems and radio engineering, Don State Technical University, DSTU, Rostov-on-Don, 344000, Russia*

[2]*Systems of automation control, Southern Federal University, SFedU, Taganrog, 347928, Russia*

[3]*Institute for Design Problems in Microelectronics of Russian Academy of Sciences, IPPM RAS, Zelenograd, 124681, Russia*

ARTICLE INFO

Article history:
Received:28 May, 2019
Accepted:04 July, 2019
Online: 11 July, 2019

Keywords:
All-Pass ARCF
Low-Pass Filter
High-Pass Filter
Band-Pass Filter
Rejection Filter
Differential Difference Amplifier
Pole Q-Factor
Pole Frequency
ARCF Transfer Ratio
Filter Parameters Trimming

ABSTRACT

We have suggested an architecture of the second-order all-pass active RC-filter (ARCF), based on three differential difference amplifiers (DDA's), which provides full set of amplitude-frequency responses (AFR's) (low pass filter (LPF), high pass filter (HPF), band pass filter (BPF), rejection filter (RF)). We have given the basic equations, which allow ARCF's. The results of computer simulation of the LPF, HPF, BPF and RF's are discussed. In case of small volume production it is possible to produce the given ARCF in a form of specialized structural arrays, in which LPF, HPF, BPF and RF characteristics are realized by using one of the seven inputs, to which the processed signal is supplied, and four outputs, from which the said signal is outputted. The ARCF is noted, because its transfer ratio and pole frequency are not changed, when the pole Q-factor is adjusted.

1. Introduction

All-pass active RC-filters (ARCF), which provide amplitude-frequency responses (AFR) of low-pass filter (LPF), high-pass filter (HPF), band-pass filter (BPF), rejection filter (RF) [1-16] at different outputs, are attractive for production in a form of special structural crystals [17,18] in conditions of small volume specific type of electronic equipment. At the same time, the application of promising electronic component base in ARCF, for example, differential difference amplifiers (DDAs) [19-33], which provides new properties of frequency selection devices of this class, is of considerable interest.

The present article's purpose and novelty are analysis of all-pass ARCF's new structure's properties [34]. The filter is provided with independent adjustment of base parameters (for example, with digital potentiometers or passive elements' digital switching. A full set of LPF, HPF, BPF, RF's characteristics are implemented by switching seven inputs, on which the signal is supplied, and four outputs, form which the said signal is outputted.

2. ARC-Filter Base Architecture's Properties

There is a circuit of the suggested all-pass ARC-filter (LPF, HPF, BPF, RF) on Fig. 1 [34]. It provides an independent adjustment of pole Q-factor, when pole frequency and transfer ratio, which depend on other elements' parameters, are constant. The independent adjustment is achieved by controlling resistors' resistance. It significantly simplifies a process of frequency selection devices' trimming and adjustment, based on the suggested ARCF circuit. The buffer amplifier (BuffAmp) in scheme Fig. 1 is implemented on a base of traditional operational amplifiers (OpAmps) or DDA, it provides a wide range of R3, R4 resistors' resistance change.

All types active RC-filters' (LPF, HPF, BPF, RF) generalized transfer function is written as (the filters are implemented on the circuit, given on Fig. 1):

*Corresponding Author: Nikolay Butyrlagin, Don State Technical University, Russia | Email: butyrlagin@gmail.com

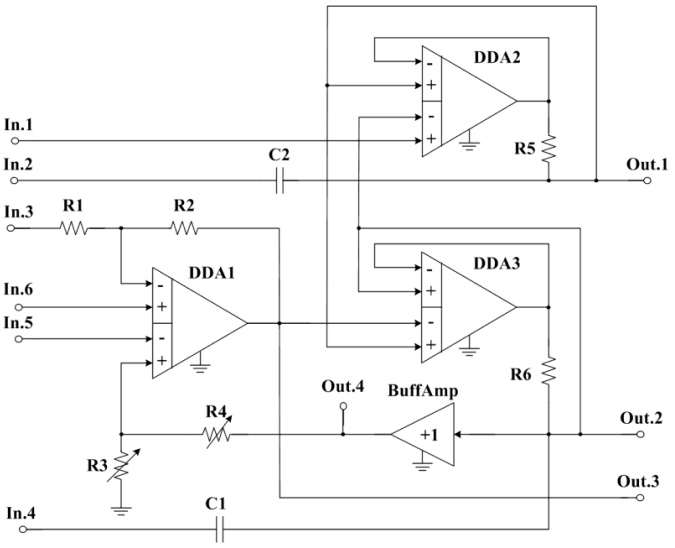


Figure 1: Suggested ARC-Filter's Circuit

$$F(p)=\frac{a_2p^2+a_1p+a_0}{b_2p^2+b_1p+b_0}, \quad (1)$$

where a_i, b_j are equation (1) numerator's and denominator's coefficients, which depend on components' parameters and inputs and outputs in circuit on Fig. 2.

A set of transfer function's numerator coefficients a_i defines a type of ARC-filter (LPF, HPF, BPF, RF).

There are transfer functions' (1) numerator coefficients ai, which are realized in particular ARCF of the Fig. 1 connection circuits, given in Table 1.

The sensitivity functions of the basic parameters of ARCF (Fig. 1) easily correspond to formulas (1) and Table 1.

In the proposed ARCF schemes, typical schemes are applicable, including digital potentiometers or digital switching passive elements, which are described in detail in [35, 36].

The implemented filters' (LPF, HPF, BPF, RF) transfer functions' (1) denominator coefficients b_j relate to elements of the circuit on Fig. 1 by the following equations:

$$b_0=\frac{1}{\tau_1\tau_2}, \quad b_1=\frac{KK_4}{\tau_1}1, \quad b_2=1, \quad . \quad (2)$$

Here the following notation are used in formulas (1), (2):

$$K=\frac{R_3}{R_3+R_4}, \quad K_1=\frac{R_2}{R_1}, \quad K_2=K_3=K_4=1+\frac{R_2}{R_1}, \quad (3)$$
$$\tau_1=R_6C_1, \quad \tau_2=R_5C_2,$$

where R_{ij} are resistance of ij-resistor, C_1, C_2 are capacitance of capacitors C1 and C2.

The active RC-filters, shown in Table 1 and defined as $LPF^{(+)}$, $HPF^{(+)}$, $BPF^{(+)}$, $RF^{(+)}$, have the pole Q-factor, transfer ratio's and pole frequency's independent adjustment properties. Here the pole Q-factor adjustment does not change the filter's transfer ratio and its pole frequency. These filters have the greatest practical interest.

The active RC-filters, defined in Table 1 as LPF, HPF, BPF, RF, do not have the Q-factor, transfer ratio's and pole frequency's independent adjustment properties. Here the transfer ratio and pole frequency coefficients may change, when the pole Q-factor is changed.

The active RC-filters, defined in Table 1 as $LPF^{(-)}$, $HPF^{(-)}$, $BPF^{(-)}$, $RF^{(-)}$, have an amplitude-frequency response slope, which corresponds the first-order transfer function. It limits application scope for these circuit solutions.

There are variants of ARCF Fig. 1 inputs and outputs application on Fig. 2-Fig. 6, they provide AFR different modifications' implementation. In this case, connecting a signal source, for example, to the first input (In.1) assumes that unused inputs are connected to a common bus.

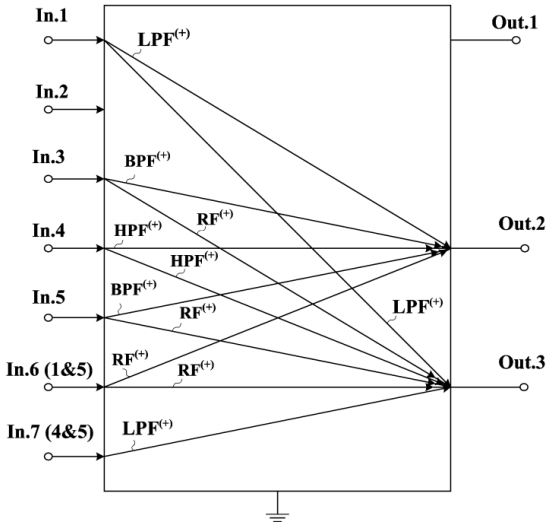


Figure 2: ARCF (Fig. 1) Inputs and Outputs Application Variants, which Provide $BPF^{(+)}$, $RF^{(+)}$, $LPF^{(+)}$,$HPF^{(+)}$ Implementation with Pole Q-Factor Independent Adjustment

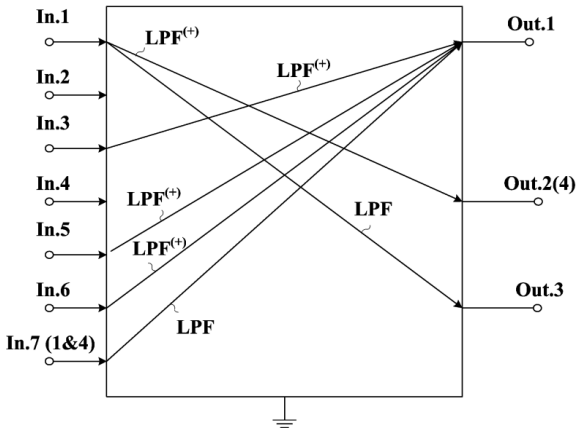


Figure 3: ARCF (Fig. 1) Inputs and Outputs Application, which Provide Low-Pass Filters $LPF^{(+)}$ and LPF Implementation

Table 1: ARC-Filters' Transfer Functions' (1) Numerator Coefficients a_i, Implemented on Base of Circuit, Given on Fig. 1

OUTPUTS / INPUTS	Out.1	Out.2(4)	Out.3
In.1 AFR curves Fig. 7b	HPF$^{(-)}$&BPF$^{(-)}$ $a_2 = \frac{KK_4}{\tau_1\tau_2}$ $a_1 = \frac{1}{\tau_2}$ $a_0 = 0$	LPF$^{(+)}$ $a_2 = a_1 = 0$ $a_0 = \frac{1}{\tau_1\tau_2}$	LPF $a_2 = a_1 = 0$ $a_0 = \frac{KK_4}{\tau_1\tau_2}$
In.2 AFR curves Fig. 8b	HPF$^{(-)}$&BPF$^{(-)}$ $a_2 = 1$ $a_1 = \frac{KK_4}{\tau_1}$ $a_0 = 0$	BPF $a_2 = a_0 = 0$ $a_1 = \frac{1}{\tau_1}$	BPF$^{(+)}$ $a_2 = a_0 = 0$ $a_1 = \frac{KK_4}{\tau_1}$
In.3 AFR curves Fig. 9b	LPF$^{(+)}$ $a_2 = a_1 = 0$ $a_0 = -\frac{K_1}{\tau_1\tau_2}$	BPF $a_2 = a_0 = 0$ $a_1 = \frac{K_1}{\tau_1}$	RF$^{(+)}$ $a_2 = -K_1$ $a_1 = 0$ $a_0 = -\frac{K_1}{\tau_1\tau_2}$
In.4 AFR curves Fig. 10b	BPF $a_2 = a_0 = 0$ $a_1 = -\frac{1}{\tau_2}$	HPF$^{(+)}$ $a_2 = 1$ $a_1 = a_0 = 0$	HPF $a_2 = KK_4$ $a_1 = a_0 = 0$
In.5 AFR curves Fig. 11b	LPF$^{(+)}$ $a_2 = a_1 = 0$ $a_0 = -\frac{K_3}{\tau_1\tau_2}$	BPF $a_2 = a_0 = 0$ $a_1 = \frac{K_3}{\tau_1}$	RF$^{(+)}$ $a_2 = -K_3$ $a_1 = 0$ $a_0 = -\frac{K_3}{\tau_1\tau_2}$
In.6 AFR curves Fig. 12b	LPF$^{(+)}$ $a_2 = a_1 = 0$ $a_0 = \frac{K_2}{\tau_1\tau_2}$	BPF $a_2 = a_0 = 0$ $a_1 = -\frac{K_2}{\tau_1}$	RF$^{(+)}$ $a_2 = K_2$ $a_1 = 0$ $a_0 = \frac{K_2}{\tau_1\tau_2}$
In.1&4 AFR curves Fig. 13b	LPF $a_2 = a_1 = 0$ $a_0 = \frac{KK_4}{\tau_1\tau_2}$	RF$^{(+)}$ $a_2 = 1$ $a_1 = 0$ $a_0 = \frac{1}{\tau_1\tau_2}$	RF $a_2 = KK_4$ $a_1 = 0$ $a_0 = \frac{KK_4}{\tau_1\tau_2}$

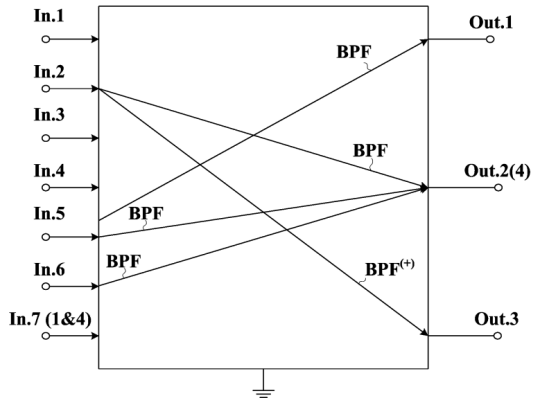


Figure 4: ARCF (Fig. 1) Inputs and Outputs Application, which Provide Band-Pass Filters $BPF^{(+)}$ and BPF Implementation

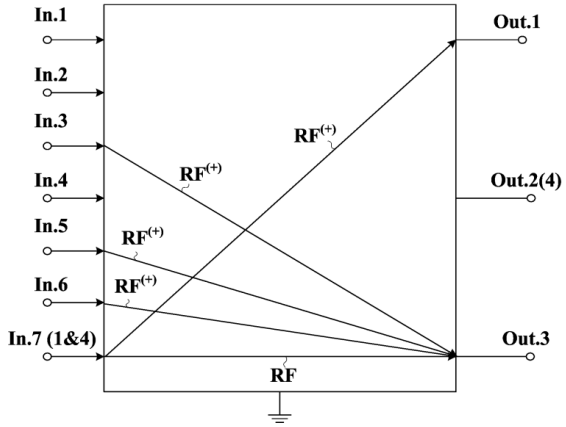


Figure 5: ARCF (Fig. 1) Inputs and Outputs Application, which Provide Rejection Filters $RF^{(+)}$ and RF Implementation

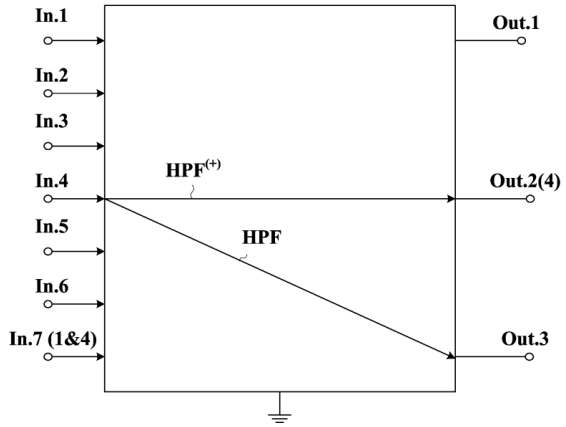
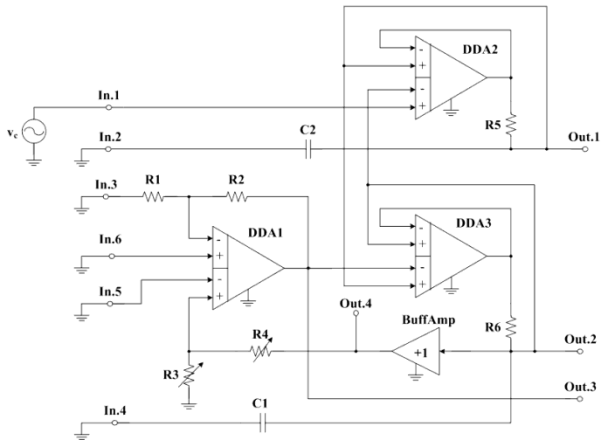


Figure 6: ARCF (Fig. 1) Inputs and Outputs Application, which Provide High-Pass Filters $HPF^{(+)}$ and HPF Implementation

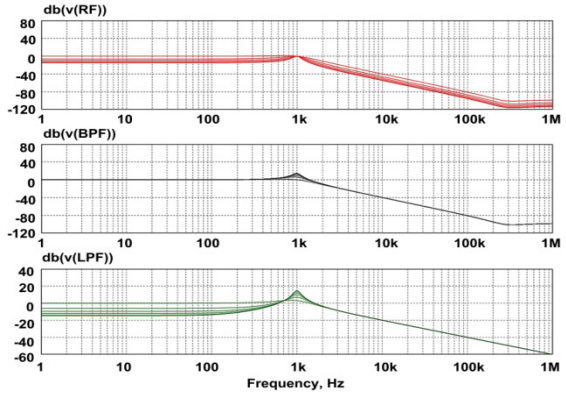
The computer simulation of the ARCF Fig. 1 different modifications, which correspond Fig. 2-Fig. 6, was made in MicroCap on AD830 DDA. It has confirmed the above properties of the suggested ARCF's circuit solution.

3. Low-Pass Filter $LPF^{(+)}$

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 7a, it provides amplitude-frequency responses of $HPF^{(-)}$&$BPF^{(-)}$, $LPF^{(+)}$, LPF on outputs Out.1, Out.2(4), Out.3.



(a)



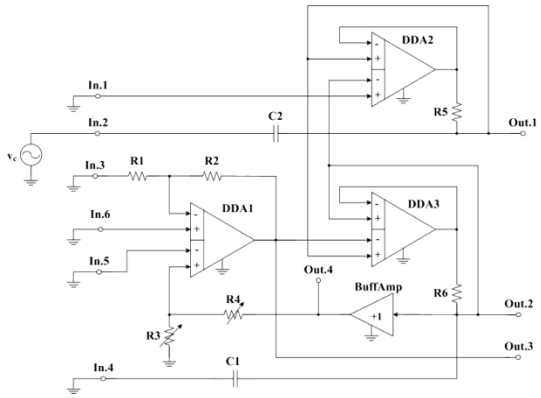
(b)

Figure 7: Special ARC-Filter's Switching Circuit (a) and its AFRs (b)

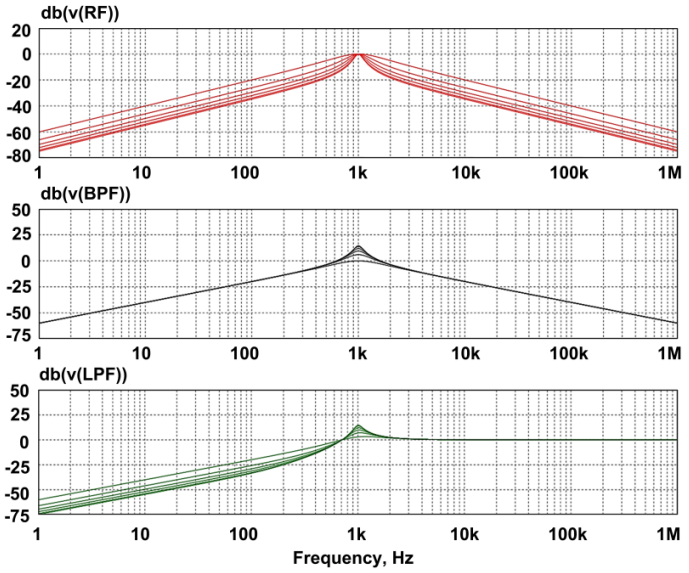
Analysis of the AFR's curves (Fig. 7b) has shown, that transfer ratio and pole frequency are not changed in $LPF^{(+)}$, implemented for output Out.2(4), when the pole Q-factor is adjusted (by changing R3 and R4 resistors resistance). At the same time, the pole frequency in this circuit can be tuned using resistors R5 and R6.

4. Band-Pass Filter $BPF^{(+)}$

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 8a, it provides amplitude-frequency responses of $HPF^{(-)}$&$BPF^{(-)}$, BPF, $BPF^{(+)}$ on outputs Out.1, Out.2(4), Out.3.



(a)



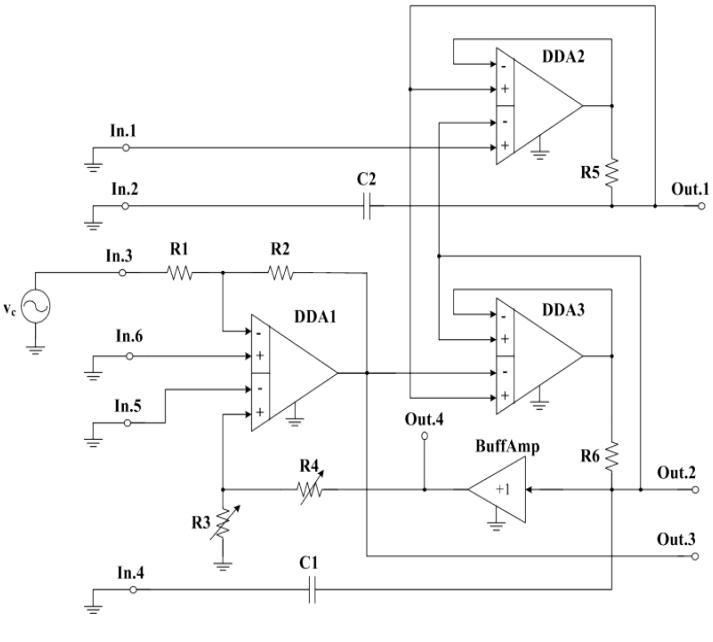
(b)

Figure 8: Special ARC-Filter's Switching Circuit (a) and its Amplitude-Frequency Responses (b)

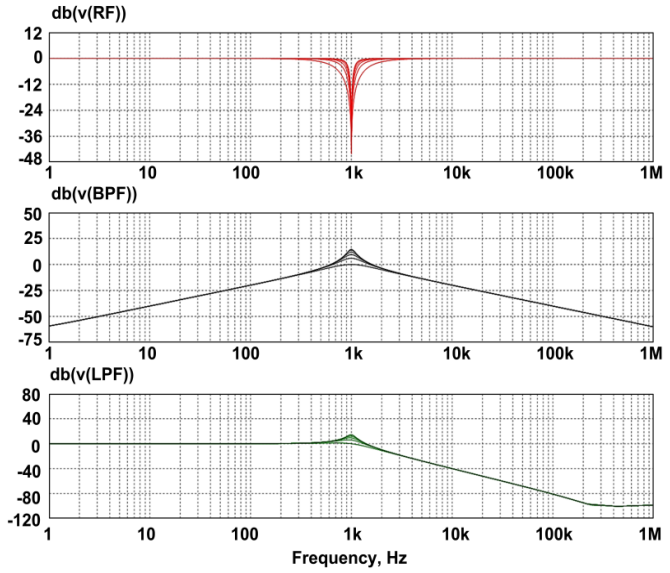
Based on the BPF(+)'s AFRs curves (Fig. 8b) for output Out.3, can be concluded, that transfer ratio and pole frequency are not changed, when the pole Q-factor is adjusted (by changing R3 and R4 resistors' resistance). In this case, the pole frequency in this circuit can be tuned with resistors R5 and R6, and the transfer ratio change with resistors R1 and R2.

5. Rejection Filter RF(+) and Low-Pass Filter LPF (+)

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 9a, it provides amplitude-frequency responses of LPF(+), BPF, RF(+) on outputs Out.1, Out.2(4) and Out.3.



(a)



(b)

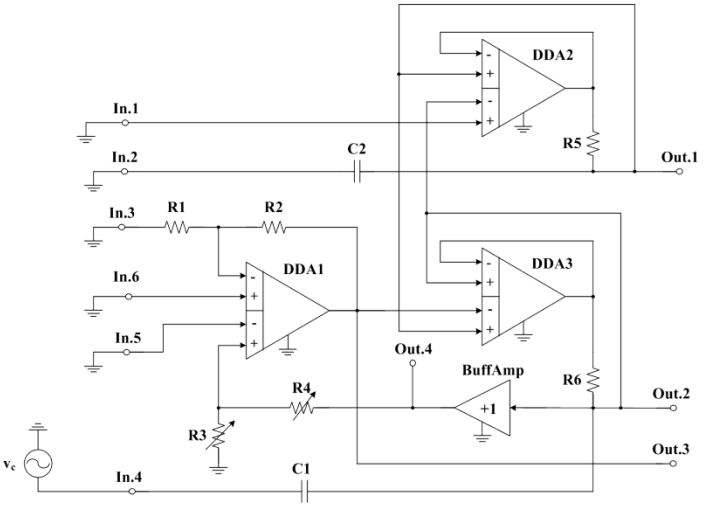
Figure 9: Special ARC-Filter's Switching Circuit (a) and its Amplitude-Frequency Responses (b)

So the ARC's curves (Fig. 9b) show, that transfer ratio and pole frequency are not changed in the implemented LPF(+) and RF(+) for outputs Out.1 and Out.3, when the pole Q-factor is adjusted by changing R3 and R4 resistors' resistance. The pole frequency in this circuit can be tuned with resistors R5 and R6, and the transfer ratio with resistors R1 and R2.

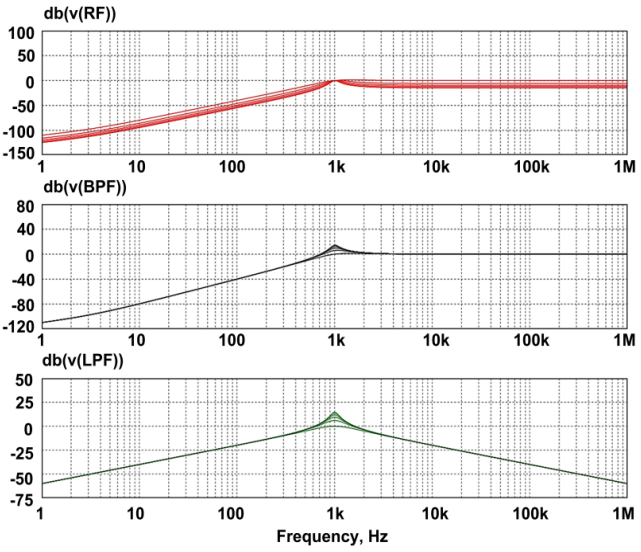
6. High-Pass Filter HPF(+)

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 9a, it provides amplitude-frequency responses of BPF, HPF(+) and HPF on outputs Out.1, Out.2(4), Out.3.

It should be noted that when the pole Q-factor is adjusted, the transfer ratio and pole frequency of HPF(+) for output Out.2(4) are not changed, which follows from AFR on Fig. 10b. At the same time, the pole frequency in this circuit can be tuned using resistors R5 and R6.



(a)



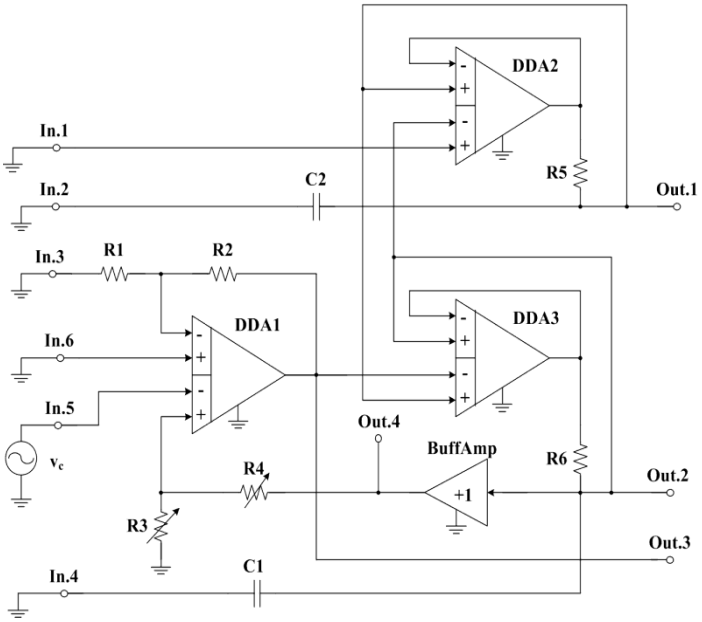
(b)

Figure 10: Special ARC-Filter's Switching Circuit (a) and Amplitude-Frequency Responses (b)

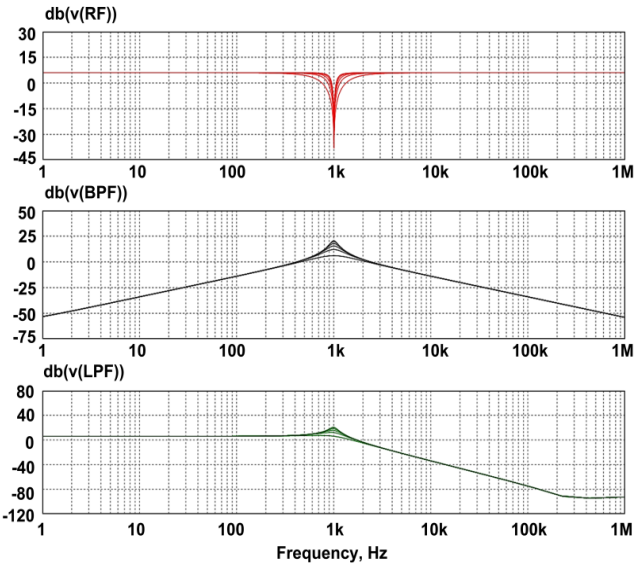
7. Low Pass Filter LPF(+) and Rejection Filter RF(+), Inverting Input Signal

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 11a, it provides amplitude-frequency responses of LPF(+), BPF, RF(+). on outputs Out.1, Out.2(4), Out.3.

The given AFRs' curves' (Fig. 11b) analysis has shown, that the transfer ratio and pole frequency in LBF(+) and RF(+) for outputs Out.1 and Out.3 are not changed, when the pole Q-factor is adjusted by changing R3 and R4 resistors' resistance. At the same time, the pole frequency in this circuit can be tuned with resistors R5 and R6, and the transfer ratio with resistors R1 and R2.



(a)



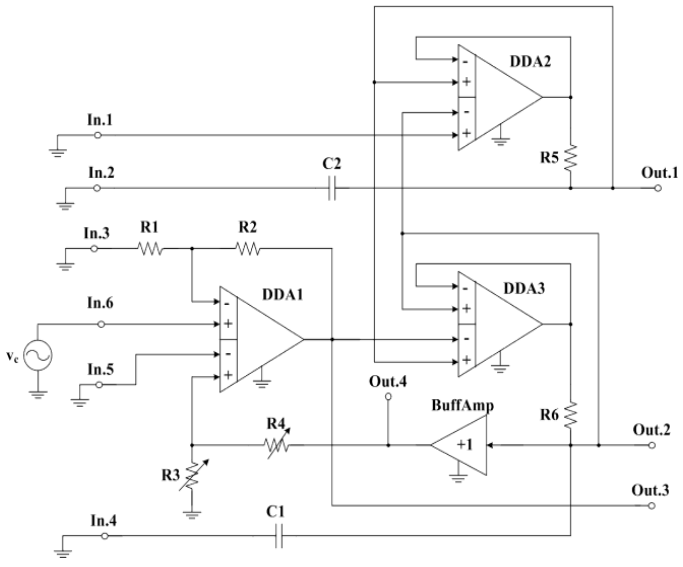
(b)

Figure 11: Special ARC-Filter's Switching Circuit (a) and its Amplitude-Frequency Responses (b)

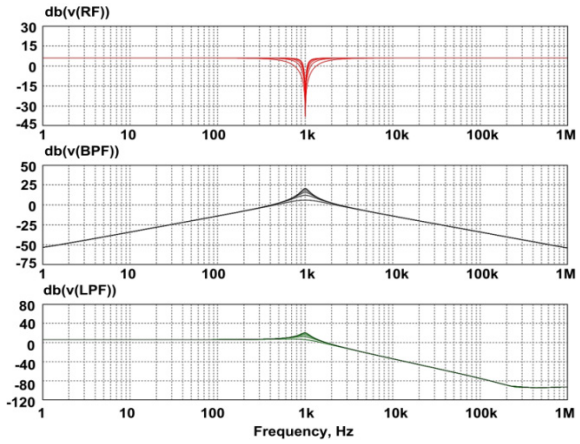
8. Low Pass Filter LPF (+) and Rejection Filters (+), Non-inverting Input Signal

ARCF's inputs (Fig. 1) are connected according to the circuit, given on Fig. 12a, it provides amplitude-frequency responses of LPF(+), BPF, RF(+) on outputs Out.1, Out.2(4) and Out.3.

When the pole Q-factor is adjusted (by changing R3 and R4 resistors resistance), the transfer ratio and pole frequency of LPF(+) and RF(+) (for outputs Out.1 and Out.3 correspondently) are not changed. At the same time, the pole frequency in this circuit can be adjusted using resistors R5 and R6, and the transfer ratio using resistors R1 and R2.



(a)

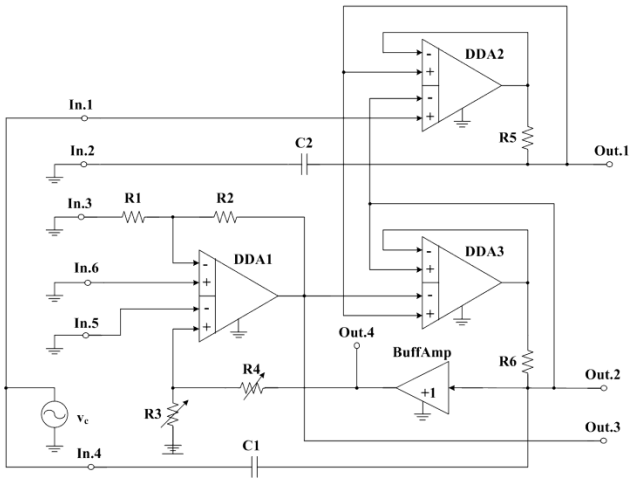


(b)

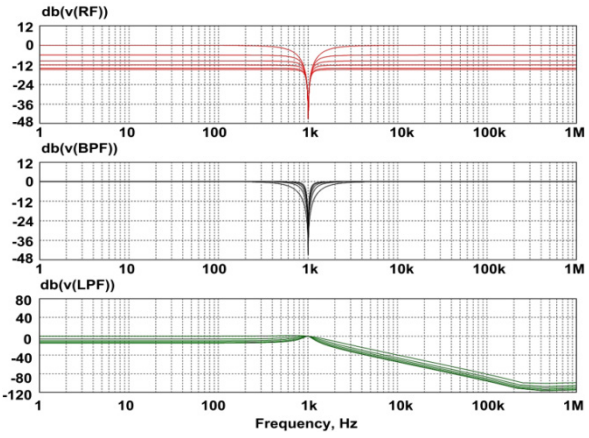
Figure 12: Special ARC-Filter's Switching Circuit (a) and its AFRs (b)

9. Rejection Filter RF(+)

ARCF's inputs (Fig. 1) are connected according to Fig. 13a, it provides amplitude-frequency responses of LPF, RF(+), RF on outputs Out.1, Out.2(4), Out.3.



(a)



(b)

Figure 13: Special ARC-Filter's Switching Circuit (a) and its Amplitude-Frequency Responses (b)

From the analysis of the AFRs' curves (Fig. 13b), it may be concluded, that the transfer ratio and pole frequency are not changed in the RF(+), implemented for output Out.2(4), when the pole Q-factor is adjusted (by changing R3 and R4 resistors resistance). At the same time, the pole frequency in this circuit can be tuned using resistors R5 and R6.

10. Conclusion

We have developed a structure of all-pass ARCF with pole Q-factor, transfer ratio and pole frequency independent adjustment for the whole spectrum of amplitude-frequency responses of the second-order filter (LPF, HPF, BPF, RF) implemented. This is a significant advantage of the considered circuit design solutions in comparison with the known ARCF of this class. We recommend using the suggested ARCF as a specialized chip of a structural array because of its universality. It is possible to implement a wide range of frequency selection devices in the said ARCF, because the signal source is connected to one of seven inputs and four outputs.

Conflict of Interest

The authors declare that there is no conflict of interests regarding publication of this paper.

Acknowledgments

The research is carried out at the expense of the Russian Science Foundation Grant (Project No. 18-79-10109).

References

[1] W. Chiu and J. Horng, "High-Input and Low-Output Impedance Voltage-Mode Universal Biquadratic Filter Using DDCCs," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 54, no. 8, pp. 649-652, Aug. 2007. DOI: 10.1109/TCSII.2007.899460

[2] Miao Huijing, Liu Xiaoyan, Tan Boxue, Shen Jin and Wang Yajing, "A universal active filter design method," 2010 International Conference on Information, Networking and Automation (ICINA), Kunming, 2010, pp. V1-18-V1-22. DOI: 10.1109/ICINA.2010.5636442

[3] C. Chang, "Universal active current filter with single input and three outputs using CCIIs," in Electronics Letters, vol. 29, no. 22, pp. 1932-1933, 28 Oct. 1993. DOI: 10.1049/el:19931286

[4] Xifeng Zhou, Shanshan Li and Jiake Wang, "An electronically tunable active-C current-mode universal filter employing CCCII and OTA," 2011 IEEE International Conference on Computer Science and Automation Engineering, Shanghai, 2011, pp. 647-651. DOI: 10.1109/CSAE.2011.5952930

[5] F. Khateb, S. B. A. Dabbous, M. Kumngern and T. Kulej, "Novel current controlled differential-input buffered output active element and its application in all-pass filter," 2015 38th International Conference on Telecommunications and Signal Processing (TSP), Prague, 2015, pp. 335-338. DOI: 10.1109/TSP.2015.7296279

[6] P. Ahmadi, B. Maundy, A. S. Elwakil, L. Belostotski and A. Madanayake, "A New Second-Order All-Pass Filter in 130-nm CMOS," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 63, no. 3, pp. 249-253, March 2016. DOI: 10.1109/TCSII.2015.2482578

[7] B. Metin, N. Herencsar, J. Koton and E. Arslan, "All-pass filter application using electronically tunable DDCC," 2014 24th International Conference Radioelektronika, Bratislava, 2014, pp. 1-4. DOI: 10.1109/Radioelek.2014.6828413

[8] L. Mohammadi and K. Koh, "Integrated C-band (4–8 GHz) frequency-tunable & bandwidth-tunable active band-stop filter in 0.13-µm SiGe BiCMOS," 2015 IEEE MTT-S International Microwave Symposium, Phoenix, AZ, 2015, pp. 1-4. DOI: 10.1109/MWSYM.2015.7167026

[9] T. S. Arora, M. Gupta and S. Gupta, "Current mode universal filter employing operational transconductance amplifier and third generation current conveyor," 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, 2016, pp. 1-4. DOI: 10.1109/ICPEICES.2016.7853305

[10] S. Gautam, P. Yunqing, M. Kashif, Y. R. Kafle, Z. Z. Hua and L. Bo, "Study on software phase locked loop for single phase active power filter," Fifth

International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2013), Bangalore, 2013, pp. 319-327. doi: 10.1049/cp.2013.2191

[11] N. Herencsar, J. Jerabek, J. Koton, K. Vrba, S. Minaei and İ. C. Göknar, "Pole frequency and pass-band gain tunable novel fully-differential current-mode all-pass filter," 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, 2015, pp. 2668-2671. DOI: 10.1109/ISCAS.2015.7169235

[12] N. Herencsar, J. Koton, K. Vrba and O. Cicekoglu, "Low-voltage fully cascadable resistorless transadmittance-mode all-pass filter," 2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS), College Station, TX, 2014, pp. 185-188. DOI: 10.1109/MWSCAS.2014.6908383

[13] A. Upadhyay and K. Pal, "A DVCC based voltage mode all pass filter using operational amplifier pole," 2014 International Conference on Power, Control and Embedded Systems (ICPCES), Allahabad, 2014, pp. 1-4. DOI: 10.1109/ICPCES.2014.7062817

[14] N. Wattikornsirikul and M. Kumngern, "Three-input one-output voltage-mode universal filter using simple OTAs," 2014 Twelfth International Conference on ICT and Knowledge Engineering, Bangkok, 2014, pp. 28-31. DOI: 10.1109/ICTKE.2014.7001530

[15] R. Sotner, J. Jerabek, J. Petrzela, K. Vrba and T. Dostal, "Design of fully adjustable solution of band-reject/all-pass filter transfer function using signal flow graph approach," 2014 24th International Conference Radioelektronika, Bratislava, 2014, pp. 1-4. DOI: 10.1109/Radioelek.2014.6828418

[16] I. Mondal and N. Krishnapura, "A 2-GHz Bandwidth, 0.25–1.7 ns True-Time-Delay Element Using a Variable-Order All-Pass Filter Architecture in 0.13 μ m CMOS," in IEEE Journal of Solid-State Circuits, vol. 52, no. 8, pp. 2180-2193, Aug. 2017. doi: 10.1109/JSSC.2017.2693229.

[17] J. Teifel, R. S. Flores, S. Pearson, C. Begay, K. K. Ma and J. Palmer, "ViArray standard platforms: Rad-hard structured ASICs for digital and mixed-signal applications," 2012 IEEE Aerospace Conference, Big Sky, MT, 2012, pp. 1-9. DOI: 10.1109/AERO.2012.6187235.

[18] "Analogue arrays from Sweden, " SLA description, Dec. 1999, pp. 1-4.

[19] V. Stornelli, L. Pantoli, G. Leuzzi, G. Ferri, "Fully differential DDA-based fifth and seventh order Bessel low pass filters and buffers for DCR radio systems, " Analog Integr Circ Sig Process 2013;75(2):305–10. https://doi.org/10.1007/s10470-013-0051-9.

[20] F. Khateb, T. Kulej, M. Kumngern and C. Psychalinos, "A compact power-efficient 0.5 V fully differential difference amplifier, " AEU - International Journal of Electronics and Communications, 105, 71–77. DOI:10.1016/j.aeue.2019.04.007

[21] F. Khateb, M. Kumngern, T. Kulej and V. Kledrowetz, "Low-voltage fully differential difference transconductance amplifier," in IET Circuits, Devices & Systems, vol. 12, no. 1, pp. 73-81, 1 2018. DOI: 10.1049/iet-cds.2017.0057

[22] H. Alzaher and M. Ismail, "A CMOS fully balanced differential difference amplifier and its applications," in IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 48, no. 6, pp. 614-620, June 2001. DOI: 10.1109/82.943332.

[23] M. Kumngern, F. Khateb, "0.8-V Floating-Gate Differential Difference Current Feedback Operational Amplifier, "2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 14-17 May 2014, pp. 1-5. DOI: 10.1109/ECTICon.2014.6839780

[24] M. Kumngern, K. Klangthan, "0.5-V Fourth-Order Low-Pass Filter, " 2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), Oct. 23, 2017, Jakarta, Indonesia, pp. 119-122. DOI: 10.1109/ICACOMIT.2017.8253398

[25] J. Wang, Zh. Zhu, S. Liu, R. Ding, "A low-noise programmable gain amplifier with fully balanced differential difference amplifier and class-AB output stage, " J. Microelectronics, 2017, no. 64, pp. 86–91.

[26] S.G. Krutchinsky, N.N. Prokopenko, E.A. Zhebrun, N.V. Butyrlagin. The Peculiarities of the Structural Optimization of the Energy-Efficient Precision ARC-Filters on the Base of Classical and Differential Difference Operational Amplifiers // IEEE East-West Design & Test Symposium (EWDTS'2015), 26 – 29 Sep. 2015. - Batumi, Georgia. DOI: 10.1109/EWDTS.2015.7493136

[27] N. N. Prokopenko, N. V. Butyrlagin, S. G. Krutchinsky, E. A. Zhebrun and A. E. Titov, "The Advanced Circuitry of the Precision Super Capacitances Based on the Classical and Differential Difference Operational Amplifiers," 2015 IEEE 18th International Symposium on Design and Diagnostics of Electronic Circuits & Systems, Belgrade, 2015, pp. 111-114. DOI: 10.1109/DDECS.2015.46

[28] D. Jana and A. K. Mal, "Design of low noise amplifier for sensor applications," 2017 Devices for Integrated Circuit (DevIC), Kalyani, 2017, pp. 451-455.doi: 10.1109/DEVIC.2017.8073990

[29] Q. Hu, L. Yang and F. Huang, "A 100–170MHz fully-differential Sallen-Key 6th-order low-pass filter for wideband wireless communication," 2016 International Conference on Integrated Circuits and Microsystems (ICICM), Chengdu, 2016, pp. 324-328. DOI: 10.1109/ICAM.2016.7813617

[30] J. S. Mincey, C. Briseno-Vidrios, J. Silva-Martinez and C. T. Rodenbeck, "Low-Power Gm-C Filter Employing Current-Reuse Differential Difference Amplifiers," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 64, no. 6, pp. 635-639, June 2017. DOI: 10.1109/TCSII.2016.2599027

[31] N. Van Helleputte, R. F. Yazicioglu, "Instrumentation amplifier and signal amplification method, " U.S. Patent 9 294 048, Mar. 22, 2016.

[32] J. Koton, N. Herencsar, J. W. Horng, "Differential second-order voltage-mode all-pass filter using current conveyors, " Elektronika ir Elektrotechnika, 2016, vol. 22, no. 5, pp. 52-57. DOI: 10.5755/j01.eie.22.5.16344

[33] A. Yesil, F. Kacar, "Band-pass filter with high quality factor based on current differencing transconductance amplifier and current amplifier, " AEU-International Journal of Electronics and Communications, 2017, vol. 75, pp. 63-69. DOI: 10.1016/j.aeue.2017.03.007

[34] D. Yu. Denisenko, N. N. Prokopenko, "All-pass Active RC Filter Based on Differential Difference Amplifiers, " Patent Appl. RU 2019107341, March 15, 2019. (In Russian).

[35] D. Y. Denisenko, Y. I. Ivanov, N. N. Prokopenko and N. A. Dmitrienko, "Digital potentiometers in the tasks of settings precision analog RC-filters taking into account the tolerances for passive components," 2017 18th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM), Erlagol, 2017, pp. 205-210. https://doi: 10.1109/EDM.2017.7981741

[36] Digital Potentiometers (DigiPOT). URL: https://www.analog.com/en/products/digital-to-analog-converters/digital-potentiometers.html

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 69-81 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Assessment of the Quality and Sustainability Implications of FIFO and LIFO Inventory Policies through System Dynamics

Phoebe Lim Ching[*, 1, 2], Jose Edgar Mutuc[1], John Anthony Jose[3]

[1]*Department of Industrial Engineering, De La Salle University, 1004, Manila, Philippines*

[2]*Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Kowloon, Hong Kong*

[3]*Department of Electronics and Communications Engineering, De La Salle University, 1004, Manila, Philippines*

ARTICLE INFO

Article history:
Received: 20 June, 2019
Accepted: 11 August, 2019
Online: 03 September, 2019

Keywords:
System dynamics
Inventory management
FIFO

ABSTRACT

Perishable inventory management contributes simultaneously to society and the economy, by reducing food wastage and capitalizing on the freshness of goods. For this reason, countless mathematical models have been developed for their effectiveness and cost-efficient management. Yet, the majority of these models can only optimize systems for a limited time frame, allowing for small gains in operations management, but failing to change the recurring patterns in inventory levels. System dynamics (SD) modelling shifts emphasizes these patterns and the recurring decisions that make them. Moreover, the framework has generated insight for other supply chain cases that could not have been derived from a short-term perspective. Thus, the current study now seeks to apply the SD framework in modelling perishable inventory systems, in designing policies that benefit the environment and the economy by reducing waste production and increasing the viability of goods reaching the customer. In particular, it evaluates the impact of opposing issuance policies (i.e. First-In-First-Out (FIFO) and Last-In-First-Out (LIFO)) on perishables to demonstrate the potential of SD in improving perishable inventory management. The simulated results share the sentiments of optimization models, that FIFO will ultimately generate less wastes and incur less material costs. Yet, the simulations also reveal implementing FIFO will result in larger fluctuations in inventory levels, which imply greater inconsistency in age-based quality. These suggest that LIFO would be preferable for quality-sensitive products, while FIFO would be preferable for cases sensitive to waste production. The current study demonstrates the efficiency of system dynamics in generating insight beyond that which can be derived from the existing mathematical models. Future studies may likewise extend this approach in the evaluation of policies on the use of technology in perishable inventory systems, which are the prevalent in present literature.

1. Introduction

In inventory models, perishability is a physical property of inventories that results in the eventual expulsion of these goods from the system as waste [1]–[3]. In terms of social and commercial impact, studies on the management of perishables have prevailing importance. Products necessary to sustain human life, such as food and healthcare, possess this characteristic [4], [5]. In particular, the case of blood banks has received much attention from studies on perishable inventories. Due to the instability of supply and short validity of blood units, a precise and quantified means of managing these inventories are much needed [4], [6]. Blood banks are a strong example of how perishability as a product characteristic can complicate inventory systems, and can be used to generate insight on how other perishable products are best to be managed [6].

Perishability complicates the traditional inventory and supply chain models. The typical components of such models are profit and sustainability measures, capacity constraints, and demand

[*]Corresponding Author Phoebe Ching, phoebe_ching@dlsu.edu.ph

distributions [7], [8]. However, policies that prove optimal for non-perishables are not necessarily optimal for perishables [5], [6]. In some models, perishability is represented as the deterioration of product quality [5], [9]. This can affect demand or the allowable pricing, thereby sharply differentiating the scenario described from inventory systems that do not differentiate between fresh and aged inventories. The transition of inventories from a state of salability to wastes is another differentiating factor. The production of wastes affects system capacity [10], [11], and any wastes generated may incur disposal costs [11], [12]. In this case, perishability affects key system constraints and parameters. Some studies on this subject have been dedicated to studying the differences between perishable and non-perishable systems. These have proven that the differences between perishables and non-perishables result in different optimal policies for each kind of inventory.

The majority of studies in this area are optimization models. From the models themselves, decisions on production quantity and product timing can be determined for a given set of parameters. In turn, managerial insights on managing inventory given similar parameters can be derived from the quantitative behavior of the variables. The optimization of inventory systems is a continuously growing body of research, with models growing in scale and complexity, or through the incorporation of new components such as technology enhancements [3], [13]. Aside from this, research efforts are also being invested in identifying the most efficient algorithms for the model [9], [14]. All in all, advancements in this area are geared towards achieving greater accuracy and numerical exactness. However, a weakness of this approach could be that the emphasis on these aspects of modelling shifts the focus away from insight-generation. While the models are becoming increasingly adept at describing specific scenarios, it may be noted that real inventory systems are constantly subject to external change. Thus, regardless of the level of exactness that is achieved by these models, its direct applicability to real systems will always be limited.

In response to these issues with current modeling approaches, applying the system dynamics methodology can augment these through (1) its applicability to large-scale systems which mathematical modeling may be inefficient for [15], (2) its emphasis on patterns of behavior over numerical exactness [16], [17], and (3) its limitation to constant system elements, as opposed to designing in response to external factors [18], [19]. The first two attributes make system dynamics apt for identifying key areas for solution implementation. Specifically, it can be used to identify the most influential variables in the system. This is insightful for future studies, which can delve more specifically in those areas. The third attribute aids in generating insight for recurrent managerial policies as opposed to period-specific decisions. These are more aligned with real managerial practices which, where rules and guidelines are favored to specific values.

In this study, the system dynamics (SD) methodology is applied to the case of perishable inventory systems. This is done for the purpose of demonstrating how the SD methodology can be applied in designing managerial guidelines, with proof that use of these guidelines will generate significant improvement for the system. As SD has already generated insights for general supply chain settings [20], [21], this study now investigates how perishability can alter the results, given that perishability has been proven to be a significant differentiating factor in previous studies [22]–[24]. Through its high-level perspective, the current study aims to identify impactful areas for solution development in the perishable inventory scenario. In this way, it can contribute to the area of study with suggestions for future studies focusing on these specific areas.

2. Literature Review

2.1. Inventory System Dynamics

There seems to be a limited research concerning inventory system dynamic (ISD) studies, thus there is potential to address multiple unexplored research gaps [19]. The challenge then is to find a purposeful gap to address, and justifying why system dynamics (SD) methodology would be more appropriate over methods that allow for more quantitative accuracy and optimization. In this respect, a review of the existing ISD studies is conducted to (1) learn the research objectives that prompted the use of SD, (2) understand how the ISD study relates to other inventory models using different methods, (3) understand how managerial and policy-making insights are derived from such studies, and (4) determine the main contribution of individual ISD studies.

In a literature review on SD applications in renewable energy supply chains, Saavedra 2018 observed that related studies seems to have objectives of either (1) improving understanding of the supply chain; or (2) developing modelling and simulation approaches. As a framework for understanding, SD is effective in depicting complex cause-and-effect chains and various sectors through its diagram-based modelling approach. From its diagramming methods alone, insight can be derived from the network of interrelationships between variables. On the other hand, as a simulation tool, the effectiveness of SD in providing insight is still being validated. This is due to its lack of quantitative accuracy and the lack of studies in this area.

Generally, a well-accepted method of applying SD is to use the diagramming approach to identify problematic areas, and apply sensitivity analysis to determine the most efficient area for solution [25]. This is the approach taken by studies seeking to apply the SD framework in a specific context. On the other hand, in studies that assess the methodology itself, SD has been coupled with various methods to augment its lack of numerical soundness, and the lack of perspective in other modeling approaches [19].

Given its frequent applications in sustainable supply chain models, there is often a part of SD studies that compares its methods with the likes of operations research and life cycle assessment (LCA), and validates that it can produce insights beyond those that can be derived from other modeling approaches. In comparing supply chain models using SD with those using other methodologies, Saavedra 2018 concluded that a major difference between SD and LCA is its focus on the behavioral patterns of variables within a model, such as the intermittency and variability of its values. On the other hand, the focus on LCA is on aggregate performance, such as the total volume of emissions [18].

The SD methodology also allows for a wide range of variables to be modeled and simulated efficiently, given its lack of numerical

exactness. In a literature review, it was observed that most SD supply chain models had a macroscopic perspective consisting of economic and environmental impact. Next in quantity were SD studies with an inter-organizational scope, and last were studies with an intra-organizational scope. On the contrary, for operations research-based studies, the order from highest to lowest volume of publications is intra-organizational, inter-organizational and macroscopic [19]. This can alter the kinds of policies that may be developed and simulated using each method.

SD can be a means of extending optimized parameters into concrete policies, particularly in cases with multiple parameters of a different nature. Song 2018 created a model to simulate the application of optimized parameters for technology, environment, energy and economy in a system where they are interrelated. These parameters were derived using data envelopment analysis (DEA), which augments the lack of numerical accuracy in most SD models. SD then strengthens the policy-making approach by highlighting the physical relationship between variables.

Scenario analysis, such as applying different parameters to represent different cases, is a quick means of validating the model's logic and policy development. In validation, it allows for multiple real-life case parameters to be applied to prove the validity of the causal relationships and the completeness of the model. In policy-making, a number of alternative solutions can also be generated and simulated, by shifting model parameters [18].

Overall, SD studies have contributed to the field of study with insight for how certain resources can be managed with optimal results [25]; knowledge on the opportunities and limitations of simulation methods [18]; and the creation of hybrid models combining SD with more math-based methodologies [19].

2.2. Perishable Inventory Systems

Perishable Inventory Systems (PIS) have received much attention in literature as early as the 1990s, up until recent time [5]. Earlier studies simply sought to incorporate 'perishability' as an attribute of inventory. Nonetheless, more recent studies seek to determine the kinds of managerial strategies and policies that may be applied to perishable inventories [2], [26]

In a literature review conducted in [5], studies from 2012 to 2015 were segregated according to the different policies being implemented in each study. It was found that the top policies being studied during that period were (1) credit and payment policies, (2) supply chain information sharing, and (3) pricing and markdown problems. There were notably few studies on the use of the introduction of advanced monitoring technology, and on classic issuance policies (i.e. FIFO, LIFO).

Since then, there has been an increase in models on the integration and use of monitoring technology in perishable inventory systems. [6] have introduced the use of RFID technology in monitoring of unit age and cost to the dual-sourcing problem for perishables. This was evaluated against fixed and exponential product lifetime scenarios, to gain insight on specific potential applications for this strategy. More recently, [3] have developed an IoT-based model for monitoring perishables in shared storage. Given the potential for cross-perishability, which refers to the increase or decrease in deterioration rate resulting from shared storage, information monitoring in this manner could then allow for lowering of deterioration costs, inventory level, and quality loss for the system. On a two-stage optimization model for pricing and replenishment of PIS, the recommended extension is also the incorporation of RFID and point-of-sales data in decision-making [7].

It seems to be reasonable for research to progress in this direction given the rise of Industry 4.0 [27], [28]. However, the implementation of advanced systems, similar to the aforementioned, are still dependent on fundamental decisions in inventory management. This is evidenced by the way that these technologies are modeled in conjunction with inventory management policies from classic literature [29]. Yet, these policies are still among the top policies being studied from 2012 to 2015, as identified by [5].

Over the years, there have been few changes to the performance functions used for perishable inventory models which enable ease of comparison. Financial measures such as profit and cost are still used in almost every PIS study. The percentage of profit growth or cost reductions resulting from the implementation of certain policies is typically used as the measure of performance in these cases [4], [6], [26]. Profit can be affected by perishability through the decline of demand alongside quality [30]. It can also manifest as disposal costs [2], [8] and wastes [7].

Wastes differ from financial measures as they are a physical outcome of the PIS policy. By quantifying waste production, a model can also be assessed for its sustainability [10], [11]. The minimization of waste, separate from its contributions to cost, is used as an objective in a number of studies [5]. In this way, the model can differentiate between financial and physical outcomes, given that the volume of wastes would not always be economically relevant in all cases.

Another physical outcome of PIS policy is the service level, or the ability to meet customer demand [28]. In the model of [7], service level is the probability of stockout in the replenishment cycle resulting from the policy. Monitoring the service level of PIS prevents over-emphasis on cost efficiency, in a more tangible way than revenues. Aside from this, the model of [4] has made use of the average age of inventories being sold, to represent the quality of the products.

The perishable nature of inventories in PIS is represented as a finite lifetime. This leads to wastes at the end of the product lifetime [11], [12], and declining demand over time for a specific unit [5], [9], [31]. In most PIS models, a scenario analysis is conducted for various rates of deterioration. Generally, the higher the deterioration rate, the greater the amount of waste [7]. However, this can also vary depending on the stochastic distribution of a product's lifetime, and the kind of policy being implemented. The model of [6] demonstrates that dual-sourcing can produce greater cost savings in cases where lifetime is fixed with a shorter average, and when lifetime is exponentially distributed with a longer average.

Generally, a number of policies have proven effective in managing PIS, with immediate improvements in cost and material efficiency. Optimization models demonstrate how such policies can be applied with optimal results for a given set of parameters.

However, the existing methodologies have proven to be inefficient for policy-to-policy comparison, given the variability in model structures and sensitivity to parameters. Thus, the basis of determining best practices based on the existing studies could still be improved.

2.3. FIFO and LIFO Issuance Policies

Issuance policies are an integral part of an inventory management problem. In general, the parameters under study in such a problem are pricing, inventory costs, and deterioration rate [7], [32]. These parameters are present in almost all studies and are used to evaluate specific policies being applied, or intended for application, on real inventory systems. These policies can include issuance policies (i.e. FIFO and LIFO), which relate to the order at which goods of the same kind at different ages are dispatched to the customer. Based on the review of Janssen (2016), issuance policies appear to be an overlooked area in literature. However, based on the current review that was conducted, it seems that most studies include issuance policies as a parameter, instead of treating it as a focus of the study. The review of literature in this area is thus segmented in two ways: (1) studies considering issuance policies as a parameter, and (2) studies explicitly focused on evaluating or identifying optimal issuance policies.

Using a policy as a parameter means that it can be changed to allow the analyst to see various scenarios. [23] sought to determine how the centralization of inventory could be beneficial in managing a perishable supply network. Specifically, the case under consideration was a blood supply chain with inventory located at a single blood bank and being supplied to multiple hospitals. FIFO was explicitly modeled for this case, as this is the standard for inventories in the healthcare service industry. [10] adapted a modeling and simulation analysis methodology determine the impact of a closing day constraint for the inventory of stores. The objective of the study was to determine if the existence of a closing day would lead to more aged inventory as closing day approach, as is the case in real grocery store inventories. It was found that the closing day constraint did result in simulated behavior being closer to the real behavior of inventories. In another study, [11] assessed the quantitative impact of micro-periodic inventory management policies, for how these may potentially improve costs and waste production. Under this approach, a single day was broken into morning, midday, afternoon and evening periods. This likewise made use of a mixed FIFO and LIFO approach. It was found that costs were reduced significantly from micro-periodic management, as compared to daily planning.

One reason why a mixed approach is taken in most studies is so that the exact impact of each policy will not have to be considered. This recognizes that there are, in fact, real differences between the effects of applying each policy, particularly on perishables. The models that explicitly analyze the issuance policies seek to determine what these effects are and how they can be managed most effectively to meet the goals of the system. [12] compared the viability of FIFO and LIFO for managing perishable inventories using a modified Economic Manufacturing Quantity. Said model was modified to give consideration for fixed order quantity and joint ordering policies. This revealed that LIFO could prove optimal under linearly decreasing price structures. This finding is relevant given that it is generally accepted that FIFO has superior performance on perishables [19], [23].

In most cases, issuance policies are modeled as a two-warehouse problem. Under LIFO, the first inventories to be sold would come from the rented warehouse; while under FIFO, the first sales are taken from the owned warehouse. A model for this problem was developed by [32], who assessed each issuance policy for its overall costs, given deterioration and holding costs. It was found that FIFO was less expensive when the cost parameters were lower for the owned warehouse than in the rented warehouse. As this is the case for most entities, it may be concluded that the study recommends FIFO under LIFO given the parameters that were considered. [9] extends the study of issuance policies for two-warehouse inventory systems to the use of modified FIFO (MFIFO) and modified LIFO (MLIFO) policies. Under the MFIFO rule, the FIFO rule is followed as far as the inventory in demand is available. When unavailable, the backorder gets assigned to the lowest priority level in the next period [28]. The model was structured in such a way that allowed for easy comparison between policies by relaxing a few constraints. It was found that when deterioration rate was different between the rented and owned warehouses, holding cost could make a significant difference in the optimality of each policy.

MFIFO and MLIFO are already evolutions of the existing issuance policies. It is understood that FIFO and LIFO each have weaknesses, and thus improvements can be made in their exact rules. [31] proposed the development of a new issuance policy entitled Allowance-in-Fraction-Out which has specific rules for when to apply FIFO and LIFO. This is likewise applied to the two-warehouse inventory problem, wherein the decision to switch is based on the fraction of rented warehouse inventory to owned warehouse inventory. Across various cases of deteriorating behavior, value of information, and perishability, it was found that the mixed method could have benefits in transportation costs and the occurrence of sub-replenishments.

In recent literature, most studies treat issuance policies as a parameter. Moreover, most studies assume a mixed issuance policy structure to protect the model from the effects of using any specific policy. However, this suggests that inventory will be taken randomly from a company, which is understandably not the case in inventory systems. The most advanced study in this area seeks to determine the appropriate rules for using FIFO and LIFO, and managing their negative repercussions. This would imply that further study is needed of the explicit behaviors of FIFO and LIFO to determine how they may be managed productively.

3. Methodology

In this study, the system dynamics (SD) methodology is used to model a perishable inventory model, and simulate the application of certain policies and methods. This method is efficient for analyzing patterns of behavior, as it emphasizes long-term patterns over quantitative accuracy [16], [17]. It is also efficient in expanding the scope of a model to encompass all variables that could have significant feedbacks on the key variables under study [20]. SD emphasizes the importance of feedbacks in a system, showing how the system's behavior is actually the result of these internal feedbacks, and not external factors [16]. In problem solving, this maintains the focus on constant factors,

rather than externalities which may or may not be present. In this way, the SD methodology is well-suited for evaluating policies in the long-term, and managing any negative repercussions that may result from their application.

The SD framework is applied with the intention of designing policies that will improve the sustainability and quality goals of the system. Most perishable inventory models attempt to address sustainability goals by having waste minimization as one of its objectives [30], [33]. The connection of waste minimization to sustainability implies that through waste reduction, resource consumption is also reduced. On the other hand, quality goals can be met in terms of age-based quality, by minimizing the time a unit spends in the system from production to sale [4]. Age or, conversely, freshness is an apt representation of quality for perishables given the fact that (1) perishables are typically fast-moving consumer goods which vary little from unit to unit, and (2) perishables will definitely degrade in some way over time, making age a significant component of quality, even if it is not the sole component.

The SD methodology classifies variables as stock, flow, auxiliary or exogenous based on their role in the model. Stock variables accumulate over time based on previous values; flow variables represent this rate of accumulation; auxiliaries serve as the transition of stocks to flows; and exogenous variables are parameters whose values are externally determined. A discussion of how the variables were classified, and of their respective equations is given in Section 3.1. Following this, the variables determined in Section 3.1 are organized in a causal loop diagram in Section 3.2. This represents the relationship of the variables in an enclosed system. The variables and loop structure are then integrated in a stock-flow model for simulation and analysis in Section 3.3.

3.1. Variables Definition

Stock variables are those which increase or decrease in relation to their value in a previous period. This can be used to represent inventory levels which physically change over time. In this model, inventory is divided into two categories based on age. Hence, there is a variable representing Fresh Finished Goods Inventory (FFGI) and Previous Finished Goods Inventory (PFGI) respectively, as levels. By logic, FFGI is a function of production, sales, and leftover inventory in the current period (t). Since it is not an option to dispose freshly-produced goods in this model, there is no difference between the equation for FFGI in the perishables and non-perishables case.

$$\text{FFGI}(t) = \text{FFGI}(t\text{-}1) + \text{Prod.}(t) - \text{Sold F. Inv.}(t) - \text{Held Inv.}(t) \quad (1)$$

PFGI represents inventory carried over from previous periods. In the perishable's scenario, a fraction of inventory becomes waste at the end of each period. Mathematically, these differences between the perishables and non-perishables case are given in equations 2 and 3 respectively.

$$\text{PFGI}(t) = \text{PFGI}(t\text{-}1) + \text{Held Inv}(t) - \text{Sold P. Inv}(t) - \text{Wastes}(t) \quad (2)$$

$$\text{PFGI}(t) = \text{PFGI}(t\text{-}1) + \text{Held Inv}(t) - \text{Sold P. Inv}(t) \quad (3)$$

Aside from variables that physically change in value, stock variables are also apt for representing human cognition in system dynamics models. Cognition is influenced by the occurrence or lack thereof of certain events. In consequence, these strengthen or weaken a certain idea, which in turn produces reactions affecting the other variables in the system. Such variables are usually not included in modeling as there is no definite metric to represent them. However, the soft emphasis on numerical accuracy in system dynamics allows for such variables to bear the same weight as tangible variables (i.e. inventory, cashflow).

In this model, there are two cognition variables, both represented as 'perception'. One of these is Perceived Quality (PQ), representing the customer's flawed knowledge on the actual quality of goods produced by the producer. System dynamics recognizes the feedback delay in perception of performance, which results in there appearing to be no effect despite efforts toward improvement. PQ changes by a positive or negative value, depending on whether the quality in recent periods has been good or bad. However, PQ itself will always be a normalized value as the change happens in relation to previous values of PQ. Conversely, Perceived Demand (PD) represents the producer's flawed knowledge on the actual volume of customer demand. In real applications, this would represent the degree of error in demand forecasts [34]. Likewise with PQ, PD changes by a positive or negative value depending on the recent demand values.

The change in the value of stock variables at each period is defined as a flow variable. A stock can have inflows that add to its value, and outflows that reduce its value. In the case of FFGI, the inflow rate is Production (Prod), while the outflow rates are Held Inventory (Held Inv) and Sold Fresh Inventory (Sold F. Inv). Prod is based on the value of PD, which represents the use of forecasts in deciding on production quantities [32], [35].

$$\text{Prod}(t) = \text{PD}(t) \quad (4)$$

Held Inv is equal to the current value of FFGI, as all fresh inventory that is not sold will be carried over to the next period as PFGI. Held Inv represents the transition of FFGI into PFGI by being an outflow of the former into the latter. The also represents the aging of products in the system. In essence, there are only two age classes of goods--fresh and previous. While this may not capture the exact age of the inventory pieces, it is nonetheless sufficient to show the impact of FIFO and LIFO. Further segmentation by age group may further complicate the model without generating additional insight to compensate for the effort.

$$\text{Held Inv}(t) = \text{FFGI}(t) \quad (5)$$

FFGI and PFGI each have outflows in the form of sales, and the equations for these vary between the FIFO and LIFO cases given that the former prioritizes selling PFGI while the latter prioritizes selling FFGI [Parlar et al, 2011]. These are Sold Fresh Inventory (Sold F. Inv) and Sold Previous Inventory (Sold P. Inv). Mathematically, the prioritized inventory class is directly influenced by Customer Demand (Cust Dem), while the deprioritized class is a function of the sales for the prioritized class. The equations for the FIFO case are given in equations 6 and 7; and those for LIFO are given in equations 8 and 9.

$$\text{Sold F. Inv}(t) = \text{MIN}(\text{MAX}(\text{Cust Dem}(t) - \text{Sold P. Inv}(t), 0), \text{FFGI}(t)) \quad (6)$$

$$\text{Sold P. Inv}(t) = \text{MIN}(\text{Cust Dem}(t), \text{PFGI}(t)) \quad (7)$$

$$\text{Sold F. Inv}(t) = \text{MIN}(\text{Cust Dem}(t), \text{FFGI}(t)) \quad (8)$$

$$\text{Sold P. Inv}(t) = \text{MIN}(\text{MAX}(\text{Cust Dem}(t) - \text{Sold F. Inv}(t), 0), \text{PFGI}(t)) \quad (9)$$

For PFGI, there is an additional outflow in the form of Wastes. This variable is only present in the perishables case, as inventory can only be accumulated in the case of non-perishables. The variable Wastes is a function of the residual PFGI and a given rate of deterioration:

$$\text{Wastes}(t) = (\text{PFGI}(t) - \text{Sold P. Inv}(t)) / \text{DelayP} \quad (10)$$

For the cognition variables Perceived Demand and Perceived Quality, it was previously mentioned that the change could be positive or negative depending on the recent values of the perceived information. These changes are the flow variables Change in Perceived Demand (CIPD) and Change in Perceived Quality (CIPQ). Mathematically, a positive change would result if the current value is greater than the perceived value, and a negative change if otherwise. In effect, the current values are normalized against previous values. A feedback delay is used to represent the limited influence on current values against long-standing perception. In equation form:

$$\text{CIPD}(t) = (\text{QOGS}(t) - \text{PQ}(t)) / \text{DelayPQ} \quad (11)$$

$$\text{CIPQ}(t) = (\text{Cust Dem}(t) - \text{PD}(t)) / \text{DelayPD} \quad (12)$$

Aside from stock and flow variables, a system dynamics model contains auxiliary variables. These are components of flow variables. It can serve as a means of breaking down potentially-complex equations for flows, both for ease of understanding and to highlight the behavior of specific components. Auxiliaries are derived from the behaviors of stock variables, and can link these to flows in the model. Unlike stock variables they are not a function of their previous value. While they do not change the overall system behavior, they are nonetheless important for clarity in representation.

The model contains three auxiliary variables: Quality Multiplier (QM), Customer Demand (Cust Dem), and Quality of Goods Sold (QOGS). QM is an example of an auxiliary that exists for ease of understanding. Its value is equal to the value of Perceived Quality (PQ), yet the conversion is made to show the logic in making PQ part of the function of Cust Dem. Hence:

$$\text{QM}(t) = \text{PQ}(t) \quad (13)$$

Cust Dem is a function of two variables, QM and a randomized demand value that is exogenous to the system (Natural Range of Customer Demand, NRCD). In effect, the perception of good quality positively influences demand by a margin, while the perception of poor quality has an inverse effect. As indicated by its usage in the previous equations, customer demand is the actual amount of purchases made by the customer in a single period. This consequently affects the sales variables, Sold Fresh Inventory and Sold Previous Inventory. The equation for this variable is given as:

$$\text{Cust Dem} = \text{NRCD} * \text{QM}(t) \quad (14)$$

QOGS represents the quality of goods sold for a current period. In this model, the quality of the perishable product is solely based on freshness. Its value is derived from the amount of Sold Fresh Inventory (Sold F. Inv) in relation to Customer Demand (Cust Dem). Higher proportions of Sold F. Inv in Cust Dem would thus equate to better quality. As information on the quality of the producer's goods is transmitted to the customer through sales, QOGS has a direct but gradual influence on PQ [30], [36].

$$\text{QOGS}(t) = 1 - \text{Sold P. Inv}(t) / \text{Cust Dem}(t) \quad (15)$$

The variables that are not directly influenced by any variable in the model are exogenous variables. The belief that the system's behavior is primarily influenced by the feedbacks between its endogenous variables, rather than exogenous variables, are at the core of the system dynamics methodology. Hence, exogenous variables are typically represented by constant parameters, as proof that their value is of minor significance to the system's behavior.

Feedback delays are a common example of an exogenous variable. These represent delays in the transmission of information across the system. In equation form, these are variables that limit the flow of information to a cognition variable. In this model, these delays exist in the flow variables for Perceived Quality and Perceived Demand, as Delay in Perception of Quality (DelayPQ) and Delay in Perception of Demand (DelayPD). Physical delays are another common exogenous variable, representing rates of transition from one state to another. In this model, the physical delay is the Delay in Perishing (DelayP). This delays the transition of inventory into wastes, according to a supposed validity period.

An exogenous variable is also used to represent additional sectors of the system that can be simplified for ease of analysis. This can be done when the variables included in that sector are not explicitly being analyzed. In this model, the Customer Demand (Cust Dem) is understandably an outcome of Perceived Quality. Yet, it is also influenced by the availability of other options in the market; the demand of the customer's own customer; and numerous other factors depending on the specific case and scenario. In effect, Cust Dem will have a degree of variability due to these other variables. To represent this, a Natural Range of Customer Demand variable is introduced as a random variable with a given range.

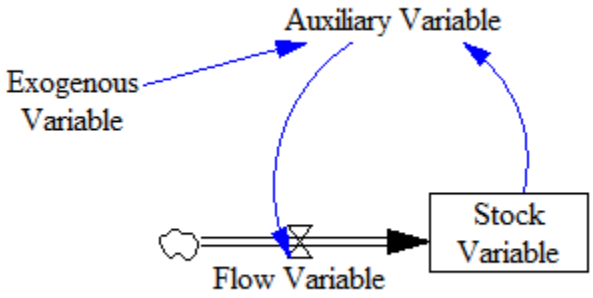


Figure 1: Basic Structure of a System Dynamics Model.

3.2. Interactions Between Endogenous Variables

The greatest differentiating factor of the SD methodology is that it organizes variables into feedback loops. This follows the principle that a system's behavior is influenced more by the system's internal reactions than any external factor. A visual representation of the role of each kind of variable in a feedback loop is given in Figure 1. All loops include at least one stock variable, which shows how the current outcomes of a system result from behavior in a previous period. This progression of cause and

effect leading from past to current behavior is demonstrated through a progression of auxiliary variables. It is shown in Figure 1 that auxiliary variables are the link from stock variables to flow variables, which mathematically means that the past values of a variable would change its next-period values. Finally, the exogenous variables are externally-determined and not influenced by any other variable in the system. Hence, there is no feedback leading to an exogenous variable.

To demonstrate the cause-and-effect relationships in a SD model, variables are arranged in causal loops without differentiation between variables. The purpose of this is to emphasize the feedbacks within the system prior to explicit modeling. In developing the causal loop diagrams for this study, it became apparent that there would be different causal loops for the FIFO and LIFO cases respectively, due to the difference in inventory type affected by sales. Specifically, demand directly impacts aged inventory in the FIFO case, and fresh inventory in the LIFO case. Aside from this, both cases have the same loops, representing a common PIS structure to enable comparison.

In PIS models, demand changes based on the age of inventory given the fact the that goods are deteriorating over time [5], [9]. In this model, a quality variable determined by age is positively linked to demand, meaning that demand will increase as quality improves. Quality is positively affected by the quantity of fresh inventory, and negatively affected by the quantity of aged inventory. This is how the reduction in demand caused by low-quality or aged goods is interpreted in the model. The resulting demand has a delayed impact on production, representing how demand from previous periods can determine production levels in the coming periods through forecasts. Production quantities then add to the level of fresh inventory, and fresh inventory becomes aged inventory, by natural progression [4], [11], [37].

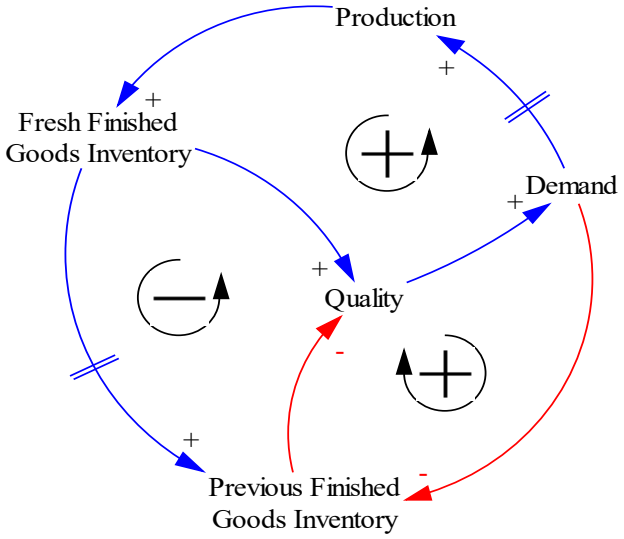


Figure 2: Causal Loop Diagram for FIFO Scenario.

The amount of positive or negative relationships in a feedback loop determine the 'positivity' or 'negativity' of that loop. Even numbers of negative relationships create a positive loop, which prompts positive exponential growth in the variables involved. Odd numbers create a negative loop with negative exponential growth. In system dynamics, systems are understood to be composed of these positive or negative loops, which balance each other out. Following this perspective, the LIFO scenario is more balanced than the FIFO scenario, as it contains equal amounts of positive or negative loops. This suggests that behavior is more consistent under the LIFO scenario, than the FIFO scenario.

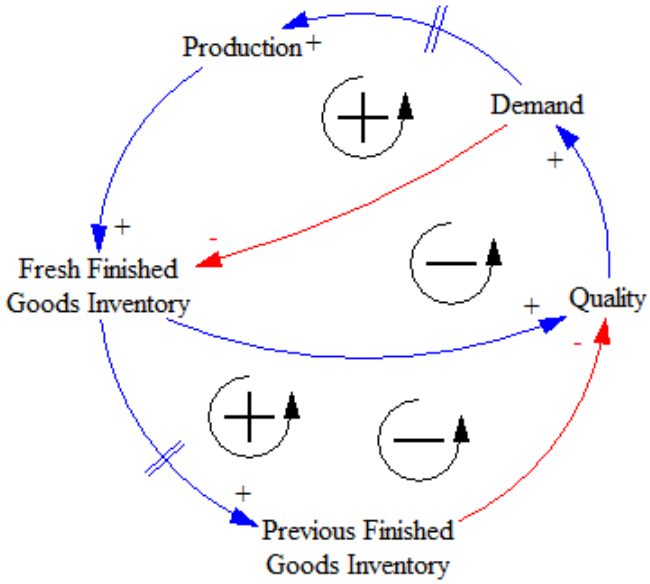


Figure 3: Causal Loop Diagram for LIFO Scenario.

3.3. Simulation Model Development

Using VensimPLE software, a system dynamics (SD) model was developed to simulate the postulated behaviors of each variable and analyze their interactions. The general form of the SD model is shown in Figure 4. This shows the variables identified in Section 3.1 in their proper form, and arranged in the feedback loops derived in Section 3.2. There are three major sections in the model: (1) the physical inventory system, which is also the main problem variable under study, (2) the perception of quality, representing the impact of the physical system on human receptors, and (3) the perception of demand, representing the human decisions which govern the physical system. These form a single loop that determines the behavior of the system.

The model parameters are feedback delays, which can be adjusted to represent various scenarios. The Delay in Perishing represents the validity period of the goods, which can vary depending on the nature of the goods. In PIS for blood banks, this parameter strongly impacts system behavior given the brief and stringent validity period of the inventory [23]. For disaster relief operations where goods are hardier, the impact of this parameter is less strong, given that the goods are practically non-perishable [2]. For the same reason, there are few disaster relief supply chain studies which consider waste and disposal.

The Delay in Perceived Quality represents how evident differences in quality may be to the customer. Smaller batch sizes and proximity to the end-customer can equate to shorter delays. In retail settings, which have this characteristic, demand is typically a function of quality [11]. On the other hand, upstream suppliers consider perishability in terms of waste and disposal costs, as opposed to impact on demand [23], [28].

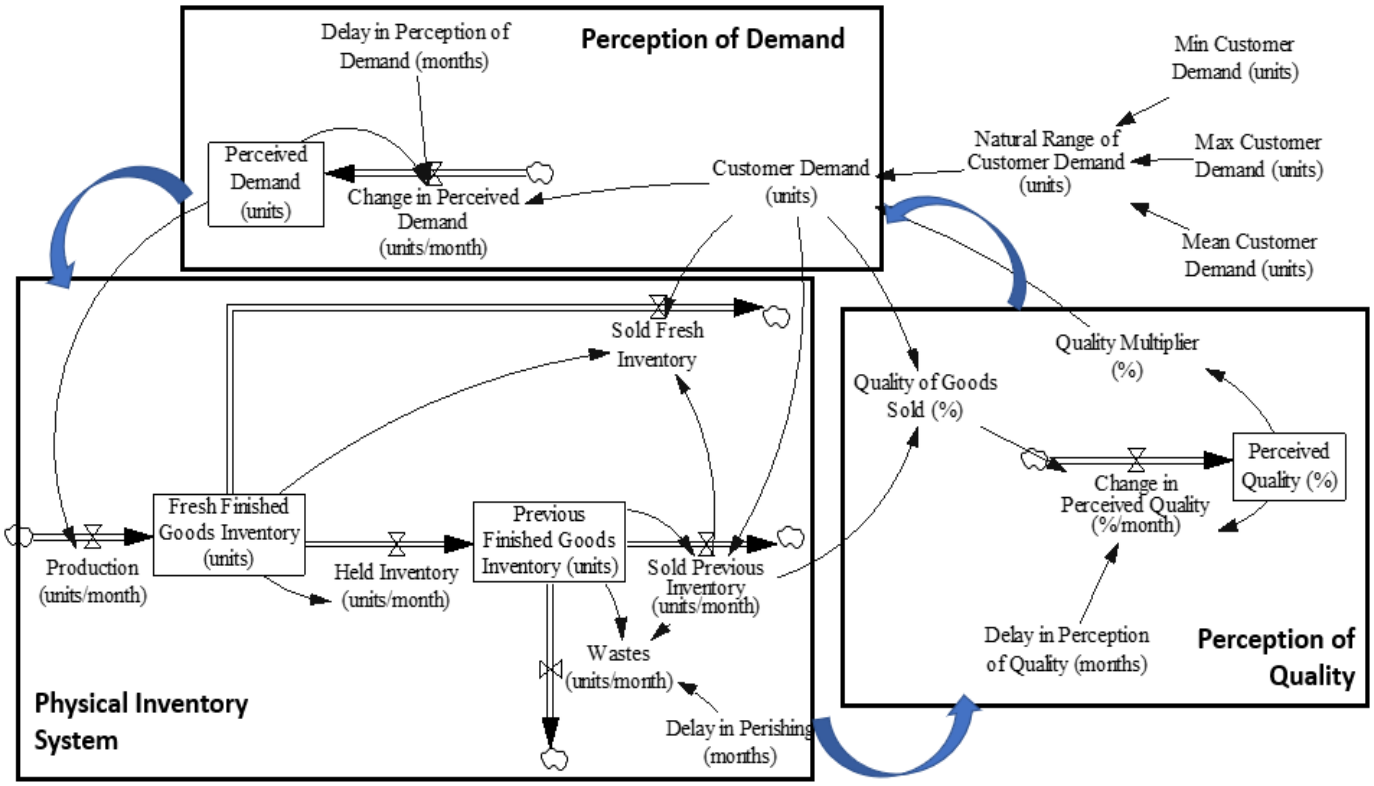


Figure 4: Structure of Inventory System.

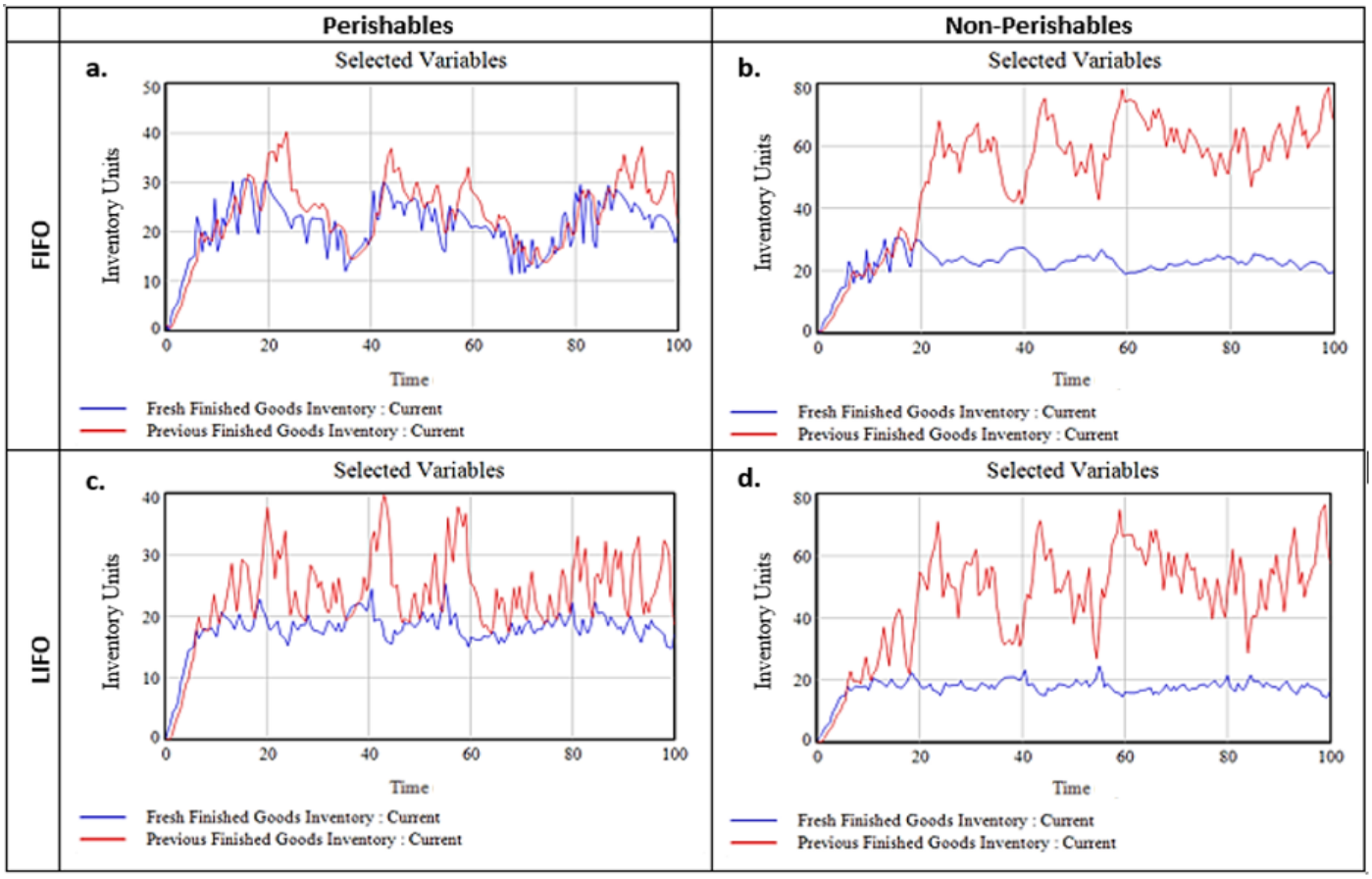


Figure 5: Simulated Fresh and Aged Inventory Levels.

The Delay in Perceived Demand represents inaccuracies in the forecasting techniques with regards to real demand. There are countless studies on forecasting techniques and replenishment policies, and each have proven effective in specific contexts and following specific demand distributions. In such cases, the policies are evaluated by the resulting service level [4], [7]. The current model extends this to examine the reactions between the performance of the policy and other components of the system.

Following the structure in Figure 4, four models were created for each of the following scenarios: (1) FIFO applied to perishables, (2) FIFO applied to non-perishables, (3) LIFO applied to perishables, and (4) LIFO applied to non-perishables. The main difference between the perishables and non-perishables scenarios is the lack of a Waste outflow for aged non-perishable inventories. In this case, the only possible outflow is through sales. Given that the emphasis of this study is on perishable inventories, the policies that may trigger the disposal of non-perishables are not considered.

By logic, FIFO and LIFO differ based on the priority of inventories to be sold. These differences were given in equation form in Section 3.1 and in diagram form in Section 3.2. It is necessary to generate a different model to represent each case given that the feedback relationships differ. Hence, this is not a case of merely adjusting parameters to represent a case, but making adjustments to the model parameter. Aside from priority of sales, the structure of the model given in Figure 4 is maintained for all versions, to allow for the results to be comparable with one another. The simulation results depict values for the behavior of variables over time. This is analyzed in graphical form to assess the model for patterns and identify areas for problem-solving.

4. Simulation Results

All inventory models have the objective of minimizing inventory levels, whether directly or indirectly. In optimization models, inventory levels are often linked to cost objectives through holding costs and disposal costs [15], [35]. In this way, minimal inventory levels becomes the most economic scenario for an inventory system. In simulation methodologies, inventory is physically monitored with the unstated understanding that it affects other performance indicators, such as cost and wastage [5], [7]. The system dynamics methodology is aligned with this perspective. This study likewise monitors physical inventory levels as a performance measure, and evaluates its effect on outcome and quality.

As mentioned in Section 3.1, the perishable aspect of inventory is modelled by the transition of goods from a fresh state (Fresh Finished Goods Inventory) to an aged state (Previous Finished Goods Inventory). The model differentiates between inventory in only two ways, fresh and aged, based on their impact on the system. The system dynamics methodology is an efficient means of monitoring and analyzing such variables for this purpose. In mathematical models, inventory is represented in a very exact manner, indexed by time of production and sale [28], [31]. This can take the focus away from the more important implications of inventory.

Four scenarios are simulated: FIFO applied to Perishables, FIFO applied to Non-Perishables, LIFO applied to Perishables, and LIFO applied to Non-Perishables. Each scenario is evaluated by its resulting inventory level. The analysis can be segmented into three parts: (1) A comparison between the Perishables and Non-Perishables scenarios for logic-checking, (2) A comparison of the FIFO and LIFO scenarios for policy evaluation, and (3) Determining the optimal issuance policy to apply to perishable inventories.

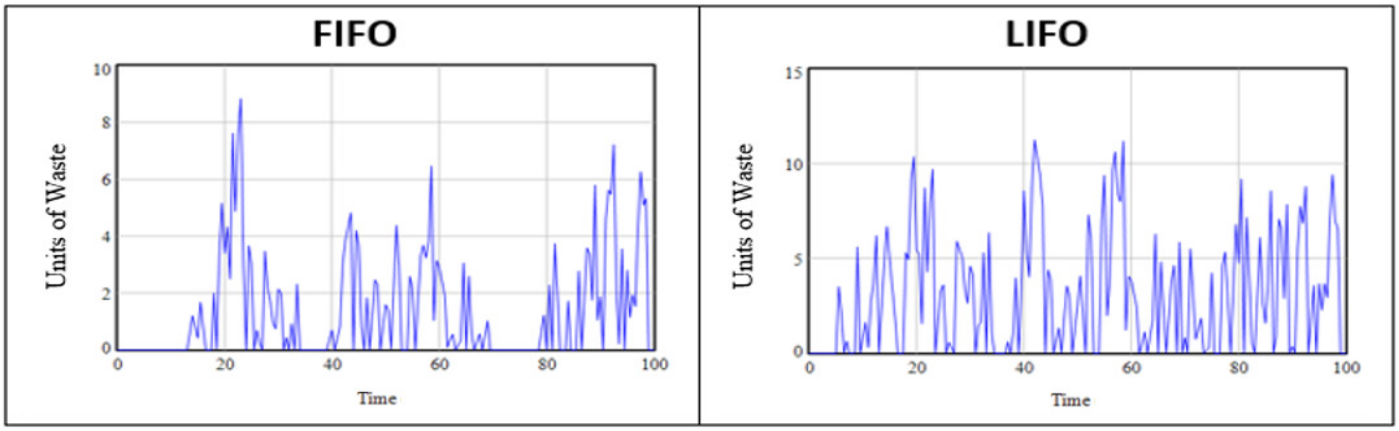
4.1. Comparison of Perishables and Non-Perishables Scenarios

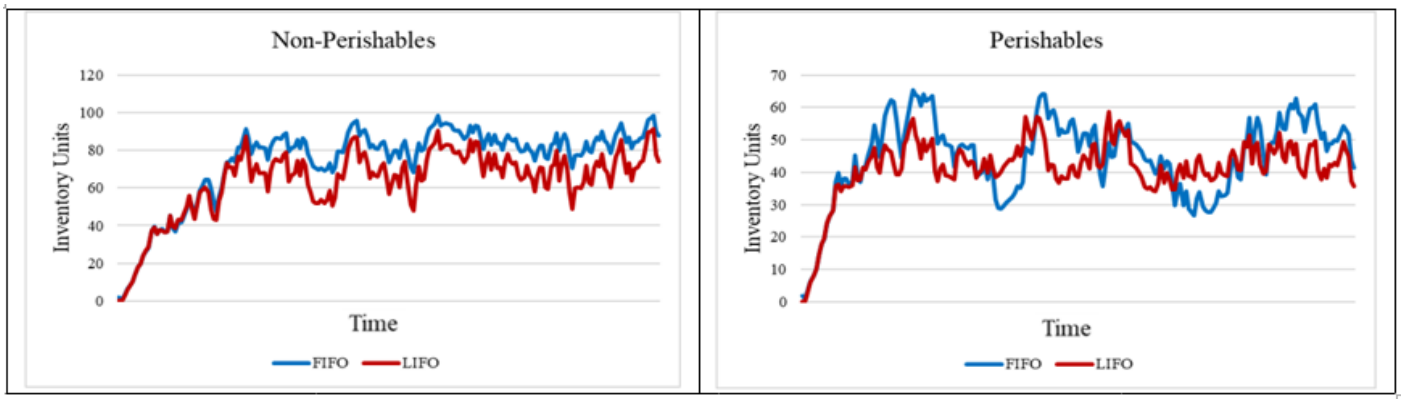
In Figure 5, sections a and c contain the simulated inventory levels for the perishables scenario, while sections b and d contain the results for non-perishables. In all cases, inventory levels exhibit oscillating behavior which is aligned with the behavior being demonstrated in other inventory models and supply chain industry studies [36]. The perishables and non-perishables scenarios differ in the long-term trend being demonstrated by inventory under each case. Under the non-perishables scenario, inventories exhibit an upwards trend, indicating the tendency for inventories to accumulate. Under the perishables scenario, inventory levels are stable in the sense that they exhibit no upwards or downwards trend. The volume of inventory in the perishables scenario is naturally regulated by the potential for wastes, which prevents accumulation. On the other hand, there is no negative feedback to control the volume of inventories in the non-perishables scenario, resulting in accumulation.

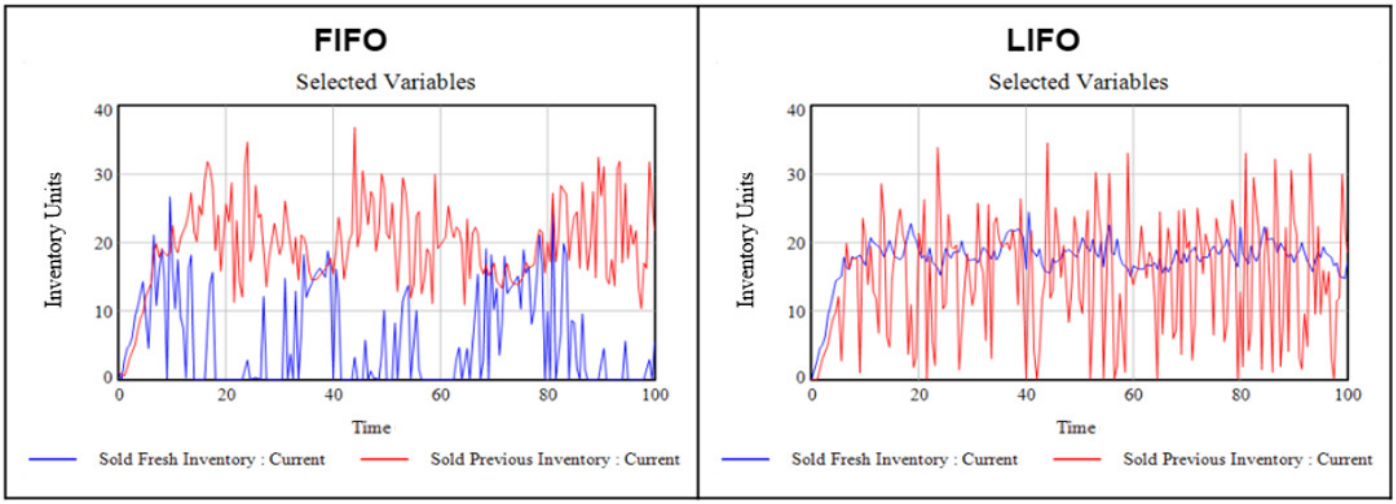
The simulated behavior for non-perishables is aligned with the behavior of most forecast-based non-perishable inventory systems. There is evidence to support this in the management of mechanical parts inventory, which has a large amount of excess [38], [39]. The accumulation may also be attributed to the fact that production quantity was based solely on sales forecasts, without any regard for current inventories. This mimics decision-making for perishables more closely. Since goods manufactured in one period do not have the same external value as goods manufactured in a different period, past inventory cannot be fully utilized to supply current demand. Therefore, decision-making parameters such as current inventory and safety stock generally carry more weight in models for non-perishable inventories.

The simulated behavior for perishable inventories follows the behavior of traditional inventory models (e.g. EOQ), which oscillate without any positive or negative trend. Inventory accumulates because of inaccuracies in forecast. Yet, because a fraction of perishable inventories become waste at the end of each period, inventories cannot accumulate. Instead, excess inventory from inaccuracies in in forecasting eventually leave the system as waste. This is not favorable given that the amount of inventory that become Wastes are just excess in a different form. It may only be said that inventory is more effectively-handled under the perishable scenario if the volume of its Wastes is not increasing in the same way excess inventory does for non-perishables. The simulated behaviors for waste, although oscillating, exhibit no increasing trend (see Figure 6). For these reasons, it may be concluded that perishable inventories do not have as strong of a tendency towards overproduction.

It may be concluded based on the simulated results for Previous Finished Goods Inventory that the main issue in the management of perishable inventory is not preventing surpluses and shortages. Under common inventory management policies, perishability as a product characteristic results in an inventory system that is less







prone to such outcomes. Inventory models with the objective of determining an optimal stockout level could thus be redundant, or even counterproductive to the system's natural ability to regulate itself. On the other hand, the system could benefit from the minimization of its peak inventory levels, or all-around variability.

As the simulated behaviors follow the trend described by real industry data and by literature, the ensuing analysis can be deemed valid. A comparison may now be conducted between the FIFO and LIFO policies given the system structure. Based on the simulated values in Figure 5, it is apparent that the choice of issuance policy hardly makes a difference on inventory levels. On the other hand, there are clear differences between the FIFO and LIFO scenarios in the perishables case.

4.2. Comparison of FIFO and LIFO Scenarios

Generally, inventory levels are higher under the FIFO case. This is true for both the non-perishables and perishables cases. Figure 7 shows that in, the non-perishables case, inventory levels

exhibit the same upwards trend. Yet, once the values stabilize, the FIFO case will begin to exhibit consistently higher inventory levels. Its highest point and lowest point upon stabilization are higher than inventory at any point in time under LIFO. This is different from the case of perishables, where inventory levels have higher maximum peaks under FIFO but also lower LIFO peaks. However, based on the simulations, it would appear that FIFO exhibits higher inventory levels more often than it exhibits lower inventory levels.

The higher inventories under FIFO may be attributed to the extent of feedback delay between production and demand. In the causal loop diagrams in Figure 2, there are multiple delays between Fresh Finished Goods Inventory and Demand under FIFO. This signifies that there are multiple delays between demand and production for forecasts based on it are actually sold. Hence, there is a higher propensity for overproduction or underproduction. In these cases, the problem is leaning towards overproduction given that current inventories are not considered in deciding on the quantity to produce in the next period.

Based on the review of studies on PIS, there are more models were FIFO proved to be optimal than otherwise. These models are optimization models with the primary objective of minimizing costs. The potential for waste production is incorporated in the form of disposal costs or deterioration costs, which align the sustainability and economic objectives for the system. It is understandable for FIFO to be optimal following this perspective, as FIFO results in the lower minimum inventory values. However, in actual inventory systems, stable inventories are preferred over fluctuations.

In literature, the case where LIFO proved to be optimal was the case where prices decreased linearly for each period that the inventory unit remained in the system. This would suggest that when quality affects revenues, LIFO becomes preferable. This is because the profit earned per unit would be more consistent. The key benefit to be derived from LIFO is consistency. In this model, quality affects revenues by acting as a multiplier to demand. Results showed that inventory levels were less variable for both perishables and non-perishables, which aligns with the suppositions from previous studies that LIFO can result in greater consistency.

4.3. Evaluation of FIFO and LIFO Applied to Perishables

For perishables, product quality is partially determined by the age of inventories. Hence, the volume of fresh and aged inventories can have an impact on demand. In this model, quality is given as the fraction of fresh inventories in relation to total inventories. Previously, it was mentioned that the levels of fresh and aged inventories that result from FIFO and LIFO respectively different greatly for perishables. This means that the use of FIFO and LIFO can also significantly alter the quality of perishable goods being sold.

It has been mentioned that LIFO may promote greater consistency in an inventory system. This also manifests in the current model where quality is more consistent under LIFO due to the fact that the levels of both fresh and aged inventory are more consistent. Consistent quality also results in greater consistency of sales. Figure 8 contains the simulated values for sales of inventory, where it may be observed that Sold Fresh Inventory, which is the primary contributor to sales in LIFO, is high consistent.

Under FIFO, sales of fresh and aged inventories both fluctuate. Specifically, the two inventories have an inverse relationship, as the sales that cannot be met with aged inventories will satisfied with fresh inventories. The fluctuations in quality also lead to fluctuations in demand, which feedbacks to inventory levels. When demand is high, it can prompt the production of new inventory, which has a negative effect on demand as the goods will be aged by the time they are sold. This problem is aggravated by the fact that there the production quantities derived from past demand will be misaligned with actual demand. When demand is rising, production will be higher, and conversely, when demand is dropping, production will be lower. This is the result of feedback delays between demand and production.

There are also feedback delays upstream, in the customer's perception of quality. The current model differentiates between actual quality and perceived quality, by the logic that there is a delay in the recognition of quality change when it comes to inventories. In most inventory models, the perception of quality and its resulting change in demand is instantaneous. However, this relationship is only descriptive of settings such as retail stores where customers see the physical units individually and its quality instantaneously affects the probability of purchase [8], [40]. This is not the case for entities where purchases are made in bulk, such as the manufacturing or warehousing settings. In such cases, the individual quality of units have little value in the system, which is evident by the miss-outs from sampling procedures [3]. There are also delays due to the fact that not all inventory units are assessed prior to production, and thus any information on quality would only be known after production or through customer returns [41].

FIFO exhibits longer cycles between the arrival and sale of an inventory unit. In the short term, this will result in (1) less wastes and lower inventory levels for the manufacturing firm, and (2) instances when sales may be lower in the periods during and following the end of an inventory cycle, when the leftover units have aged. The trade-offs can be managed with a classical optimization model that can balance waste generation with freshness. On the other hand, the long-term impact of FIFO in the supply chain is inconsistent quality. This complicates reordering decisions downstream in the supply chain, by adding variability to product lifetimes. This in turn heightens the amplitudes of bullwhips being exhibited by the inventory levels of a supply chain. Therefore, in the long term, FIFO promotes variability and greater efforts being made by members of a supply chain to manage it.

LIFO has the opposite effect on inventory levels in the short term; and will result in more consistent quality in the long term. While this lessens the complexity of decision-making for the members of the supply chain, there is nonetheless the issue of waste to be addressed. The consistent generation of waste would require greater efforts in waste management. On a positive note, the consistent waste generation would also regularize the demand for waste management, and through that require the standardization of waste management procedures. Currently, the efforts to minimize waste and the natural variability of production add complexity to decisions on when and where to set up waste management

facilities. Given the high investment required by such facilities, this generally results in there being insufficient infrastructure for this purpose. In this sense, the consistency promoted by implementations of LIFO would be beneficial, by giving a clearer perspective on the volume of waste management activity that would be needed.

5. Conclusion

In this study, classic FIFO and LIFO issuance policies were assessed for their impact in a perishable inventory system. The performance of these policies was evaluated in terms of their ability to meet sustainability and quality goals. Specifically, it aimed to improve sustainability through waste reduction and quality by maximizing the freshness of inventories in the system. While similar studies had been made in the past, the present study viewed the problem from a different approach, through the development of a system dynamics (SD) framework. Using the SD framework emphasizes the patterns of the inventory levels, rather than its exact values, allowing for the generation of insights that can be generalized regardless of lifetime variability.

A simulation model was developed comprising of three major sections that eventually form a general single loop: (1) The physical inventory system, (2) The perception of quality, and (3) The perception of demand. Furthermore, feedback delays such as delays in perishing, perceived quality, and perceived demand were also considered. FIFO and LIFO models for both perishables and non-perishables products were then formed with the presence of Waste outflow variable as a distinguishing factor between non-perishables and perishables model.

Comparing the perishables and non-perishables scenario yielded that for all cases, inventory levels exhibit an oscillating behavior. Non-perishables showed upward trend due to accumulation of inventory since quantity to be produced relies solely on sales forecasts and that current leftover usable inventory was not factored in as to how much quantity is to be produced for the next period. This is to emulate perishables inventory systems wherein external values of goods differ per period. On the other hand, perishables seem to have no upward for downward trend and seem to follow the behavior of traditional inventory models like EOQ. Perishable inventories do not have strong tendency towards overproduction. Wastes is not increasing the same way excess inventory does for non perishables. Based on these observations, it may be counterproductive to determine the optimal stockout level of perishable goods since it seems to have its way of regulating itself.

A comparison of the FIFO and LIFO scenarios showed that LIFO is preferable when revenues are heavily influenced by quality since LIFO promotes greater consistency in inventory system. Consistency in quality brings consistency in profit earned per unit and this was explicitly observed in Sold Fresh Inventory which is a primary contributor to the sales in LIFO.

On the other hand, FIFO exhibits higher inventory levels which is attributed due to feedback delays between production and demand. Furthermore, FIFO system tends to lean towards overproduction since current leftover inventories were not factored into the decision for the amount or quantity to be produced for the next period. Also, fluctuations were observed in FIFO for both sales in fresh and aged inventories - as evidenced by its inverse relationship - lead to fluctuations in demand. Although FIFO may be more optimal for minimizing costs and inventory levels once wastes is treated as disposal costs or deterioration costs, actual inventory systems still prefer stable inventories rather than constant fluctuations. FIFO is preferable for cost-sensitive or waste-sensitive cases, while LIFO is preferable in quality-sensitive cases.

The results of this study demonstrate the ability of SD to produce insights in addition to those that can be derived from optimization models. While the current study applied SD on classic inventory management policies (i.e. FIFO and LIFO), literature on perishable inventory models comprise various new policies and technologies that can be assessed more extensively using the SD framework. In particular, information technologies such as RFID and IoT have been proposed as means of monitoring the age-based quality of perishables with better accuracy, yet there is still the question of how to use this feedback effectively in decision-making. The current study serves as a starting point for future studies applying SD in evaluating and improving policies that will prove useful in a wide variety of cases.

One of the major contributions of this study is its demonstration of the SD framework as a means of modelling and simulating policy outcomes that are difficult to, or cannot be specifically quantified. This is very relevant to the field as it recent literature centers on the application of information technology in inventory management. The effects of greater accuracy and shortened feedback granted by information technology are not directly quantifiable, making it a suitable extension of this policy. The study also contributes by providing insight on what the long-term implications of issuance policies could be, particularly in the area of sustainability. While it may appear that FIFO is the most efficient in terms of cost and waste generation, LIFO actually has more positive long-term implications by reducing complexity in supply chain decisions, and standardizing waste management efforts.

References

[1] J. Blackburn and G. Scudder, “Supply Chain Strategies for Perishable Products : The Case of Fresh Produce,” Prod. Oper. Manag., vol. 18, no. 2, pp. 129–137, 2009.

[2] G. Ferreira, E. Arruda, and L. Marujo, “Inventory management of perishable items in long-term humanitarian operations using Markov Decision Processes,” Int. J. Disaster Risk Reduct., vol. 31, pp. 460–469, 2018.

[3] Y. Yang, H. Chi, O. Tang, W. Zhou, and T. Fan, “Cross perishable effect on optimal inventory preservation control,” Eur. J. Oper. Res., vol. 276, no. 3, pp. 998–1012, 2019.

[4] M. Dehghani and B. Abbasi, “An age-based lateral-transshipment policy for perishable items,” Int. J. Prod. Econ., vol. 198, pp. 93–103, 2018.

[5] L. Janssen, T. Claus, and J. Sauer, “Literature review of deteriorating inventory models by key topics from 2012 to 2015,” Intern. J. Prod. Econ., vol. 182, pp. 86–112, 2016.

[6] C. Kouki, M. Z. Babai, and S. Minner, “International Journal of Production Economics On the bene fi t of dual-sourcing in managing perishable inventory,” Intern. J. Prod. Econ., vol. 204, no. May, pp. 1–17, 2018.

[7] Z. Azadi, S. D. Eksioglu, B. Eksioglu, and G. Palak, “Stochastic optimization models for joint pricing and inventory replenishment of perishable products,” Comput. Ind. Eng., no. November, pp. 1–18, 2018.

[8] B. Afshar-nadja, “The influence of sale announcement on the optimal policy of an inventory system with perishable items,” J. Retail. Consum. Serv., vol. 31, no. C, pp. 239–245, 2016.

[9] X. Xu, Q. Bai, and M. Chen, “two-warehouse inventory systems for deteriorating items,” Appl. Math. Model., vol. 0, pp. 1–16, 2016.

[10] L. Janssen, Development and simulation analysis of a new perishable inventory model with a closing days constraint under non-stationary stochastic demand. 2018.
[11] L. Janssen, A. Diabat, J. Sauer, and F. Herrmann, "A stochastic micro-periodic age-based inventory replenishment policy for perishable goods," Transp. Res. Part E, vol. 118, no. January, pp. 445–465, 2018.
[12] Y. Lee, S. Mu, Z. Shen, and M. Dessouky, "Computers & Industrial Engineering Issuing for perishable inventory management with a minimum inventory volume constraint," Comput. Ind. Eng., vol. 76, pp. 280–291, 2014.
[13] G. Gaukler, M. Ketzenberg, and V. Salin, "International Journal of Production Economics Establishing dynamic expiration dates for perishables : An application of RFID and sensor technology," Int. J. Prod. Econ., vol. 193, no. November 2016, pp. 617–632, 2017.
[14] A. Hiassat, A. Diabat, and I. Rahwan, "A genetic algorithm approach for location-inventory-routing problem with perishable products," J. Manuf. Syst., vol. 42, pp. 93–103, 2017.
[15] R. Dominguez, B. Ponte, S. Cannella, and J. M. Framinan, "On the dynamics of closed-loop supply chains with capacity constraints," Comput. Ind. Eng., 2018.
[16] M. Song, X. Cui, and S. Wang, "International Journal of Production Economics Simulation of land green supply chain based on system dynamics and policy optimization," Intern. J. Prod. Econ., no. July, pp. 0–1, 2018.
[17] X. Wang, S. M. Disney, and J. Wang, "Int . J . Production Economics Exploring the oscillatory dynamics of a forbidden returns inventory system," Intern. J. Prod. Econ., vol. 147, pp. 3–12, 2014.
[18] M. R. S. M, C. Hora, D. O. Fontes, F. Gaudêncio, and M. Freires, "Sustainable and renewable energy supply chain : A system dynamics overview," Renew. Sustain. Energy Rev., vol. 82, no. June 2017, pp. 247–259, 2018.
[19] T. Rebs, M. Brandenburg, and S. Seuring, "System dynamics modeling for sustainable supply chain management : A literature review and systems thinking approach," J. Clean. Prod., 2018.
[20] M. Kalverkamp and S. B. Young, "In Support of Open-Loop Supply Chains :," J. Clean. Prod., 2019.
[21] H. Jalali, R. Carmen, I. Van Nieuwenhuyse, and R. Boute, "Quality and Pricing Decisions in Production/Inventory Systems," Eur. J. Oper. Res., vol. 272, no. 1, pp. 195–206, 2018.
[22] S. Kumar and A. Nigmatullin, "Simulation Modelling Practice and Theory A system dynamics analysis of food supply chains – Case study with non-perishable products," Simul. Model. Pract. Theory, vol. 19, no. 10, pp. 2151–2168, 2013.
[23] Z. Hosseinifard and B. Abbasi, "Computers & Operations Research The inventory centralization impacts on sustainability of the blood supply chain," Comput. Oper. Res., pp. 1–7, 2016.
[24] P. M. Ching and J. A. Wu, "An Assessment of FIFO and LIFO Policies for Perishable Inventory Systems Using the System Dynamics Approach," in 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 2019, pp. 251–255.
[25] R. Poles, "Int . J . Production Economics System Dynamics modelling of a production and inventory system for remanufacturing to evaluate system improvement strategies," Intern. J. Prod. Econ., vol. 144, no. 1, pp. 189–199, 2013.
[26] Y. Qiu, J. Qiao, and P. M. Pardalos, "Optimal production, replenishment, delivery, routing and inventory management policies for products with perishable inventory," Omega, vol. 82, no. C, pp. 193–204, 2019.
[27] P. Moser, O. Isaksson, and R. W. Seifert, "Inventory Dynamics in Process Industries : AC SC," Int. J. Prod. Econ., 2017.
[28] J.-S. Song and Y. Zhao, "The Value of Component Commonality in a Dynamic Inventory System with Lead Times," Manuf. Serv. Oper. Manag., vol. 11, no. 3, pp. 493–508, 2009.
[29] M. Gaggero and F. Tonelli, "ScienceDirect Optimal Control of Distribution Optimal Control of Distribution Optimal of Distribution Perishable Goods Optimal Control Control of Distribution Perishable Goods Perishable Goods Perishable Goods Chains Chains Chains Chains for," IFAC-PapersOnLine, vol. 48, no. 3, pp. 1049–1054, 2015.
[30] Y. Duan, G. Li, J. M. Tien, and J. Huo, "Inventory models for perishable items with inventory level dependent demand rate," Appl. Math. Model., vol. 36, no. 10, pp. 5015–5028, 2012.
[31] A. A. Alamri and A. A. Syntetos, "Beyond LIFO and FIFO : Exploring an Allocation-In-Fraction-Out (AIFO) policy in a two-warehouse inventory model," Intern. J. Prod. Econ., vol. 206, no. September, pp. 33–45, 2018.
[32] C. C. Lee, "Two-warehouse inventory model with deterioration under FIFO dispatching policy," vol. 174, pp. 861–873, 2006.
[33] Y. Li, F. Chu, Z. Yang, and R. W. Calvo, "A Production Inventory Routing Planning for Perishable Food with Quality Consideration," IFAC-PapersOnLine, vol. 49, no. 3, pp. 407–412, 2016.
[34] J. Huber, A. Gossmann, and H. Stuckenschmidt, "Cluster-based hierarchical demand forecasting for perishable goods," Expert Syst. Appl., vol. 76, pp. 140–151, 2017.
[35] L. A. San-josé, J. Sicilia, and J. García-laguna, "Optimal lot size for a production – inventory system with partial backlogging and mixture of dispatching policies," Intern. J. Prod. Econ., pp. 1–10, 2013.
[36] A. C. Braz, A. M. De Mello, L. Augusto, D. V. Gomes, P. Tromboni, and D. S. Nascimento, "The bullwhip effect in closed-loop supply chains : A systematic literature review," J. Clean. Prod., vol. 202, pp. 376–389, 2018.
[37] G. Dobson, E. J. Pinker, and O. Yildiz, "An EOQ model for perishable goods with age-dependent demand rate," Eur. J. Oper. Res., vol. 257, no. 1, pp. 84–88, 2017.
[38] Q. Hu, J. E. Boylan, H. Chen, and A. Labib, "OR in spare parts management : A review," Eur. J. Oper. Res., vol. 266, no. 2, pp. 395–414, 2018.
[39] S. Rahimi-ghahroodi, A. Al Hanbali, I. M. H. Vliegen, and M. A. Cohen, "International Journal of Production Economics Joint optimization of spare parts inventory and service engineers staffing with full backlogging," Intern. J. Prod. Econ., vol. 212, no. February, pp. 39–50, 2019.
[40] H. Hvolby and K. Steger-jensen, "Managing cannibalization of perishable food products in the retail sector," Procedia - Procedia Comput. Sci., vol. 64, pp. 1051–1056, 2015.
[41] K. Hun and H. Hwang, "A returns policy for distribution channel coordination of perishable items," vol. 152, pp. 770–780, 2004.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 69-74 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Multi-Step Iteration Algorithm of Total Asymptotically Quasi-Nonexpansive Maps

Salwa Salman Abed*, Zahra Mahmood Mohamed Hasan

Department of Mathematics, college of Education for Pure Sciences (Ibn Al-Haitham) / University of Baghdad, Baghdad, Iraq

ARTICLE INFO

Article history:
Received: 14 February 2019
Accepted: 28 April, 2019
Online: 21 May, 2019

Keywords:
Banach space,
total asymptotically quasi-nonexpansive map,
weak convergence,
strong convergence,
common fixed points

ABSTRACT

In Banach spaces an iteration algorithm for two finite families of total asymptotically quasi-nonexpansive maps is introduced. Weak and strong convergence theorems of this algorithm to approximation common fixed points are proved by using suitable conditions. As well as, numerical example by using Mat-lab is given.

1. Introduction and Preliminaries

This paper was originally published in the Conference: 2018 International Conference on Advanced Science and Engineering (ICOASE), Iraq [1]. It is well known that the concept of asymptotically nonexpansive introduced by Goebel and Kirk [2]. Additionally, every asymptotically nonexpansive map of a Banach space has a fixed point is proved. In [3], Petryshyn and Williamson proved the weak and strong convergence for quasi-nonexpansive map by using a sufficient and necessary condition. Alber [4], a new class of asymptotically nonexpansive is introduced. As well as, approximating methods for finding their fixed points are studied. In 2014, G. S. Saluja [5] established the strong and weak convergence for approximating common fixed point for generalized asymptotically quasi-nonexpansive maps in a Banach space.

Very recently, In [6], the authors proposed an implicit iteration for two finite families of generalized asymptotically quasi-nonexpansive maps. As well as, some strong convergence theorems are established. It is useful to point out our findings in this area which appeared in [7].

Let B be a non-empty closed convex subset of a real Banach space M and T be a self-map of B. The set of all fixed points denoted by F(T). A self-map T from B into M is called nonexpansive map [2] if

$$\|Ta - Tb\| \leq \|a - b\| \; for \; all \; a, b \in B$$

and is called quasi- nonexpansive map [6] if $F(T) \neq \emptyset \; and$

$$\|Ta - a^*\| \leq \|a - a^*\|$$

for all $a \in B$ and for all $a^* \in F(T)$.

A Banach space M is satisfying:

- "Opial's condition if for each sequence (a_n) in M, is weak convergence to a implies that $\lim_{n\to\infty} inf\|a_n - a\| < \lim_{n\to\infty} inf \; \|a_n - b\|$ for all $b \in M \; with \; a \neq b$".
- "Kadec-Klee property if for each sequence (a_n) in M is weak convergence to (a) together with $\|a_n\|$ converges strongly to $\|a\|$ imply that (a_n) is strong convergence to a point $a \in M$ [7]".

The aim of this paper, an iterative scheme for two families of total asymptotically quasi-nonexpansive maps is established. The strong and weak convergence theorems of this scheme for approximation of common fixed points in Banach space by using suitable conditions are established. For this purpose, let us recall the following definitions and lemmas.

Definition (1.1): "A map T is named asymptotically nonexpansive [1] if there is a sequence (f_n) in $[0, +\infty)$ with $\lim_{n\to\infty} f_n = 0$ and $\|T^n a - T^n b\| \leq (1 + f_n)\|a - b\|$, for all $a, b \in B, n = 1,2,..$.

*Salwa Salman Abed, Email: salwaalbundi@yahoo.com

If $F(T) \neq \emptyset$ there is a sequence (f_n)in $[0, +\infty)$ with $\lim_{n\to\infty} f_n = 0$ and $\|T^n a - a^*\| \leq (1 + f_n)\|a - a^*\|$,
for all $a \in B, a^* \in F(T)$ and $n = 1, 2, \ldots$.Therefore, T is named asymptotically quasi-nonexpansive map [10]".

Definition (1.2): "A map T is named total asymptotically nonexpansive map [4] if there are null sequences of postive real numbers $(f_n)_{n=1}^{\infty}, (g_n)_{n=1}^{\infty}, n \geq 1$ and nondecreasing continuous function $\psi: [0, \infty) \to [0, \infty)$ with $\psi(0) = 0$ such as $\forall a, b \in B$

$$\|T^n a - T^n b\| \leq \|a - b\| + f_n \psi\|a - b\| + g_n".$$

T is named total asymptotically quasi-nonexpansive map if $F(T) \neq \emptyset$, there are null sequences of postive real numbers $(f_n)_{n=1}^{\infty}$, $(g_n)_{n=1}^{\infty}, n \geq 1, \sum_{n=1}^{\infty} f_{in} < \infty$ and $\sum_{n=1}^{\infty} g_{in} < \infty$, and nondecreasing continuous function $\psi: [0, \infty) \to [0, \infty)$ with $\psi(0) = 0$ such as $\forall a \in B$ and $a^* \in F(T)$
$\|T^n a - a^*\| \leq \|a - a^*\| + f_n \psi\|a - a^*\| + g_n$.
If $g_n = 0, \forall n = 1, 2, \ldots$ therefore T is asymptotically quasi-nonxpansive map.

Definition (1.3)[11]: "Let B be a nonempty closed convex subset of a Banach space M. A self-map T is named uniformily K-Lipschitzain if there exists a constant $K > 0$ such that $\|T^n a - T^n b\| \leq K\|a - b\|, \ \forall a, b \in B$.

Definition (1.4)[12]: "A map $T: B \to M$ is named demi-closed with respect to $b \in M$ if for each sequence (a_n) in B, (a_n) is weak convergence to a and $T(a_n)$ is strong convergence to b. Hence $a \in B$ and $T(a) = b$. If $(I - T)$ is demiclosed which means if (a_n) is weak convergence to a in B and $(I - T)$ is strong convergence to 0. Therefore $(I - T)(a) = 0$".

Note: Now to explain the relation between the above definitions:

$$\begin{array}{ccccc} Nonexpansive & \Rightarrow & Asymptotically\ nonexpansive & \Rightarrow & Total\ asymptotically\ nonexpansive \\ \Downarrow & & \Downarrow & & \Downarrow \\ Quasi-nonexpansive & \Rightarrow & Asymptotically\ quasi-nonexpansive & \Rightarrow & Total\ asymptotically\ quasi-nonexpansive \end{array}$$

Lemma (1.5)[13]:" Let M be a uniformly convex Banach space and $0 < L \leq t_n \leq K < 1, \forall\, n \in N$. Presume that (a_n) and (b_n) are two sequences of M such as:

$$\lim_{n\to\infty} \|a_n\| \leq r, \lim_{n\to\infty} \|b_n\| \leq r \ \ and \ \ \lim_{n\to\infty} \|t_n a_n + (1 - t_n) b_n\| = r$$

hold for some $r \geq 0$. Thus $\lim_{n\to\infty} \|a_n - b_n\| = 0$".

Lemma (1.6)[14]: "Let $(\mu_n)_{n=1}^{\infty}, (\sigma_n)_{n=1}^{\infty}$ and $(e_n)_{n=1}^{\infty}$ be sequences of postive numbers accomplishing the following inequality:

$$\mu_{n+1} \leq (1 + \sigma_n)\mu_n + e_n \ , \forall\, n \geq 1$$

if $\sum_{n=1}^{\infty} \sigma_n < \infty$ and $\sum_{n=1}^{\infty} e_n < \infty$, then $(\mu_n)$is bounded and $\lim \mu_n$ exists. In additional if, $\liminf_{n\to\infty} \mu_n = 0$ then $\lim_{n\to\infty} \mu_n = 0$".

Lemma (1.7)[15]: "Let B be a nonempty convex subset of a uniformly convex Banach space. Therefore there exists a strictly nondecreasing continuous function $f: [0, \infty) \to [0, \infty)$ with $f(0) = 0$ such as for each Lipschitzain map $T: B \to B$ with Lipschitz constant K:

$$\|tTa + (1 - t)Tb - T(ta + (1 - t)b\| \leq Kf^{-1}\left(\|a - b\| - \frac{1}{K}\|Ta - Tb\|\right),$$

$\forall\, a, b \in B$ and $\forall t \in [0,1]$".

Lemma (1.8)[12]: "Let M be a uniformly convex Banach space and its dual M^* accomplishing the Kadec-Klee property. Presume that (a_n) bounded sequence in M such as $\lim_{n\to\infty} \|ta_n + (1 - t)p_1 - p_2\|$ exists $\forall\, t \in [0,1]$ and $p_1, p_2 \in W_w(a_n)$, thus $p_1 = p_2$".

2. Main Results

Let B be a nonempty closed convex subset of a Banach space M and $\{T_j, S_j, \forall j = 1, 2, \ldots, k\}$ be two families of total asymptotically quasi-nonexpansive self-maps. We define the iteration algorithm (a_n) as follows:

$$\begin{aligned} &a_1 \in B \\ &a_{n+1} = (1 - \alpha_{jn}) S_j^n a_n + \alpha_{jn} T_j^n b_{jn} \\ &b_{jn} = (1 - \alpha_{jn}) S_j^n a_n + \alpha_{jn} T_j^n b_{(j-1)n} \\ &b_{(j-1)n} = (1 - \alpha_{(j-1)n}) S_{j-1}^n a_n + \alpha_{(j-1)n} T_{j-1}^n b_{(j-2)n} \\ &\cdot \\ &\cdot \\ &\cdot \\ &b_{2n} = (1 - \alpha_{2n}) S_2^n a_n + \alpha_{2n} T_2^n b_{1n} \\ &b_{1n} = (1 - \alpha_{1n}) S_1^n a_n + \alpha_{1n} T_1^n b_{0n} \end{aligned} \qquad (1)$$

Where $b_{0n} = a_n$ and $(\alpha_n)_{n=1}^{\infty}$ are sequences in [0, 1].

Lemma (2.1): Let B be a nonempty closed convex subset of a normed space M and $T_j, S_j, j = 1, 2, \ldots, k$ be two family of total asymptotically quasi-nonexpansive self-maps of B. Presume that $F(T_j, S_j) \neq \emptyset$ and the sequence (a_n) be as shown in step (1). Then:
i- There are sequences (u_n) and (v_n) in $[0, \infty)$ such as $\sum_{n=1}^{\infty} u_n < \infty, \sum_{n=1}^{\infty} v_n < \infty$ and
$\|a_{n+1} - a^*\| \leq (1 + u_n)^{j+1}\|a_n - a^*\| + v_n^{j+1}$,
$\forall\, a^* \in F(T_j, S_j)$ and $\forall\, n$.
ii- There exist constants $J_1, J_2 > 0$ such as
$\|a_{n+p} - a^*\| \leq J_1\|a_n - a^*\| + J_2$, $\forall\, a^* \in F(T_j, S_j)$
and $n, p = 1, 2, \ldots$..
If there is $Z > 0$ such that $\psi(\lambda_j) \leq Z\lambda_j, j = 1, 2, \ldots, k$.
Proof: i- Let $a^* \in F, u_n = \max_{1\leq j\leq k} f_{jn}$ and $v_n = \max_{1\leq j\leq k} g_{jn}$.

Now, we have

$$\begin{aligned} \|b_{1n} - a^*\| &= \|(1 - \alpha_{1n}) S_1^n a_n + \alpha_{1n} T_1^n a_n - a^*\| \\ &\leq (1 - \alpha_{1n})\|S_1^n a_n - a^*\| + \alpha_{1n}\|T_1^n a_n - a^*\| \\ &\leq (1 - \alpha_{1n})\{\|a_n - a^*\| + f_{1n}\psi\|a_n - a^*\| + g_{1n}\} \\ &\quad + \alpha_{1n}\{\|a_n - a^*\| + f_{1n}\psi\|a_n - a^*\| + g_{1n}\} \\ &\leq (1 - \alpha_{1n})(1 + f_{1n}Z)\|a_n - a^*\| + (1 - \alpha_{1n}) g_{1n} \\ &\quad + \alpha_{1n}(1 + f_{1n}Z)\|a_n - a^*\| + \alpha_{1n} g_{1n} \\ &\leq (1 + f_{1n}Z)\|a_n - a^*\| + g_{1n} \\ &\leq (1 + u_n)\|a_n - a^*\| + v_n \end{aligned}$$

Assume that $\|b_{jn} - a^*\| \leq (1 + u_n)^j\|a_n - a^*\| + v_n^j$
Therefore,

$$\|b_{(j+1)n} - a^*\| = \|(1 - \alpha_{(j+1)n}) S_{j+1}^n a_n + \alpha_{(j+1)n} T_{j+1}^n b_j - a^*\|$$

$$\leq (1-\alpha_{(j+1)n})\|S_{j+1}^n a_n - a^*\| + \alpha_{(j+1)n}\|T_{j+1}^n b_j - a^*\|$$
$$\leq (1-\alpha_{(j+1)n})(1+f_{(j+1)n}Z)\|a_n - a^*\| + (1-\alpha_{(j+1)n})g_{(j+1)n} + \alpha_{(j+1)n}(1+f_{(j+1)n}Z)\|b_j - a^*\| + \alpha_{(j+1)n}g_{(j+1)n}$$
$$\leq (1-\alpha_{(j+1)n})(1+u_n)\|a_n - a^*\| + v_n + \alpha_{(j+1)n}(1+u_n)(1+u_n)^j\|a_n - a^*\| + \alpha_{(j+1)n}(1+u_n)v_n^j$$
$$\leq (1+u_n)^{j+1}\|a_n - a^*\| + v_n^{j+1}$$

Thus, by induction, we obtain

$$\|b_{jn} - a^*\| \leq (1+u_n)^j\|a_n - a^*\| + v_n^j \quad (2)$$

for all $j = 1,2,\dots,k$.

Now, by (2), we get

$$\|a_{n+1} - a^*\| \leq (1-\alpha_{jn})\|S_j^n a_n - a^*\| + \alpha_{jn}\|T_j^n b_{jn} - a^*\|$$
$$\leq (1-\alpha_{jn})(1+u_n)\|a_n - a^*\| + v_n + \alpha_{jn}(1+u_n)(1+u_n)^j\|a_n - a^*\| + \alpha_{jn}(1+u_n)v_n^j$$
$$\leq (1+u_n)^{j+1}\|a_n - a^*\| + v_n^{j+1}$$

ii- By using part (i), we get

$$\|a_{n+p} - a^*\| \leq (1+u_{n+p-1})^{j+1}\|a_{n+p-1} - a^*\| + v_{n+p-1}^{j+1}$$
$$\leq e^{(1+u_{n+p-1})^{j+1}\|a_{n+p-1}-a^*\|} + e^{v_{n+p-1}^{j+1}}$$
$$\leq e^{(j+1)u_{n+p-1}\|a_{n+p-1}-a^*\|} + e^{(j+1)v_{n+p-1}}$$
$$\leq e^{(j+1)\sum_{k=1}^{n+p-1}u_k}\|a_n - a^*\| + e^{(j+1)\sum_{k=1}^{n+p-1}v_k}$$
$$\leq J_1\|a_n - a^*\| + J_2.$$

Setting $J_1 = e^{(j+1)\sum_{k=1}^{n+p-1}u_k}$ and $J_2 = e^{(j+1)\sum_{k=1}^{n+p-1}v_k}$.

Lemma (2.2): Let B be a nonempty closed convex subset of a normed space M and T_j, S_j , $j = 1, 2, \dots, k$ be two families of total asymptotically quasi-nonexpansive self-maps of B. Presume $F(T_j, S_j) \neq \emptyset$ and (a_n) be as shown in step (1). Therefore, $\lim_{n\to\infty}\|a_n - a^*\|$ exists for all $a^* \in F(T_i, S_i)$.

Proof: By Lemma (2.1.i)

$$\|a_{n+1} - a^*\| \leq (1+u_n)^{j+1}\|a_n - a^*\| + v_n^{j+1}$$
$$\leq (1+u_n)\|a_n - a^*\| + v_n$$

and $\sum_{n=1}^{\infty}u_n < \infty, \sum_{n=1}^{\infty}v_n < \infty$. So by Lemma (1.6.i), we get $\lim_{n\to\infty}\|a_n - a^*\|$ exists for all $a^* \in F(T_j, S_j)$.

Lemma (2.3): Let B be a nonempty closed convex subset of a Banach space M and T_j, S_j , $j = 1, 2, \dots, k$ be two families of Lipschitzain and total asymptotically quasi-nonexpansive self-maps of B. Let (a_n) be as shown in step (1). Therefore, for all $a_1^*, a_2^* \in F(T_j, S_j)$, the limit $\lim_{n\to\infty}\|ta_n + (1-t)a_1^* - a_2^*\|$ exists for all $t \in [0,1]$.

If there is $Z > 0$ such that $\psi(\lambda_j) \leq Z\lambda_j, j = 1, 2, \dots, k$.

Proof: By using Lemma (1.6),we have $\lim_{n\to\infty}\|a_n - a^*\|$ exists $\forall a^* \in F(T_j, S_j)$ and (a_n) is bounded. Let $\gamma_n(t) = \|ta_n + (1-t)a_1^* - a_2^*\|$, $\forall t \in [0,1]$.

Therefore, $\lim_{n\to\infty}a_n(0) = \|a_1^* - a_2^*\|$ and $\lim_{n\to\infty}a_n(1) = \|a_n - a_2^*\|$ exist by Lemma (2.2).

Then, for $t \in [0,1]$ and for all $a \in B$, we define the map $R_n: B \to B$ by:

$$b_{1n} = (1-\alpha_{1n})S_1^n a_n + \alpha_{1n}T_1^n b_{0n}$$
$$b_{2n} = (1-\alpha_{2n})S_2^n a_n + \alpha_{2n}T_2^n b_{1n}$$
$$\vdots$$
$$b_{jn} = (1-\alpha_{jn})S_j^n a_n + \alpha_{in}T_j^n b_{(j-1)n}$$
$$R_n a = (1-\alpha_{jn})S_j^n a_n + \alpha_{jn}T_j^n b_{jn}$$

Now,

$$\|R_n a - R_n c\| \leq (1-\alpha_{jn})\|S_j^n a - S_j^n c\| + \alpha_{jn}\|T_j^n b_{jn} - T_j^n d_{jn}\|$$
$$\leq (1-\alpha_{jn})(1+f_{jn}Z)\|a-c\| + (1-\alpha_{jn})g_{jn} + \alpha_{jn}(1+f_{jn}Z)\|b_{jn} - d_{jn}\| + \alpha_{jn}g_{jn}$$
$$\leq (1+u_n)\|a-c\| + v_n$$
$$\leq (1+f_nZ)\|a-b\| + g_n$$

with $\sum_{n=1}^{\infty}u_n < \infty$, $\sum_{n=1}^{\infty}v_n < \infty$ and $s_n = 1 + u_n$, it follows that $s_n \to 1 \; as \; n \to \infty$.

Setting $W_{n,m} = R_{n+m-1}R_{n+m-1}\dots R_n$ and $b_{n,m} = \|W_{n,m}(ta_n + (1-t)a_1^* - (tW_{n,m}a_n + (1-t)a_1^*)\|$.

Thus,

$$\|W_{n,m}a - W_{n,m}b\| = \left\| \begin{matrix} R_{n+m-1}R_{n+m-2}\dots R_n(a) \\ -R_{n+m-1}R_{n+m-2}\dots R_n(b) \end{matrix} \right\|$$
$$\leq s_{n+m-1}\left\| \begin{matrix} R_{n+m-2}\dots R_n(a) \\ -R_{n+m-2}\dots R_n(b) \end{matrix} \right\| + v_{n+m-1}$$
$$\leq \prod_{j=n}^{n+m-1}s_j\|a-c\| + \sum_{j=n}^{n+m-1}v_j$$
$$= A_n\|a-c\| + \sum_{j=n}^{n+m-1}v_j$$

for all $a, c \in B$, where $A_n = \prod_{j=n}^{n+m-1}s_j$, $W_{n,m}a_n = a_{n+m}$ and $a^* = a^*$ for all $a^* \in F(T_i, S_i)$.

Hence,

$$\gamma_{n+m}(t) = \|ta_{n+m} + (1-t)a_1^* - a_2^*\|$$
$$= \left\| \begin{matrix} tW_{n,m}a_n + ((1-t)a_1^* - a_2^* + W_{n,m}(ta_n + (1-t)a_1^*) - a_2^* \\ +a^* - a^* \\ -W_{n,m}(ta_n + (1-t)a_1^*) - a_2^* \end{matrix} \right\|$$
$$\leq b_{n,m} + \|W_{n,m}(ta_n + (1-t)a_1^*) - a_2^*\|$$
$$\leq b_{n,m} + A_n\gamma_n(t) + \sum_{j=n}^{n+m-1}v_j$$

By using Lemma (1.7), we have

$$b_{n,m} \leq Kf^{-1}(\|a_n - a^*\| - \frac{1}{K}\|W_{n,m}a_n - W_{n,m}a^*\|)$$
$$\leq Kf^{-1}(\|a_n - a^*\| - \frac{1}{K}(\|a_{n+m} - a^*\| - \|W_{n,m}a_n - a^*\|)$$

and $(b_{n,m})$ converges uniformly to zero. Since $\lim_{n\to\infty}A_n = 1$ and $\lim_{n\to\infty}v_n = 0$, we get

$$\lim_{n\to\infty}\lim_{m\to\infty}sup\gamma_{n+m} \leq \lim_{n\to\infty}\lim_{m\to\infty}b_{n,m} + \lim_{n\to\infty}inf\gamma_n(t) = \lim_{n\to\infty}inf\gamma_n(t)$$

Thus, $\lim_{n\to\infty}\gamma_n(t)$ exists for all $t \in [0,1]$.

Theorem (2.4): Let B be a nonempty closed convex subset of a uniformly convex Banach space M, T_j, S_j , $\forall j = 1, 2, \dots, k$ be two families of Lipschiztain and total asymptotically quasi-nonexpansive self-maps of B and the sequence (a_n) be as shown in step (1). If there is $Z > 0$ such that $\psi(\lambda_j) \le Z\lambda_j, j = 1, 2, \dots, k$. Then, $\lim_{n\to\infty}\|T_j^n a_n - a_n\| = 0 = \lim_{n\to\infty}\|S_j^n a_n - a_n\|$, $\forall j$.

Proof: By Lemma (2.2), $\lim_{n\to\infty}\|a_n - a^*\|$ exists. Assume that

$$\lim_{n\to\infty}\|a_n - a^*\| = e, \qquad \forall e \ge 0.$$

If e=0, the proof is straight forward.

Now, suppose $e > 0$. We get

$$a_{n+1} = (1 - \alpha_{jn})S_j^n a_n + \alpha_{jn}T_j^n b_{jn}$$

and

$$e = \|a_{n+1} - a^*\| = \|(1 - \alpha_{jn})(S_j^n a_n - a^*) + \alpha_{jn}(T_j^n b_{jn} - a^*)\|.$$

Since $\|S_j^n a_n - a^*\| \le K\|a_n - a^*\| \le e$

and $\|T_j^n b_{jn} - a^*\| \le K\|b_{jn} - a^*\| \le K\{(1 + u_n)^j\|a_n - a^*\| + v_n^j\} \le e.$

Therefore, by Lemma (1.5), we get

$$\lim_{n\to\infty}\|S_j^n a_n - T_j^n b_{jn}\| = 0$$

Next,

$$\begin{aligned}\|b_{jn} - a_n\| &\le (1 - \alpha_{jn})\|S_j^n a_n - a_n\| + \alpha_{jn}\|T_j^n b_{(j-1)n} - a_n\| \\ &\le (1 - \alpha_{jn})(1 + f_{jn}Z)\|a_n - a_n\|(1 - \alpha_{jn})g_{jn} \\ &\quad + \alpha_{jn}(1 + f_{jn}Z)\|b_{(j-1)n} - a_n\| + \alpha_{jn}g_{jn} \\ &\le g_{jn} + \alpha_{jn}(1 - \alpha_{(j-1)n})(1 + f_{jn}Z) \\ &\quad \|S_{j-1}^n a_n - a_n\| + \alpha_{jn}\alpha_{(j-1)n}(1 + f_{jn}Z) \\ &\quad \|T_{j-1}^n b_{(j-2)n} - a_n\| \\ &\le g_{jn} + \alpha_{jn}(1 - \alpha_{(j-1)n})(1 + f_{jn}Z)g_{jn} \\ &\quad + \alpha_{jn}\alpha_{(j-1)n}(1 + f_{jn}Z)(1 \\ &\quad + f_{(j-1)n}Z)\|b_{(j-2)n} - a_n\| \\ &\quad + \alpha_{jn}\alpha_{(j-1)n}(1 + f_{jn}Z)g_{(j-1)n} \\ &\quad . \\ &\quad . \\ &< \infty\end{aligned}$$

Since $Z > 0, \sum_{n=1}^{\infty} f_{jn} < \infty$ and $\sum_{n=1}^{\infty} g_{jn} < \infty$, hence $\lim_{n\to\infty}\|b_{jn} - a_n\| = 0$.

Then,

$$\begin{aligned}\|S_j^n a_n - a_n\| &\le \|S_j^n a_n - T_j^n b_{jn}\| + \|T_j^n b_{jn} - a_n\| \\ &\le \|S_j^n a_n - T_j^n b_{jn}\| + K\|b_{jn} - a_n\| \to 0 \\ &\quad as\ n \to \infty.\end{aligned}$$

And

$$\begin{aligned}\|T_j^n a_n - a_n\| &\le \|T_j^n a_n - S_j^n a_n\| + \|S_j^n a_n - a_n\| \to 0 \\ &\quad as\ n \to \infty.\end{aligned}$$

Theorem (2.5): Let B be a nonempty closed convex subset of a Banach space and T_j, S_j , $j = 1, 2, \dots, k$ be two families of total asymptotically quasi-nonexpansive self-maps of B. Presume that $F(T_j, S_j) \ne \emptyset$ and (a_n)be as shown in step (1)is strong convergence to a common fixed point of T_j and S_j iff $\lim_{n\to\infty} inf\ d(a_n, F) = 0$, where $d(a, F) = inf_{a^* \in F}\|a - a^*\|$.

Proof: To show $\lim_{n\to\infty} inf\ d(a_n, F) = 0$ implies that (a_n) is strong convergence to a common fixed point of T_j , $S_j, j = 1, 2, \dots, k$, since by (2)

$$\begin{aligned}d(a_{n+1}, F) &\le (1 + u_n)^{j+1}d(a_n, F) + v_n^{j+1} \\ &\le (1 + u_n)d(a_n, F) + v_n\end{aligned}$$

By Lemma (1.6), we get $\lim_{n\to\infty} a_n$ exists and $\lim_{n\to\infty} inf a_n = 0$.

Hence, $\lim_{n\to\infty} a_n = 0$.

Next, to prove the sufficiency, firstly show that (a_n) Cauchy sequence. By using Lemma (2.1.ii), we get

$$\|a_{n+p} - a^*\| \le J_1\|a_n - a^*\| + J_2 \qquad (3)$$

$\forall a^* \in F(T_j, S_j), n = p = 1,2, \dots$

Since $\lim_{n\to\infty} a_n = 0, \forall \epsilon > 0, \exists N$ such that

$$d(a_n, F) \le \frac{\epsilon}{3J_1} - \frac{J_2}{J_1}, \qquad \forall n \ge N$$

therefore, there is $l \in F(T_i, S_i)$ such that

$$\|a_N - l\| \le \frac{\epsilon}{2J_1} - \frac{J_2}{J_1} \qquad (4)$$

From (3) and (4), $\forall n \ge N$, we get

$$\begin{aligned}\|a_{n+p} - a_n\| &\le \|a_{n+p} - l\| + \|a_n - l\| \\ &\le J_1\|a_N - l\| + J_2 + J_1\|a_N - l\| + J_2 \\ &\le J_1\frac{\epsilon}{2J_1} - J_2 + J_2 + J_1\frac{\epsilon}{2J_1} - J_2 + J_2 \\ &= \epsilon\end{aligned}$$

Then, (a_n) is a Cauchy sequence and converges to $c \in M$.

Lastly, to show that $c \in F(T_j, S_j)$, for any $\epsilon^* > 0$, there is N_1such that

$$\|a_n - c\| \le \frac{\epsilon^*}{2(2 + f_jZ)} - \frac{3g_j}{(2 + f_jZ)}, \forall n \ge N_1 \qquad (5)$$

Since $\lim_{n\to\infty} a_n = 0$ implies that $N_2 \ge N_1$ such that $d(a_n, F) \le \frac{\epsilon^*}{3(4+3f_jZ)}, \forall n \ge N_2$.

Thus, $\exists\ l_1 \in F(T_j, S_j)$ such that

$$\|a_N - l_1\| \le \frac{\epsilon^*}{2(4+3f_jZ)} \qquad (6)$$

From (5) and (6) for any $T_j, \forall j = 1, 2, \dots, k$, we obtain

$$\begin{aligned}\|T_j c - c\| &\le \|jc - l\| + 2\|T_j a_{N_2} - l_1\| + \|a_{N_2} - l\| \\ &\quad + \|a_{N_2} - c\| \\ &\le \|c - l_1\| + f_j\psi\|p - l_1\| + g_j + 2\|a_{N_2} - l_1\| \\ &\quad + 2f_j\psi\|a_{N_2} - l\| + 2g_j + \|a_{N_2} - l_1\| \\ &\quad + \|a_{N_2} - c\| \\ &\le (1 + f_jZ)\|c - l_1\| + 2(1 + f_jZ)\|a_{N_2} - l_1\| + 3g_j \\ &\quad + \|a_{N_2} - l_1\| + \|a_{N_2} - c\| \\ &\le (1 + f_jZ)\|a_{N_2} - c\| + (1 + f_jZ)\|a_{N_2} - l_1\| \\ &\quad + 2(1 + f_jZ)\|a_{N_2} - l_1\| + 3g_j + \|a_{N_2} - l_1\| \\ &\quad + \|a_{N_2} - c\| \\ &\le (2 + f_jZ)\|a_{N_2} - c\| + (4 + 3f_jZ)\|a_{N_2} - l_1\| \\ &\quad + 3g_j \\ &\le (2 + f_jZ)\frac{\epsilon^*}{2(2 + f_jZ)} - 3g_j + 3g_j \\ &\quad + (4 + 3f_jZ)\frac{\epsilon^*}{2(4 + 3f_jZ)} \\ &= \epsilon^*\end{aligned}$$

Therefore, $\|T_j c - c\| = 0\ \forall i$ which means $T_j c = c, \forall j = 1, 2, .., k$.

By using the same above argument we can prove $\|S_j c - c\| = 0, \forall j$.Thus $c \in F(T_j, S_j)$.

Theorem (2.6): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, j = 1, 2, \dots, k$ be two families of Lipschitzain and total asymptotically quasi-nonexpansive self-maps of B. If M accomplishes Opial 's condition and the maps $I - T_j$ and $I - S_j, j = 1,2, \dots, k$ are demiclosed to zero, therefore (a_n) be as shown in step (1) is weak convergence to a common fixed point of T_j and $S_j, j = 1, 2, .., k$.

Proof: Let $a^* \in F(T_j, S_j)$. By Lemma (2.2), $\lim_{n\to\infty} \|a_n - a^*\|$ exists.

By Theorem (2.4), we have

$$\lim_{n\to\infty} \left\|T_j^n a_n - a_n\right\| = 0 = \lim_{n\to\infty} \left\|S_j^n a_n - a_n\right\|$$

for j=1, 2,…, k. Since by the supposition the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demiclosed to zero, therefore $T_j a^* = a^*$ and $S_j a^* = a^*$, that means $a^* \in F(T_j, S_j)$.

Next, to prove (a_n) converges weakly to a^*. Assume there is other subsequence $(a_{ni})\ of\ (a_n)$ is weak convergence to $b^* \in F(T_j, S_j)$ and $a^* \neq b^*$. By using the same argument as above we can show that $b^* \in F(T_j, S_j)$.

Now, to prove the uniqueness, assume $a^* \neq b^*$. Therefore, by using Opial 's condition, we obtain:

$$\begin{aligned}\lim_{n\to\infty} \|a_n - a^*\| &= \lim_{nj\to\infty} \left\|a_{nj} - a^*\right\| \\ &< \lim_{nj\to\infty} \left\|a_{nj} - b^*\right\| \\ &= \lim_{ni\to\infty} \|a_{ni} - b^*\| \\ &< \lim_{ni\to\infty} \|a_{ni} - a^*\| \\ &= \lim_{n\to\infty} \|a_n - a^*\|\end{aligned}$$

This is contradiction, therefore $a^* \neq b^*$. Hence, (a_n) is weak convergence to a^*.

Theorem (2.7): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, j = 1, 2, \dots, k$ be two families of Lipschitzain and total asymptotically quasi-nonexpansive self-maps of B. If the dual space $M^*\ of\ M$ has the Kadec-klee property and the maps $I - T_j, I - S_j, j = 1,2, \dots, k$ are demi-closed to zero, therefore, (a_n) be as shown in step (1) is weak convergence to a common fixed point of T_j and S_j.

Proof: As showed by Lemma (2.2), that $\lim_{n\to\infty} \|a_n - a^*\|$ exists. Since (a_n) is bounded in B and M is reflexive. Therefore, there exists a subsequence $(a_{ni})\ of\ (a_n)$ which is weak convergence to a point$a^* \in B$. By Theorem (2.4)

$$\lim_{n\to\infty} \left\|T_j^n a_n - a_n\right\| = 0 = \lim_{n\to\infty} \left\|S_j^n a_n - a_n\right\|$$

∀j=1, 2, …, k .

Since by the supposition the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demi-closed to zero. Thus, $a^* \in F(T_j, S_j)$. Now, to prove (a_n) is weak convergence to a point a^*. Presume that (a_{nk}) is other subsequence of (a_n) that is weak convergence to a point $b^* \in F(T_j, S_j)$. By using the same argument as above, we obtain $b^* \in F(T_j, S_j)$.

Therefore, by Lemma (2.3) $\lim_{n\to\infty} \|ta_n + (1-t)a^* - b^*\|$ exists for all $t \in [0,1]$.

By Lemma (1.8) $a^* = b^*$. As a result, the sequence (a_n) is weak convergence to the point $a^* \in F(T_j, S_j)$.

The following corollaries are special cases

Corollary (2.8): Let B be a nonempty closed convex subset of a Banach space and $T_j, S_j, j = 1, 2, \dots, k$ be two families of total asymptotically nonexpansive self-maps of B. Presume that $F(T_j, S_j) \neq \emptyset$ and $\sum_{n=1}^{\infty} f_n < \infty, \sum_{n=1}^{\infty} g_n < \infty$.

Presume that (a_n) be as shown in step (1) is strong convergence to a a common fixed point of T_j and S_j iff $\lim_{n\to\infty} inf\ d(a_n, F) = 0$, where $d(a, F) = inf_{a^* \in F} \|a - a^*\|$.

Corollary (2.9): Let T_j, S_j, B, f_j and $g_j,\ j = 1, 2, \dots, k$ be as in corollary (2.8). Therefore (a_n) be as shown in step (1) is strong convergence to $a^* \in F(T_j, S_j)$ iff $(a_{ni})\ of\ (a_n)$ that converges to a^*.

Corollary (2.10): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, \forall j = 1, 2, \dots, k$ be two families of Lipschitzain and total asymptotically nonexpansive self-maps of B. If the dual space $M^*\ of\ M$ has the Kadec-klee property and the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demi-closed to zero, therefore (a_n) be as shown in step (1) is weak convergence to a common fixed point of $T_i\ and\ S_i$.

Corollary (2.11): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, \forall j = 1, 2, \dots, k$ be two families of total asymptotically nonexpansive self-maps of B. If M accomplishes Opial 's condition and the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demi-closed to zero, therefore, (a_n) be as shown in step (1) is weak convergence to a common fixed point of $T_j, S_j, \forall j = 1, 2, .., k$.

Corollary (2.12): Let B be a nonempty closed convex subset of a Banach space and $T_j, S_j, j = 1, 2, \dots, k$ be two families of asymptotically quasi-nonexpansive self-maps of B. Presume that $F(T_j, S_j) \neq \emptyset$ and $\sum_{n=1}^{\infty} f_n < \infty$. Presume that (a_n) be as shown in step (1) is strong convergence to a common fixed point of T_j and S_j iff $\lim_{n\to\infty} inf\ d(a_n, F) = 0$, where $d(a, F) = inf_{a^* \in F} \|a - a^*\|$.

Corollary (2.13): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, \forall j = 1, 2, \dots, k$ be two families of Lipschitzain and asymptotically quasi-nonexpansive self-maps of B. If the dual space $M^*\ of\ M$ has the Kadec-klee property and the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demi-closed to zero, therefore, (a_n) be as shown in step (1) is weak convergence to a common fixed point of T_j and S_j.

Corollary (2.14): Let B be a nonempty closed convex subset of a uniformly convex Banach space and $T_j, S_j, \forall j = 1, 2, \dots, k$ be two families of Lipschitzain and asymptotically quasi-nonexpansive selfmaps of B. If M accomplishes Opial 's condition and the maps $I - T_j$ and $I - S_j, \forall j = 1,2, \dots, k$ are demi-closed to zero, therefore, (a_n) be as shown in step (1) is weak convergenve to a common fixed point of T_j and $S_j, \forall j = 1, 2, .., k$.

3. Numerical Example

We illustrate our results by the following

Example (3.1): Let $T_j, S_j: R \to R, \forall j = 1, 2, \ldots, k$ be two maps such that $T_j a = \frac{2a}{3^j}$ and $S_j a = \frac{46a}{2^j}, \forall\, a \in R$. Choose $\alpha_{jn} = \frac{n}{4(n+1)}\ \forall\, n$ with initial value $a_1 = 15$. Let (a_n) be the sequence. According to Table 1 and Figure 1, it eases to sight that (a_n) converges to the fixed point $a^* = 0$.

Table 1: Numerical results corresponding to $a_1 = 15$ for 36 steps.

n	Iteration (1)	n	Iteration (1)
1	15.0000	**13**	0.0184
2	9.4401	**14**	0.0102
3	5.6594	**15**	0.0056
4	3.3084	**16**	0.0031
5	1.9044	**17**	0.0017
6	1.0849	**18**	0.0009
7	0.6134	**19**	0.0005
8	0.3448	**20**	0.0003
9	0.1930	**21**	0.0002
10	0.1076	**22**	0.0001
11	0.0599	**23**	0.0000
12	0.0332	**24**	0.0000

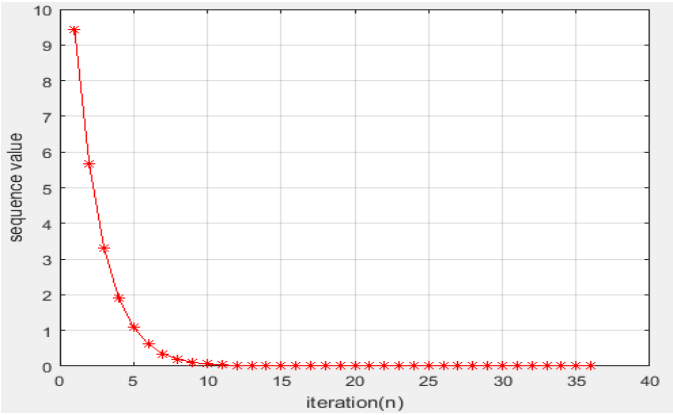


Figure. 1. Convergence behavior corresponding to $a_1 = 15$ for 36 steps.

4. Conclusion

We study the strongly and weakly convergence of new type of finite-step iteration processes under total asymptotically quasi-nonexpansive maps, see Theorems (2.4)-(2.6). Our results are generalizing and unifying the results of others who have been referred to in the references.

5. Open Problem

Recently, S.S. Abed has been defined as the following type of generalizations of total asymptotically quasi-nonexpansive[16]: Let $\mathcal{A}$ be a subset of real Banach space a set–valued map $G: \mathcal{A} \to 2^{\mathcal{A}}$ is called the general asymptotic set-valued if for each $x \in \mathcal{A}$ there exists null non- negative real sequences $\{a_n\}$ and $\{b_n\}$ such that

$$\mathfrak{D}(G^n x, G^n y) \leq \|x - w\| + a_n\, \mu(\|x - w\|) + b_n \qquad (7)$$

for any $y \in \mathcal{A}$, $w \in G^n y$ and $\mu: R^+ \to R^+$ with $\mu(0) = 0$. One can study convergence theorems in (1) and in [17, theorem (11)] for families of general asymptotic set-valued maps. As well as possible to demonstrate new results in the case of other spaces as a modular space [18].

Acknowledgment

The authors wish to express his thanks to the referees for their helpful advice.

References

[1] S.S.Abed, Z.M Mohamed Hasan "Convergence theorems of a finite-step iteration algorithm under two finite families of total asymptotically quasi-nonexpansive maps" IEEE Xplore Digital Library Internationl Conference on Advanced Science and Engineering (ICOASE), , 2018.

[2] U K.Goebel, W. A. Kirk, "Fixed point theorem for asymptotically nonexpansive mappings" Proc. Amer. Math. Soc. 35, 171-174, 1972.

[3] W. V. Petryshyn, T. E. Williamson, "Strong and weak convergence of the sequence of successive approximations for quasi-nonexpansive mappings", J. Math. Anal. Appl. 43, 459-497, 1973.

[4] YA. I. Alber, C. E. Chidume, H. Zegeye, " Approximating fixed points of total asymptotically nonexpansive mappings" Fixed point theory Appl. 2006, article ID 10673.

[5] G. S. Saluja, "Convergence to common fixed points for generalized asymptotically quasi-nonexpansive mappings" Bull. Soc. Math. 4, 69-79, 2014.

[6] G. S. Saluja, "Strong convergence theorems for two finite families of generalized asymptotically quasi-nonexpansive mappings with application" Math. Moravica, 22(1), 1-14, 2018.

[7] S.S.Abed, Z.M Mohamed Hasan,."Common fixed point of a finite-step iteration algorithm under total asymptotically quasi-nonexpansive maps" is accepted in Baghdad Science Journal.

[8] J. B. Diaz, F. B. Metcalf, "On the structure of the set of subsequential limit points of successive approximation" Bull. Amer. Math. Soc., 73, 516–519, 1967.

[9] D. R, Sahu, D. ORegan, R. P. Agarwal ,"Fixed point theorey for Lipschitzain-type mappings with applications, Topological fixed point theory and its applications" doi:10.1007/978-387-75818-3-1.

[10] Q. H. Liu, "Iterative sequences for asymptotically quasi-nonexpansive mappings" J. Math. Anal. Appl. 259(1), 1–7, 2001.

[11] A. Abkar, M. Shekarbaigi,"A novel iteration algorithm applied to totally asymptotically nonexpansive mappings in CAT(0) spaces" Mathematices 5,14, 2017, doi:10.3390/math5010014.

[12] F.E.Browder," Semicontractive and semiaccretive nonlinear mappings in Banach spaces" Bull. Amer. Math. Soc. 74, 660-665, 1968.

[13] A.Sharma, M.Imdad, "Approximating fixed points of generalized nonexpansive mappings Via faster iteration schemes" Fixed point theory, 4(4), 605-623, 2014.

[14] C. E. Chidume, E.U. Ofoedu, "Approximation of common fixed points for finite families of total asymptotically nonexpansive mappings" J. Math. Anal. Appl. 333, 128-141, 2007,doi:10.1016/j.jmaa.2006.09.023.

[15] J.G. Falset, W. Kaczor, T. Kuczumow, S. Reich, " Weak convergence theorem for asymptotically nonexpansive mappings and semigroups", Nonlinear Anal. 43, 377- 401, 2001.

[16] S.S.Abed, "Approximating fixed points of the general set –valued asymptotically mappings " to appear.

[17] S.S.Abed, Z.M Mohamed Hasan "Convergence comparison of two schemes for common fixed points with an application" accepted in Ibn Al Haitham journal for pure and applied sciences.

[18] S.S.Abed, K. E. Abdul Sada, " Common fixed points in modular spaces", IHSCICONF 2017, Ibn Al Haitham journal for pure and applied sciences.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 73-93 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Intelligent Foundations for Knowledge Based Systems

Mukundan Kandadai Agaram*

Delta Dental of Michigan, 4100 Okemos Road, Okemos, MI 48864, U.S.A,

ARTICLEINFO

Article history:
Received: 18 April, 2019
Accepted: 23 June, 2019
Online: 15 July, 2019

Keywords:
Knowledge Based Systems
Rule Based Information Systems
Intelligent Discovery
Data Visualization
Enterprise Decision Management
Master Data Management
Software Architecture
Enterprise Architecture
Domain Driven Design
Healthcare Adjudication

ABSTRACT

Knowledge Based Systems play a very important role, within Healthcare, with a primary goal of supporting a high quality service at an optimal cost. A widely accepted knowledge acquisition technique is through the use of Business Rules in a natural language like format. As clinical terminology is centric to HealthCare, a large percentage of these rules use industry standard codes to describe clinical concerns. Usually there is a large number of codes to denote the various clinical processes and procedures within Healthcare. These can translate to a combinatorially explosive number of Business Rules within the HealthCare IT system. It is common practice to leverage a Rule Engine to execute these rules and produce decisions within the system. A Rule Engine accepts a collection of rules called a ruleset. Within Healthcare Insurance, these rules and rulesets embody regulatory, policy and contractual concerns. To effectively manage this huge body of rules and rulesets, it is typical for Knowledge Based Systems to reuse rules across rulesets. Further, the knowledge required to author these rules and constructing comprehensive rulesets is specialized and requires deep expertise within the domain. Further, this also requires an expertise with authoring unambiguous business rules and operating the Rule Management Systems or Knowledge Based Systems. There is a need for the continuous Governance of these rule based knowledge assets for the organization. Typically knowledge experts in this field have a decade or more of experience. An emergent challenge or trend witnessed within the industry is that the experienced knowledge workers are retiring and their positions are being replaced by people not as experienced. This paper proposes techniques to build a layer of intelligent capabilities that can transform the methods for the automation, creation and maintenance of knowledge artifacts, that can aid and support the inexperienced knowledge workers with the effective management and administration of these assets within a Healthcare system.

1 Introduction

Healthcare domains are characterized by a rich knowledge accumulated through a combination of science, technology, experience, innovation within the industry and standardization through regulatory entities. Given the high cost of administration of health care, it is common practice to be reliant on a strong infrastructure of insurance to fund the high costs of administering healthcare. Further, there is a strong demand and pressure for absorbing scientific and technological advances within the field into operational practice within a short period of time. There is a high risk of adverse practice effects, fueling the necessity for standardization of practice, within the industry. This paper is an extension of the paper titled *Intelligent discovery features for EDM and MDM Systems* originally presented at 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW) [1]. This paper further extends the concepts beyond the discovery features described in the original paper. It further formalizes these concepts into a set of foundational patterns, called *intelligent foundations*, with a view to providing automation towards hard to scale and hard to automate human labor intensive processes around business rule mining and knowledge manage-

*Mukundan Kandadai Agaram, Delta Dental of Michigan, 4100 Okemos Road, Okemos, MI 48864, U.S.A, Phone: +1 (517) 381-4736, Email: magaram@deltadentalmi.com

ment. This paper further lays down a formalized foundation of ingredients to construct recipes for building intelligent virtual knowledge working software agents to aid a new breed of digitally savvy, but functionally inexperienced workforce.

1.1 Healthcare in the United States

In the US, healthcare can be broadly categorized into the *Provider* i.e. the clinical entity viz., hospitals, doctor's offices, clinics etc. that provides the clinical services and the *Payor* i.e. the funding entity a.k.a insurance companies that pays for these services. The cost of healthcare is expensive and exorbitant to the general population. In addition it is highly susceptible to litigation, which further acutely drives up the cost. Healthcare insurance infrastructure is predominantly operated by the private sector, through a combination of "for profit" and "not for profit" entities. Stereotypically, healthcare insurance within the US is offered as a benefit, to employees of companies. The employers purchase health insurance for their employees through customized benefit plans, from healthcare payor companies. The health insurance costs are more affordable when bought in bulk for larger groups of people and is often referred to as the "group discount". This is due to the fact that the risk of "adverse selection" is significantly lowered. The cost of insurance is high for the average citizen seeking to buy it [2], [3], [4], [5]. Recent reforms in healthcare have attempted to introduce competitive pricing through the notion of healthcare exchanges at competitive rates. The general gist of this is to be able to provide insurance for healthcare to citizens, who otherwise cannot participate in employer health benefit plans either due to lack of employment or lack of availability of the benefit from the employer. In the US, there is a specialization of insurance products for *Medical* healthcare, *Dental* healthcare and *Vision* healthcare.

1.2 Background

The work presented in this paper and the original foundational paper is part of an evolving case study in the practical application of data science, analytics and visualization techniques that can automate the process of knowledge analysis and mining at a Dental Insurance company. Delta Dental of Michigan, Indiana and Ohio is the leading provider of dental benefits, serving the mid-western part of the US, with a subscriber base of nearly 15 million members throughout the states of Michigan, Indiana, Ohio and it's growing base of other affiliated states in other parts of the US. The company has witnessed a strong growth over the last decade of it's 60+ years of history. A key factor for this growth is it's highly automated insurance and benefits administration software platform, that achieves an industry enviable 95% auto adjudication rate (the percentage of claims that are entirely processed by the software platform with no human intervention). The cornerstone of the software platform is a sophisticated rule management system called BRIDE, which enable the domain *knowledge workers* or Subject Matter Experts (SMEs), to author, test and manage the operational rules, using highly configurable *Domain Specific Languages* and a semantically rich *Business Vocabulary* [6]. The BRIDE platform is an innovative homegrown product which is the result of original collaborative work by the author and other fellow contributors, which is elaborated in a published conference paper [7]. The knowledge externalized into BRIDE as operational *Business Rules*, can be translated and deployed to any forward chaining production rule engine format [8].

1.3 Motivation

The process of deciding to pay a claim a.k.a *adjudication* in the healthcare claims arena, requires a deep understanding of healthcare clinical knowledge. Within the dental healthcare space, it is not uncommon to manually adjudicate complex and costly dental procedures, by dental experts who have had prior hands on experience with the clinical aspect of the science. Further a second dimension of expertise is to understand, the payor's business policy as it pertains to the interpretation of allowances and coverages. As a consequence, a typical knowledge worker has decades of experience within the industry and/or company. However, these experts are now reaching retirement age and will be transitioning their responsibilities to younger and lesser experience functional specialists. This provides strong motivation to build *"intelligent capabilities"* into the platform, that can automate laborious manual processes and to provide a decision support system for generating or recommending rules to the knowledge workers as they author these rules. This will be elaborated further in subsequent sections.

1.4 Organization

This paper is organized as follows. In 2, we will explore the characteristics of knowledge sources, their nature and consequences, within the context of healthcare claims processing and adjudication. We will cast a wider net to describe how the patterns presented by these problems can be applied to other functional verticals with similar characteristics. This will then provide a segue to 3, where the subject of *intelligent capabilities* will be explored with a viewpoint of automating manually intensive tasks and to provide seamless decision support to the act of *rule/knowledge mining*. In 4, we will explore the subject of building intelligent foundations and capabilities with the backdrop of the BRIDE [7] case study, that can support a growing and diverse eco-system of artificial intelligence. In 5 we will discuss inherent challenges in addressing the intelligent automation of knowledge/rule mining. This will naturally set the stage for discussion of future research/work, in 6. We will finally provide a meaningful conclusion for the work in 7.

2 Knowledge Characteristics

Knowledge/Rule Mining through traditional means can be a manually intensive process. Traditional Rule-based Information Systems, rely on the ability of human experts, to author high quality, non-ambiguous system enforceable rules. Further, knowledge rich areas such as clinical domains, yield a high number of such rules. These rules exhibit symptoms of combinatorial explosion, when certain influencing parameters change. These parameters are commonly influenced by external forces, such as customer needs, advances in science, market trends, regulatory mandates , adherence to standards etc. Second, the complexity of the rules can be high. In certain cases automation of the business logic, through procedural programming techniques can yield, very large (in the order of 10K lines of code), extremely complex programs [9]. The maintenance cost of such programs increase dramatically over time. In the face of such complexity, a system of externalized rules provides significant cost savings and improved agility to change. However, the system of rules that are mined out from large legacy procedural language programs is iterative in nature and is best achieved in a phased manner.

The original paper [1], outlines the complexities of the problem, especially around the mining, discovery and management of business domain knowledge as externalized business rules, into six different categories viz., *Volume and Growth Complexity*, *Organizational and Reuse Complexity*, *Rule Logic Complexity*, *Domain and Subject Matter Complexity*, *Healthcare Complexity*, *Healthcare Consolidation Complexity*. In this section we will contrast these with other intrinsic characteristics surrounding *Knowledge* and their relevance within surrounding eco-systems.

Healthcare systems operate around a hub of knowledge. This hub is surrounded by a triad of pillars, *People*, *Process* and *Technology*, which is depicted in 1. The *People* realm represents the Clinical knowledge workers, Business User Community, Claims/Insurance Practitioners etc., and their collective knowhow. The *Process* realm represents the organization's conscious act of documenting the knowhow with a view to standardizing, documenting and repeating them in a predictable fashion. The *Technology* realm reflects the collective collage of automation software, which in the case of the Healthcare Payor business, could be the Claims Processing System, Benefits Administration System, Accounting System etc. The company's knowledge assets live in the hub of these three pillars. We will now proceed to examine the characteristics of knowledge which intersects with each of these pillars. We will further examine the influences of introducing *intelligent foundations* to *Knowledge Management* and how it helps to preserve the most valuable assets of the organization i.e *knowledge*.

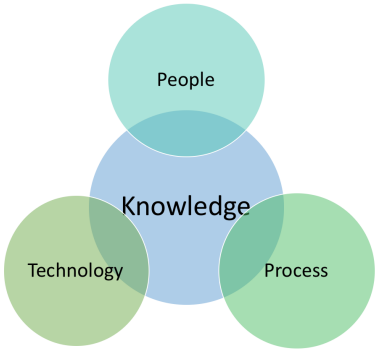


Figure 1: Knowledge Hub

2.1 Documented Knowledge

Documented Knowledge as the name suggests, represents all documented artifacts describing business processes, policies, contracts etc. Documented knowledge, exists in a variety of formats for different audiences. Healthcare is a highly regulated domain and is subject to a variety of audits. Non-compliance with regulation and audit findings present dire consequences, in the shape of costly penalties and potential litigation. Hence, it is crucial to maintain detailed documentation of processes, contracts and such. The basis of *Documented Knowledge* is depicted in 2.

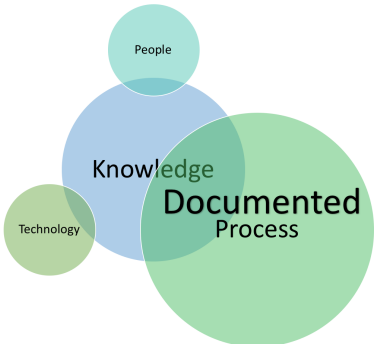


Figure 2: Documented Knowledge

Documented Knowledge, has the following consequences.

1. Footprint of *Documented Knowledge* can be huge and verbose. As it grows over time, the maintenance and upkeep of it becomes increasingly labor intensive and costly. As a result it is necessary to have quick and correlated discovery available to navigate the knowledge bases.

2. *Documented Knowledge* through traditional means is not *interactive* and *adaptable* to different audiences and contexts. Unlike *tribal* knowledge where knowledge workers can decide to provide a detailed explanation or high level overview depending on the appropriate audience and context, documented knowledge needs to be captured at varying levels of abstraction.

3. The knowledge can become outdated quickly, i.e. the knowledge bases if not updated periodically

can lose its accuracy, and hence its efficacy and value. To derive continual value from this source of knowledge, proper controls need to be put in place to effect the timely update of these knowledge resources.

4. Being textual and descriptive, semantics can get distorted. To mitigate this, the documented knowledge bases need to be reviewed and updated periodically by knowledge workers.

But if done right, this is an important source of knowledge and as such an asset to the enterprise. We will further explore how building a foundation of intelligent capabilities within the Knowledge Based Systems will capture the documented knowledge at the source of knowledge, through explicit means i.e. where the users can document aspects of pertinent contextual knowledge and implicit means where the underlying system learns, predicts and creates knowledge content.

2.2 Tribal Knowledge

Tribal knowledge is any undocumented information that is not commonly known by others within a company. It can also be thought of as the collective wisdom of the organization. It is the sum of all the knowledge and capabilities of all the people [10]. In a highly specialized functional domain such as healthcare, detailed nuances of claims adjudication procedures, interpretation of high level organizational policy or adaptation of external regulatory stances or standards can reside in this domain. Typical knowledge workers, with decades of experience in the field and more importantly experience administering the knowledge, within the domain within the same company are a very rich source of such knowledge. The origins of *Tribal Knowledge* in the *People* domain is depicted in 3.

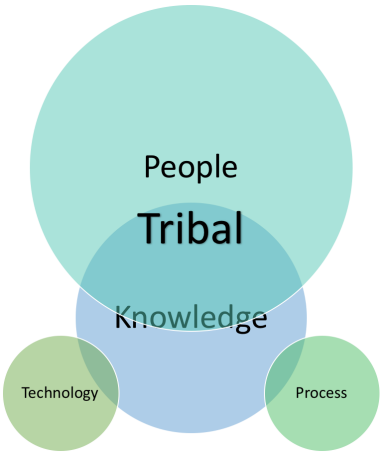


Figure 3: Tribal Knowledge

This form of knowledge has the following consequences.

1. There is a high risk of erosion, as the people holding them leave the organization through attrition or retirement. This presents a strong motivation to build intelligent systemic capabilities that can tap into this rich source of tacit knowledge and be able to capture and leverage where possible, within the organization.

2. Traditional methods for capturing or documenting the knowledge i.e. transforming it into *Documented Knowledge*, in entirety is challenging at the least.

3. Tribal Knowledge can be clouded by subjective influences.

4. Some aspects of tribal knowledge pertains to the expertise or experience of the knowledge workers and their innate ability to leverage different sources of knowledge. This can be very contextual and situational by nature and unless well integrated into the knowledge management platform can become very hard to assimilate and correlate contextually. Hence, there is a very strong need to not only capture knowledge, but to artfully disseminate, correlate and index it for easy and fuzzy discovery.

The industry at large has been undergoing a significant generational workforce transition [11], which puts a heavy burden on the organization to effectively transition the tacit tribal knowledge held by the seasoned "baby boomer" generation, to a younger inexperienced generation. Special attention needs to be paid towards building a foundation of *Artificial Intelligence*, that can aid in the seamless capture from the older generation while effecting a smooth transition with interactive decision support capabilities back to the newer generation joining the workforce. On the other side of the coin, there is a strong need to attract and retain digitally-savvy, millennial and post-millennial generations of workers [13], while training them to be future knowledge workers.

2.3 Knowledge Systemization

We will now examine the concept of *systemization* of knowledge. This is the act of capturing the knowledge into a software system with a view of automating it. This can be achieved with one of two approaches.

1. In this approach, the knowledge is programmed and automated into software applications that capture the essence of the knowledge. We will call this the *Black Box* approach, for the obvious reason that the knowledge is now buried deep within the implementation of a software system, as depicted in 4.

   This approach while being non-transparent to the non-programming business user community, at least does captures the knowledge within the bounds of a software. In theory this knowledge could be mined out (albeit painfully) and documented. However, there are some drawbacks.

(a) The organization might not own the software system, which implements the business logic. In this scenario the organization has limited control on the mining effort at a later stage and has to entirely depend on the owning vendor of the software to mine out their business logic as metadata.

(b) The software system is programmed in a legacy technology, with limited technical resources who understand the technology. This presents the same consequences challenges of erosion as posed by *Tribal Knowledge*, described earlier in 2.2

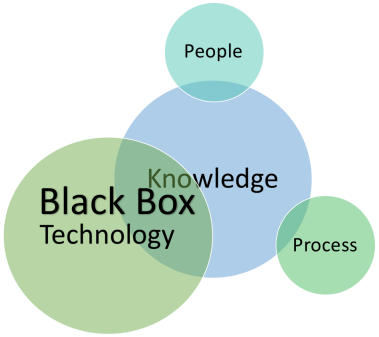


Figure 4: Blackbox Knowledge

2. This approach can be described as the act of formally capturing the knowledge in a *Knowledge Management System*. The most tangible forms of knowledge are *business rules* and *business policies* as these tend to be precise statements describing aspects of the business. Within the healthcare payor space, these are typically captured as operational rules that can be automated. Such knowledge is *externalized* into a Rule Management System a.k.a Decision Management System as operational rules intended for consumption by a software system for automation, but more importantly externalized to be human readable and comprehendible by non-programming savvy users. We will call this the *White Box* approach, as this knowledge is *externalized* and is now transparently available to be understood and manipulated by the organization. The knowledge is externalized and managed under *Knowledge Management*, as shown in 5.

   Externalization of the knowledge, presents the following benefits.

   (a) Externalization as described above, makes the organization more *agile*, as they now have the ability to change the operational business rules without making systemic changes which require a software release cycle. Further these changes can be directly controlled by the end business users, rather than the Information Technology community within the organization.

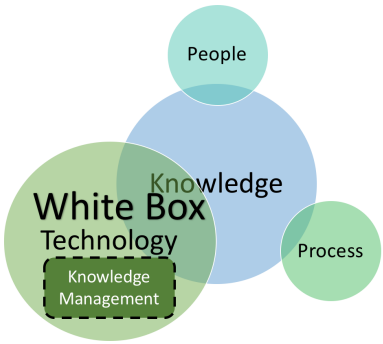


Figure 5: Whitebox Knowledge

3. Externalization of knowledge makes the organization's operation more transparent, but more importantly explainable as opposed to the *Black Box Approach*.

   However, these are some consequences of knowledge externalization that need to be considered.

   (a) The number of externalized operational rules and policies could be large. Further, these rules are combined into large rulesets can present a combinatorially explosive [7], difficult to manage scenario over time. It takes highly experienced rule authors and knowledge workers well versed in the science and the organization's business operations, to effectively manage this, over time. This exposes the organization to the risk of knowledge erosion.

   (b) While the externalized knowledge is more transparent and explainable, to be able to comprehend far reaching cause and effect, the underlying system needs rich traceability capabilities [12], to provide detailed comprehensive explanations to the business users. This provides the deep motivation for *Intelligent Foundations* and providing an intelligent way to manage knowledge and sets up the business case for a *Virtual Knowledge Worker*, that is capable of answering deep *knowledge* questions, something that can aid and support the new younger generation of *digitally savvy* knowledge workers.

2.4 Intelligent Foundations

As discussed in 2.2, transition of the labor force [11], and catering to a diverse workforce with different knowledge assimilation requirements [13], presents a big challenge. This bolsters the case to build sophisticated and *intelligent capabilities* to address these needs. The intelligent capabilities will be elaborated further in 3. In this section, we will introduce the concept of *intelligent foundations*. Building such a foundation can then enable *intelligent capabilities*, that can rapidly train a new generation of young workforce, on the deep knowledge expertise. This is depicted in 6. The idea is as *Tribal Knowledge* erodes in an organization, the

new business users can learn rapidly, thereby restoring the tribal knowledge held by the people who leave the organization.

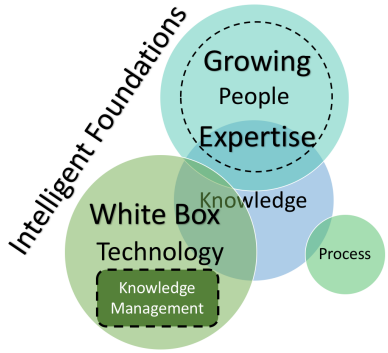


Figure 6: Intelligent Foundations

These *Intelligent Foundations*, can also be articulated as *capability* statements. We define a capability as a simple statement that describes an *ability* or feature of the system. Here are the ten foundational capabilities, at a high level, that enable building more specific intelligent capabilities around knowledge analysis, management and intelligent automation. These will be discussed in more depth in 4, against the backdrop of a case study, within BRIDE [7].

1. *Discovery:* Ability to fuzzily discover diverse artifacts
2. *Tagging:* Ability to *tag* or decorate knowledge artifacts with manual and generated metadata.
3. *Visualization:* Ability to graphically visualize *affinity* between artifacts
4. *Classification:* Ability to classify artifacts based on characteristics
5. *Correlation:* Ability to build correlation between different artifacts
6. *Event Processing:* Ability to generate, process and react to system and user events, within the platform.
7. *Recommending:* Ability to recommend action to take based on correlation between artifacts
8. *Language Processing:* Ability to absorb, disseminate and score keywords from textual content/documentation at detailed inflection points
9. *Generation:* Ability to generate new content
10. *Learning:* Ability to learn from data patterns and from human feedback

2.5 Knowledge Governance

Healthcare as discussed earlier is a highly regulated, knowledge rich domain. There is a strong need to formalize, document and more importantly *govern* the knowledge, which refers to choosing structures and mechanisms that can influence the processes of sharing and creating knowledge [14], surrounding the detailed aspects of administration, business rules, business policy and processing. Within healthcare the business rules around administration are driven by different motivations, varying from regulatory, industry standardization, organization policy to client contractual customizations. Understanding and cataloging the various business rules and associated knowledge artifacts, contributes to better organizational management of the same. This consequently enhances the documented process knowledge surrounding them. By building intelligent capabilities around integrating and automating aspects of *Knowledge Governance* into the Knowledge Management System, aids in better administration of the governance process. Further it centralizes and promotes the use of the externalized knowledge across all three pillars i.e. *People*, *Process and Technology*, in a seamless fashion. This is shown in 7 .

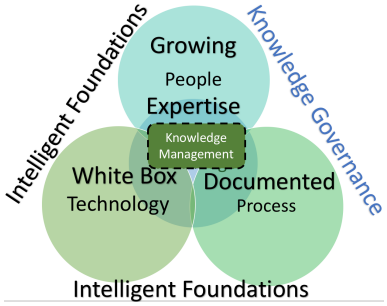


Figure 7: Intelligent Foundations with Governance

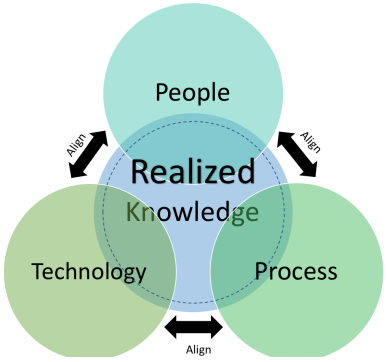


Figure 8: Realized Knowledge

A big goal is to better *align* knowledge sources across the three pillars of knowledge and unify them as a net sum of *Realized Knowledge*, as shown in 8. The introduction of *Intelligent Foundations*, coupled with *Knowledge Governance* strives to achieve that [15].

3 Intelligent Capabilities

In this section we will describe a subset of *Intelligent Capabilities* that are built on top of the set of ten foundational intelligent capabilities i.e. *intelligent foundations*

described in 2.4. These *Intelligent Capabilities* cover the vast gamut of functions needed to effectively manage a complex, vast, functionally diverse and growing footprint of externalized knowledge, as seen in healthcare. A logical viewpoint of this framework is shown in 9.

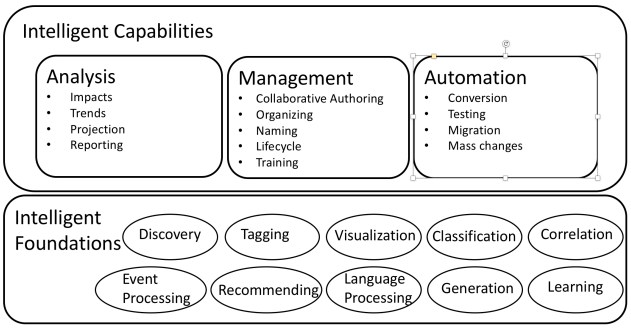


Figure 9: Intelligent Knowledge Management Framework

The capabilities are broken up into these broad categories viz., *Analysis*, *Management* and *Automation*. We will elaborate each of these broad categories of capabilities.

3.1 Analysis

Analysis covers the most important and far reaching capability across the organization. Practically every *business user* would need access to some or all aspects of this capability to perform their job effectively. As mentioned earlier, authoring operational *business rules* is a central part of knowledge externalization. With healthcare, some of the most complex rules are around *clinical terminology*.

To illustrate this further, we will contrast this with a Dental Insurance use case. With dental health insurance, clinical rules are specified using ADA (American Dental Association) mandated CDT (Current Dental Terminology) Procedure Codes. The knowledge workers are making changes to these rules as contractual benefits with customers change, or when the ADA releases a new version of the CDT, each year. Each Dental Procedure has a nomenclature that describes the procedure using key clinical terminology. When a certain aspect of dental clinical science changes, the ADA makes appropriate changes to the CDT to -

1. Introduce new Procedure Codes to reflect new clinical procedural advances
2. Retire obsolete Procedure Codes

In the United States, all healthcare business participants are mandated to be HIPAA (Health Insurance Portability and Accountability Act) compliant. In Dental Insurance, among other things, this means that Dental Providers (Dentists, Dental Offices) and Payors (Dental Insurance companies) are CDT compliant. This translates to the Dental Insurance adjusting thousands of dental clinical rules to reject or adjudicate claims based on the correct/appropriate dental procedure code being submitted.

These are some of the more fine-grained use cases within the *Analysis* capability.

1. *Impact Analysis:* Sometimes there are business rule changes across segments or the whole book of business, to specific areas of treatment, caused by a variety of reasons. These reasons, for example can be one of the following:
   (a) Periodic ADA CDT version changes that are introduced on a yearly basis.
   (b) Changes to a large client's contractual benefits on renewal, on a periodic basis.
   (c) Introduction of a new line of business or product.
   (d) Regulatory mandates.
   (e) Corporate directional changes.

   There is a strong need, for an ability to analyze and make micro-level operational rule changes based on macro-level shifts in function. To be able to do this the system needs the ability to correlate and analyze impacts across varying granulation of knowledge assets.
2. *Trend Analysis:* Trend Analysis can offer up useful insights to the organization and is intrinsic to the collective knowledge it holds. Here are some examples of trend analysis.
   (a) *Reuse Trending:* As mentioned in 2, certain knowledge rich domains, can have combinatorially explosive rules under certain conditions. Therefore there is a need to provide appropriate tools for managing the analysis of trends to minimize the explosion of such rules through intelligent optimization and reuse.
   (b) *Metric Trending:* The collection of accurate metrics leads to better knowledge governance and quality of rules/knowledge that yield better business decisions. For example, within the dental claims adjudication arena, we would like to analyze the most popular rules and be able to see seasonal or long term trends in their usage. Another useful analysis is to study the trending of the auto adjudication metric, which is the percentage of adjudicated claims that do not require any form of manual expert intervention, as this directly ties to the cost of administration.
   (c) *Volatility Trending:* It is important to understand how frequently a knowledge artifact or rule changes. These can lead to deeper questions that would help analyze - are there seasonal influences to the change ? Are these changes reflective of changes to the book of business ? Can we use this as a future predictor of like phenomena?

(d) *Variance Trending:* When looking at a collective body of knowledge artifacts for example, an entire set of rules, there is typically a variant component and an invariant component. This form of trending is again useful as a future predictor to study correlational causal effects with underlying books of business. For example the variance trending for Government Contracts can differ from Contracts, underwritten for small commercial clients, or clients within a certain vertical industry like car manufacturers or the demographic makeup of a certain group of underwritten people within a group.

(e) *Complexity Trending:* Are knowledge artifacts (i.e. rules, rulesets, policy, key terms etc.) within a certain sub-domain of the business growing in complexity ? This form of trend analysis helps identify and proactively address the management of such artifacts.

(f) *Obsolescence Trending:* It is not uncommon to *retire* knowledge artifacts i.e. business rules or policies, over time and this form of trending can be an indicator in correlating the obsolescence of knowledge with evolving the book of business. In a certain sense this is similar to the volatility trending, discussed above, however obsolescence of knowledge can provide certain key insights and indicators around directional change for the organization, especially around obsolete business models or outdated processes, which are valuable.

3. *Projection Analysis:* Projection Analysis is the ability to analyze existing knowledge artifacts, to generate or "project" inductively, through inferred heuristics, gained from historical correlation of associated historical contextual changes to the corresponding changes of knowledge artifacts. This technique of analysis provides an approach for both the system and humans within the knowledge economy to project or generate knowledge artifacts based on the correlated data. This correlated data can further be used to generate features, to develop future state capabilities such as machine learning based on the generation and validation of projections based on correlation between characteristics of knowledge artifacts and high-level characteristics.

4. *Reporting Analysis Results:* With *Intelligent Capabilities*, built on AI foundations of fuzzy discovery, classification and correlation, business users can now be empowered to gather insights with deep drill down capabilities. It is important however to be able to package the results of ad-hoc analysis sessions into interactive reporting that can be exported and shared to a broader audience or community within the organization. It is also important to be able to provide interactive *data visualization* features, packaged into this capability. By combining multiple *Intelligent Foundation* concepts, new users can be onboarded and trained on complex interrelated concepts in a more effective, timely and continual manner.

3.2 Management

Traditionally, management of knowledge assets within an organization requires deeply experienced multifaceted knowledge workers, well versed in the domain terminology and the functional vertical. In addition, the knowledge workers tasked with authoring externalized business rules, need to possess traits that can translate complex functional subject matter into clearly articulated, unambiguous business rules and policies. Lastly, they need to be adept at navigating a complex knowledge management system, replete with a rich interconnected set of business vocabularies and complex features for other life cycle functions of knowledge management. The *Intelligent Capabilities* around management can be broken up into these following broad areas:

1. *Collaborative Knowledge Authoring:* This capability is at the heart of any Knowledge Management System, and to a large extent determines the success and efficacy of such a system. With traditional means i.e. with no *intelligent foundational* capabilities, a big part of the success and efficacy of the external knowledge hinges on the skills of the knowledge workers to deliver the quality. By building this capability on an intelligent foundation, new possibilities abound. For example here are a few plausible use cases, that leverage intelligent management capabilities. With a *collaborative* approach, the system collaborates seamlessly, with the users and facilitates collaboration amongst participating knowledge workers

   (a) *The authoring tool provides a guided wizard like experience for complex rules.* For example, for new business rule authors who are not familiar with the intricacies of the detailed business vocabulary of a complex domain, the act of authoring a complex business rule that achieves a granular operational business decision can be a daunting task. In such scenarios, a high level, questionnaire driven wizard can step the business rule author through the act of authoring.

   (b) *The authoring tool provides impromptu training sequences for guided authoring.*An extension of 1a, described above could show a step by step progression of an authored business rule to the new business rule author. This will help train that business rule author to learn the intricacies of the business domain vocabulary and it's associated key terms. Further, such sessions could be

recorded for future playback as interactive help topics.

(c) *The authoring tool provides help based on correlation with keywords used in key business vocabulary.* This feature is another interactive training feature for the authoring tool to train new users. Further, while this use case is centric to the day to day management of knowledge, it can be used as a key analysis feature, that can help the business rule authors to do integrated impact and correlation analysis as they author business rules, policies and other knowledge artifacts.

(d) *The authoring tool disseminates complex authored rules to catalog and correlate keywords to contextual help topics.* As new knowledge assets i.e. Business Rules, Policies etc., are created inside the Knowledge Management System, new business terms and language could be introduced. A consequence of this is that only small segments of the business community are aware of the semantics and impact of the newly introduced terminology. This bolsters the case for cataloging and correlating this information at the point of entry, into the management system. This could make the system smarter through the continual layering of correlational metadata.

(e) *The authoring tool provides a means for a community of knowledge workers, to collaboratively author related knowledge artifacts, concurrently by providing push notifications of changes and preventing conflicting or inconsistent changes to occur.* This capability leverages the intelligent foundations in 6, (the seminal book [16] provides a sound introduction to the science of events). Using this, the platform is capable of sending push notifications of user edit events, to different users, analyzing sequences of events, to make inferential decisions about dynamically constraining certain situational edit operations for certain knowledge assets, by certain participating knowledge workers.

2. *Knowledge Organization:* To manage the combinatorially explosive growth of knowledge assets, there is a compelling need for a sophisticated organization method that promotes the optimal reuse of knowledge assets. To be able to effectively reuse common knowledge assets like business rules, the first barrier to overcome is the ability to discover (rather fuzzily) that such a business rule or a *similar* business rule does exist. In the absence of this, a lot of duplicate or near duplicate knowledge assets are created. Organization of knowledge is key to enabling knowledge recall, management, mining and last but not the least Governance. We will now consider the organization of knowledge from the following vantage points.

(a) *Structure and Granulation:* Here we need to consider how the knowledge is structured and abstracted. Typically the knowledge artifacts can be structured in a hierarchical fashion, from most granular to least granular (i.e. high-level to low-level). Further each level of granulation can have several viewpoints. For example in dental insurance, the highest level of knowledge can be emblematically represented in a client/customer contract, which in turn, can be represented as a ensemble of *business policy* statements, which are high level statements of business purpose, that are driven through a combination of regulatory, contractual and administrative concerns.

These business policies are implemented through sets of operational *business rule* statements, that enforce the intent of the policy, within the Claims Administration platform. The breakdown of artifacts from a high-level perspective to a low-level perspective has to be well thought through and it is desirable to follow some form of a structured dissemination, based on some pre-established categorization. For example, the clinical adjudication policies are administered through a set of well established *treatment categories* within BRIDE [7]. This helps in managing combinatorial explosion of rule artifacts through a controlled reuse of pre-existing configurations of sets of rules. In the absence of such a categorized structure, there would be significant redundancy and duplication in structure as each hierarchical structure would end up duplicating, sub-structural elements of it's hierarchy. However, in spite of this form of structured organization duplication and redundancy can be introduced due to the lack of *discovery*, as discussed in 2.4.

Lastly, to achieve these hierarchical structures of granulated knowledge artifacts with reuse, the underlying domain structures share a many to many correlation amongst each other, which are elaborated in more length in [7].

(b) *Ownership* This is an important aspect of *Knowledge Governance*, that can influence the level of reusability discussed in 2a. With a growing consciousness and need for privacy, customers and clients are increasingly unwilling to share common data. Within, the consolidating healthcare payor space, this presents some interesting consequences.

Within the dental insurance space, the collective knowledge i.e. policies and rules

gained to preempt Fraud, Waste and Abuse, can be a shared asset, as these knowledge assets are gained through collective experiences. But to effectively leverage the shared operational knowledge, the historical claims *data*, across customer boundaries are needed to effectively adjudicate them. But privacy considerations and constraints can explicitly govern and prevent us from leveraging cross-customer data and knowledge, across boundaries, thereby limiting the potential to act on opportunities, that can minimize claims utilization.

Another consequence of ownership is that there are strict rules, that can be driven by a contractual agreements that dictate, who *owns* the knowledge and who has the rights to view, use or change them, under specific conditions. This poses interesting challenges to sharing and pooling of knowledge assets

(c) *Correlation and Connectivity* To effectively organize knowledge, we need the ability to *correlate* and *connect*, pieces of knowledge assets. This introduces the important notion of *traceability* [12], which provides the users the ability to perform key activities such as root cause analysis, answer audit questions and explain processing and/or adjudication reasoning to customers and other stakeholders. Correlation can be achieved through the following three means or mechanisms.

i. *Structural* This is the strongest and most basic form of correlation and connectivity that can be built to correlate different knowledge assets, explicitly correlated through well defined data and object models relationships. Within the BRIDE eco-system [7], different knowledge assets such as Rule Stores, Rule Packages, Rule Packets, Rules etc. are correlated in this fashion, which lends to easy organization.

ii. *Manual or Explicit* This form of correlation is weaker, but equally important. This allows the business users to explicitly tag or store metadata about various knowledge assets. These can be used to connect and organize knowledge assets based on the explicitly defined metadata.

iii. *Derivative or Implicit* This form of correlation builds on the above two forms of correlation. Here, we leverage smart algorithms to derive, infer, classify, tag, index, store, visualize and report additional correlations.

3. *Knowledge Artifact Naming:* As mentioned earlier, the detailed or granular knowledge assets i.e. operational business rules and policies can grow combinatorially. The business users usually would like to introduce some mnemonic naming conventions to better organize and catalog the knowledge base. Hence, introducing capabilities to automate naming of newly introduced knowledge assets that adhere to user introduced naming conventions make sense. Further smart automated algorithms that manage the naming and cataloging of the knowledge assets especially when knowledge assets split, morph or converge, make the continued governance of such assets, over time, error and conflict free. Here are some of the explicit advantages or automation opportunities that can be realized by automated enforced mnemonic naming conventions.

(a) Infer stereotypical characteristics of knowledge assets based on the mnemonic name

(b) Projection and generation of *new* knowledge content based on mnemonic name of close affinity.

(c) Proactive discovery of semantically like knowledge assets based on mnemonic name similarities

4. *Knowledge Lifecycle Management:* From a systemic standpoint knowledge assets need to honor constraints placed by systems. In addition, knowledge externalization and development processes need to synthesize knowledge asset development and lifecycle with systemic software development, especially for emergent areas of knowledge externalization and management. In doing so, at a very high level, knowledge assets goes through the following main lifecycle phases:

(a) *Creation:* This is where the knowledge asset is generated.

(b) *Change:* This is when the semantics or content of the asset morphs.

(c) *Culmination:* This is where the knowledge asset loses relevance and is retired or terminated.

Within an environment, certain knowledge assets, such as business rules and rulesets have their own miniature lifecycle, before they are utilized by the automated system. For example BRIDE [7], has the following lifecycle

(a) *Author:* The business rule is authored or edited within the BRIDE Guided Editor.

(b) *Translate:* The authored business rule is translated to a format that can be interpreted by an *Rule Engine* [8]. This action is implicit for the standard use cases within BRIDE i.e. the Business User does not need to explicitly translate the rule, but it is tied into the Authoring phase of the lifecycle.

(c) *Deploy to Sandbox* The overarching *Rule Package*, that contains the *Rule Packet*, that contains the *Rule*, is deployed to a safe sandbox, where the business users can test the rules.

(d) *Deploy to System* The overarching *Rule Package* is deployed to the System, where the operational business processes can consume the decisions, made by these rules.

It is a common practice for knowledge based systems in organizations to have different environments. Broadly the environments can be categorized into:

(a) *Production:* This is where the knowledge is applied to day to day operational business processes. New business rules and policies for established areas i.e. business decisioning services can still be developed here by the business users without any software or systemic development impacts. The BRIDE system supports this type of configuration, wherein the Business Rule authors enjoy such autonomy, wherein they can fully develop, test and deploy new operational knowledge assets *Rule Packages* [7], to the running production Claim Administration Platform.

(b) *Non-Production:* This is where the organization develops, tests and validates the effects of newly introduced knowledge on the overall system. For emergent areas of knowledge externalization, the process entails the development of a standardized set of *Business Vocabulary* [6], that are mapped to an underlying implementation model, that integrates well with the other systemic components. Non-production environments can vary in number and purpose depending upon the organization's software and automation capability development processes. For example, here is a list of plausible non-production environments:

i. *Production Dry-run and Training Environment:* Intended for the Business Users to run live simulations, where they can do large scale application of knowledge without impacting live production business processes.

ii. *User Acceptance Testing Environment:* Intended for Business Users to run full scale end to end software release testing regiments, applying emergent areas of knowledge externalization and applying existing large knowledge bases with large scale impact changes.

iii. *Integration Testing Environment* Intended for component level integration testing of new and existing knowledge bases with systemic components.

iv. *Development Environment* Intended for the development of business vocabulary and unit testing it's effects in a limited sense.

Finally, as knowledge assets take birth in different environments i.e. *Production* or *Non-Production*, there is a strong need to be able to migrate these assets from one environment to another in a safe and secure manner, without loss of integrity and fidelity. The BRIDE system [7], supports these capabilities in full, which enable the business knowledge workers to autonomously migrate and administer these knowledge assets.

5. *Knowledge Training:* It is crucial to have succession planning amongst knowledge workers, to avoid large scale dissipation of knowledge. With this capability the Knowledge Management System, uses intelligent foundations 1, 2, 4, 5, 8 and 10 described in 2.4 to absorb knowledge topics about the use and context of knowledge assets, to fuzzily discover, classify, correlate, disseminate and analyze language, learn and recommend actions to the knowledge worker in a proactive fashion.

3.3 Automation

As the externalized knowledge bases, grow, evolve and diversify within the organization, there is a strong need to *automate*, aspects of analysis, management and lifecycle management. We will now explore the intelligent capabilities needed to support this.

There are several tasks in the lifecycle that the knowledge workers have to do to externalize and create knowledge assets and related collateral. The biggest task is the process of authoring business rules and policies. Business Rules in particular are more challenging, for the following reasons:

1. The business logic captured by a single business rule could be complex and if not authored properly will not enforce the policies accurately or worse, introduce errors in claims processing and adjudication. Another consequence is they could introduce unforeseen loopholes that can be exploited by Fraud, Waste and Abuse scenarios.

2. There could be a high number of business rules due to combinatorially explosive sets of conditions. The knowledge workers have to expend significant analysis effort to analyze if there are gaps or conflicts in the logic and implementation. As a consequence, there could be a lot of duplicate, equivalent or subsumed business rules that get created over time, by knowledge workers, as there were no sophisticated discovery mechanisms present to discover them.

3. The business rules need to be carefully assembled into sets of rules, to form executable rulesets.

Many of the rules in a ruleset can have overlapping conditions and prioritization of rules within a ruleset matter. Between business rules in a ruleset there could be partial or complete overlaps of logic called subsumption. The knowledge workers need to make sure that these are minimized and restricted, only allowing some legitimate partial overlaps.An error in configuration of these rulesets causes errors in processing. The BRIDE system [7], provides a sophisticated system of configurations such as *Rule Packages* and *Rule Packets*. The Rule Packages are deployed as executable rulesets, that can be executed by a Rule Engine [8].

Here are some of the more fine-grained use cases within the *Automation* capability.

1. *Automating Conversion:* Consolidation and Digital Transformation are driving organizations to consolidate and transform their business operations. Within the healthcare payor space, healthcare insurance companies are seeking to move their operations to either:

   (a) Homegrown healthcare administration platforms of healthcare insurance companies.

   (b) Commercially available healthcare administration platforms.

   The process of mining, converting and automating the knowledge i.e. Business Rules and Policies, of an insurance company on to a new benefits administration platform, is a risky, time-consuming and expensive process. There is a high risk of failure for the business and if not implemented accurately, there are hidden and ongoing costs introduced due to sub-optimal claims utilization, which can run the company out of business. Traditional methods for mining the knowledge from legacy systems [9] are manual, subjective and error-prone. The conversion process is dependent on multiple documented knowledge sources and is heavily reliant on knowledge workers who have had prior experience with the systemized knowledge, within the legacy platform. The challenges posed by traditional manual methods are:

   (a) It is dependent on knowledge workers intimately familiar with knowledge sources in the legacy platform.

   (b) It is ad-hoc and situational to the conversion project or customer.

   (c) It is time-consuming, typically a couple of years or more.

   (d) It is labor intensive, i.e. needs focused effort from a dedicated team of skilled knowledge workers.

   (e) It is not scalable to handle conversions of large magnitude or multiple parallel conversion efforts.

Partially automated approaches to *mining* information from legacy platforms entail, writing customized conversion programs/scripts that perform extraction and transformation of the knowledge to a mapping schema or file. The knowledge workers can then, manually analyze the data in the mapping data source and manually author the knowledge assets, into the target platform. Unfortunately, this again poses the same challenges except for challenge 1a, which is automated through the conversion program.

This is where more comprehensive and intelligent automation capabilities help solve the problem. As a case-study, using intelligent automation capabilities, the conversion effort was dramatically reduced from two or more years to under six months. The case study leveraged building automation capabilities that relied on an ensemble of *intelligent foundations* viz., 1, 2, 3, 8, 5 and 9, from 2.4.

Figure 10, depicts the steps to automate the mining and conversion of knowledge.

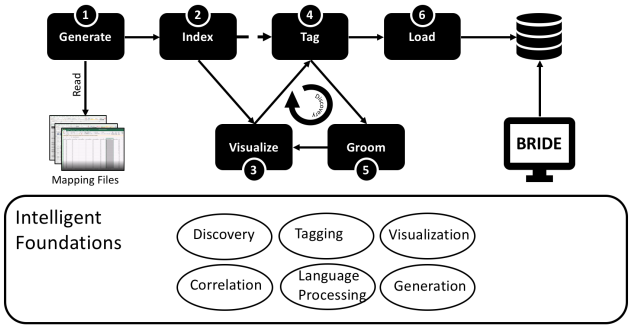


Figure 10: Automated Conversion Process

The conversion automation process is described through a sequence of six steps described below:

(a) *Generate:* This is the first step, that analyzes the conversion data in the Mapping File and generates business rules, organizing them into rulesets, which for BRIDE [7] entail, organizing them into *Rule Packages* made up of reusable *Rule Packets*, containing the *generated* business rules. The generation algorithm follows the structural organizational rules mandated by BRIDE to stereotype the packets and rules with the appropriate *Treatment Category*. The generated artifacts are assigned a generated systemic naming convention in this step. The complex normalized data in the mapping files are put through a sequence of file transformations.

(b) *Index:* The intermediate data that is generated by the above step is then loaded in an ElasticSearch database, which is a search engine based on the Lucene library, to enable

them for easy and fuzzy discovery. This enables the intelligent foundation of *Discovery*, for the steps detailed below.

(c) *Visualize:* Once loaded, custom data drilldown visualization programs aid the knowledge workers to analyze and proof the generated knowledge assets. In this step we use intelligent foundations of Visualization, Discovery and Correlation..

(d) *Tag:* In this step, specialized algorithms analyze the generated business rules and elicit important parametric data to tag the knowledge asset. The two fundamental intelligent foundations used here are Tagging and Language Processing.

(e) *Groom:* This is the phase of feedback for the algorithm to learn from knowledge workers, more specialized scenarios, to enable more accurate auto generation of knowledge assets. In addition, this is a process that is well integrated into the overall conversion capability to allow the knowledge workers to work collaboratively with the automated conversion process.

(f) *Load:* This is the phase, where the generated rules are loaded into the BRIDE knowledge platform, into the *authored representation* [8]. Techniques such as rule parametrization (i.e. the variant information from rules are externalized as rule parameters) and mnemonic rule template selection are used to select the appropriate rule template and the parameters are substituted using macro substitution. Alternatively, more sophisticated intelligent builder approaches can be employed, wherein the authored representation for the rule can be generated based on structure of the generated rule, which can be gleaned by a command pattern generated during the *Generate* phase. Finally, the knowledge workers can resume normal knowledge management activity with the generated rules within BRIDE.

Figure 11, depicts the anatomy of an index for a Rule. This is created during the *Index* process, from the outputs generated from the *Generate* step. In addition to the generated rule, the additional metadata generated during the *Tag* phase sets the stage for the subsequent grooming and accurate and automated loading into BRIDE knowledge base.

Figure 12 depicts the anatomy of an index for a Rule Packet. The basic structural data for the Rule Packet is created during the the *Generate* phase. The intelligent conversion algorithm blends in manually authored rules by knowledge workers with the generated rules, to construct a fully functional subset of rules that constitute a Rule Packet for a given Treatment Category. Further during the *Tag* phase, there is an abstraction phase of the algorithm that filters only relevant or important tags from rules that are emblematic of characterization for the Rule Packet. Lastly, the algorithm generates a descriptive name for the Rule Packet that facilitates easy discovery for knowledge workers for higher level activities such as contract implementations.

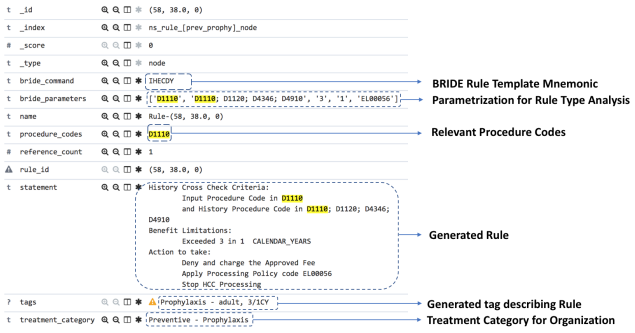


Figure 11: Anatomy of a Rule Index

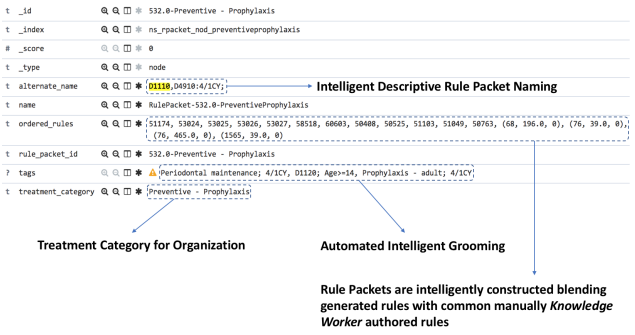


Figure 12: Anatomy of a Rule Packet Index

2. *Automating Testing* With rapid growth of knowledge assets within the platform, especially in the form of operational business rules, there is a compelling necessity for having an intelligent automation technique that can analyze rules within a ruleset and generate appropriate test scenarios. For example within the BRIDE [7] framework, there are over 17,000 test cases in just one area of adjudication rules. These are test cases that have been created by knowledge workers during their testing phase of knowledge management. A significant chunk of these test cases were generated from live claim processing data scenarios. However, with a strong and steady growth of operational business rules, more sophisticated techniques are needed to conquer the scale problem. Here are some use cases that explain this capability further:

(a) An automated test scenario creation algorithm intelligently interprets semantic tags about the rule parameters collectively con-

tained within a ruleset to generate test data and scenarios.

(b) As new rules are authored and included into a ruleset, an algorithm amends or augments test scenarios to test out new semantic consequences of the change.

(c) As knowledge workers decide to edit rules, an algorithm that can analyze the impacts from a test scenario perspective interactively communicates consequences of the change to the knowledge worker. This is a capability that can grow in the degree of sophistication.

3. *Automating Migration:* As we discussed in 4, it is typical for knowledge workers to manage knowledge assets across multiple *production* and *non-production* environments. Within BRIDE, select knowledge workers can elevate their role to an administrator, to perform critical administrative functions of knowledge asset management. *Migration* of knowledge assets across environments is one such function, that needs to be done with care.

   Here are a few challenges with administering this function, where the administrator is performing these tasks manually.

   (a) The number of physical environments can grow within an organization, over time. With consolidation the number of knowledge tenants in a multi-tenant environment, further impact the number of environments to manage.

   (b) The number of knowledge assets to migrate across environments grow. This puts an increased burden on the knowledge administrator to keep knowledge assets consistent across environments.

   (c) Not all environments, especially not all tenants may share the same assets i.e. business rules, policies, business vocabulary, test scenarios etc.. This introduces increased complexity for the migration path of these assets.

   (d) Maintenance in target environments can impede or impact the migration of knowledge artifacts. This problem is very similar to supply chain inventory and inventory movement, within Supply Chain Management.

   (e) Associated management functions such as deployment need to occur in target environments, post migration. This puts an extra burden on the administrator to ensure that these activities occur in a timely fashion.

   Strong intelligent foundations can help with the automation of pieces of this important function. Here are a few of the scenarios, that can help intelligently automate this function and reduce the dependency and burden on the knowledge worker.

   (a) Tag and inventory knowledge assets with migration specific metadata that is human and more importantly system consumable. This can enable intelligent algorithms to proactively audit, report and govern knowledge assets, migrating to target environments. Here intelligent foundation 2 is primarily leveraged.

   (b) Restrict the migration of knowledge artifacts to target environments that are undergoing maintenance or is being reserved for some focused activity. This prevents undesirable collisions and interference with activities in the target environment. Here the intelligent foundation 6 is leveraged.

   (c) Automatically include related artifacts impacted by the inclusion of a knowledge asset within a migration manifest. Knowledge Assets are intricately related, sometimes explicitly and structurally, but other times through correlation as explained in 2a. Hence movement of such knowledge assets should do the necessary impact analysis to ensure consistency and integrity of the knowledge assets. Here intelligent foundation 5 is primarily used. In addition, intelligent foundation 7 can be leveraged when the inclusion needs to be reported or recommended to the knowledge worker performing the administrative role.

   (d) The *Automated Migration Assistant* participates as a *virtual knowledge worker* in the target environment, as described in 1e. Here the intelligent foundation 6 is leveraged.

4. *Automating Mass Changes:* Sometimes it is necessary to make large scale or *mass* changes to knowledge assets. For instance, a large number of business rules need to be changed because a key business policy changed or the semantics of a key business term changed.

   Through traditional methods, this would mean knowledge workers need to track down every single relevant knowledge asset that are impacted by the change and modify them to reflect the macro-change. A classic example of this are new yearly versions of the *Common Dental Terminology*, released by the *American Dental Association*.

   Here by leveraging all the 10 intelligent foundations, intelligent capabilities that automate making mass changes can significantly reduce human labor, erroneous edits and significantly reduce cost.

4 BRIDE Case Study

4.1 Background

Delta Dental of Michigan, Indiana and Ohio is the leading provider of dental benefits within the midwestern part of United States. The company has been in business for over 60 years and has witnessed a dramatic growth within the last decade. The reason for this dramatic growth is a highly automated and sophisticated benefits administration software platform, powered by BRIDE [7], a powerful homegrown, elegant Enterprise Decision Management System. BRIDE houses the company's growing pool of complex, externalized operational business rules which allows the company to automate the vast majority of it's claims adjudication process allowing it to scale it's business footprint to unprecedented levels, while keeping operation costs at an optimal level. The current metric for measuring this automation level is called *"Drop to pay"* or *"Auto Adjudication Rate"*, which is currently hovering at around 96%, which is an industry leading score.

To keep up with the fast growth, it was necessary for the BRIDE platform, at inception to introduce industry leading capabilities for surfacing full life cycle rule management functions, such as authoring, testing, deployment, migration, such that non-technical knowledge workers and subject matter experts could entirely manage these with little to no involvement form the IT Department. Over the last decade the BRIDE platform has kept up with innovation, by expanding the capabilities and staying a step ahead of the business needs.

The BRIDE ecosystem organizes its knowledge artifacts in a structured, highly reusable and componentized fashion. A logical view of the domain model for BRIDE at a high level is depicted in refbrideDomainModelFig.

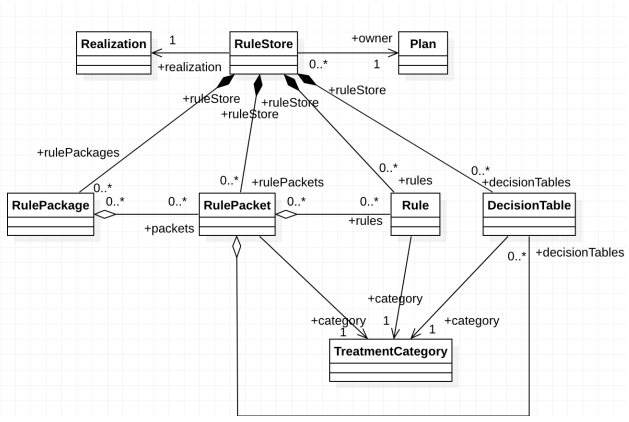


Figure 13: Highlevel BRIDE Domain Model

The core BRIDE constructs [7] are:

1. *Plan:* Represents a Dental Insurance Owner, that owns the rule artifacts.

2. *Rule Store:* A repository for storing rule artifacts, each rule store holds the reusable rule artifacts that can be reused within it, in a structured and constrained manner.

3. *Realization:* Represents an area of business functionality that require specialized terminology to externalize and articulate operational business rules using a structured vocabulary. There are over 30 areas of business functionality for the dental insurance domain so far, ranging from several detailed aspects of claims administration to billing, rating and actuarial applications. This is a growing body of knowledge as opportunities for automation and externalization are identified.

4. *Rule Package:* Represents an *executable* set of rules, that can be executed in a forward chaining, fact inferential rules engine [8]

5. *Treatment Category:* A mechanism to categorize finely granular rule artifacts. Within the clinical domains, the treatment category construct allows the business to categorize specialized clinical rules to literal clinical treatment categories as recommended by clinical standard bodies such as American Dental Association. For non-clinical domains, the treatment category construct is repurposed to category sub functional constituent components of processing.

6. *Rule Packet:* Represents a subset of rules for a given treatment category. Each Rule Packet is stereotyped for a treatment category and should only contain Rules of that same treatment category.

7. *Rule:* Represents a reusable authored production rule (i.e. a rule containing conditions followed by actions) authored in a "Domain Specific Language" using an english-like vocabulary. It is also the most

The next wave of innovation within the BRIDE platform introduces the application of *Artificial Intelligence* techniques to effectively address a new, growing breed of current and anticipated challenges. The work done on developing some of these capabilities has gone towards shaping and developing the concepts discussed in this paper.

The approach to building this new wave of innovation, has been to implement a strong foundational architecture i.e. *"Intelligent Foundations"*. We will discuss these foundational capabilities, their motivation, the architectural choices and application, subsequent sections.

4.2 Discovery

"Ability to fuzzily discover diverse artifacts" As the number of rule artifacts have grown over time, in spite of the ability to reuse rules a lot of unneeded redundancy and duplication have proliferated causing the knowledge base of operational rules and rulesets to

bloat. A major reason for this has been attributed to *discoverability*.

Within the original architecture, the BRIDE system provided SQL based searches against a relational database backdrop of the BRIDE schema. With this, the knowledge workers could search for specific rules with key terms and vocabularies and see their association with rule packets and rule packages. This was a bottom up search. They could also search for Rule Packages with specific Rule Packets in their configuration. With this they could search for full or partial rule package configurations, generate a report and then drill down. In spite of these rich search capabilities, discovery was not sufficient to prevent the proliferation of redundant rule assets overtime. Here are the main reasons why this alone is not sufficient:

1. *Layered Reuse Complexity:* BRIDE uses a layered reuse paradigm for constructing Rule Packages that eventually translate to executable rulesets. While this approach in and off itself is a sound from a component reuse vantage point, the major complexity is the inability of the traditional search functions to get a comprehensive and 360°view of the Rule Package.

2. *Naming Complexity:* With the componentized approach promoted by BRIDE, the knowledge workers set about building reusable Rule Packets, which in essence were combinations of rules, within treatment categories. In the interest of providing meaningful names that described Rule Packets, the knowledge workers resorted to lengthy descriptive names that sought to summarize the semantics of emblematic rules contained within the Rule Packet. However, this approach is highly manual, lacks predictable standardization and is not sufficient to describe huge and complex Rule Packets.

3. *Variational Complexity:* The characteristics of specific clinical domains require blending of specific clinical terminology that by the very nature can lead to a combinatorial explosion of sorts. With the nuances and subtleties in books of business and the semantic authoring flavors introduced by different knowledge workers there were subtle variations of sequences of rules, individual rules that were introduced over time. Some of these were made deliberately to support the variational needs of the requirements. Yet, others were made unknowingly introducing inconsistencies in approach. These were challenging to discover.

ElasticSearch, is a search engine built on the Lucene library. Nowadays, it is a common trend with modern IT architectures to leverage this tool to enable and enhance fuzzy discoverability. Data can be loaded or *indexed* into ElasticSearch as JSON (JavaScript Object Notation) documents. Subsequently this data can be discovered through a textual search of terms or fields contained in the JSON documents loaded and the tool will return all documents that match or nearly match textual patterns searched.

With complex natural language rules, this enables easy discoverability of these assets in a fuzzy fashion, that traditional relational queries do not allow. But this alone is not sufficient as the knowledge workers not only need to discover individual rules based on key business terms and vocabularies, but they also need to be able to discover the connections between rule artifacts, their levels of similarity and the characteristics of their similarities. With that there are five intelligent foundations that need to be applied in conjunction to fully realize the benefits of *discovery*. How these intelligent foundations collaborate is shown in 14. To understand these relationships:

1. *Language Processing* is used to disseminate the textual data in policies and rules, filtering key business terms and builds important metadata for *Correlation*.

2. *Classification* runs advanced algorithms, leveraging data created by *Correlation* to build more detailed correlation data.

3. *Tagging* is the default mechanism for storing metadata generated during *Correlation* and *Classification*

4. *Classification* and *Correlation* builds the foundation for *Discovery*.

5. *Visualization* helps makes sense out of *Discovery*, especially when there is a large amount of interdependent, connected data.

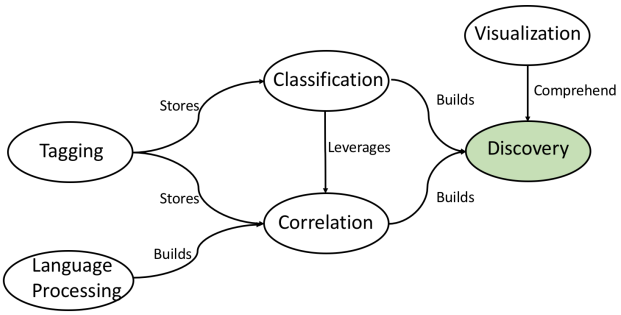


Figure 14: Collaborative Foundations for Discovery

4.3 Tagging

"Ability to tag or decorate knowledge artifacts with manual and generated metadata."

The basic implementational notion of a tag is to assign a key and description tuple to an artifact of interest. However sometimes, there can be more than one tag to decorate the artifact with. Therefore, it makes sense to introduce these key concepts:

1. *Tag Aggregation (Tagregation):* A Tag Aggregation is a composite structure of sub aggregations and Tags.

2. *Tag:* A Tag is simple key and description.

3. *Taggable Artifact (ITagArtifact)* A Taggable Artifact is any entity that needs tagging metadata.

The logical domain models for the basic tag constructs and taggable artifact constructs are shown in 15 and 16. The actual implementation of tags with the artifacts is more relevant when indexed into ElasticSearch for discoverability. A JSON representation of a Rule with a sample of extracted tags, used for discovery, is shown in 17. In this example, the tags for dental procedure codes used in the rule are, extracted to the tag "procedureCodes". Similarly, other useful correlational metadata could be extracted.

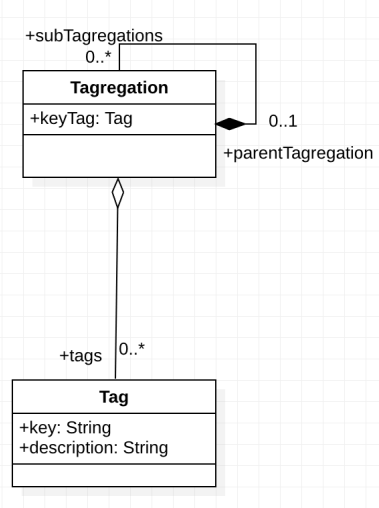


Figure 15: Tagging Domain

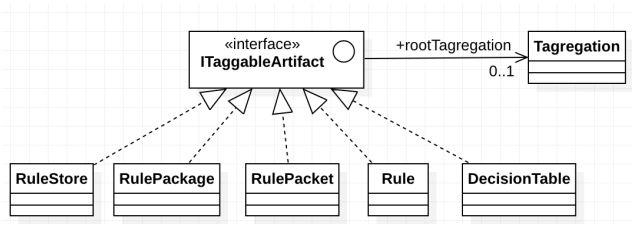


Figure 16: Taggable Artifact Domain

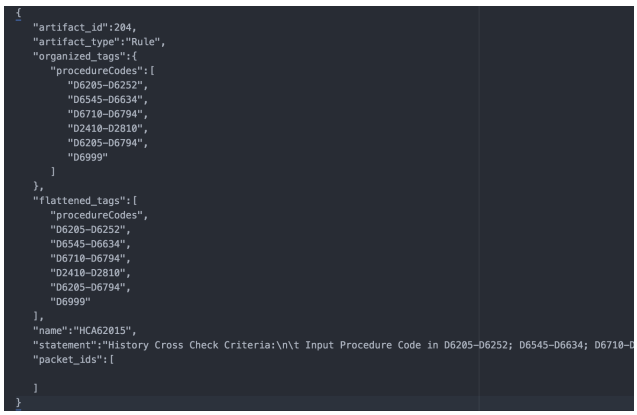


Figure 17: JSON of a Tagged Artifact

4.4 Visualization

"Ability to graphically visualize affinity between artifacts"

To enhance the discoverability of the knowledge assets data visualization is an effective method of analysis. Here a few different visualization techniques have been leveraged, with the correlation and classification data collected through tagging. Knowledge assets as a graph visualization with color coded edges to depict the different types of classifications, is depicted in 18. This is an effective way to discover common trends and patterns. The original workshop paper [1], outlined six different affinity classification measures, which are shown in this visual with different colors. This kind of visual is very useful for identifying common clusters of affinitive rules within an area and the nature of their affinity. There has been significant research and work on the efficacy of applying topological data analysis techniques, to use cases within healthcare [17]



Figure 18: Color Coded Clustered Affinity Visualization

An edge configuration for the same visual is shown in 19. This shows the nodes on the circumference and different colored edges criss-cross connecting the different nodes which represent the rule artifacts. The advantage of this visual is that useful, metadata can be shown on the edges, on hover that aid immensely during knowledge analysis.

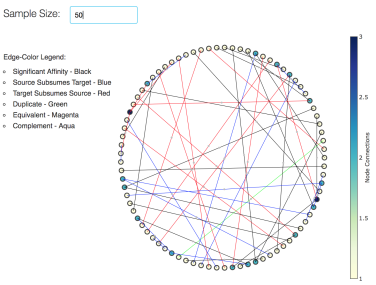


Figure 19: Color Coded Edge Affinity Visualization

A common themed rule asset, is picked out from a crowd of assets, in 20. This form of visualization is useful in identifying an emblematic asset for a certain area.

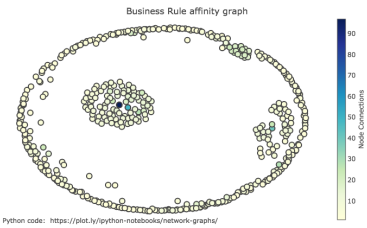


Figure 20: Clustering to identify baseline rule

A term cloud visualization is shown in 21. This is a very useful visualization for performing impact analysis based on key term correlation. In conjunction with correlation of key terms across rule and knowledge assets, it is possible to do end to end impact analysis.

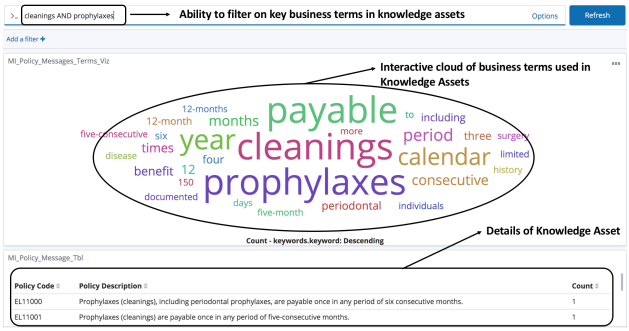


Figure 21: Interactive Term Cloud Visualization

A few observations when building data visualizations are:

1. It is useful to provide a way to filter based on category. This helps with narrowing sample sets and separation of concerns.

2. It is useful to provide a way to configure the size of a sample set. With certain types of analysis the knowledge workers are seeking understand common trends, to make informed decisions about management and governance.

3. It is useful to provide a mechanism for searching or filtering for combinations of terms. This aids in narrowing down the analysis. The visualization in 21, provides this mechanism.

4. It is useful to provide a detailed drill down capability on visualizations where possible. This allows the knowledge worker to identify the concrete knowledge assets with a single visual.

4.5 Classification

"Ability to classify artifacts based on characteristics"

Several sophisticated algorithms were developed to study the affinities between business rules and to classify them. Techniques to classify rule and knowledge assets ranged from naive simple bag-of-words to more sophisticated variations that combined bag-of-words with correlation to relevant master data assets. In the original workshop paper six different classifications were developed, comparing business rules, which were:

1. *Duplicate:* When two rules were exact duplicates with different names. These were top candidates for elimination and consolidation.

2. *Equivalent:* When two rules were semantically the same but had conditions that were in a different order. The classification algorithm here was shallow and only checked for conditions or actions in a different order. Deeper semantic analysis of rules will be able to identify equivalent rules beyond these narrow confines, creating the opportunity for better cleanup of these redundant rules.

3. *Complement:* These were specific to this area of rules where clinical procedures were complemented across the input and history procedures, to provide a complete adjudication coverage. Often times, claims are submitted out of order and these kind of rules analyze logical completeness. The scope for these can be expanded to cover other clinical criteria such as teeth, tooth surfaces etc. in the future. Further this pattern can be applied to other functional verticals to analyze logical completeness.

4. *Subsumption:* Here there were captured as two classifications i.e. Source subsumes Target and Target subsumes Source. Here again deeper semantic analysis can analyze more subtle aspects of subsumption, like analyzing inclusive procedure patterns etc. In the absence of these deeper analysis capabilities, come subsumptions could miss this classification.

5. *Significant Affinity:* Here the algorithm used a threshold of score similarity to identify rules that were closely affinitive with each other. However this was pretty effective and did capture rules that were significantly affinitive with each other.

4.6 Correlation

"Ability to build correlation between different artifacts"

This has been a centric intelligent foundation, for connecting different concepts together, to perform associative analysis and impact analysis. The implementation of correlation relied on algorithms, that analyzed, indexed and decorated knowledge assets with metadata in the form of tags and/or first class correlative attributes, in specific cases. Within the BRIDE case study we implemented correlation for three forms of correlation and connectivity as explained in 2c. Here are a few architectural aspects to consider, while integrating this intelligent foundation within the Knowledge based system.

1. Correlative algorithms need to run close to point of knowledge. There is an inherent risk of running correlation algorithms in a batch fashion

as morphing data conditions could either miss certain temporal correlation or miss transient changes that can be critical information to capture.

2. The initial correlations should be captured at well designed intermediate sink points and should be verbose enough to not miss seemingly inconsequential details, that might add value at a later point. Intermediate logging could be an option for such an architectural implementation. Generating and storing changes as events is another option.

3. The correlation algorithms should be designed in an idempotent fashion where possible, especially with the dimension of temporality. There are use cases that might require the regeneration and storage of correlation, especially during software upgrades and patching.

4. Even though, for the most part correlation of knowledge assets do not include PHI (Personal Health Information) or PII (Personal Identification Information), there are some corner cases where indirect correlation of operational data that have such concerns might crop up. One example of this is the reverse engineering capability of BRIDE for test cases from live data. In such cases de-identification and regular purge of such correlation need to occur to address security concerns.

4.7 Event Processing

"Ability to generate, process and react to system and user events, within the platform"

This intelligent foundation forms the basis of a very important infrastructural capability. As stated earlier, it can be leveraged for Collaborative Authoring. But it can also be leveraged for Correlation and Connectivity. Within the BRIDE case study, this is currently on the roadmap but has not yet been implemented. However, a lot of work is projected on the roadmap to innovate this intelligent foundation, especially for capturing fine grained operational execution metrics to aid correlation, for machine learning and machine training purposes.

4.8 Recommending

"Ability to recommend action to take based on correlation between artifacts"

Within the BRIDE case study this was implemented as an explorative proof of concept, wherein the system could recommend close rules based on the current authored content. While this could be foreseen as a useful feature in and off itself, this is barely scratching the surface. The BRIDE case study, hopes to leverage this intelligent foundation for deeper more sophisticated scenarios, which can be the subject of another paper.

4.9 Language Processing

" Ability to absorb, disseminate and score keywords fro textual content/documentation at detailed inflection points"

Within the BRIDE Case Study, this intelligent foundation has been used to disseminate policy descriptions, and procedure code descriptions, to correlate business rules and other artifacts such as Rule Packages, Rule Packets, Rules, Benefit Plans etc, with key term correlation. The Language Processing capabilities with Python and NLTK give a good starting point to filter out key business terminology, which can then be used to correlate. Again there is not more to report here from a case study perspective as we are in very early stages of fully realizing and leveraging this intelligent foundation.

4.10 Generation

"Ability to generate new content" Within the BRIDE Case Study, this intelligent foundation was used to automate conversion as discussed in 1. However, this is just one use case or application of this intelligent foundation. However, there are some other scenarios where Generation can help.

1. Perform mass changes to rules. As mentioned earlier the ADA, publishes a new version of the CDT every year. With the ability to generate new or change existing content, based on high level heuristics, it will be possible to automate the mass changes to Rules.

2. Generate new rule templates based on mnemonic naming patterns that are generated based on high level heuristics. Within the BRIDE Case Study, this was exercised in a limited purview, as new rules had to be generated based on the conversion data.

3. Automatically generate test cases based on rule content. Based on semantic analysis of knowledge and rule artifacts, intelligent generators can automate the act of building test cases on the fly and saving them for regression purposes.

4.11 Learning

"Ability to learn from data patterns and from human feedback"

This has not yet been implemented within BRIDE, but it is there in the roadmap. This is an area of future work and builds on the other intelligent foundations.

5 Automation challenges for Knowledge-intensive systems

In this section we will consider some of the challenges around automation, for knowledge intensive systems. In 3.3, automating the task of authoring was identified as a the biggest opportunity for automation. Within

the BRIDE case study, automating conversion was one of the biggest opportunities. Automating the process of knowledge acquisition, reduced the time effort from 2 years to a few months. However, there are some inherent challenges with automation of the rule mining and authoring process. They are:

1. *Standardization :* For an automated rule conversion process to work repeatedly, the data for that area of functionality needs to be standardized into a well thought out canonical format.

2. *Specificity:* It is not uncommon for the quality of knowledge mined out from legacy blackbox systems to contain a lot of noise and inconsistencies that are specific to the legacy system. Further the nature of the inconsistencies can vary across different target systems and conversion initiatives. The automation processes while being streamlined should allow for custom transformation rules that can be externalized and customized for each conversion effort.

3. *Semantics:* The interpretation of standard benefit clauses implemented in rules in the legacy source system could have different semantics and might need specialized handling or transformation into the target system's knowledge base.

4. *Inconsistencies:* Sometimes there are misinterpretations of documented high level policies and contracts in the source legacy system, that are being converted. Technically one point of view amongst knowledge workers is to rectify these during the conversion process. Yet, this can have dire behavioral regressions on the target system, post-conversion. These kind of situations, need to identified during the automated conversions, traced back to the source and have a good explanation facility. Whether this method *can* be formalized consistently to bolster the efficacy of such automation efforts, remains a tantalizing proposition.

5. *Universal applicability:* For the BRIDE case study, the automated mining and generation process is fairly well established for a specific area of adjudication rules, that had the most number of rules and artifacts, that could not have been authored manually. But, whether this automated process can be streamlined and applied universally to other areas of rule mining is the subject of future work. This could be an area of research in the future.

6 Future Work

The paradigm of *intelligent foundations* to knowledge based systems holds great promise and potential. Especially within healthcare these techniques are making some significant differences. However there are areas that need further research and elaboration. Current state of the art AI techniques involve an ensemble of methods. Intelligent Foundations, strives to formalize and document a set of reusable patterns, language and processes to effectively apply them on a large and universally diverse setting. This paper was born out of a case study to address seemingly unsolvable and hard to automate and humanly unscalable problems. Here are some high level areas for future work.

1. *Universal application of techniques:* One area of future work is to continue the rigorous formalization of the techniques described here for universal application to other diverse verticals.

2. *Identifying more intelligent foundations:* While this paper attempted to do a reasonable job in identifying key intelligent foundational capabilities and building a framework for intelligent knowledge base systems, there is a significant opportunity for more research and elaboration, to further refine and enhance this concept and framework. Future work should continue to investigate other intelligent foundations that can introduce other intelligent capabilities that make the reality of a *virtual knowledge worker* real.

3. *Exploration of intelligent capabilities:* While this paper, outlined a few of the capabilities that can be built atop these intelligent foundations, there are more to be discovered and documented. This will motivate AI researchers to apply techniques to effectively solve the one real problem i.e. make healthcare cost effective to the general populace.

4. *Automation of conversion and generation of knowledge content:* Lastly, this is a major challenge as current methods are fairly ad-hoc, human and expert intensive. While prevalent data science methods and machine learning techniques seek to solve *"well defined"* problems, the challenge of replacing or augmenting the loss of a human subject matter expert by *emergent* system capabilities is an alien concept. The automation of conversion, rule mining and generation holds great promise and should open up rich areas of research.

7 Conclusion

As stated earlier, this paper came out of a case study, but introduces concepts that hold great potential. The work motivating this paper is a project to modernize and improve the *state of the art* of the BRIDE platform. The case study for automation and generation of rules and knowledge artifacts into the BRIDE platform has been instrumental in practically implementing and proving out a majority of the *intelligent foundations* discussed in this paper. The primary goal has been to, introduce elements of Artificial Intelligence at a grass root level that can provide an ambitious shift in capabilities. *Intelligent Foundations for Knowledge*

Based Systems, provide a perfect foil for solving the giant problem of generational shift, which seems to be driving Digital Transformation.

Conflict of Interest The author declares no conflict of interest.

Acknowledgment The author would like to acknowledge the executive leadership at Delta Dental for sponsoring and fueling the innovation, and for their continual commitment and support for the BRIDE platform, which remains a source of inspiration and innovation to the organization.

References

[1] M. K. Agaram, "Intelligent Discovery Features for EDM and MDM Systems," 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW), Stockholm, 2018, pp. 135-144. doi: 10.1109/EDOCW.2018.00028

[2] A. Finkelstein, et. al. "Moral Hazard in Health Insurance", Columbia University Press, 2015.

[3] D.M.Cutler, R.J.Zeckhauser "Adverse Selection in HealthCare", 1997

[4] D.M.Cutler, S.J.Reber "Paying For Health Insurance: The Trade-off Between Competition And Adverse Selection"

[5] T. Hall, "'Monte Carlo Estimation of the Impact of Adverse Selection on Multiple Option Dental Plans", 2010.

[6] M. K. Agaram and C. Liu, "Vocabulary Model Requirements for Production Rule Systems," 2012 IEEE 16th International Enterprise Distributed Object Computing Conference Workshops, Beijing, 2012, pp. 132-139. doi: 10.1109/EDOCW.2012.26

[7] M. K. Agaram and B. Laird, "A Componentized Architecture for Externalized Business Rules," 2010 14th IEEE International Enterprise Distributed Object Computing Conference, Vitoria, 2010, pp. 175-183. doi: 10.1109/EDOC.2010.26

[8] M. K. Agaram and C. Liu, "An Engine-Independent Framework for Business Rules Development," 2011 IEEE 15th International Enterprise Distributed Object Computing Conference, Helsinki, 2011, pp. 75-84. doi: 10.1109/EDOC.2011.20

[9] M. K. Agaram "The Phased Approach to Mining Business Rules" Business Rules Journal Vol. 8, No. 5, (May 2007)" URL: http://www.brcommunity.com/a2007/b344.html

[10] L.F.Bertain, G.Sibbald "The Tribal Knowledge Paradigm", 2012

[11] Adam Starks PhD (2013) The Forthcoming Generational Workforce Transition and Rethinking Organizational Knowledge Transfer, Journal of Intergenerational Relationships, 11:3, 223-237, DOI: 10.1080/15350770.2013.810494

[12] M.K.Agaram "Traceability Aspects for Business Rules Management" Business Rules Journal Vol. 9, No. 1, (Jan. 2008)" URL: http://www.brcommunity.com/a2008/b384.html

[13] Cahill, Terrence F., EdD, FACHE; Sedrak, Mona, PhD, PA Leading a Multigenerational Workforce: Strategies for Attracting and Retaining Millennials, Frontiers of Health Services Management: October 2012 - Volume 29 - Issue 1 - p 315

[14] Foss et. al. "Knowledge Governance Processes and Perspectives", Oxford University Press, 2009.

[15] Attila Bruni, Silvia Gherardi, Laura Lucia Parolin Knowing in a System of Fragmented Knowledge, Mind, Culture, and Activity, 14:1-2, 83-102, 2007

[16] D.C.Luckham "The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems", Addison-Wesley Longman Publishing Co., 2001

[17] F.X.Campion, G. Carlsson "Machine Intelligence for Healthcare", 2017

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 75-87 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Development of Smart Technology for Complex Objects Prediction and Control on the Basis of a Distributed Control System and an Artificial Immune Systems Approach

Samigulina Galina[1], Samigulina Zarina[2,*]

[1]*Intellectual control systems and forecasting, Institute of information and computing technologies, 050010, Kazakhstan*

[2] *Faculty of Information Technologies, Kazakh British Technical University, 050000, Kazakhstan*

A R T I C L E I N F O

Article history:
Received: 07 February 2019
Accepted: 03 May, 2019
Online: 21 May, 2019

Keywords :
Intellectual control systems
Complex object
Artificial Immune System
Distributed control system
Multi-agent Smart system

A B S T R A C T

This paper is an extension of work originally presented in 2018 Global Smart Industry Conference (GloSIC). Researches are devoted to the development of Smart technology for complex objects control and prediction on the basis of a distributed Honeywell DCS control system of the TengizChevroil enterprise using the example of a technological process of medium pressure gas cleaning. The article describes how on the basis of the multi-algorithm approach there was developed a modified algorithm based on modern artificial intelligence methods in order to select informative features (principal component method, Random Forest algorithm, particle swarm algorithm) and artificial immune systems (clonal selection) solving the image recognition problem and predicting the state of a complex control object. There was conducted a comparative analysis of the simulation results using the example of real production data (daily data of sensors from an average pressure absorber).

1. Introduction

Nowadays, the efficiency of industrial enterprises depends on many factors: material and technical support (installation of modern production equipment, modernization and reconstruction of the existing technical base), economic factors (growth of production and raw materials costs), management factors (information support for decision-making), and etc [1]. Therefore, modern researches in the field of development of automated complex objects control systems are aimed at increasing the reliability and at reducing the influence of various negative factors on the production process.

The implementation of modern Smart technology enables the possibility to improve production by effective prediction of the technical state of the equipment and the timely resolution of problems at the stage of their occurrence in real time. A characteristic feature of the complex objects control is the need to process a huge array of production data, as well as prediction and decision-making under the condition of parameters uncertainty.

In order to solve such problems at the complex objects control process, there were proven modern methods of artificial intelligence with the following advantages: the ability to process multidimensional data, self-organization and relative ease of implementation.

Widespread use has control systems based on neural networks (NN). For example, in [2] there is presented the use of a decentralized robot control system based on a neural network. There were proposed four decentralized control systems which can identify the dynamics of the robot's movement. Training of each neural network is performed online using the advanced Kalman filter. The researches [3] are devoted to the development of a combined adaptive neural network and to the nonlinear prediction model for production processes control.

The effectiveness of the proposed method was tested on a system of reactors with continuous mixing. In work [4] there is presented a PID controller based on a neural network for the pH neutralization process control system. The classical PID control cannot adapt to changes in the process, and therefore a PID controller with a signal neuron was developed. The simulation results showed the reliability and adaptability of the proposed control system.

In addition to neural networks there are used swarm intelligence (SI) algorithms for complex objects control. For example, in work [5] there is presented the application of the

*Corresponding Author Zarina Samigulina, e-mail: zarinasamigulina@mail.ru
www.astesj.com
https://dx.doi.org/10.25046/aj040312

particle swarm algorithm (PSO) for the optimal choice of process control parameters. In researches [6] there was used an ant colony optimization at setting the PID controller for a conical tank. The proposed approach has shown better results compared to setting of a closed loop based on the Ziegler-Nichols method and to setting of a controller in a closed loop system. In work [7] there is presented the development of a control system for autonomous mobile robots based on a bee colony algorithm and a genetic algorithm. Researches [8] are devoted to applying the method of bacteria movement optimization at the development of a PID controller by a two-wheeled robot with a two-way movement mechanism for a working environment with limited space.

Also there is widely used an approach based on fuzzy logic for complex objects control. Another study a control system for mobile robots with nonholonomic wheels moving on a plane on the basis of a fuzzy controller in order to achieve a given trajectory of motion. The problem of rounding obstacles is considered. Parameters of PID controller are calculated on the basis of fuzzy logic. There are presented simulation results based on a real vehicle. The researches are devoted to cascade control based on a fuzzy PID controller for a conveyor for weighing materials. The article presents a non-linear model of a conveyor system of continuous weighing, taking into account the delay time of material transportation. There has been developed a PID controller based on fuzzy logic with two control loops: an internal speed control loop and an external mass flow control loop. The results of simulation and experiment show that the proposed algorithm for fuzzy PID control can improve the quality of material mass flow control in an industrial conveyor system of continuous weighing using a programmable logic controller.

In work [9] there is proposed an innovative approach for fuzzy control systems setting with low parametric sensitivity based on the Gray Wolfs Optimization Algorithm (GWO). The servo system is controlled on the basis of the Takagi-Sugeno-Kang fuzzy PI controller (TSK PI-FC). The GWO algorithm is used in solving the optimization problem. Article [10] is devoted to an optimized fuzzy-genetic algorithm (F-GA) for controlling the process of production of aluminum foam with open pores. The application of the F-GA method allows to take into account the uncertainty associated with the model and the statistical variability of the production process within one model.

Nowadays, the researches in the field of the use of the bio-inspired approach of artificial immune systems (AIS) for complex objects control are also promising. In work [11] there is presented AIS application at PID controller parameters setting for nonlinear control systems. Researches [12] are devoted to the development of a complex object control algorithm based on an adaptive artificial immune system. The simulation was carried out on the example of a reactor with a mixing function. The article [13] provides an AIS application for control a group of robots. The work [14] is devoted to the use of a meta-heuristic AIS for signal evaluation in identifying the contour of the angle control of an unmanned aerial vehicle. In the article [15] there is used the AIS algorithm for preliminary signal processing at detection of failures for autonomous underwater vehicles. Article [16] is devoted to applying the AIS approach (negative selection algorithm) in order to control the combustion process in the chamber with the implementation on the microcontroller. In researches [17] there was developed an immune reconfigurable controller based on four modules for the immune system modeling: the observation process, response, mechanism of memory and self-training. The proposed algorithm is able to detect drive failures and to provide failure-tolerant control for the systems with multiple inputs and outputs.

In addition to bioinspired methods, machine learning algorithms for complex objects control and solving the problem of identifying informative features have become widespread. The most promising is the Random Forest (RF) algorithm. For example, in [18], RF is used to detect defective plates by chemical characteristics in the production of semiconductors. The article [19] presents the use of algorithms for machine learning of RF, decision trees, etc. in heavy industry for analysis of large amounts of production data. The work [20] is devoted to the use of the Random Forest algorithm in sensor failures diagnostics. The algorithm is used to reduce memory requirements in the process of assessing the state of the equipment in real time.

Therefore, the analysis of the literature proves the relevance of the development of modern Smart technologies and prediction of complex objects control based on industrial enterprises using artificial intelligence methods.

The following structure of the article is proposed. The second section contains the formulation of the research problem. The third section presents the development of Smart technology for complex objects control and prediction based on the Honeywell Experion Process Knowledge System (PKS) distributed control system (DCS) and the approach of artificial immune systems; reviewed the operation of the DCS at the Tengizchevroil enterprise, and presented the general structure of the Smart technology. The fourth section is devoted to the development of a modified algorithm of artificial immune systems. The fifth section describes the multi-agent Smart system for complex object prediction and control under industrial operation conditions. The sixth section contains the simulation results. Description of a real complex medium-pressure absorber, its principle of operation, readings from sensors, etc. The simulation results are presented in the form of graphs, as well as a comparative analysis of the effectiveness of the proposed algorithms. The conclusion describes the main findings and a list of references.

2. Problem statement

Since the complex objects control with the presence of parameters uncertainty in the conditions of industrial operation is a time-consuming task due to the need to process multidimensional data in real time, the use of promising heuristic algorithms of artificial intelligence is important.

The research problem statement is formulated as follows: it is necessary to develop a modern Smart technology for complex industrial objects control based on a multi-algorithm approach. Within the framework of the multi-algorithm approach to synthesize a modified algorithm of artificial immune systems in order to solve the problem of prediction and control. In order to organize an autonomous, flexible and decentralized control system based on the proposed technology, it is necessary to develop a multi-agent Smart system for implementation in real production based on TengizChevroil enterprise, with the aim of working with Honeywell PKS distributed control systems.

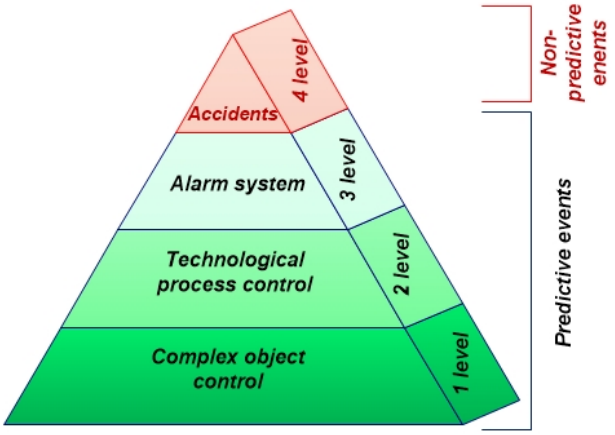


Figure 1: Honeywell DCS process control system with independent protection levels

3. Development of Smart technology for complex objects prediction and control on the basis of a distributed control system and an artificial immune systems approach

3.1. Distributed control system at TengizChevroil enterprise

TengizChevroil is the largest company engaged in the development and production of oil in the territory of the Republic of Kazakhstan together with the American company Chevron. In TengizChevroil there is used modern industrial equipment from Honeywell, in particular, Honeywell PKS distributed control systems (DCS). Nowadays, the Experion PKS system is the best in the field of distributed control systems and received an award from the magazine "Control Engineering". The company created a consortium for emergency situations control, aimed at exploring the causes of emergency situations and developing solutions aimed at reducing them. Therefore, the researches on improving the quality of equipment at the enterprise are relevant.

The Honeywell PKS distributed control system contains several independent levels of protection (Figure 1), designed to work with predictable and unpredictable events [23].

Predicted events include the following levels of independent protection:

1 level. Maintains "normal" operating conditions of the process. The distributed control system is engaged in the collection and processing of technological parameters, as well as their adjustment to ensure a given mode.

2 level. This level requires decision-making from the operator if the alarm is triggered and the first level is unable to maintain normal operating conditions of the control object.

3 level. This level indicates that the control object has left the "normal conditions" of operation mode and a controlled stop is required. The Safety Instrumented Systems (SIS) is activated. The SIS security system also includes: Emergency Shutdown System (ESD), Fire & Gas System (F&G), Safety Shutdown System (SSD) и т.д.

4 level. The level refers to unpredictable events, such as the occurrence of fires, gas leaks and other major accidents. A distributed system cannot predict the occurrence of such accidents, but it is designed for quick response and minimization of damage.

Information on the operation of the technological process is stored in the historical module (History Module, HM) of Honeywell DCS [23]. Developed Smart technology is aimed at working with the predicted events of the Honeywell DCS, with obtaining data for processing using the HM.

3.2. Smart technology based on Honeywell PKS distributed control system

Let consider the general structure of the proposed Smart-technology for complex objects prediction and control. The Honeywell PKS distributed control system consists of nodes described by controllers, operator consoles, control devices, data storage equipment, etc. The information transfer between nodes occurs through a local and universal control network. Information is collected in a historical module that is connected to a local control network, and a process history database is formed here.

The developed Smart-technology is designed to work with the predicted events of Honeywell PKS (Figure 1). The process history database containing current information on the behavior of a complex control object (readings from sensors, monitoring systems, etc.) is processed on the basis of a multi-algorithm approach (Figure 2).

Then a database of optimal parameters is formed for prediction and control on the basis of AIS algorithm, clonal selection. Depending on the nature of the processed data in the modified algorithm based on the AIS method, the most effective algorithm is selected that shows the best predictive result.

Further in the paper, as an example, we will consider the complex control object of the Tengizchevroil enterprise: a fragment of the Installation 300 for cleaning gases from acidic components [21].

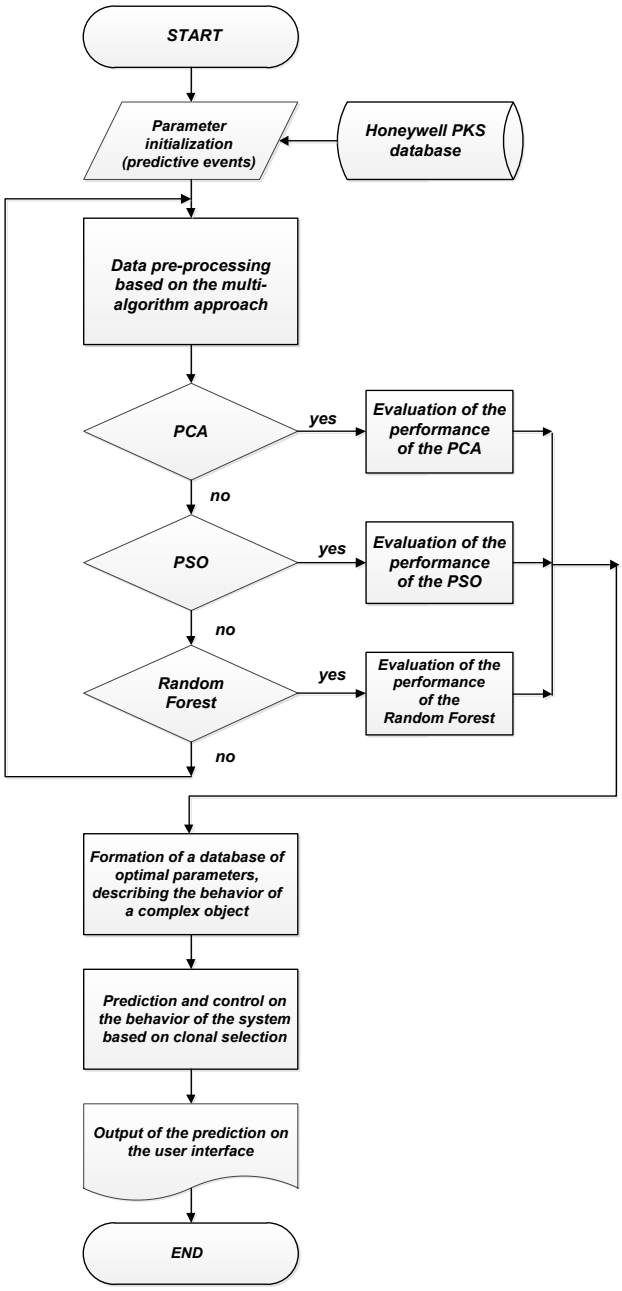


Figure 2: The structure of the developed Smart technology for complex objects prediction and control based on Honeywell PKS

4. Modified artificial immune algorithm based on clonal selection

Nowadays, the approach of artificial immune systems is a promising area of research. This method has the following advantages: adaptability, memory, dynamic arrangement of elements, ability to self-organization, etc. There are several main areas of AIS research: negative selection algorithms (NSA), immune system algorithms (INA), danger theory algorithms (DTA), and clonal selection algorithms (CSA) [22].

Let consider in more detail the clonal selection algorithm [23-25]. When a foreign antigen enters the body, special cells B-lymphocytes produce antibodies A_b. Each cell produces only one type of antibody that is relatively specific for the antigen. At binding antibodies (receptors), as well as using T-helper cells, the antigen stimulates B cells to multiply and to mature into plasma cells. Different cell divisions produce clones, by which there is meant a set of cells that are descendants from a single cell. Plasma cells most actively produce antibodies B-lymphocytes that divide rapidly. Antibodies A_b divide much more slowly. Lymphocytes can also differentiate in addition to plasma cells into long-lived memory cells - B - memory cell. Memory cells circulate through the body through the blood, lymph and tissue. In a secondary immune response, they can turn into large lymphocytes capable of producing antibodies with high affinity. These cells are pre-selected for the specific antigen that stimulated the primary response. The clonal selection algorithm is based on the following operations: selection, cell cloning and mutation. The selection procedure is carried out for cells with higher affinity and subsequent cloning, while other cells undergo a mutation process, which helps to improve their quality.

The clonal selection algorithm is presented as follows [23]:

Algorithm 1

Step 1. Generation of a set of possible solutions P, composed of a subset of memory cells M, added to the remaining population P_r.

$$P = P_r \cdot M \tag{1}$$

Step 2. Determination of the best individuals n, the population P_n based on the value of affinity (affinity).

Step 3. Cloning of n best individuals of a population, with the generation of a temporary population of clones C.

Step 4. Representation of a clone population to a hypermutation procedure, where hypermutation is proportional to the affinity of the antibody and antigen. Creation of a population of mature antibodies C^*.

Step 5. Repeated selection of improved individuals from the population C^* in order to form a set of memory M. Some members of the population P may be replaced by other improved population individuals C^*.

Step 6. Replacement of antibodies d by the new ones. Cells with lower affinity have a higher probability of replacement.

Modern methods of artificial intelligence are well established for the development of complex objects control systems, but unfortunately these approaches are not universal, and need to be customized for specific cases, taking into account the features of the object and the system parameters. The quality of forecasting and management directly depends on the processed data. The higher the quality of the data is, the more accurate is the forecast. Therefore, the application of the multi-algorithm approach is actual, within the framework of which the procedure of preliminary production data processing is carried out on the basis of modern statistical and intellectual algorithms. As a result of

research several methods are considered. Algorithms that will give the best predictive result are used for further research. Preliminary production data processing within the framework of the proposed Smart technology is carried out on the basis of the following approaches:

- Statistical approach - factor analysis and principal component method;

Statistical approaches are often used to process production data. In order to solve the problem of reduction of non-informative parameters during the study of complex industrial facilities, factor analysis (principal component method) has proven itself well. Among the advantages of factor analysis we can note the ability to analyze hidden (latent) relationships between parameters, as well as the relative simplicity of implementation [26]. As a disadvantage there should be noted the difficulties in the interpretation of the obtained results.

The algorithm of factor analysis consists of the following steps. The entire data set must be rotated counterclockwise, with the first axis associated with the maximum dispersion, and each subsequent one with the others. According to the results of the data rotation, the parameters that lie closer to the origin of coordinates are subject to reduction, as uninformative.

- Swarm intelligence - particle swarm optimization algorithm;

Let consider the concept of swarm intelligence (SI) [27]. The term SI was proposed by Beni and Wang in application to cellular robots and implied the collective behavior of a decentralized self-organizing system. Then it was used in the theory of artificial intelligence and describe the social behavior of living organisms (particle swarm method, ant algorithm, bee algorithm, wolf pack algorithm, bat algorithm, etc.).

In the proposed Smart technology, in the framework of the multi-algorithm approach, a particle swarm optimization algorithm is considered in detail.

A feature of the PSO method is that during each iteration, the solution is evaluated using the objective function to determine its suitability [28]. Each solution is represented as a particle in the search space. Particles move in the search space in order to find the maximum value of the objective function. As the particles move, each particle has a position in the search space, speed, and an individual best position. The PSO algorithm consists of the following steps [26, 27]:

Algorithm 2 Particle Swarm Optimization

Step1 . Initialization of a particles swarm.

Step 2. Evaluation of the suitability of each particle.

Step 3. Calculation of historically optimal positions of individuals.

Step 4. Calculation of historically optimal positions of the swarm.

Step 5. Speed and position update of particles according to the equation of speed and position. Repeat of this iteration until the stop condition is reached.

The speed of each particle is calculated by the formula:

$$\upsilon_i(t+1) = \varpi\upsilon_i(t) + c_1 \cdot r_1 \cdot (\hat{x}_i(t) - x_i(t)) + \\ + c_2 \cdot r_2 \cdot (g(t) - x_i(t)) \quad (2)$$

where i - particle index; ϖ - inertia coefficient $0.8 \le \varpi \le 1.2$; c_1, c_2 - acceleration coefficients ($0 \le c_1, c_2 \le 2$); r_1, r_2 - random values lying in the range $0 \le r_1, r_2 \le 1$; $\upsilon_1(t)$ - particle speed in time t; $x_1(t)$ – particle position in time t; $g(t)$ – the best swarm solution in time t.

The change in the position of each particle of the swarm is calculated by the following formula:

$$x_i(t+1) = x_i(t) + \upsilon_i(t+1) \quad (3)$$

The advantages of the PSO algorithm include: a simple implementation; small amount of adjustable parameters; parallel computations are possible; with a certain setting has the ability to fast convergence; no matches and mutations; has small requirements for computer memory and fast calculation speed; can be used in case when the representation of a mathematical model is impossible.

- Random Forest (RF) algorithm, etc.

Nowadays, the RF algorithm is a promising method of machine learning, capable of solving the problems of image recognition, prediction and selection of informative features [28]. A random forest is created on the basis of decision trees and has the same set of hyperparameters, while in RF the process of searching for the root node and the separation of object nodes is performed randomly [29, 30]. The more trees in the forest, the more accurate the simulation result are.

The advantages of the RF algorithm are following: there is no risk of retraining of the model; the algorithm has flexibility and high accuracy; no need to scale data; individual decision trees can be trained in parallel; data processing ability with a large amount of attributes, etc.

The operation of the Random Forest algorithm is presented below:

Algorithm 3 Random Forest

Step 1. Arbitrary selection of f signs from the total number of F signs.

Step 2. Calculation of a *node* among f features using the best separation point.

Step 3. Separate a *node* on child nodes using the best partition.

Step 4. Repeat steps 1-3 until reaching the value of the number one.

Step 5. Creation of a forest by n repeating of the steps 1-4 in order to create the n number of trees.

Based on the previously discussed methods, the modified AIS algorithm based on clonal selection can be described as follows (Figure 2).

Algorithm 4 Modified artificial immune algorithm

1. Connection to the historical module of the Honeywell PKS distributed system.
2. Data collection (readings from sensors) describing the behavior of a complex control object.
3. Formation of a database of system parameters.
4. Pre-processing of the obtained database, rationing, centering and filling in the missing data.
5. Formation of the optimal set of parameters due to the reduction of non-informative data based on the multi-algorithm approach.
5.1 Reduction of non-informative parameters based on the statistical approach, the principal component method.
5.1.1 Evaluation of the performance of the PCA algorithm.
5.1.2 Formation of a database of optimal parameters DB_PCA, describing the behavior of a complex object.
5.2 Reduction of non-informative parameters based on swarm intelligence, particle swarm algorithm.
5.2.1 Evaluation of the performance of the PSO algorithm.
5.2.2 Formation of a database of optimal parameters DB_PSO, describing the behavior of a complex object.
5.3 Reduction of non-informative parameters based on the Random Forest algorithm.
5.3.1 Evaluation of the performance of the Random Forest algorithm.
5.3.2 Formation of a database of optimal parameters DB_RF, describing the behavior of a complex object.
6. The solution of the problem of image recognition on the basis of clonal selection. Processing the received databases (DB_PCA, DB_PSO, DB_RF) and prediction.
6.1 Selection of an efficient algorithm that shows the best prognostic result.
6.2 Decision-making on the complex object control.

The proposed technology in industrial operation is most conveniently implemented on the basis of a multi-agent approach. Further there is presented a developed multi-agent Smart system for complex objects prediction and control.

5. Multi-agent object control Smart-system

Nowadays, multi-agent systems (MAS) are widely used in the field of industrial automation and artificial intelligence. A multi-agent system consists of independent agents with autonomy, decentralization, and flexibility. At developing complex industrial systems using MAS, the process of modeling and control is greatly facilitated. The paper [31] presents the MAS technology for organization of a flexible production process of customized products, through cooperation between companies, for the integration of heterogeneous production systems. Article [32] presents the use of a multi-agent artificial immune system for detecting and classifying network intrusions. Researches [33] are devoted to a decentralized multi-agent control system based on artificial immune systems. This system consists of the following target agents: the agent of self-organization, the agent of interaction, the agent of following specific goals and has the properties of flexibility, the ability of agents to work autonomously, the ability to optimally distribute the system resource, multifunctionality and self-organization.

Based on the foregoing, the development of a multi-agent Smart system for complex objects control at TengizShevroil for the integration of the developed modified artificial intelligence algorithms into production is relevant. Figure 3 shows the structure of the multi-agent Smart system.

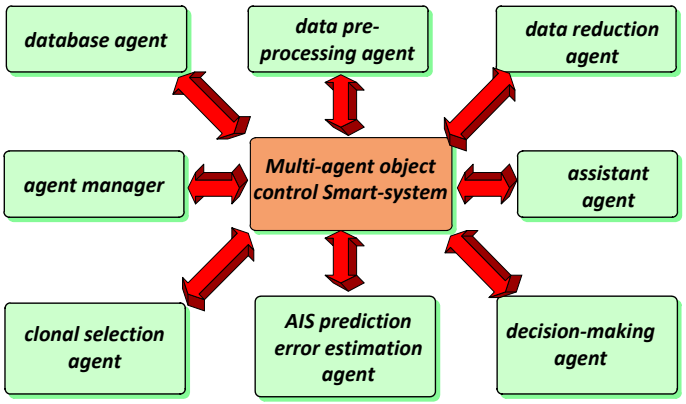


Figure 3: Structure of multi-agent smart system for complex objects control

Table 1 describes the main agents and their functions.

Table 1: Functions of agents of multi-agent smart system

Agent name	Function of the agents
Database agent	A database agent is needed to organize data readout from the Honeywell PKS historical module of the distributed control system. The generated database consists of sensors and monitoring systems, which allow to assess the state of the technological control object (normal operation, work with deviations, an accident, etc.).
Data preprocessing agent	The agent of preliminary data processing is responsible for rationing, centering, filling in the missing data in the database of parameters of the control object. The agent interacts directly with the database agent.
Data reduction agent	The data dimension reduction agent performs information processing based on the multi-algorithm approach for reducing non-informative parameters describing the behavior of a complex object.
Clonal selection agent	The agent implements the approach of artificial immune systems, the clonal selection algorithm for prediction of the behavior of a complex object.
AIS prediction error estimation agent	This agent is designed to select the algorithm with the best predictive result.
Decision-making agent	The decision support agent generates recommendations for process control based on the obtained data on the results of the

	prediction based on the modified AIS algorithm.
Assistant agent	The agent contains reference information about the algorithms and principles of data processing. Assistant functions are necessary for correct input of information into the system.
Agent manager	Agent Manager coordinates the work of other agents.

6. Simulation results

6.1. Technological process of oil gases cleaning from acidic components at installation 300

Researches were conducted on the real production data of the TengizChevroil company. The enterprise consists of several complex technological lines (CTL) and installations (Figure 4).

As the control object we will consider the Installation 300, which is intended for the petroleum gases cleaning from acid components: hydrogen sulfide, carbon dioxide and carbon dioxide. Acidic components are removed using a solution of diethanolamine [21]. The raw materials of the U300 are high and medium pressure oil gases coming from Installation 200 (U200) intended for gas separation. The final products of the U300 are: purified high-pressure gas entering the Installation 700 for fractionation; refined medium pressure gas returned to the U200 for compression; acid gas with a high content of hydrogen sulfide, is sent to the installation 400 for the production of sulfur. In general, the U300 consists of the following main components: high-pressure gas cleaning, medium-pressure gas cleaning, dietalomine regeneration, amine filtration, chemical supply, etc.

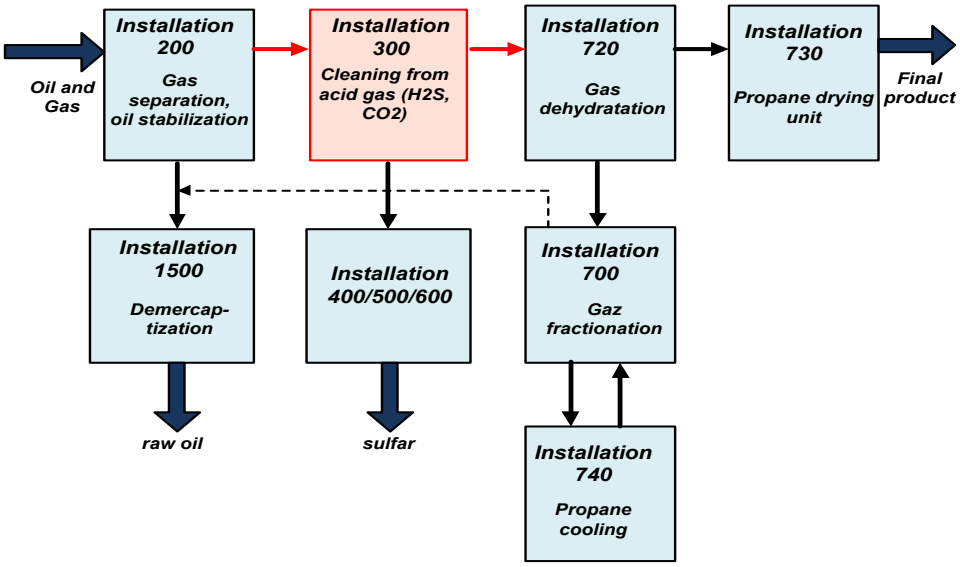


Figure 4: The structure of the Tengizchevroil enterprise

Table 2: Specification of medium pressure gas cleaning equipment

Designation on the diagram	Unit Name
D-302	Medium Pressure Absorber
F-328	Medium Pressure Separator and Coalescer
F-324	Medium Pressure Filter-Separator
F-320	Separator
F-319	Separator
F-303	Saturated Amine Tank
F-203	Low pressure separator
GC-200	First stage line
GC-201	Second stage suction line
EA-302	Regenerator reboolers (4 heat exchangers)
EA-304	Regenerated / saturated amine heat exchanger

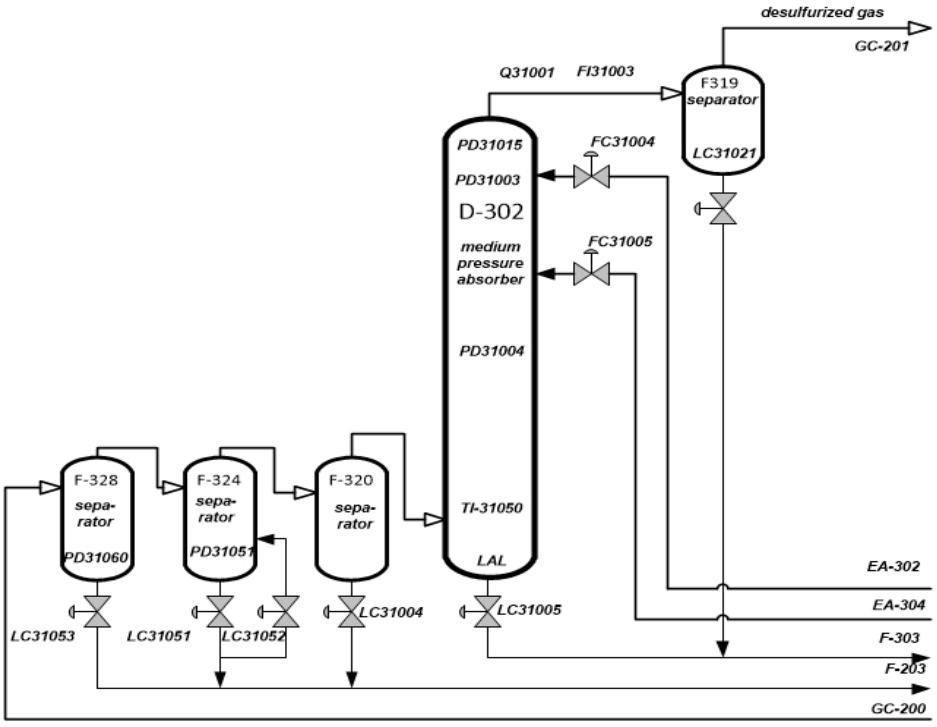


Figure 5: Fragment of the U300 technological scheme, medium pressure gas cleaning

The technological process of medium pressure gas cleaning in the absorber D-302 (medium pressure absorber) was chosen as a complex control object.

6.2. Medium Pressure Gas Cleaning of the Installation 300

Let consider a fragment of the U300, a medium pressure gas cleaning process (Figure 5, Table 2). The flow of raw gas enters the medium pressure absorber D-302 from the U200. In separators F-328, F-324, F-320 the hydrocarbon condensate is removed from the raw gas and transferred to the low-pressure separator F-203. This process is controlled by level controllers: LIC31053, LIC31051, LIC31052 and LIC31004.

Next, gas from the F-320 enters D-302. In the medium pressure absorber, the absorption process of H_2S and CO_2 also takes place due to the contact with two streams of the solution of regenerated diethanolamine [21], the temperature of which should not be lower than 540° in order to prevent condensation of hydrocarbons. In a medium pressure absorber, a pressure is approximately 24.5 bar. Gas temperature is measured using TI31050.

The absorber D-302 in the upper part consists of 11 valve plates, the pressure drops on which are measured using PDI31015, PDI31003 and PDI31004. In the case of a high pressure differential, plates may be blocked or foamed in the absorber. In order to prevent foaming, the antifoam injection points are located on the supply line of the regenerated amine in the column to the flow regulators: FIC31004 and FIC31005. The output of saturated dietalomine to the F-303 receiver is regulated by the LIC31005 regulator.

Next, the purified gas passes through an F-319 separator, where about 99% of the amine solution is removed in droplet particles of 10 microns and more in size and discharged to F-303. The output of the amine solution in F-303 is controlled by the LIC31021 level controller. Sulfured gas from F-319 is directed to the suction line of the second stage of GC-201. The consumption of diethanolamine in the upper section is regulated by the flow regulator FIC31004, and in the lower section by the FIC31005. The consumption of purified gas from D-302 is measured using FI31003, the concentration of H_2S is measured by QRAH31001. A total of 457 sensors are located on the U300, information from which is collected on the basis of the distributed Honeywell DCS control system. Sensor specifications are presented in Table 3.

As an example, let consider daily measurements from the U300 medium-pressure gas cleaning unit. A fragment of the parameters database is presented in Table 4.

6.3. Simulation of the developed Smart technology on the example of a complex object of the medium pressure absorber

Modern distributed control systems, such as the Honeywell DCS during the operation collect a huge amount of production data, the processing of which in industrial operation conditions is a complex task. Particular attention is paid to industrial safety, in connection with which there are different priorities of the alarms: emergency, high, low, logging and sending to the printer, writing to the log, sending to the printer and no action [21]. In this case, there are three alarms states: on, off, disabled. Therefore, for the application of the developed Smart technology there are selected 5 classes, depending on the state of the object for an alarm:

1 class. The alarm priority - emergency, the alarm state is on;
2 class. The alarm priority - high, the alarm state is on;
3 class. The alarm priority - low, the alarm state is off;

Table 3: Specification of sensors for the U300 circuit fragment, medium pressure gas cleaning block

№	Number of position	Purpose of the device	Measure	Measurement range	Manufacturer
D1	*LIC31053*	buoy level gauge	kg/1007 мм	0,046-0,75	Fisher 2390B-249BF
D2	*LIC31051*	buoy level gauge	kg/559 мм	0,051-0,843	Fisher DLC3010-249BF
D3	*LIC31052*	buoy level gauge	kg/631 мм	0,058-0,949	Fisher 2390B-249BF
D4	*LIC31021*	buoy level gauge	kg/500 мм	0,029-1,054	Fisher 2390B-249BF
D5	*LIC31004*	buoy level gauge	kg/500 мм	0,034-0,724	Fisher 2390B-249BF
D6	*LIC31005*	difference converter	mbar	0-294,2	Rosemount 3051CD2
D7	*TI31050*	temperature converter	°C	0-100	Honeywell STT350-0-EP0
D8	*PDI31003*	pressure drop	mbar	0-280	Rosemount 3051CD3A
D9	*PDI31004*	pressure drop	mbar	0-180	Rosemount 3051CD2A
D10	*FIC31004*	consumption regulator with indicator	mbar	0-600	Rosemount G1151DP4E
D11	*FIC31005*	consumption regulator with indicator	mbar	0-1124	Rosemount G1151DP5E
D12	*QRAH31001*	analyzer of H_2S in gas	ppm H_2S	0-50	Tracor Atlas 722R/102
Dn	...	…	…	…	…

Table 4: Fragment of the parameters database of medium pressure gas purification process of U300

№	*TI31050*	*LIC31053*	*PDI31003*	*…*	*QRAH31001*
1	85,563	41,815	229,829	…	44,509
2	57,315	7,017	243,282	…	43,887
3	96,944	92,453	186,916	…	47,597
4	85,257	44,324	232,445	…	237,979
5	85,844	47,465	233,306	…	46,529
6	87,093	50,434	247,162	…	45,659
7	83,714	33,813	215,779	…	49,894
8	86,189	40,041	228,184	…	46,105
9	85,688	41,387	228,406	…	49,168
10	83,381	32,606	213,857	…	45,197
11	82,75	31,564	212,562	…	44,564
12	84, 76	30,187	211,865	…	43,765
…	…	…	…	…	…
800	74,379	51,453	220,456	…	25,648

4 class. The alarm priority - writing to the log and sending to the printer, the state of the alarm - prohibited;
5 class. The alarm priority - no action, the state of the alarm is prohibited.

Remark. The creation of classical mathematical models for complex control objects is a time consuming task. As a mathematical model for research, there is considered a time series consisting of records from the medium-pressure absorber sensors.

As a modeling environment, modern Rapid Miner software products and WEKA operators were used. Let consider a fragment of the DB_D302 database describing the state of the complex control object of the absorber D-302 at medium pressure gases cleaning on the U300 consisting of 12,000 copies of data. Figure 6 shows the visualization of results from the DB_D302 database sensors According to the 4th algorithm, the database DB_D302 is pre-processed.

Then, within the framework of the multi-algorithm approach, the PCA algorithm is applied in order to reduce the low-informative features that characterize the behavior of the complex control object *D-302.* Figure 7 shows the simulation results in the form of a graph of attributes ranking by degree of importance, as well as the visualization of the *DB_D302* database after the non-informative parameters reduction. Similarly, the reduction of low-information descriptors is performed using the Random Forest algorithm (Figure 8) and the PSO swarm intelligence algorithm (Figure 9).

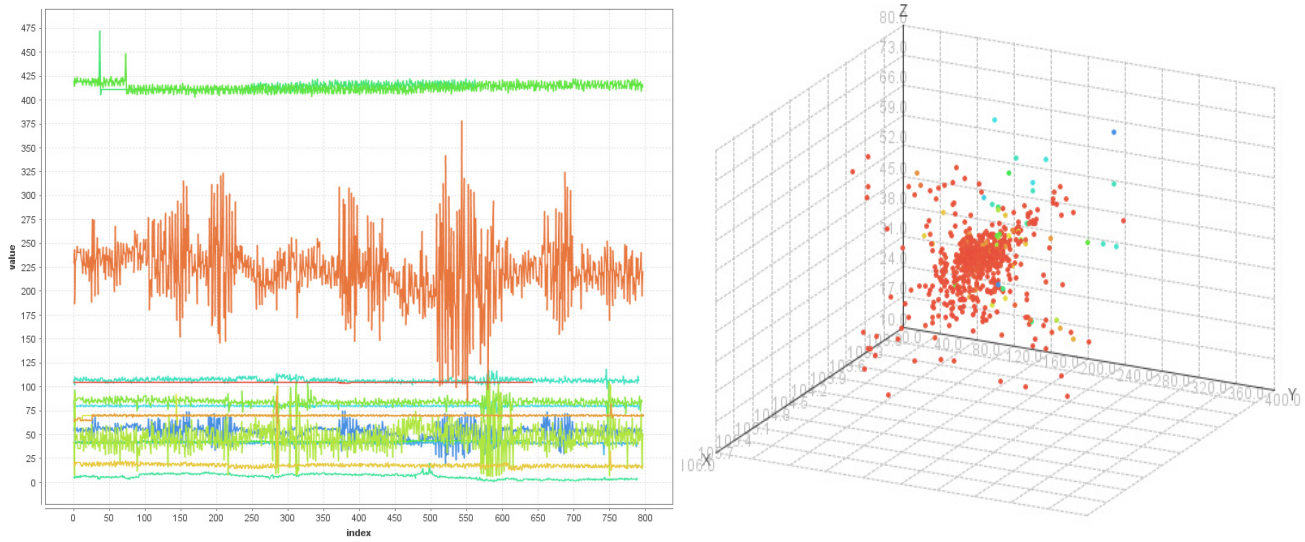


Figure 6: Fragment of *DB_U300_D302* database visualization in 2D and 3D form

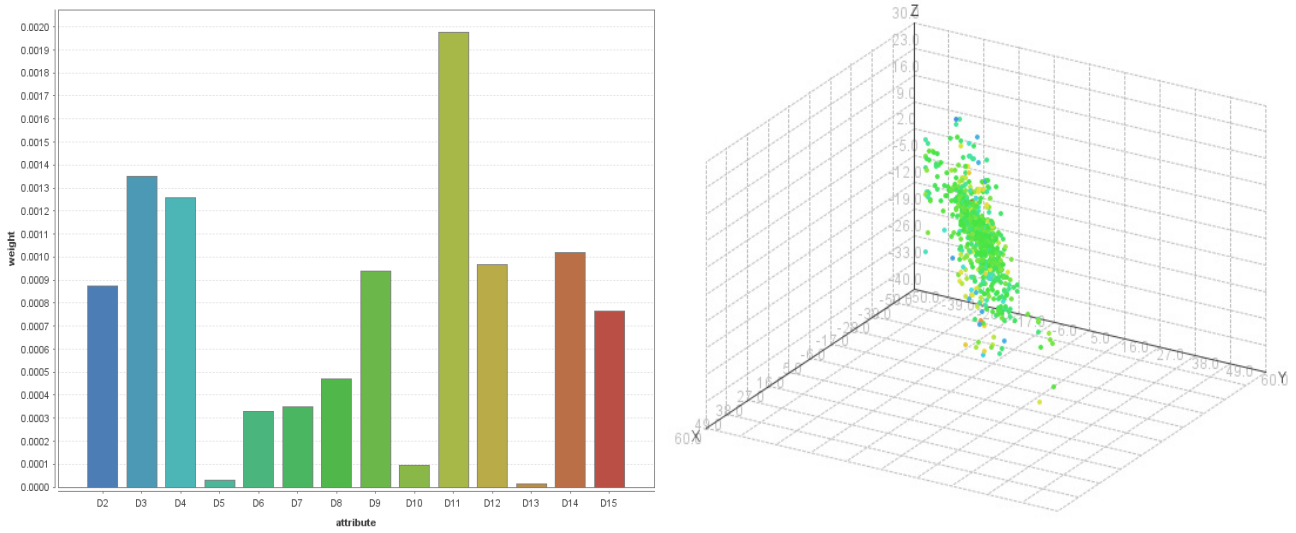


Figure 7: Low-informative parameters reduction using PCA

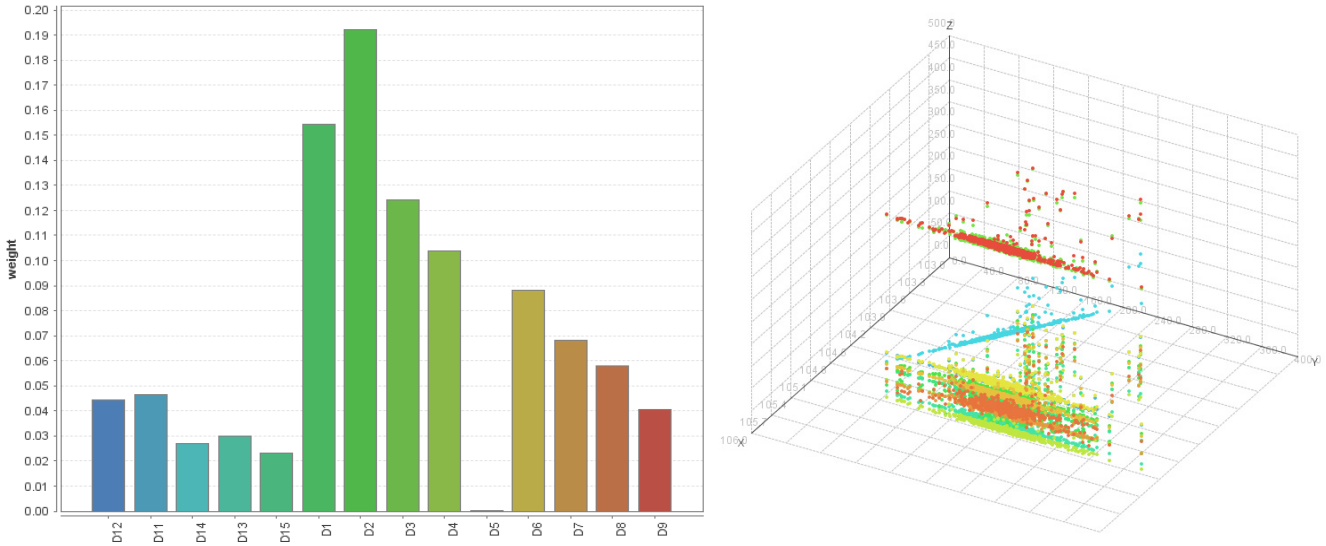


Figure 8. Low-informative parameters reduction using Random Forest

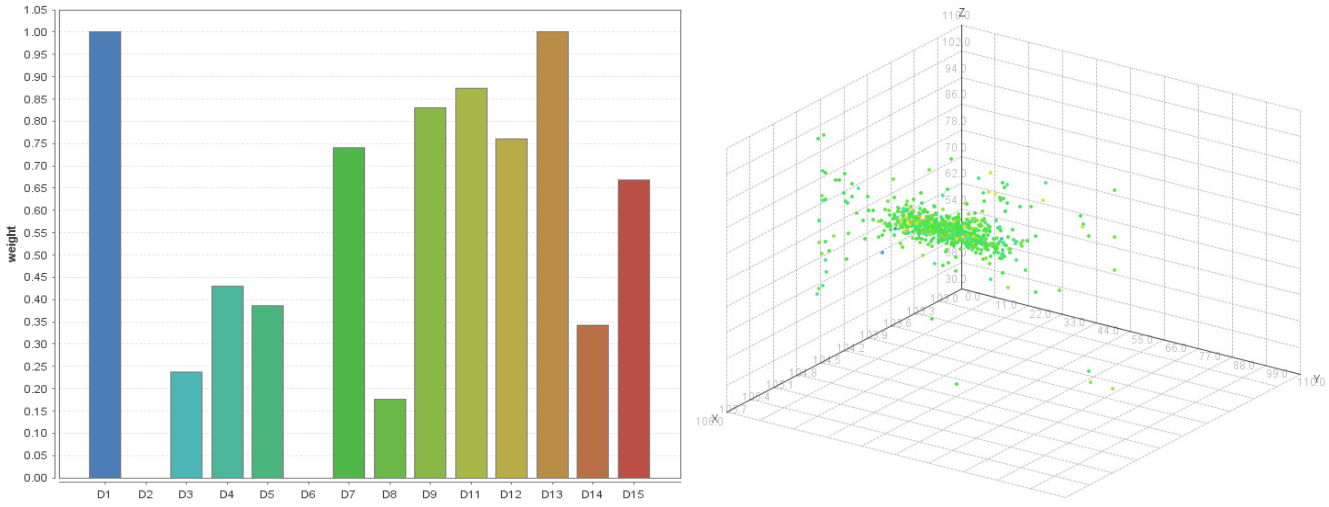


Figure 9. Low-informative parameters reduction using PSO

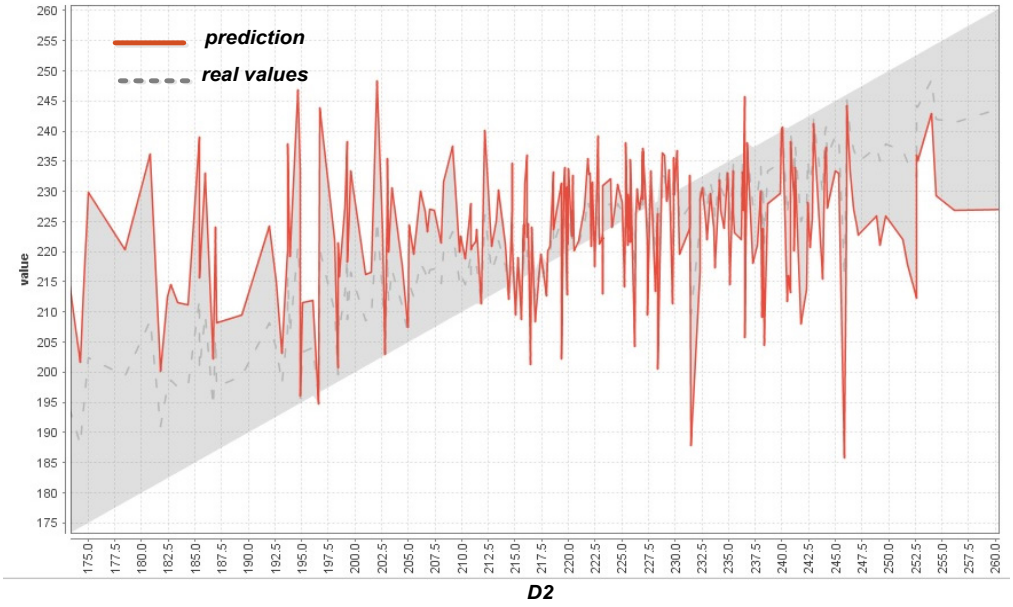


Figure 10. Prediction based on modified algorithm *PSO - Clonal Selection*

Further, according to the 4th algorithm, after the selection of informative features, the task of classification is solved using the AIS algorithm, based on clonal selection. A comparative analysis of the simulation results is presented in Table 5.

Table 5: Comparative analysis of image recognition results based on clonal selection

Feature selection method	PCA	RF	PSO
Performance of feature selection	87,67%	95,45%	94,56%
Runtime of feature selection	1 min 25 sec	7 min 34 sec	1 min 4 sec
Performance of clonal selection classification	85,53%	86,74%	93,67%
Runtime of feature selection	56 sec	2 min 59 sec	34 sec

Because the Honeywell DCS distributed control system from the Installation 300 collects a huge amount of production data, the required time to simulate multidimensional data is a necessary criterion for evaluating the effectiveness of the modified algorithm.

In the case of the use of intelligent algorithms in the development of control systems, it is impossible to use classical sustainability criteria in order to assess the effectiveness of the developed Smart technology. The effectiveness of the heuristic algorithms is carried out using the calculation of such parameters as [35]:

- *accuracy* - accuracy (relative amount of correctly classified examples, percentage of correct predictions);
- *classification_error* - the percentage of incorrect predictions;
- *kappa* - kappa statistics calculation;

- *weighted_mean_precision* - weighted average value for each class of measurement accuracy;
- *spearman_rho* - rank correlation between actual and predictable labels using Spearman's rho measure, which shows a linear relationship between two variables;
- *kendall_tau* - rank correlation using Kendall's tau measure, showing the strength of the relationship between two variables;
- *absolute_error* - average absolute deviation of the predicted values from the actual value;
- *relative_error* - average absolute deviation of the forecast from the actual value divided by the actual value;
- *relative_error_lenient* - average relative error, which shows the average absolute deviation of the forecast from the actual value divided by the maximum of the actual value and the forecast;
- *relative_error_strict* - average severe relative error;
- *normalized_absolute_error* - normalized absolute error;
- *root-relative-squared error* - averaged relative quadratic error, etc.

Table 5 shows the final accuracy parameter showing the percentage of correct predictions by the proposed methods.

Therefore, according to the research results within the multi-algorithm approach, the best prognostic result for solving the problem of prediction and control of the process of medium-pressure gases cleaning based on real data of the control object D-302 is a modified algorithm *PSO - Clonal Selection*. A visualization of a fragment of the prediction results based on the *PSO - Clonal Selection* algorithm is presented on Figure 10.

7. Conclusion

As a result of the research, there was proposed a Smart technology for complex objects control based on Honeywell's distributed control system and an artificial immune system approach. The scientific novelty of the proposed technology is following:

1. The technology allows the prediction and control of complex objects of the oil and gas industry based on the intellectual analysis of historical data obtained from the distributed control system of Honeywell DCS.

2. There was developed a modified algorithm of artificial immune systems on the basis of a multi-algorithm approach in order to improve the quality of prediction of the behavior of complex objects under industrial operation.

3. There has been created a multi-agent Smart system that allows the integration of the proposed modified artificial intelligence algorithms into real industrial control systems.

4. Presented the simulation results on the basis of a real complex facility of the TengizShevroil enterprise (medium pressure absorber) of a fragment of the installation line 300 for gas cleaning from acid components.

Therefore, the proposed Smart technology allows the implementation of artificial intelligence algorithms based on real industrial production, in case when the classical methods of control theory are not efficient and unsuitable for working with large amount of data.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The work was carried out under the grant of the KN MES of the Republic of Kazakhstan on the topic: AP05130018 "Development of cognitive Smart-technology for intellectual complex objects control systems based on artificial intelligence approaches" (2018-2020).

References

[1] D. O'Brien, "Smart operations", Pulp and Paper Canada, 2018.
[2] D. Yongsheng, Z. Lei, H. Kuangrong, Bio-Inspired clloaborative intelligent control and optimization, Springer, 2018.
[3] R. Hernandez, M. Lopez-Franco, E. Sanchez, A. Alanis, J. Ruz-Hernandez, Decentralized neural control, application to robotics, Springer, 2017.
[4] T. Wang, H. Gao, J. Qiu, "A combained adaptive neural network and nonlinear model predictive control for multirate networked industrial process control" IEEE Transactions on Neural Networks and Learning Systems, 27(2), 416-425, 2016. http://doi.org/10.1109/TNNLS.2015.2411671
[5] G. Yan, H. Fang, X. Chen, L. Shi, "Process control parameters optimal selection based on particle swarm algorithm", in 8th International Symposium on Computational Intelligence and Design, Hangzhou, China, 2015. http://doi.org/10.1109/ISCID.2015.195
[6] K. Lakshami Narayana, V. Naveen Kumar, M. Dhivya, R. Prejila Raj, "Application of ant colony optimization in tuning a PID controller to a conical tank", 8(s2), 217-223, 2015. http://doi.org/10.17485/ijst/2015/v8iS2/61635
[7] A. Ramzy, A. Aldair, A. Almousawi, "Design an optimal PID controller using artificial bee colony and genetic algorithm for autonomous mobile robot", International Journal of Computer Applications, 100(16), 8-16, 2014. http://doi.org/10.5120/17607-8016
[8] K. Goher, S. Fadlallah, "Bacterial foraging-optimized PID control of a two-wheeled machine with a two-directional handling mechanism", Robotics and Biomimetics, 4(1), 2-19, 2017. http://doi.org/10.1186/s40638-017-0057-3.
[9] R. Precup, R. David, E. Petriu, "Grey wolf optimizer algorithm – based tuning of fuzzy control systems with reduced parametric sensitivity", IEEE Transactions on Industrial Electronics, 64(1), 527-534, 2017. http://doi.org/10.1109/TIE.2016.2607698
[10] G. Ponticelli, S. Guarino, V. Tagliaferri, O. Giannini, "An optimized fuzzy-genetic algorithm for metal foam manufacturing process control", The International Journal of Advanced Manufacturing Technology, 1-12, 2018. http://doi.org/ 10.1007/s00170-018-2942-5
[11] S. Mustapha, K. Faycal, S. Mohammed, "Application of artificial immune algorithm-based optimisation in tuning a PID controller for nonlinear systems", Int. J. Automation and Control, 9(3), 186-200, 2015. http://doi.org/ 10.1504/IJAAC.2015.070955
[12] Y. Peng, X. Luo, W. Wei, "A new control method based on artificial immune adaptive strategy", Electronika ir elektrotechnika, 19(4), 3-8, 2013. http://doi.org/10.5755/j01.eee.19.4.1246
[13] J. Daudi, "An overview of application of artificial immune system in swarm robotic systems", Advances in Robotics & Automation, 4(127), 1-6, 2015. http://doi.org/10.4172/2168-9695.1000127
[14] H. Rezende, M. Silva, M. Santos, L. Honório, L. Silva, V. Vidal, J. Ribeiro, A. Cerqueira, A. Pancoti, B. Reginax, "Signal Estimation for UAV Control Loop Identification Using Artificial Immune Systems", in 22nd International conference on system theory, control and computing, 579-584, 2018. http://doi.org/ 10.1109/ICSTCC.2018.8540706
[15] F. Yao, F. Wang, M. Zhang, "Weak thruster fault detection for autonomous underwater vehicle based on artificial immune and signal pre-processing", Advances in mechanical engineering, 10(2), 1-13, 2018. https://doi.org/10.1177/1687814018758739
[16] J. Marciniak, K. Wawryn, P. Widulinski, "An Artificial Immune Negative Selection Algorithm to Control Water Temperature in the Outlet of the Chamber", in International Conference on Signals and Electronic Systems, 236-241, 2018. https://doi.org/10.1109/ICSES.2018.8507293
[17] Y. Ding, Nan Xu, S. Dai, L. Ren, K. Hao, B. Huang, "An immune system inspired reconfigurable controller", IEEE Transactions on Control Systems

Technology, 24(5), 1875 – 1882, 2016. https://doi.org/10.1109/TCST.2016.2514847
[18] L. Puggini, J. Doyle, S. McLoone, "Fault detection using random forest similarity distance", IFAC-PapersOnLine, 48(21), 583-588, 2015. https://doi.org/10.1016/j.ifacol.2015.09.589
[19] K. Hanson, S. Yella, M. Doughert, H. Fleyeh, "Machine learning Algorithms in Heavy Process Manufacturing", American Journal of Intelligent Systems, 6(1), 1-13, 2016. https://doi.org/10.5923/j.ajis.20160601.01
[20] L. Yang, J. Zhang, F. Deng, J. Chen, "Sensor fault diagnosis based on on-line random forests", in 35th Chinese Control Conference, 1934-1768, 2016. https://doi.org/10.1109/ChiCC.2016.7553992
[21] Technological regulations for the technological process of hydrocarbon gases cleaning at the plant 300. TengizShevroil LLP, 2016.
[22] J. Brownlee, "Clonal selection algorithms", CIS Technical Report 070209A, 2007, 1-13.
[23] L. de Castro, F. von Zuben, "The Clonal Selection Algorithm with Engineering Applications", Proceedings of Workshop on Artificial Immune Systems and Their Applications, 36-37, 2000.
[24] U. Berna, S. Kulturel-Konak, "A review of clonal selection algorithm and its applications," Artificial Intelligence Review, Springer, 36(2), 117–138, 2011. https://doi.org/10.1007/s10462-011-9206-1
[25] J. Brownlee, "Clever Algorithms, Nature-Inspired Programming Recipes", 2011.
[26] J. Kennedy and R.C. Eberhard, "Particle swarm optimization", in IEEE International Conference on Neural Networks, 4, 1942-1948, 1995. https://doi.org/10.1109/ICNN.1995.488968
[27] J. Blondin, "Particle Swarm Optimization Applications in Parameterization of Classifiers", Armstrong Atlantic State University, 1-33.
[28] G.A. Samigulina, Z.I. Samigulina, "Modified immune network algorithm based on the Random Forest approach for the complex objects control", Artificial Intelligence Review, 2018. https://doi.org/10.1007/s10462-018-9621-7
[29] L. Breiman, A. Cutler, "Random Forest", Berkeley, 2005.
[30] S.P. Chistyakov, "Random forests: a review", in proceedings of the Karelian Scientific Center of the Russian Academy of Sciences, 1, 117-136, 2013.
[31] L. Salazar, F. Mayer, D. Schütz, B.Vogel-Heuser, "Platform Independent Multi-Agent System for Robust Networks of Production Systems", IFAC Proceedings Volumes, 51(11):1261-1268, 2018. https://doi.org/ 10.1016/j.ifacol.2018.08.359.
[32] A. Sayed, S. Hanafi, A. Hassanien, "Multi-agent Artificial Immune System for Network Intrusion Detection and Classification", Proceeding of International Joint Conference SOCO'14, 145-154.
[33] S. German, S. Shin, A. Tsourdos, "Immune-system inspired approach for decentralized multi-agent control," in IEEE 24th Mediterranean Conference on Control and Automation (MED), 1020 – 1025, 2016. https://doi.org/10.1109/MED.2016.7536048
[34] G.A. Samigulina, Zh.A. Massimkanova, "Multi-agent system for forecasting based on modified algorithms of swarm intelligence and immune network modeling," in 12th International Conference Agents and Multi-agent Systems: Technologies and Applications (KES-AMSTA-18), 199-208, 2018. https://doi.org/10.1007/978-3-319-92031-3_19
[35] M. Hofmann, R. Klinkenberg, "RapidMiner: Data Mining Use Cases and Business Analytics Applications", 2014, 48-49.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 82-87 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Study on Development of Evaluation Metrics for Learners in Physical Computing

Woochun Jun*

Seoul National University of Education, Department of Computer Education, 06639, Korea

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 11 August, 2019
Online: 03 September, 2019

Keywords:
Physical Computing
Evaluation Metrics
ICT Education
Software Education

ABSTRACT

Physical computing is important for ICT (information and communication technology) Education and other informatics education such as software education since physical computing can provide learning-by-doing education for students. It is also a strong tool to increase students' programming ability using various type of physical computing tools like a robot. In physical computing, it is necessary to have evaluation metrics for learners in order to test students' progresses and environmental safety, etc. However, in the literature, there is little work for evaluation metrics for learners. In this paper, for the proposed evaluation metrics for learners developed in the previous research, the usefulness and validity are discussed and justified with statistical analysis. The final 23 evaluation metric will be very useful in actual physical computing education.

1. Introduction

This paper is an extension of work originally presented in ICTC2018[1].

Physical computing refers to downloading data from the real world to a digital device, processing it in the form of software, and then outputting the results to a monitor, LED or various devices. To put it more simply, computers and the real world are talking through data. The main content of the physical computing was developed by artists dealing with New Media Art. This is because it was necessary to understand computers in order to create works in the field of media art using images or light.

Physical computing is the foundation of the Internet of Things field. The basic nature of the Internet of Things is that it recognizes its surroundings through sensors and shares or processes the necessary information over the network, so it can be said to be an important technology that forms the Internet of Things.

On the other hand, for students who are new to programming, physical computing can be a good way to learn how computers and programs work. For example, students can learn about computers by assembling blocks directly, controlling robots by assembling and controlling them, or using sensors and outputs that can be easily detached to measure the quality and temperature of air which makes it easier for them to understand the principles of machine and programming.

Simplified coding programs and computer devices help students learn about physical computing. Instead of advanced programming languages such as C, students can learn programming simply by using block-coding software called Scratch. And instead of dealing with complex computer parts, students can understand the principles of computers by assembling microcomputer boards such as Arduino and Raspberry Pies, as well as various sensors and outputs to complete their own machine.

With this characteristic that the virtual world and the real world can exchange information, it is expected that the software industry will play an important role in the future thanks to the fact that it is the basis of the Internet of Things field that recognizes the surrounding environment and processes the necessary information, and that it can be approached with a simple and active image in terms of programming education.

In physical computing, evaluation metrics for learners are very important. Those metrics are very useful in evaluating learners' progresses, teaching-learning models, class materials, etc. However, in the literature, there is little research work on evaluation of learners in physical computing. The purpose of this paper is to develop evaluation metrics of learners in physical computing and justify them based on rigorous statistical analysis.

The rest of this paper is organized as follows. In Chapter 2, related works for physical computing are introduced. In Chapter3, evaluation metrics of learners are proposed and discussed. In Chapter 4, statistical analysis is presented for those evaluation

*Corresponding Author: Woochun Jun, Seocho Dong, Seocho Gu, Seoul, Korea, wocjun@snue.ac.kr

metrics. Finally, in Chapter 5, conclusions and further research issues are presented.

2. Related Works

2.1. Evaluation standards of Physical Computing

In this section, literature review works for evaluation standards for robot computing and physical computing are introduce.

In [2], achievement criteria of robot computing are proposed as below. In their work, criteria of robot computing is classified into 5 categories, understanding of robot, robot experience, robot motor control and sensor experience, control of sensor and creation of works, project in daily life, respectively.

[Step 1] Understanding Robot

(1) Definition of robots

I can understand and explain the basic elements of robot definition.

(2) Robot type and composition

I can understand the kinds of robots used in real life and the parts and applications that make up them.

(3) Robot operating principles

Robots understand how they react through external input or recognition and know how they work.

[Step 2] Robot Experience

(4) Robotic rules and safe use

I can understand the basic rules for building robots and how to use them safely.

(5) Understanding robot behavior

It is able to know necessary procedures for robot operation such as connecting devices and turning them on with variety of robots.

(6) Fabricate simple motion robot

Robotic behavior can be created in the order presented through simple block programming. It can also be created by modifying robot movements by applying various themes.

(7) Brief description of operation robot

Simple block programming can create robot behavior and explain the contents of the procedure.

[Step 3] Robot Motor Control and Sensor Experience

(8) Understanding Rotational Motion

Understanding the motor, learning about the operating procedure, principle of rotation, and component control required for member operation can be understood and simple robot rotation operation can be created and understood.

(9) Various robot-driven work production

Based on the principle of motion and understanding of the robot's driving procedure, various robots can be used to make a drive. I can also learn how to control the precise movement of a robot.

(10) Utilization of robot drive works

I can explain to my friends how the robot drives, how the robot is programmed in order, and its functions. It develops social relationship cooperation and communication skills through sharing of works and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

(11) Creation of simple sensor robot works

I can create a robot or device that can control behavior combined with simple sensors. It can make robots or devices using one of several sensors, including light sensors, ultrasonic sensors and sound sensors.

(12) Brief description of sensor robot work

I can identify the types of sensors used in my life, understand the characteristics of sensors, and describe robots or devices with sensors. I can express to my friends the design, function and programming procedures of a robot or device-driven work combined with one sensor. It develops social relationship cooperation and communication skills through sharing of works and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

[Step 4] Control of Sensors and Creation of Works

(13) Produce various sensor robot works

I can create and program controllable robot or device works by combining various sensors that can solve problems found in my life. It can also create works and explain them logically according to the programming sequence according to the purpose and purpose of production.

(14) Description of the various sensor robot works

I can describe the function, use and programming procedures and methods of robot or device work combined with various sensors. It also develops social relationship cooperation and communication skills through works sharing and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

(15) Rule Design and Robot Fabrication

Understanding the conditions for compliance with the rules, understanding the algorithms that can perform them, and designing the parts and devices needed to drive can be designed to operate through programming that applies algorithms to create robots or devices and perform them.

[Step 5] Project in Daily Life

(16) Designing Robot Works in Life

I can think about topics that can solve life problems and design algorithms and creative works that can define necessary parts and functions and control them by individuals or teams.

(17) Making Robot Works in Life

A creative work can be created through programming with algorithms applied to create robots or devices designed by individuals or teams and to perform defined functions according to topics that can solve life problems.

(18) Sharing and Expression of Robot Works

To solve problems found in life, creative works created by individuals or teams can be shared with friends, and better improvements can be found and expressed. In addition, better algorithms can be found, supplemented and shared to perform functions.

2.2. Educational Effects of Physical Computing

In [3], they presented EPL and physical computing curriculum using Entry and sensor boards that can be easily learned and used by elementary school students. The results of this study showed that both the comparative groups applying the problem-oriented EPL curriculum and the groups applying the additional problem-oriented EPL curriculum and the physical computing curriculum have significant effects on the improvement of students' computational thinking skills. Specifically, the problem-oriented EPL curriculum and the physical computing curriculum using the sensor board contributed to increasing learner's problem-solving skills, and showed that they are more effective in terms of improving satisfaction and computational thinking-related problem-solving skills when the physical computing education is combined with the programming training alone

In [4], high school students in the industrial field were trained in programming Android applications that control Arduino using app computers and analyzed the changes in students' creative and fusion thinking abilities. To this end, they designed a teaching course based on a creative problem-solving model and an integrated thinking model. Studies have resulted in significant improvements in diffuse thinking and motivational factors among the subcomponents of creative problem solving. In addition, according to the results of a survey by students on integrated thinking, many students answered that they can devise IoT systems that can be applied to their daily lives based on the knowledge learned through this class. Thus, the education of physical computing using app computers and Aduinos could be seen to have a positive effect on students' ability to think creatively and diversely.

On the other hand, other research works showed benefits of physical computing in ICT education [5-7].

3. Development of Evaluation Metrics for Learners

3.1 Design Principles

In order to development of evaluation standards of learners in physical computing, the following principles are adopted [1].

First, student-oriented activity is emphasized. Second, process-oriented activity is emphasized. Third, cooperative activity is encouraged rather than individual study in physical computing. Fourth, attitudinal and emotional experience or achievement are included.

In overall, evaluation standards are developed based on educational philosophy called constructivism [8]. Constructivism has the following major characteristics. First, it is assumed that there is no objective knowledge in the real-world. It means that every can make his or her own knowledge depending on his or her way of thinking or view. Second, it encourages social interaction. In other words, cooperative works are encouraged for study. Third, it also emphasizes problem-solving ability for authentic problems. Authentic problems mean that problems can happen in our daily life.

3.2 Evaluation Metrics

The following evaluation standards of physical computing are proposed in [1].

Table 1. Evaluation Standards of Learners

Division	Standards
Study Guidance	-Do students know study objective in advance or in the beginning of the activity? -Do students know any prerequisite or background information necessary for the activity? -Do students know the necessary equipment or teaching aids for the activity? -Do students know study materials for the physical computing activity?
Cooperation	-Does each student have a proper role in physical computing activity? -Is any leader in the group? -Are ways of communication good enough during activity?
Safety	-Do students know safety instruction for the activity? -Do students break or ruin any parts in the activity?
Output	-Do students finish the activity and produce output? -Is output good enough as expected or incomplete? -Are students supposed to have a change to expression their review or impression of the activity?
Report	-Do students write a memo or report during the activity? -Do students produce a final report after end of the activity?
Attitude	-Are students polite in the process of activity? -Does any student cause a brawl and stop the activity?
Time	-Do students spend right time for the whole class? -Do students spend too much time in specific activity of the class?
Emotion	-Are students satisfied with the activity? -Are students interested in physical computing activity? -Do students have some confidence on the subject after activity?
Theory	-Do students know the principle or theory of the physical computing activity? -Can students link the theory with physical computing activity?

4 Statistical Analysis

4.1 Design of Analysis

Verification of the reliability of the tool to measure the importance of evaluation metrics for learners in physical computing showed that Cronbach α was 0.90 and was reliable, as shown in Table 2.

Table 2. Reliability of Measurement Tool

Item Number	Cronbach's α
23	0.90

The data collected in this study were analyzed using the Statistical Package for the Social Science (SPSS) WIN 25.0 program. The analyzer method calculated Cronbach α to verify the reliability of the measuring instrument. In addition, the average and standard deviation were obtained in order to find out the importance of the evaluation standards of learners in the physical computing, and the correlation was conducted to understand the relationship between the sub-areas of the evaluation metrics of learners in the physical computing.

4.2 Analysis Results

-Descriptive Statistics

The average and standard deviation of questions asked about the importance of evaluation metrics of learners in physical computing are as shown in Table 3.

Table 3. Results of Descriptive Statistics

Evaluation Standard	AVG.	SD
1. Do students know study objective in advance or in the beginning of the activity?	3.66	1.10
2. Do students know any prerequisite or background information necessary for the activity?	3.16	1.14
3. Do students know the necessary equipment or teaching aids for the activity?	2.91	1.17
4. Do students know study materials for the physical computing activity?	2.84	1.11
5. Does each student have a proper role in physical computing activity?	3,88	1.04
6. Is any leader in the group?	3.22	1.21
7. Are ways of communication good enough during activity?	4.22	1.01
8. Do students know safety instruction for the activity?	4.56	0.72
9. Do students break or ruin any parts in the activity?	3.50	1.30
10. Do students finish the activity and produce output?	3,41	0.95
11. Is output good enough as expected or incomplete?	3.16	0.92
12. Are students supposed to have a change to expression their review or impression of the activity?	4.06	0.91
13. Do students write a memo or report during the activity?	2.97	1.23
14. Do students produce a final report after end of the activity?	2.94	1.19
15. Are student polite in the process of activity?	4.13	0.91
16. Does any student cause a brawl and stop the activity?	3.69	0.90
17. Do students spend right time for the whole class?	4.34	0.70
18. Do students spend too much time in specific activity of the class?	3.31	1.09
19. Are students satisfied with the activity?	4.28	0.68
20. Are students interested in physical computing activity?	4.41	0.67
21. Do students have some confidence on the subject after activity?	4.19	0.93
22. Do students know the principle or theory of the physical computing activity?	3.56	1.11
23. Can students link the theory with physical computing activity?	3.66	1.04

Where AVG and SD stand for average and standard deviation, respectively.

4.3 Sub-area of Descriptive Statistics

The results of the average and standard deviation of the sub-area of the importance of the evaluation metrics of learners in physical computing showed that the average of the importance of the evaluation metrics of learners in physical computing was 3.65, as shown in Table 4.

Among the sub-categories of the importance of learners' evaluation metrics for physical computing, the average was 4.29, followed by "safety" 4.03, "attitude" 3.91, "time" 3.83, "cooperation" 3.77, "theory" 3.61, "output" 3.54 and "student guidance" 3.14 and "report" 2.95 were the lowest.

Table 4. Results of Descriptive Statistics of Sub-area

Evaluation Standard Area	AVG.	SD
Student Guidance	3.14	0.97
Cooperation	3.77	0.92
Safety	4.03	0.86
Output	3.54	0.67
Report	2.95	1.17
Attitude	3.91	0.78
Time	3.83	0.68
Emotion	4.29	0.58
Theory	3.61	1.02
Overall Evaluation Metrics	3.65	0.56

Where AVG and SD stand for average and standard deviation, respectively.

As shown above in Table 4, among the sub-areas of the learners' evaluation metrics in physical computing, the importance of 'emotion' is highest and the importance of 'report' is lower than that of other areas.

4.4 Correlation Analysis

The results of the correlation between the sub-area of evaluation metrics of learners in physical computing and the entire evaluation metrics are shown in Table 5.

Where * p<.05, *** p<.001

As shown above in Table 5, student guide(r=.661, p<.001), cooperation(r=.802, p<.001), safety(r=.716, p<.001), output(r=.419, p<.05), report (r=.764, p<.001), attitude(r=.696, p<.001, time(r=.614, p<.001), emotion(r=.616, p<.001), and theory (r=.606, p<.001) showed a statistically significant static correlation with evaluation metrics of learners in physical computing. Therefore, the higher the importance of student guidance, cooperation, safety, output, report, attitude, time, emotion, and theory, the higher the importance of evaluation metrics of learners in physical computing.

Table 5. Results of Correlation Analysis

Area	Correlation Value
Student Guidance	0.661***(0.000)
Cooperation	0.802***(0.000)
Safety	0.716***(0.000)
Output	0.419*(0.017)
Report	0.764***(0.000)
Attitude	0.696***(0.000)
Time	0.614***(0.000)
Emotion	0.616***(0.000)
Theory	0.606***(0.000)

5 Conclusions and Further Research Issues

Physical computing is very important in current ICT education and software education since it can provide learning-by-doing activity for learners. In other words, learners can enjoy the physical computing activity and have interests and motivation in physical computing class. Physical computing becomes very popular as ICT education is extended in the future.

In physical computing, evaluating learners is very essential in checking learners' study progress. Also, evaluating learners is also important in developing teaching-learning models in classrooms, study contents, etc. In the literature, there is little work on evaluation standards or metrics of learners. In the previous research work, new evaluation standards are developed and proposed. Through statistical analysis, overall evaluation standards are justified to use in physical computing.

In this research, based on previous research work, the purpose of the work is to justify the results one more time. Using descriptive statistics and correlation analysis, overall 23 evaluation metrics of learners are verified.

The immediate research issues are as follows. First of all, it is necessary to develop more detailed evaluation metrics depending on physical computing tools such as robots, toy, etc. Second, it is also necessary to develop evaluation standards or metrics for study materials. It is not known yet how to develop study materials in physical computing. That is, we need the general design principles and detailed guidelines.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] W. C. Jun, "Development of Evaluation Standards of Learners for Physical Computing", in 2018 International Conference on Information and Communication Technology Convergence, Jeju Island, Korea, 2018.

http://10.1109/ICTC.2018.8539466

[2] C. Kim, "A Study on the Achievement Criteria of Robot Computing Curriculum for Elementary School" Journal of the Korean Society of Information Education, **21**(1), 97-104, 2017. http://dx.doi.org/10.14352/jkaie.2017.21.1.97

[3] J. H. Kim, D. H. Kim, "Development of Physical Computing Curriculum in Elementary Schools for Computational Thinking" Journal of the Korean Society of Information Education, **20**(1), 69-82, 2016. http://dx.doi.org/10.14352/jkaie.2016.20.1.69

[4] S. Y. Choi, S. M. Kim, "Effects of Physical Computing Education Using App Inventor and Arduino on Industrial High School Students' Creative and Integrative Thinking" Journal of the Korean Association of Computer Education, **19**(6), 45-54, 2016.

[5] J. K. Shim, H. C. Kim, W. G. Lee, "A Study on Gender Differences in Programming Attitude and Achievements on Physical Computing Education in Informatics Curriculum Revised 2015" Journal of the Korean Association of Computer Education, **19**(4), 1-9, 2016.

[6] I. H. Yoo, "The Effects on Flow at Using Robots of Introductory Programming Course" Journal of the Korean Society of Information Education, **17**(3), 329-337, 2013. http://dx.doi.org/10.14352/jkaie.2013.17.3.329

[7] J. B. Song, S. H. Baik, T. W. Lee, "The Effect of Robot Programming Learning Considered Gender Differences on Female Middle School Student′s Flow Level and Problem Solving Ability" Journal of the Korean Association of Computer Education, **12**(1), 44-55, 2009..

[8] Constructivism, https://www.learning-theories.com/ constructivism.htm

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 88-99 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Analysis of Fuzzy Data for the Selection of Ingredients of the Polymer Composition by Technical Functions in Quality Management

Ilya Vasilyevich Germashev[*,1], Tatyana Petrovna Mashihina[1], Evgeniya Vyacheslavovna Derbisher[2], Vyacheslav Evgenyevich Derbisher[3], Mikhail Alexeevich Kharitonov[4]

[1] *Department of Mathematical Analysis and Function Theory, Volgograd State University, 400062, Russia*

[2] *Department of Analytical, Physical Chemistry and Polymers Physico-Chemistry, Volgograd State Technical University, 400005, Russia*

[3] *Department of Technology of Macromolecular and Fibrous Materials, Volgograd State Technical University, 400005, Russia*

[4] *Department of Fundamental Informatics and Optimal Control, Volgograd State University, 400062, Russia*

ARTICLE INFO

Article history:
Received: 07 April, 2019
Accepted: 08 May, 2019
Online: 21 May, 2019

Keywords:
Data analysis
Fuzzy numbers
Polymer composites
Mathematical support
Intelligent system
Quality
Optimal choice
Management
Physical and chemical properties

ABSTRACT

The task of fuzzy mathematical modeling of the composition of a polymer composition is singled out and formulated, as one of the types of chemical-technological systems for creating the theoretical foundations for managing the physicochemical properties of the material produced from this composition. Fuzzy data on the technical functions and concentration of individual ingredients that included in the composition in the form of fuzzy numbers are used as the initial formalized information at modeling the polymer composition. As a result, a parametric space is formed that describes the multicomponent system under study. The index of compliance of each parameter of the polymer composition with the requirements for the physicochemical properties and quality of the polymer material is calculated. This allows you to go to the relative dimensionless real values. Weighted voting procedure is used as an aggregate function. The obtained numerical value reflects the integral measure of the compliance of the chemical-technological system with the requirements for the polymer composition.

An example of the choice of ingredients of the polymer composition by calculating using fuzzy numbers is presented. In addition to the formulation of the problem, the study develops the mathematical software for the intellectual system supporting the research and production activities of the chemical technologist to substantiate the decision to choose a polymer composition.

Accepted abbreviations: PC — polymer composition, CTS — chemical-technological system, PCM — polymer composite material, FS — fuzzy set, IS — information system, DNA — Deoxyribonucleic acid

1. Introduction

This study is an extension of work originally presented in International Russian Automation Conference [1]. The study gives a new vector of application of fuzzy mathematics methods in the analysis of complex multifactor technical systems, including chemical-technological system (CTS), for their identification and pre-experimental research. These systems, as a rule, are described by fuzzy information: the presence of linguistic descriptions, intervals, lack of data, incompatible research results, conflicting expert opinions, etc. Under these conditions, mathematical modeling using this information also has a fuzzy nature and a multiplicity of solutions. Nevertheless, the use of fuzzy mathematics, fuzzy modeling allows at the preliminary stage to choose the optimal and economical strategy of the experiment and evaluate the expected technical result, which is directly related to product quality. It should be noted that the use of fuzzy set (FS) for the preparation of initial information in the simulation of CTS has recommended itself from the positive side, as a form of theoretical

[*]Ilya Vasilyevich Germashev, Department of Mathematical Analysis and Function Theory, Volgograd State University, Prospect Universitetskii, 100, Volgograd, 400062, Russia, germashev@volsu.ru

analysis of polymer composition (PC) and finding the relationship between the components and properties of the polymer system as a whole. Here, a prerequisite is the identification and ranking of the technical functions of the ingredients of the PC. Managing the properties of polymer composite material (PCM) is an important task of chemical technology, it can be extended to: oil and gas processing (primary separation, refining, age hardening), petrochemical and organic synthesis, biotechnology, the technology of high-molecular substances, the technology of processing plant and animal raw materials and etc. [2].

The formation of the PC optimization problem and its solution is based on the "composition – property – quality – application" dependence and provides for the selection of a certain number of ingredients with the necessary technical functions using fuzzy modeling. In this paper, this idea is extended to the PCM analysis.

Raw PC, in other words, mixtures, are complex systems. They are based on polymer matrices in which various ingredients are added. The interaction between such matrices, their structure and composition also depend on the chemical and technological processes of their processing. PC-derived composites are also complex heterogeneous systems and can be divided into five leading classes:

1. Matrix, consisting of a continuous phase (matrix) and a dispersed phase (discrete particles) (dispersed — hardened composite materials).
2. Compositions with fibrous fillers.
3. Mixed compositions.
4. Compositions having an interpenetrating structure of two or more continuous phases.
5. Layered PCMs.

Considering the special interest of researchers and the market of consumers of polymer products to nanotechnology, let us mention the polymer matrix of the "polymer matrix – additive (filler)" composition, in which 2D and 3D nano-ingredients are used as fillers — graphene and carbon nanotubes. Here, mathematical modeling can be conducted based on the analysis of ingredients in the direction of predicting the level of physical and mechanical properties, depending on the number and size of nanoparticles and the parameters of the polymer matrix of a stable configuration. This direction is at the beginning of development, therefore, in most cases, problems are solved on the basis of experiments [3, 4]. Here the applicability of fuzzy mathematics is quite possible when calculating and analyzing results. Thus, the PC is ultimately a multi-component, inhomogeneous system, the characteristics of which do not allow to calculate the exact values of technical effects by its component. The versatility of the situation emphasizes the fact that in addition to a large set of active and inactive fillers, modern technology with a different purpose gives rise to the use of new modifying additives, a brief description of which is proposed below. In parentheses are their possible concentrations in the PC.

Light stabilizers allow for a long time to maintain the original color of the product and its mechanical strength, providing protection from ultraviolet rays. Here you can point to derivatives of benzonitrizol, benzophenol, nickel and cobalt salts of substituted phenols, etc. It has been established experimentally that they are effective even at very low concentrations (0.1 –5.0%).

Antioxidants prevent thermo-oxidation and slow down destruction during extrusion and injection molding. Also, these additives are used to protect the polymer when processing in aggressive environments. The mixtures compounds of phenols and phosphides (0.1 – 2.0%) are the most popular here.

Plasticizers are low molecular weight organic substances that, when introduced into the composition at the preparation stage, reduce the interaction between adjacent macromolecules. At the same time, common plasticizers include esters of organic and inorganic acids, such as dibutyl phthalate, di-(2-ethylhexyl) phthalate, dibutyl sebacate, tricresyl and tributyl phosphate, refined products, (petroleum oil), coal and wood materials (cumarone indene resin, rosin, etc), epoxidized vegetable oils, etc. The temperature of decomposition of plasticizers should be higher than the processing temperature of PC (0.1 – 45.0%).

Combined additives (light stabilizer + antioxidant + dye, etc.) are widely used in the production of agricultural films. They allow to use the polymer for its intended purpose after recycling (0.1 – 45.0%).

Flame retardants are the most important additive limiting the flammability of PCM. Special flame retardants are used for the production of non-combustible and low-flammable films, sheets and molded products. They are divided into three large groups:

- chemically interacting with the polymer;
- intuminescent;
- additives mechanically miscible with polymer.

Flame retardants are most often halogen-containing and phosphorus-containing compounds, as well as metal hydroxides and other inorganic compounds (0.1 – 3.0%).

Antistatics can get rid of the static effect. For example, in the manufacturing of cases for audio and video equipment. Commonly used antistatic agents include alkylamines (up to 3.0%).

Slip agents are an internal lubricant in a polymer matrix that reduces melt viscosity, improves extrusion performance, makes the surface smoother, shinier and glossy, and reduces the coefficient of friction of products. Active ingredients are derivatives of higher fatty acids (depending on technology).

Anti-blocking additives prevent sticking, that is especially important in the production of polypropylene films. The packages made of a film with such additives are easily opened. Package walls do not stick together. Active ingredients are silicic acid or amide waxes (0.1 – 0.3%).

Antislips are additives that are similar in properties to antiblocking. They give the surfaces of the films roughness. They are used in the manufacturing of packaging, bags for packing bulk products and other products. One of the antislips is ultra high molecular weight polyethylene (0.1 – 0.3%).

Antiseptics prevent the biological destruction of PCM (up to 2.0%).

Nucleates affect the supramolecular structure of the polymer matrix, making it more fine-grained. At the same time, the

crystallization time is significantly reduced, the product is cooled faster. In addition, the effect of greater transparency of the films is created. Nucleates are used exclusively for polypropylene (0.1 – 0.3%).

Foaming additives are used to manufacture porous polypropylene and polyethylene products. The structure of the matrix is loosened, the extensions disappear with the introduction of additives. They are designed to produce decorative films used in the packaging of gift sets, as decorative ribbons, gaskets in cork and other purposes (depending on the requirements of technology).

Modifier in the production of stretch films. It gives them the effect of additional elasticity, and for some polymers, including polypropylene, impact resistance (no information).

This list is rapidly replenishing and expanding today, for example, so-called processing additives that improve the economy and technological parameters of processing activities (in each case individually) can be singled out separately.

Today, PC theory is an attempt to solve the problem of how the final physicochemical properties of finished products depend on the composition, chemical-technological and physical-technical parameters. The ultimate goal here is to obtain materials with desired properties. To achieve this goal it is necessary, setting the final properties of the designed material or product, to know the parameters of their manufacture.

General requirements for PC ingredients are as follows:

a) thermodynamic compatibility with the polymer matrix should be ensured, preferably before the formation of a molecular solution;

b) should not be volatile;

c) do not bleed out of the polymer matrix;

d) should be nontoxic, chemically and thermally resistant.

As an example (Table 1) [5], we consider the composition (in weight parts) of an electrically insulating PC from the Russian patent which is well illustrating the multifactorial nature of the identification problem and the versatility of PCM ingredients. And this is not the most difficult PC.

Therefore, for PCM ingredients, their effect on the polymer matrix is empirically determined. And to obtain a PC with the desired set of properties, both polymer matrices and individual ingredients are selected. The complexity of the problem is obvious, as is the complexity of the system.

It should be noted that for particular cases of the theoretical analysis of the properties of PCM, many CTS analysis methods are used: discriminant, factorial, regression, etc., to identify the control actions and calculate their consequences, while operating in numerical spaces [6 – 8]. But in the presence of nonnumeric, for example, linguistic parameters, such methods become of little use and require additional efforts to harmonize the source data and mathematical methods. Comparing and critically analyzing the existing mathematical approaches [2, 6 – 9] we can conclude that an important part of this problem is overcome by using FS for describing both numerical and nonnumeric parameters. The solution of the problem itself is carried out by methods of fuzzy mathematics. Important elements of the latter are increasingly being applied in the scientific analysis of the CTS [10 – 13].

Table 1: Composition of a Polymeric Composition (in Weight Parts)

Components	No.						
	1	2	3	4	5	6	7
PVC S-70*	*100*	*100*	*100*	*100*	*100*	*100*	*100*
Dioctyl phthalate	*40*	-	-	-	*70*	-	-
Diisononyl phthalate	-	*45*		*40*	-	*80*	-
Trioctyltrimeth yl metat	-	-	*50*	-	-	-	*50*
Isodecyldephen yl phosphate	-	-	-	*20*	-	-	*15*
Calcium carbonate	*40*	*60*	*80*	*100*	*120*	*140*	*160*
Tribasic lead sulfate	*3*	-	*5*	-	*4*	-	*6*
Dibasic lead phthalate	-	*4*	-	*3*	-	*5*	-
Antimony trioxide	*3*	*4*	*5*	*6*	*7*	*6*	*8*
Zinc oxide	*2*	*3*	*4*	*6*	*8*	*2*	*4*
Zinc borate	*2*	*5*	*6*	*3*	*3*	*8*	*4*
Diphenylolprop ane	*0.20*	*0.25*	*0.30*	*0.35*	*0.4*	*0.35*	*0.30*
Magnesium hydroxide	*20*	-	*40*	*25*	*60*	-	*80*
Aluminium hydroxide	-	*30*	*20*	*50*	-	*70*	-
Metallic silver in the form of nanoparticles	*0.001*	*0.005*	*0.01*	*0.015*	*0.02*	*0.025*	*0.03*

* PVC S-70 – polyvinyl chloride polymer matrix, Russian brand S-70.

In connection with the above, let us point out that the management of PCM properties is possible in various theoretical and further practical ways, such as:

1) pre-experimental heuristic identification,

2) the choice of the polymer type,

3) changes in production technology of polymer used as a matrix (polymerization, polycondensation, modification, crystallization, structuring),

4) regulation of the concentration and technical functions of the ingredients,

5) variation of composition and structure.

Among the newer modern experimental approaches to the management of PCM properties through the "composition – properties" relationship, for example, layer-by-layer deposition of polymers can be specified to control the conductivity of ophthalmic preparations through lenses [11], the regulation of technology to control the physicochemical properties of PCM [12 – 16], or increase, in the particular case, of the structural stability of such a biopolymer as DNA, when counter-ions are neutralized with the help of polyamines [15]. From what has been said, it is clear that controlling the properties of the polymer matrix, whether it be an olefin or DNA, is a separate scientific and technological task. Here, we have not found any works on the application of FS for the theoretical analysis of the polymer matrix — a full-scale experiment is fully used. So we have attributed points 4, 5 to the above list to our problems.

Here, the use of fuzzy analysis for the study of polymer systems and their particular case of PCM and comparison of mathematical procedures of fuzzy analysis [16, 17] made it possible to formulate the task of managing the properties of PCM and to obtain particular methods of its solution. Let us make the restrictions related to the subject area of this article and introduce it into the framework of the general task of studying CTS, while considering the PC as a system consisting of an undefined set and their number. Mathematical modeling and optimization of this CTS will be carried out under the conditions mentioned above, multicriteriality, incompleteness and vagueness of the initial information. This information can include the knowledge, experience and intuition of a specialist expert, which provides the basis for managing these systems — one of the main modern vectors of their development. We repeat that the identification of optimal and compromise technical solutions at the early stages of the formation of a plan for laboratory research and pilot product releases in world practice is known [18, 19], but it is still an urgent task focused both on the creation of new equipment and technology, and on the savings of intellectual and material resources. In principle, the problem of analyzing fuzzy data is relevant for analyzing the internal problems of the intelligent systems themselves, which are oriented toward interaction with both researchers and the consumer. As an example, one can cite a system for analyzing linguistic information for predicting the commercial attractiveness of goods [20] and other areas using problems with fuzzy data [21, 22].

The work of creating a fuzzy model is multidimensional and complex. It can be noted that in chemistry and technology, PCM tasks with fuzzy data are most of all:

- selection of compositions,
- evaluation of new materials properties,
- design of competitive technologies,
- consideration of possible environmental damage,
- safety accounting, etc.

Describing the problem of modeling CTS in this part, it is necessary to proceed from the fact that the very formulation of the problem of fuzzy modeling, computer implementation of calculations and their effectiveness can have, as discussed above, the nature of the set, which is largely determined by the formalization of the source data (parameters and conditions) and the adequacy of the test results obtained during the simulation. At the same time, depending on the level of fuzziness and the type of model obtained, the solutions may have different significance [22].

Let us add that modern research focused on processing fuzzy information uses the following route: "raw data – fuzzification – analysis – defuzzification – result". At the same time, they formulate variables, including linguistic ones, and, using conventional methods of fuzzing terms, go to FS and further, as for example, in [23] analyze the resulting data model in a mathematical package [24]. Also fuzzy components are explored using intelligent systems. Since the considered approach focuses on certain subject areas, the formalization of information is carried out in accordance with the requirements of these areas using a mathematical apparatus. This is often beyond the power of ordinary users of intelligent systems. In such cases, it is offered either detailed manuals, or to fully automate the process.

We performed the real work, which is the software of the information system (IS) summarizing the information available in this part and using our own experience in solving such fuzzy problems. Here it will be possible, firstly, to automate the initial processing of the raw data (to carry out the fuzzification process) and, secondly, to analyze the solutions for the information support of researcher work. The second part is focused on a wide range of chemical-technological tasks, such as: classification, identification, examination and management of the CTS. The specific analysis methods discussed below are focused on relative data and have no substantive substantiation; therefore, they can be applied in other fields of science and technology. We also point out that to solve the problems of our profile, we used the latest advances in the field of FS analysis [25 – 29], interval mathematics [30], and others [31], however, a universal method for processing fuzzy data, we have not found neither in Russian nor in foreign practice, although research in this direction is being actively conducted, as an analysis of scientific and technical literature of recent years shows.

This paper takes into account the complexity of the system and the PCM analysis and is aimed at developing a methodology for formalizing the source data in the form of fuzzy numbers, defining a library of mathematical models for selecting ingredients from a given set. Ultimately, information is provided for the quantitative support of the decision to choose PCM. Also, the typing of the source data will allow to unify and algorithmize the process of automated analysis of the results of the experiment.

2. Research objective

Now we will set a more specific task of forming the initial data, and then the mathematical method of calculation. We point out the conceptual fuzziness that exists and arises when describing the PCM:

- the technical functions of the ingredients are not presented;
- in the description of the ingredients of their concentration in the PC are given in the form of intervals;
- the level of action of the technical function of additives is characterized linguistically and is not related with their concentration in PCM;

- the possible relationship between the physicochemical parameters of the additives themselves and PCM and the positive or negative chemical interaction between them is not indicated.

Considering these fuzzinesses, which are more practical in practical technologies, the description of the data about the PCM by fuzzy numbers and operations with them by the methods of fuzzy mathematics will be adequate techniques. In general, the quality management of PCM will be carried out according to the scheme proposed in Figure 1 using fuzzy data analysis techniques.

Note that the main requirements for the PCM are formulated by consumers (customers) and process engineers, taking into account industrial capabilities and needs, on their part introduce additional requirements. If the former determine the range and tolerances of physicochemical, mechanical, ergonomic, environmental and other operational characteristics; the latter proceed from technological limitations, polymer matrix synthesis processes (polymerization, polycondensation, modification) and processing: (melting, mixing, casting, molding, stamping, extrusion, etc.) The formulation of such requirements is made by experts in the relevant fields, and, in many cases, is cited in reference data sources.

We assume that the basic consumer requirements already exist. It is necessary to construct the composition of the PCM that best meets these requirements and select ingredients with certain technological and technical functions.

The management of the design and analysis process of the PCM itself is carried out by assessment and selection of:

- polymer matrix,
- ingredients (including fillers),
- coordination of concentrations,
- justification of processing technology.

These actions need to be carried out in concert, since the nature of each polymer matrix determines a unique combination of possible ingredients and their content. You can get far from optimal low-potential low-quality PCM even choosing a polymer matrix with good physico-chemical characteristics and using an unsuccessful set of ingredients.

The following route is proposed for creating an original technical solution:

1. The choice of the polymer matrix is carried out on the basis of the analysis of the achieved level of technology by a variable method from the available kits, most often from the brand kits.
2. Determination of technical functions and concentrations of ingredients and their combinations is carried out through technical expertise.
3. Specific representatives are selected from the known ingredients, taking into account the available possibilities, and the compliance index for each of them is calculated.
4. PC is numerically evaluated based on the properties of the selected components by calculating the PC compliance index and, if necessary, proceeds to step 1.

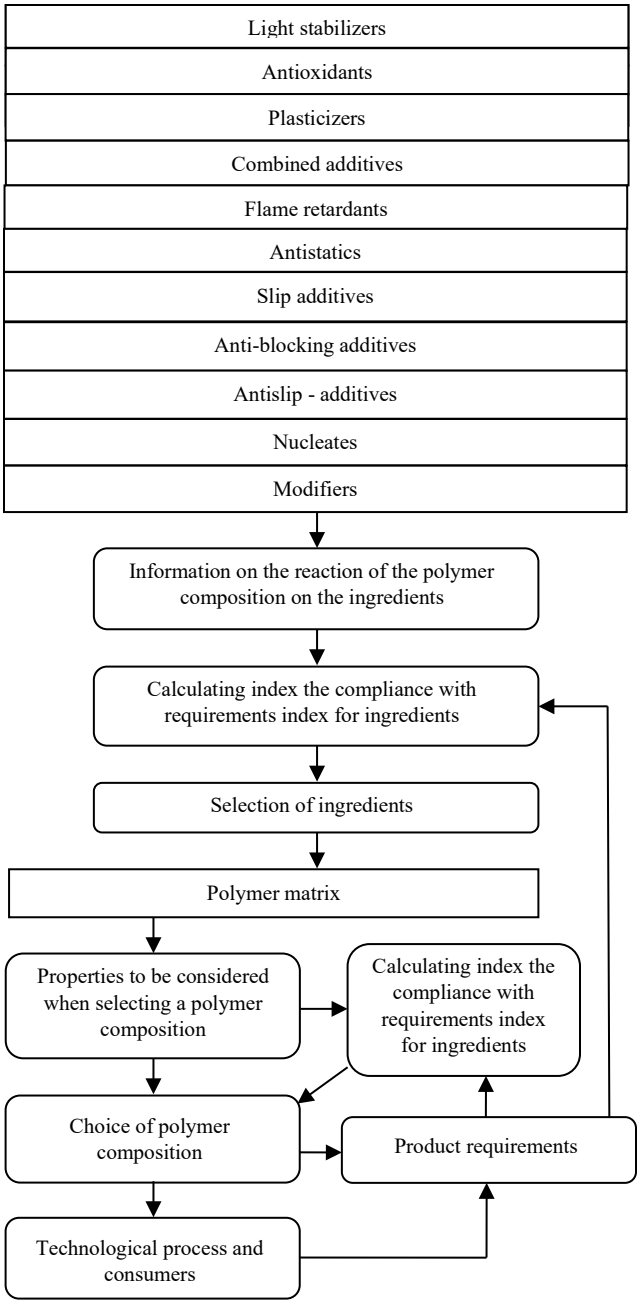


Figure 1: Control Scheme of the Physicochemical Properties of PCM and the Quality of Polymer Products

5. The decision is made on further actions: the choice of the polymer matrix, the specific composition of the PC, the individual ingredient, changes in the technological regime, etc.

Actions under item 1 are related to the experience of the technologist and researcher, the availability of information and material resources. They are obvious and, we believe, comments do not require. The actions under item 2 can be implemented either by designers or directly by process engineers — experts in the field of chemistry and PCM technology. It can also use reference books, databases and knowledge bases or knowledge systems created on the basis of generalization of subject information. In any case, the technology of these actions goes far beyond the scope of this article and is not discussed in detail here.

This article discusses in more detail the steps in paragraphs 3, 4 and 5 related with solving the problems of evaluating the technical functions of ingredients in a given set for designing a PC, in the context of the decision to choose its composition. A solution to the problem of optimal PC selection using a given polymer matrix and variability of additives is also proposed. In the simplest case, it is CTS type "one polymer matrix + one functional additive" or "polymer matrix + filler + functional additive". In the latter case, the "polymer matrix + filler" is considered inseparably. Then we will complicate the task.

We formulate the general task of determining compliance with the requirements of an abstract technical object and present the main points of its solution method, to assess the quality of the additive used to form a PC, and then the quality of PCM.

Let $Q = (Q_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ be the matrix of input data describing the results of measurements or calculations of m characteristics of n representatives of a certain class of chemical systems $s_1, \ldots, s_n$, $Q_{ij} \in Q_j$, where Q_j is a set of values representing numbers, sets, words, etc.

Here and below, we will illustrate the reasoning with an example of the choice of a flame retardant for a polymer matrix presented in the form of a fibrous substrate (Table 2). At the same time, since this is not significant for the description of the mathematical methodology, the names of chemical compounds, as well as the dimension of physicochemical parameters will be omitted. The point is that it is necessary for those given in Table 2 data to evaluate the PC system being formed and in some sense choose the best option. The choice of flame retardant as an illustration of FS application in this case is related to its technical function (see above) and cannot be considered accidental or abstract. The fact is that in the practice of chemistry and chemical technology of high-molecular compounds, imparting noncombustibility of PCM or a significant decrease in combustibility is one of the most urgent tasks, along with increasing the durability of polymer products.

Table 2: Abstracted Characteristics of Industrial and Laboratory Flame Retardants [25]

Compound no., i	Q_{i1}	Q_{i2}	Q_{i3}
0 (Reference, q/δ)	*1.0/8*	*1.5/8.5*	*21/10*
1	*19*	*12.1*	*9*
2	*18*	*10.7*	*9*
3	*18*	*3.7*	*14*
4	*6 – 10*	*6 – 10*	*15*
5	*18*	*2*	*9*
6	*3*	*3*	*10.6*
7	*1.0 – 1.2*	*2.67*	*15*
8	*1.0 – 1.2*	*1.5 – 2*	*9 – 11*

It is clear that the method discussed below can be applied to any flame retardant, and in general, to an ingredient, both from the list discussed above and any other using any polymer matrix. The solution of the problem is divided into several stages, including: analysis of the initial data Q, construction of fuzzy numbers U, formalization of requirements in the form of fuzzy numbers, comparison of U with formal requirements, choice of PC, decision making.

3. Analysis of the source data

3.1. Method Description

The purpose of this stage is to bring different types of data to one type in the form of a segment. We use the operator proposed earlier [31], i.e. define the operator A, for this:

$$A(Q_{ij}) = [q_{ij1}; q_{ij2}],\ q_{ij1}, q_{ij2} \in \mathbb{R},\ q_{ij1} < q_{ij2},$$

where $\mathbb{R}$ — set of real numbers.

It is not possible to define this operation precisely, since in practice Q_{ij} can be represented in any form, but first you can describe the solution for the most typical forms by creating a library of data models, and later, if necessary, expand the library, which will allow you to adapt the proposed approach to almost any type of source data. Consider the most common cases.

1. Q_{ij} — real number.
2. Q_{ij} — set of real numbers.
3. Q_{ij} — linguistic value.

For all these cases, it is proposed to use the following operator.

$$A(x) = \begin{cases} A_1(x), \text{ если } x \in \mathbb{R} \times \mathbb{R}^+ \setminus \{0\}, \\ A_2(x), \text{ если } x \in \Re, \\ A_3(x), \text{ если } x \in t, \end{cases}$$

where

$$A_1: \mathbb{R}\times\mathbb{R}^+\setminus\{0\} \to \Im,\ \Im = \{[a; b] |\ a, b \in \mathbb{R},\ a < b\},$$

namely,

$$A_1(x, y) = [x - y; x + y],\ x \in \mathbb{R},\ y \in \mathbb{R}^+\setminus\{0\};$$

$$A_2: \Re \to \Im,\ \Re = \{X |\ X \subset \mathbb{R}\},$$

namely,

$$A_2(X) = [\inf X; \sup X],\ X \subset \mathbb{R};$$

$$A_3: t \to \Im_1,\ \Im_1 = \{[a; b] |\ [a; b] \subset [0; 1]\},$$

namely,

$$A_3(t_k) = [e_{k-1}; e_k].$$

Operator A_3 requires additional explanations on the set of terms t. Let be $Q_{ij} \in t = \{t_k |\ k = 1, \ldots, r\}$, and terms t_k are such that $t_1 \prec \ldots \prec t_r$, where $\prec$ is an ordering relation. We split the segment $E = [0; 1]$ into r parts in accordance with the terms t_k, while maintaining the same order, i.e.

$$E = \bigcup_{k=1}^{r} E_k ,$$

where $E_k = [e_{k-1}; e_k]$, $k = 1, \ldots, r - 1$, $E_r = [e_{r-1}; e_r]$, $0 = e_0 < e_1 < \ldots < e_r = 1$.

Illustrating with compound no. 1 from the example under consideration (Table 2), we obtain, for example, for parameter no. 1

$$Q_{11} = 19, \delta_{11} = 1,$$

$$q_{111} = 18, q_{112} = 20,$$

i.e.

$$A_1(19, 1) = [18; 20].$$

Further, similar to [25], fuzzy numbers are given for q values:

$$u_{ij}(x) = \exp\left(-\frac{\ln 2}{\delta_{ij}^2}(x - q_{ij})^2\right),$$

where $q_{ij} = (q_{ij1} + q_{ij2})/2$, $\delta_{ij} = (q_{ij2} - q_{ij1})/2$, $x \in \mathbb{R}$. Main provisions of fuzzification described in [25] are given in a more generalized form in clause 3.2.

As a result, a matrix of fuzzy numbers $U(x) = (u_{ij}(x))$, $i = 1, \ldots, n, j = 1, \ldots, m$ (Table 3) is obtained.

Further it is necessary to formalize the requirements for representatives of the class of systems. In the context of our example it is necessary to formalize the operational requirements for PCM, for which can serve properties defined in statics and in the dynamics, for example degradation, such as durability and biodegradability. The latter is a time challenge. Characteristics must be expressed quantitatively and / or qualitatively. For further actions, we proceed in the same way as [31], namely, let us set a hypothetical fire retardant s_0, possessing such characteristics that fully meet the operational requirements (Table 2, fire retardant 0). At the same time, the characteristic q_{0j} reflects the required value, and δ_{0j} is the tolerance value, within which the value of the corresponding characteristic satisfies the requirements.

Table 3: Values of the Elements of the Matrix $U(x)$ Calculated from the Table 2

No. compound, i	q_{i1}/δ_{i1}	q_{i2}/δ_{i2}	q_{i3}/δ_{i3}
0 (Reference, q/δ)	*1.0/8*	*1.5/8.5*	*21/10*
1	*19/1*	*12.1/0.1*	*9/1*
2	*18/1*	*10.7/0.1*	*9/1*
3	*18/1*	*3.7/0.1*	*14/1*
4	*8/2*	*8/2*	*15/1*
5	*18/1*	*2/0.1*	*9/1*
6	*3/1*	*3/0.1*	*10.6/0.1*
7	*1.1/0.1*	*2.67/0.01*	*15/1*
8	*1.1/0.1*	*1.75/0.25*	*10/1*

3.2. Computational experiment

To evaluate systems from a given class, we use the method described in [25], which consists of the following.

Let $S = \{s_i | i = 1, \ldots, n\}$ is a set consisting of n objects. For each object s_i, m characteristics $Q_{ij}, j = 1, \ldots, m$ are defined. Depending on the conditions of use, the set of considered characteristics may expand and contract, up to the introduction of chemical, biological, ergonomic and other properties. We select the reference object s_0 and its characteristics Q_{0j}, and s_0 can be either an element of the set S or not, it depends on the specific conditions of the problem and the preferences of the research or expert opinion. The characteristics of Q_{0j} are selected in such a way that s_0 is optimal in terms of the requirements for the quality of the object.

It is required to rank the elements of the set S according to the given m characteristics for compliance with the standard.

For each characteristic Q_{ij} we construct a FS, $\hat{Q}_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. To do this, we define the variables x_j with the range of values of G_j. Let the values of the variable x_j, satisfying the characteristic Q_{ij}, make up the segment X_{ij}. Then

$$X_{ij} = A(Q_{ij}).$$

We select the membership function μ_{ij} for the criterion Q_{ij}. Based on the construction of the set X_{ij} we find that in the q_{ij}, function has a maximum point, within the set X_{ij} the membership function takes values greater than 0.5, and outside X_{ij} it is less than:

$$\mu_{ij}: G_j \to [0; 1],$$

$$\mu_{ij}(q_{ij}) = 1,$$

$$\mu_{ij}(x_j) \geq 0.5 \Leftrightarrow x_j \in X_{ij}.$$

As a membership function, you can, for example, use the following functions:

$$\mu_{ij}(x_j) = \exp\left(-\frac{\ln 2}{\delta_{ij}^2}(x_j - q_{ij})^2\right), i = 0, \ldots, n, j = 1, \ldots, m.$$

As a result, we get FSs

$$\hat{Q}_{ij} = \{x_j | \mu_{ij}(x_j)\}, i = 0, \ldots, n, j = 1, \ldots, m.$$

We'll calculate the degree of equality ν_{ij} of the corresponding FS [26] to determine to what extent the characteristic of the object si is close to the characteristic of the reference object s_0:

$$\nu_{ij} = \max_{G_j} \min(\mu_{ij}(x_j), \mu_{0j}(x_j)), —$$

and the value 1 will correspond to absolute equality, and 0 to absolute inequality.

Further using the aggregate function (as such, we can take, for example, weighted voting), we'll obtain an integral estimate ν_i of the correspondence of the set of object characteristics s_i of the set of object characteristics s_0.

We'll calculate the degree of equality ν_{ij} of the corresponding FSs [32], using this method in our experiment, in order to determine to what extent the characteristics of flame retardant no. i is close to the characteristic reference no. 0:

$$\nu_{ij} = \max_{\mathbb{R}} \min(u_{ij}(x_j), u_{0j}(x_j)).$$

For the membership functions of the reduced form [33], we have obtained that the maximin value is reached at the point:

$$\nu_{ij} = u_{0j}(x_{ij}^*),$$

where $x_{ij}^* = \left(q_{ij}\delta_{0j} + q_{0j}\delta_{ij}\right)/\left(\delta_{0j} + \delta_{ij}\right)$, $i = 1, \ldots, n, j = 1, \ldots, m$.

For example,

$$x_{11}^* = \left(q_{11}\delta_{01} + q_{01}\delta_{11}\right)/\left(\delta_{01} + \delta_{11}\right) =$$

$$= \left(19 \cdot 8 + 1 \cdot 1\right)/\left(8 + 1\right) = 17,$$

$$\nu_{11} = u_{01}(x_{11}^*) = \exp\left(-\frac{\ln 2}{\delta_{01}^2}(x_{11}^* - q_{01})^2\right) =$$

$$= \exp\left(-\frac{\ln 2}{14^2}(16.867 - 1)^2\right) \approx 0.0625,$$

similarly, $\nu_{12} \approx 0.349$, $\nu_{13} \approx 0.438$.

Having conducted a weighted vote, we'll obtain an integral assessment ν_i of the correspondence of the set of object characteristics s_i to the set of object characteristics s_0:

$$\nu_i = \sum_{j=1}^{m} \alpha_j \nu_{ij},$$

where $\alpha_j \geq 0$, $\sum_{j=1}^{m} \alpha_j = 1$. The weights α_j are selected in accordance with the level of importance of each of the characteristics and are determined by the expert in the subject area so that the PC fully meets consumer requirements, technological process, etc. It is accepted here $\alpha_1 = \alpha_2 = 0.33$, $\alpha_3 = = 0.34$. The results of calculations based on the data Table 2 are presented in Table 4.

Table 4: Evaluation of Fire Retardants to Compliance with Operational Requirements

Fire retardants no., *i*	ν_i
1	*0.28*
2	*0.33*
3	*0.60*
4	*0.76*
5	*0.51*
6	*0.81*
7	*0.93*
8	*0.83*

Thus, the no. 7 fire retardant is preferable from the point of view of the presented requirements.

The proposed computational model can form the basis for automated data analysis when choosing the optimal CTS, including with the participation of experts. The approximate route of operation of such an IS is shown in Figure 2. Here, the requirements for the evaluated objects (PCM) are set directly in the IS interface when formulating the initial task, allowing specialists to quickly adjust its formulation in the context of the research.

When generating generic CTS evaluation tasks, it is possible based on known experimental data to form a library (database) of expert assessments (along with the data model library), which will relieve the user of the IS from having to obtain or collect expert assessments and fully automate the evaluation process of a specific CTS.

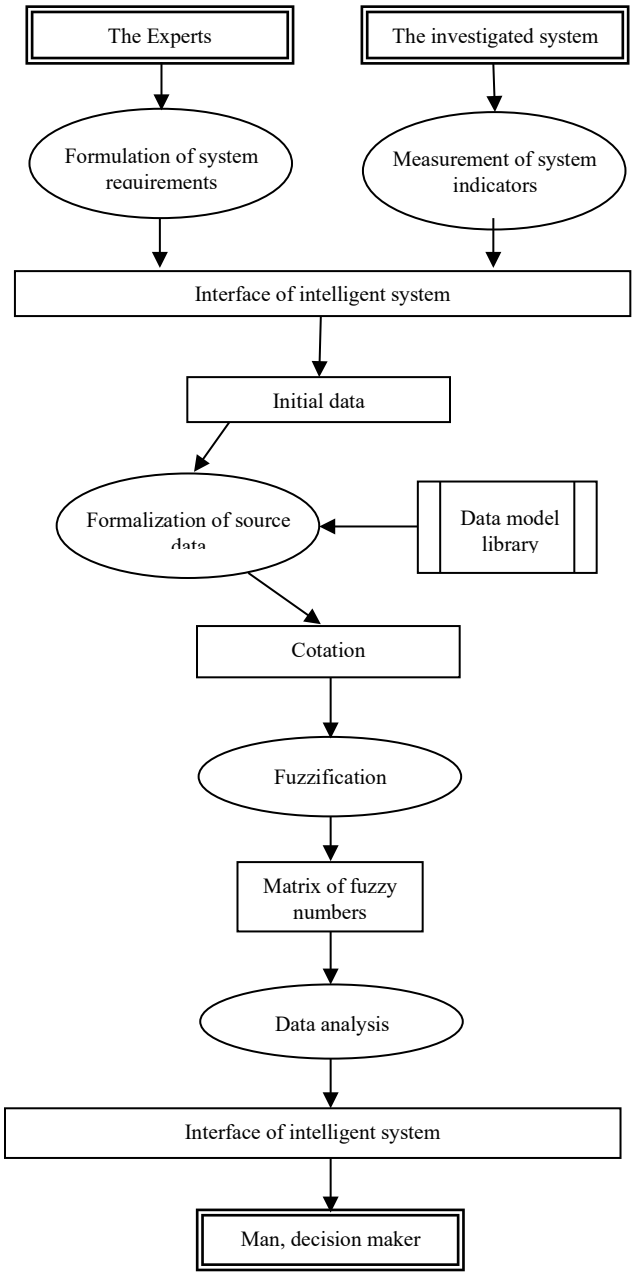


Figure 2: Scheme of the Information Flow in the IS

4. The optimal choice of a single additive for the polymer composition

4.1. Method Description

Let p_k, $k = 1, \ldots, l$ — be polymeric matrices, s_{ik}, $i = 1, \ldots, n$ — the additives for p_k. As mentioned above, by Q_{0j} we denote the value of the j-th property, which the PC should possess.

After modifying the polymer p_k by adding s_{ik} its properties become equal Q_{ikj}, $i = 1, \ldots, n$, $k = 1, \ldots, l$, $j = 1, \ldots, m$.

Further we get fuzzy numbers

$$u_{ikj} \text{ based on } X_{ikj} = A(Q_{ikj}),$$
$$u_{0j} \text{ based on } X_{0j} = A(Q_{0j}), \quad (1)$$

where $i = 1, \ldots, n$, $k = 1, \ldots, l$, $j = 1, \ldots, m$.

To select one additive and one polymer matrix we calculate

$$\nu_{ikj} = u_{0j}(x^*_{ikj}), \quad (2)$$

where $x^*_{ikj} = \left(q_{ikj}\delta_{0j} + q_{0j}\delta_{ikj}\right) / \left(\delta_{0j} + \delta_{ikj}\right)$, $i = 1, \ldots, n$, $k = 1, \ldots, l$, $j = 1, \ldots, m$ and

$$\nu_{ik} = \sum_{j=1}^{m} \alpha_j \nu_{ikj} . \quad (3)$$

In the end, choose the additive $s_{i_0 k_0}$ and polymer matrix p_{k_0}, where $(i_0, k_0) = \arg \max_{\substack{(i,k),\\ i=1,\ldots,n,\\ k=1,\ldots,l}} \nu_{ik}$.

4.2. Computational experiment

We present a study of PC with alternative polymer matrices of polyurethane elastomer and polycaproamide and use a set of fire retardants presented in Table 5, for an extended demonstration of the capabilities of the fuzzy data analysis method. The number of flame retardants for simplicity of notation is the same as in the Table 2, but this is not essential for the method and, in general, the amount of additives for each polymer matrix can be different. In addition, the reference value, that is, additive number 0, is left the same as in Table 2. This is essential for the considered mathematical method of PC analysis.

Table 5: Background Characteristics of Fire Retardants for Elastomer Matrix [34]

Compound no., i	Q_{i1}	Q_{i2}	Q_{i3}
0 (Reference, q/δ)	*1.0/8*	*1.5/8.5*	*18/10*
1	*12.7*	*17.1*	*20*
2	*8.5*	*8.5*	*18*
3	*20.6*	*20.7*	*20*
4	*18.7*	*18.7*	*17*
5	*6.2*	*6.2*	*16*
6	*18.0*	*22.1*	*17*
7	*28.1*	*27.0*	*20*
8	*4.4*	*4.4*	*10*

Let p_1 be polycaproamide, and p_2 — polyurethane elastomer, data on additives to which are given in Table 2 and Table 5 respectively. After the fuzzification of the initial data using (1), we obtain the fuzzy numbers u_{i1j} (Table 3) and u_{i2j} (Table 6) for p_1 and p_2, respectively.

Table 6: Parameters of Elements of the Matrix $U(x)$ Calculated on the Data of Table 5

Compound no., i	q_{i1}/δ_{i1}	q_{i2}/δ_{i2}	q_{i3}/δ_{i3}
0 (Reference, q/δ)	*1.0/8*	*1.5/8.5*	*18/10*
1	*12.7/0.1*	*17.1/0.1*	*20/1*
2	*8.5/0.1*	*8.5/0.1*	*18/1*
3	*20.6/0.1*	*20.7/0.1*	*20/1*
4	*18.7/0.1*	*18.7/0.1*	*17/1*
5	*6.2/0.1*	*6.2/0.1*	*16/1*
6	*18.0/0.1*	*22.1/0.1*	*17/1*
7	*28.1/0.1*	*27.0/0.1*	*20/1*
8	*4.4/0.1*	*4.4/0.1*	*10/1*

Applying to the data Table 3 ($k = 1$) and Table 6 ($k = 2$) formula (2) we obtain the values ν_{ikj}, $i = 1, \ldots, 8$, $k = 1, 2$, $j = 1, 2, 3$. Next, using (3), we obtain ν_{ik}, $i = 1, \ldots, 8$, $k = 1, 2$, are given in Table 4 for $k = 1$ and in Table 7 for $k = 2$.

Table 7: Evaluation of Elastomer Fire Retardants in Compliance with Operational Requirements

Fire retardants no., i	ν_{i2}
1	*0.45*
2	*0.71*
3	*0.35*
4	*0.34*
5	*0.81*
6	*0.33*
7	*0.34*
8	*0.77*

Among the data Table 4 and Table 7 find the maximum value of *0.93*, which corresponds to the addition of s_{71} and the polymer matrix p_1.

Thus, a specific recommendation was received on the flame retardant, which is part of the PC. Extensive work is needed to formulate a recipe based on these recommendations. This, for example, the definition of other components or their part in the PC, which allows you to take the first step in developing a new PCM in an automated way. Here we note that the proposed analytical method also has the possibility of choosing additional components for PC design. This is covered below.

Above (see Table 1) we have already indicated that thanks to the growth and expansion of consumer requirements, the complexity of polymer systems are increasing simbatically. Therefore, identification tasks are also becoming more complex.

Consider the complicated task of choosing several ingredients (additives) for PC design.

5. Selection of several additives for PC

5.1. Method description

Let it be necessary to choose one polymer matrix from the set p_k, $k = 1, \ldots, l$ and $r \le n$ additives for it from s_{ik}, $i = 1, \ldots, n$.

To do this, we define r properties $Q_{j_1}, \ldots, Q_{j_r}$, to improve which the choice of r additives is directed.

For each polymer matrix p_k, we define a set of additives $s_{i_s k}$ ($i_s \in \{1, \ldots, n\}$, $s = 1, \ldots, r$, $k = 1, \ldots, l$) such that the additive $s_{i_s k}$ provides the best property value Q_{j_s} among the additives $s_{1k}, \ldots, s_{nk}$. However, some numbers may coincide with each other if the same additive provides the best values for several properties at once. In this case, for the modification, not r, but a correspondingly smaller amount of additives is used, but the result will correspond to the set task — the same r properties are improved.

Next, we obtain ν_{ik}, $i = 1, \ldots, n$, $k = 1, \ldots, l$ as it was described in Section 4 and define

$$\nu_k = \sum_{s=1}^{r} \alpha_{j_s} \nu_{i_s k} . \quad (4)$$

Determine the number

$$k_0 = \arg \max_{k=1,\ldots,l} \nu_k$$

and choose a modification of the polymer matrix p_{k_0} a set of additives $s_{i_1 k_0}, \ldots, s_{i_r k_0}$.

5.2. Computational experiment

To demonstrate the capabilities of the method presented here, we will use the same data (Table 2 and 5) and the same notation as above. Since the formulas for calculating the same, the values of ν_{ik} are also given in Table 4 and 7.

Table 8: Values of the Degree of Compliance of the Characteristics of Additives to Consumer Requirements to PCM

Compound no., i	$k = 1$			$k = 2$		
	ν_{i11}	ν_{i12}	ν_{i13}	ν_{i21}	ν_{i22}	ν_{i23}
1	*0.0625*	*0.349*	*0.438*	*0.235*	*0.102*	*0.994*
2	*0.0843*	*0.452*	*0.438*	*0.552*	*0,.632*	*0.950*
3	*0.0843*	*0.956*	*0.755*	*0.0173*	*0.0316*	*0.994*
4	*0.712*	*0.767*	*0.814*	*0.0365*	*0.0625*	*0.912*
5	*0.0843*	*0.998*	*0.438*	*0.752*	*0.813*	*0.867*
6	*0.966*	*0.979*	*0.480*	*0.0472*	*0.0187*	*0.912*
7	*1.0*	*0.987*	*0.814*	*0.000427*	*0.00226*	*0.994*
8	*1.0*	*0.999*	*0.5*	*0.885*	*0.924*	*0.5*

Let $r = 2$, and $j_1 = 1$, $j_2 = 2$. To determine i_1 and i_2, ν_{ikj} values are needed that were not of immediate interest previously and were omitted from the examples of calculations. Now we bring these data to Table 8

To find i_1 for $k = 1$, we find the largest among the ν_{i11} values. This is *1*. This value is achieved for two supplements. We take $i_1 = 7$. Similarly, we get $i_2 = 8$. Acting in the same way, for $k = 2$ we define $i_1 = 8$, $i_2 = 8$.

Next, we calculate by (4)

$$\nu_1 = \alpha_1 \nu_{71} + \alpha_2 \nu_{81} = 0.33 \cdot 0.93 + 0.33 \cdot 0.83 \approx 0.58,$$

$$\nu_2 = \alpha_1 \nu_{82} + \alpha_2 \nu_{82} = 0.33 \cdot 0.77 + 0.33 \cdot 0.77 \approx 0.51.$$

The value $\nu_1 \approx 0.58$ indicates that for $k = 1$, the PC has a higher degree of compliance with the requirements. As a result, we choose a PC based on textile fiber ($k = 1$) with additives no. 7 and 8 (Table 2). Thus, a specific recommendation on the main additional component of the PC is also received here. The formation of a complete recipe based on these recommendations has the peculiarity that the proposed composition is already a multi-component system, but still has an incomplete character and requires more design and experimental work to implement the full "task-practice" algorithm. This, as well as noted above, can be the definition of additional components, their concentration in PC, technological and operating conditions. Calculations should be considered as a quantitative justification for the decision to develop a new PC. With the development of these approaches it is possible to create an automated system.

6. Conclusion

Summarizing the above, we note that often the CTS design problem arising at the initial stages is the task of choosing the best or, in a certain sense, optimal variant from the set of admissible ones. They must meet the specified requirements on the "composition – property – application" platform. When you try to get a solution in terms of incomplete and fuzzy information about the object of design, there is a high probability of incorrect or inaccurate multiple solutions. Therefore, the use of a clear strategy can minimize this probability, also reducing the cost of design and practical implementation of a technical solution.

The approach proposed in this article to mathematical modeling using fuzzy information in the field of chemical technology is essentially the mathematical support of a profiled IS. The formation of a database of models within the system allows us to bring the linguistic resources of the user interface to a new level of interaction using language environments more similar to natural languages and setting the input data in a form that is quite familiar to the researcher.

We should also note that only one of the most important aspects of using fuzzy simulation of PCM is shown here. It is aimed at using information in the form of fuzzy numbers when choosing PC components from a uniform set. As can be seen from the above results, the analysis of experimental data is well algorithmized and can be implemented as an IS of a chemist-researcher and a chemist-technologist. The above mathematical calculations are illustrated by the example of PCM design in general methodical terms, so they can be transferred to other natural and engineering systems in which multicomponentity is the main feature.

The development of the proposed computational methods creates theoretical foundations for the formalization and automation of the process of quality management of technical systems. In this case, it is possible to attract highly qualified experts only once to customize methods for solving specific tasks. The necessary step in this case is the formation of a library of settings and the solution of basic ("zero") tasks. This will further allow you to work offline.

At the same time, the expert has several roles in stock:

- to identify the links between informal entities of the subject domain and formal entities of the theory of FS;
- the formation of an optimal set of requirements that a quality product must meet;
- applied to the PCM — selection of the PC composition according to the selected requirements.

This information will further allow to formalize the initial data of the subject area in terms of the theory of FS and use them in calculations, when upon receipt of an order from the consumer a set of requirements for PCM is automatically generated, and then estimates of formed requirements for specific candidates for PC components are calculated. Then a decision is made on the technical implementation of the composition, which ensures high quality of PCM. The analysis given here is given as a possible prospect for the development of the applied use of fuzzy mathematics tools in engineering. The stated essence of the method shows its role and value in solving the problem of quality management of PCM and other technical systems.

At the same time, we selectively point out the nearest additional applied work-analogues and areas in which methods for analyzing fuzzy information are being developed. These are economics [35], management systems [27], analysis of leasing projects [36], traffic optimization [37], occupational studies [38]. As a prospect, it can be viewed as technical compositions of medicinal and cosmetic forms, building and drilling mixtures and solutions, alloys, and some colloidal systems. As a nontrivial and nontechnical one, it is possible to propose to consider personnel recruitment tasks, for example, for diplomatic or other complex, for example, artistic work, from a certain number of applicants with a quantitative assessment of personal data and interview results. The basis for this is [39, 40]. True, the fully successful outcome of the future work of candidates cannot be guaranteed — one thing is a technique, the other is people-specialists. The risk factor, as in polymer and other systems, remains, but it is much higher.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The reported research was funded by Russian Foundation for Basic Research and the government of Volgograd region, grant № 18-48-340011 "Quality management of polymer products based on the optimization of formulation of composite materials"

References

[1] I. V. Germashev, T. P. Mashihina, V. E. Derbisher, "Analysis of Fuzzy Data for Polymer Production Quality Management" in 2018 International Russian Automation Conference, RusAutoCon, Sochi Russia, 2018. https://doi.org/10.1109/RUSAUTOCON.2018.8501653

[2] J.-o. Kim, C. W. Mueller, Factor Analysis: Statistical Methods and Practical Issues, SAGE Publications, 1978. https://dx.doi.org/10.4135/9781412984256.

[3] Y. S. Vygodsky, T. V. Volkova, O. N. Zabegaeva, Z. Y. Chistyakova, V. A. Shanditsev, M. I. Buzin, Y. V. Zubavichus, O.V. Sinitsyna, G. G. Nikiforova, A. P. Krasnov et al., "Synthesis and characteristics of the composites based on poly(caproamide) and multiwalled carbon nanotubes" Polym. Sci. Ser. C+, 51(1), 63–73, 2009. https://doi.org/10.1134/S1811238209010111

[4] P. D. Shestakov, A. V. Galtsev, A. S. Tarasov, "Hardening Polymer Composites with Carbon Nanotubes" International Scientific Research Journal, 6(6), 79–81, 2012.

[5] M. G. Svistunov, V. E. Golovin, K. K. Koshelev, G. A. Kosheleva, Y. N. Nesterov, O. K. Barashkov, E`lektroizolyatsionnaya polimernaya kompozitsiya, Patent RU 2636383, 2017.

[6] W. R. Klecka, Discriminant Analysis, SAGE Publications, 1982. https://dx.doi.org/10.4135/9781412983938.

[7] M. S. Aldenderfer, R. K. Blashfield, Cluster Analysis, SAGE Publications, 1984. https://dx.doi.org/10.4135/9781412983648.

[8] A. Faure, Perception et Reconnaissance des Formes, Nombreux schémas N&B, 1985.

[9] A. G. Blokhnin, "Fuzzy-set algebra" J. Comput. Sys. Sc. Int+, 37(5), 747–754, 1998.

[10] M. G. Voskoglou, I. Y. Subbotin, "Application of the triangular fuzzy model to assessment of analogical reasoning skills" American Journal of Applied Mathematics and Statistics, 3(1), 1–6, 2015. https://doi.org/10.12691/ajams-3-1-1.

[11] D. Silva, L. F. V. Pinto, D. Bozukova, L. F. Santos, A. P. Serro, B. Saramago, "Chitosan/alginate based multilayers to control drug release from ophthalmic lens" Colloid Surface B, 147, 81–89, 2016. https://doi.org/10.1016/j.colsurfb.2016.07.047.

[12] T. Ikeda, H. Tamura, T. Sakurai, S. Seki, "Control of optical and electrical properties of nanosheets by the chemical structure of the turning point in a foldable polymer" Nanoscale, 8(30), 14673–14681, 2016. https://doi.org/10.1039/c6nr01066k.

[13] A. Kobayashi, T. Nakaza, T. Hirano, S. Kitagawa, H. Ohtani, "Variation in the chromatographic, material, and chemical characteristics of methacrylate-based polymer monoliths during photoinitiated low-temperature polymerization" J. Sep. Sci., 39(13), 2459–2465, 2016. https://doi.org/10.1002/jssc.201600344.

[14] B. Strachota, J. Hodan, L. Matejka, "Poly(N-isopropylacrylamide)-clay hydrogels: Control of mechanical properties and structure by the initiating conditions of polymerization" Eur. Polym. J., 77, 1–15, 2016. https://doi.org/10.1016/j.eurpolymj.2016.02.011.

[15] L. Y. T. Chou, F. Song, W. C. W. Chan, "Engineering the structure and properties of DNA-nanoparticle superstructures using polyvalent counterions" J. Am. Chem. Soc., 138(13), 4565–4572, 2016. https://doi.org/10.1021/jacs.6b00751.

[16] I. V. Germashev, V. E. Derbisher, T. M. Kokorina, I. A. Murashkina, "Assessing the quality of knitted fabrics using the theory of inexact sets" Izvestiya Vysshikh Uchebnykh Zavedenii, Seriya Teknologiya Tekstil'noi Promyshlennosti, (4), 7–10, 1999 (article in Russian with an abstract in English).

[17] V. E. Derbisher, I. V. Germashev, G. G. Bodrova, "Fuzzy-set-based quantitative estimates of the efficiency of thermo- and photostabilizing additives in polymeric compositions" Polym. Sci. Ser. A+, 39(6), 630–633, 1997.

[18] B. A. Khidhir, W. Al-Oqaiel, P. M. Kareem, "Prediction models by response surface methodology for turning operation" American Journal of Modeling and Optimization, 3(1), 1–6, 2015. https://doi.org/10.12691/ajmo-3-1-1.

[19] M. R. S. Emami, "Fuzzy logic applications in chemical processes" The J. of Math. and Comput. Sci., 1(4), 339–348, 2010. https://doi.org/10.22436/jmcs.001.04.11.

[20] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction" Expert Syst. Appl., 42(7), 3751–3759, 2015. https://doi.org/10.1016/j.eswa.2014.12.044.

[21] K. Nasrollahzadeh, M. M. Basiri, "Prediction of shear strength of FRP reinforced concrete beams using fuzzy inference system" Expert Syst. Appl., 41(4), 1006–1020, 2014. https://doi.org/10.1016/j.eswa.2013.07.045.

[22] A. Salski, F. Bartels, "A fuzzy approach to land evaluation" IASME Trans., 5(2), 774–780, 2005.

[23] M. I. Kizhisseri, M. M. A. Mohamed, "Fuzzy-based wastewater quality indices for pollution classification: a case study in the United Arab Emirates" Environment Systems and Decisions, 36(1), 62–71, 2016. https://doi.org/10.1007/s10669-015-9579-9.

[24] P. H. Chandra, S. M. S. T. Kalavathy, A. M. I. Jayaseeli, J. P. Karoline, "Mechanism of fuzzy ARMS on chemical reaction" in V. Snášel, A. Abraham, P. Krömer, M. Pant, A. Muda (eds) Innovations in Bio-Inspired Computing

and Applications. Advances in Intelligent Systems and Computing, 424, 43–53, 2016. https://dx.doi.org/10.1007/978-3-319-28031-8_4.
[25] I. V. Germashev, V. E. Derbisher, T. F. Morozenko, S. A. Orlova, "Otsenka kachestva tehnicheskih ob``ektov s ispol`zovaniem nechyotkih mnozhestv" Industrial laboratory. Diagnostics of materials, 67(1), 65–68, 2001.
[26] I. V. Germashev, V. E. Derbisher, "Fuzzy Optimization of Polymer Compositions" Theor. Found. Chem. Eng., 35(4), 418–421, 2001.
[27] A. A. Uskov, I. A. Kiselev, "Comparing fuzzy arithmetic with quaternion arithmetic and applying the former to control system analysis" Large-scale Systems Control, (48), 59–70, 2014 (article in Russian with an abstract in English).
[28] A. I. Orlov, E. V. Lutsenko, Sistemnaya Nechyotkaya Interval`naya Matematika, KubGAU, 2014.
[29] A. I. Orlov, Nechislovaya Statistika, MZ-Press, 2004.
[30] E. Egrioglu, C. H. Aladag, M. A. Basaran, U. Yolcu, V. R. Uslu, "A new approach based on the optimization of the length of intervals in fuzzy time series" J. Intell. Fuzzy Syst., 22(1), 15–19, 2011.
[31] I. V. Germashev, V. E. Derbisher, E. V. Derbisher, E. A. Markushevskaya, "The Fuzzy Data Analysis for the Chemical-Technological Systems Optimization" Information Technologies, 24(3), 153–160, 2018 (article in Russian with an abstract in English).
[32] V. P. Meshalkin, E`kspertny`e Sistemy` v Himicheskoy Tehnologii, Chemistry, 1995.
[33] I. V. Germashev, V. E. Derbisher, "Properties of unimodal membership functions in operations with fuzzy sets" Russian Mathematics, 51(3), 72–75, 2007. https://doi.org/10.3103/S1066369X07030115.
[34] I. V. Germashev, V. E. Derbisher, S. A. Orlova, "Evaluation of activity of the fireproofing compounds in elastomer compositions by means of fuzzy sets" Kauchuk i Rezina, (6), 15–17, 2001 (article in Russian with an abstract in English).
[35] Y. I. Davydov, "Primenenie teorii nechyotkih mnozhestv k e`konomicheskim zadacham" News of St. Petersburg State Agrarian University, (26), 261–266, 2012.
[36] Y. I. Davydov, "Teoriya nechyotkih mnozhestv v otsenke e`konomicheskoy e`ffektivnosti lizingovy`h proektov v usloviyax neopredelennosti i riska" News of the International Academy of Agrarian Education, (14-2), 32–39, 2012.
[37] V. A. Romanenko, "Fuzzy optimizing of transfer air transportation system parameters" Bulletin of the Samara State Technical University, Series: Technical Sciences, (4), 73–80, 2012 (article in Russian with an abstract in English).
[38] G. A. Kryzhanovsky, V. V. Kupin, "Analiz vazhnosti professional`ny`h zadach na osnove nechetkih chislovy`h e`kspertny`h otsenok" Bulletin of St. Petersburg State University of Civil Aviation, (2), 43–50, 2012.
[39] S. V. Skorokhod, "Otbor personala v usloviyah nechyotkoy informatsii na osnove e`kspertny`h otsenok" News of SFU. Engineering Science, (9), 125–130, 2008.
[40] S. V. Skorokhod, "Primenenie nechyotkih chisel dlya otsenki kvalifikatsii personala" News of TRTU, (3), 214–216, 2005.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 2, 183-189 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Effects of Educational Support Robots using Sympathy Expression Method with Body Movement and Facial Expression on the Learners in Short- and Long-term Experiments

Yuhei Tanizaki[*,1], Felix Jimenez[2], Tomohiro Yoshikawa[1], Takeshi Furuhashi[1], Masayoshi Kanoh[3]

[1]*Graduate School of Engineering, Nagoya University, 464-8603, Japan*

[2]*Information Science and Technology, Aichi Prefectural University, 480-1198, Japan*

[3]*Engineering, Chukyo University, 466-8666, Japan*

ARTICLEINFO

Article history:
Received: 19 January, 2019
Accepted: 08 March, 2019
Online: 28 March, 2019

Keywords:
Educational-Support Robot
Sympathy Expression Method
Body Motion

ABSTRACT

Recently, educational-support robots have been attracting increasing attention as studying-support gadgets. Previous studies used the sympathy expression method in which the robot expressed emotions in sympathy with the learners; however, the robots considered in those studies expressed only facial emotions. Presently, there is no study that uses body movements together with facial expressions in the sympathy expression method. Thus, in this paper, we examine the effects of two types of robots that have different method of expressing emotions on learners in two experiments.

1 Introduction

Recent developments in robotics have prompted an increase in educational-support robots that assist in studying. For instance, a robot supports students at school [1] or helps them to learn the English language[2]. Koizumi introduced the robot as a "watching over" in situations where children learn while discussing the programming that controls the assembly and motion of the car robot by the Lego block [3]. Our research focus on educational support robots to learn with people. The serious problem that these robots face is that the learner gets tired of the robot because they find the behavior of the robots to be monotonous [4], [5]. To address this problem, previous researches developed a sympathy expression method that allows a robot to express emotions sympathetic to learners [6]. Through the subject experiments, the previous research reported that the robot using empathy expression method reduced the monotonicity of learning and reduced tiredness. However, sympathy expression method has been studied only for a robot that expresses emotions by changing facial expressions. In the area of human-robot interactions, conventional research stated that body motions of robots are useful in the interactions between humans and robots as well as in expressing emotions [7]. Thus, sympathy expression method using both facial expression change and body motion may be better than one based only on facial expression change. This paper reported the Experiment1 to measure the short-term impression and the Experiment2 to measure studying effects and long-term impression.

2 Sympathy expression method

The sympathy expression method is based on Russell's circumplex model of affect (shown in Fig.1). This method can be used to express emotions in the circumplex model using the correct answer vector $\vec{A}$ and the incorrect answer vector $\vec{B}$; thus, the learners believe that the robot sympathizes with them. If learner correctly answers the question, the robot expresses emotions using $\vec{A}$; in contrast, the robot expresses emotions using $\vec{B}$ for incorrect answer. $\vec{A}$ moves in the area $0 \leq L_A \leq 1.0$ and $-90° \leq \theta_A \leq 90°$. $\vec{B}$ moves in the area $-1.0 \leq L_B \leq 0$ and $-90° \leq \theta_B \leq 90°$. $L\cos\theta$ refers to the axis of "Pleasure-Displeasure." $L\sin\theta$ refers to the axis of

[*]Yuhei Tanizaki, Room #309, Building No.3, Furo-cho, Chikusa-ku, Nagoya, Aichi, +81-52-789-2793 & tanizaki@cmplx.cse.nagoya-u.ac.jp

https://dx.doi.org/10.25046/aj040224

"Arousing-Sleep." The emotion vectors move as follows [6].

if (learners solve the problem correctly)
$L_A \leftarrow L_A + 0.2$
$L_B \leftarrow L_B + 0.2$
else
$L_A \leftarrow L_A - 0.2$
$L_B \leftarrow L_B - 0.2$
if (the time of answer $<$ reference time)

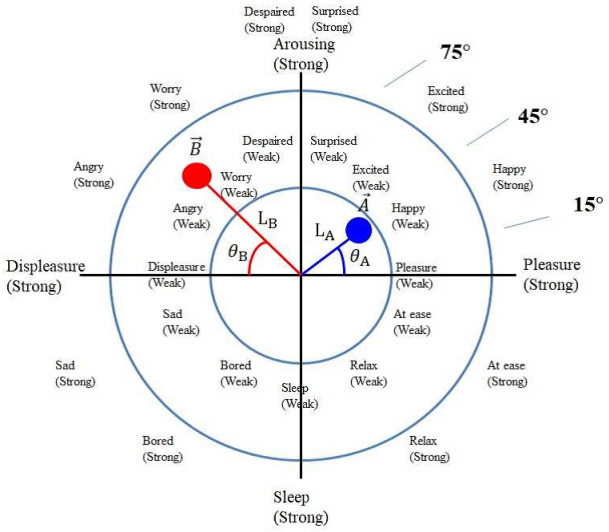


Figure 1: Sympathy expression method

if (learners solve the problem correctly)
$\theta_A \leftarrow \theta_A + 15$
else
$\theta_B \leftarrow \theta_B + 15$
else
if (learners solve the problem correctly)
$\theta_A \leftarrow \theta_A - 15$
else
$\theta_B \leftarrow \theta_B - 15$

In these experiments, the reference time is decided by the answer time in the earlier question. In the 1st question of studying, reference time is decided by the average answer time before studying. However, in first studying, reference time is 60. All robots utilized in these experiments were assumed to have the same sympathy expression method to enable a fair comparison.

3 Robot

3.1 Overview

In this study, we used a tablet robot ,"Tabot", whose head consists of tablets. as shown in Fig. 2. Tabot can express a lot of facial expressions by expressing the agent on the screen of the tablet.. It has 14 degrees of freedom: the neck has 3 degrees of freedom, the arms have 10 degrees of freedom, and the legs have 1 degree of freedom. So, it can do many body motions. It can express many emotions by combination of body motions and facial expressions. In these experiments, 2 Tabots were utilized for reducing the effects of the shape of robot.



Figure 2: The Tabot used in the experiment

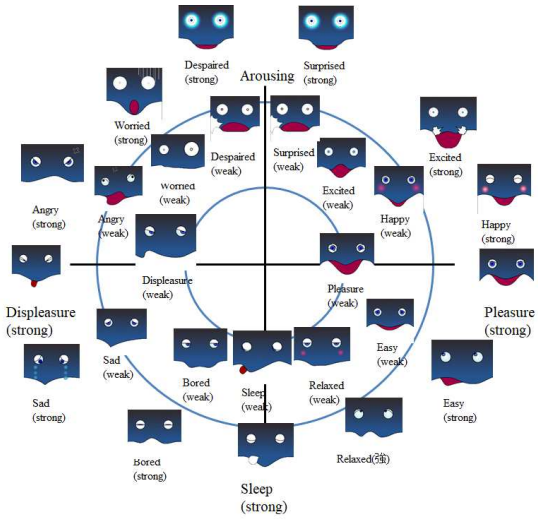


Figure 3: Facial expression

3.2 Facial expressions and body motions

The face displayed in its tablet and the changes in Tabot's facial expressions were developed based on a specific design. Two types of facial expressions were used to express the emotions corresponding to each emotion in the circumplex model. Fig. 3 shows example of facial expression of the Tabot. Conversely, Tabot's body motions were created on the basis of [8] and [9], which discussed the relation between human emotion and physical activity. Two types of body motions can be used to express each emotion and the corresponding emotions in the circumplex model. The body motions and facial expressions that expressed the same emotions were combined. We implemented

facial expression as well as a combination of facial expressions and body motions such that emotions can be communicated precisely to the learners. Fig. 4 shows example of combination of facial expression and body motion of the Tabot.

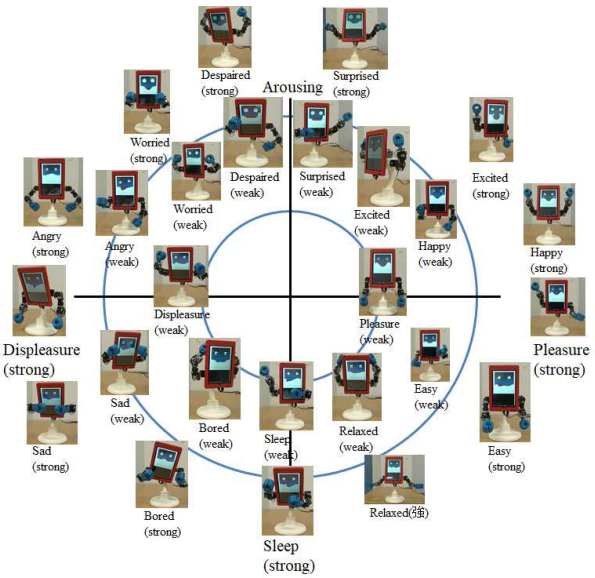


Figure 4: Combination of facial expression and body motion

4 Experiment 1

4.1 Method

The subjects learned with the robot comprising the sympathy expression method and a studying system. When a subject answered a question, the studying system displayed a correct or a incorrect response. Subjects learned by answering the questions displayed in the studying system. 4 university students and 8 graduate students in the science stream were gathered for this experiment. The participants had never used a Tabot previously. The twelve learners used two types of robots. The first type of robot was called "Robot-Facial-Group", which expressed emotions using only facial expressions. The second type of robot was called "Robot-Combination-Group", which used both facial expressions and body motions. The robots were used in a random manner. The subjects answered 20 questions in each group.

The upper side of Tabot's tablet displayed facial expressions. The lower side of the robot's tablet displayed studying systems. Tabot expressed emotions when the studying system gave either a correct or a incorrect response.

4.2 Overview of the studying system

We used the studying system (Fig. 5) for math questions called "Synthetic Personality Inventory 2 (SPI2)." SPI2 is used as a recruitment test for employment. Moreover, it uses junior high school level math questions. Therefore, university students do not require knowledge for SPI2.

Subjects first log in with an account number. Fig. 5(a) shows a menu of the study contents (i.e.,math questions). The column used for choosing the number of questions is shown under the study items. When a learner selected "20," 20 questions were shown randomly. Subsequently, repeating the same selection would display different twenty questions that could be repeated until all 100 questions are displayed. Thus, it enables learners to answer all the questions within the chosen study item. When a subject selected a study content and the number of questions, the studying screen (Fig. 5(b)) appeared and the studying process began. The subject answered the question from the selection list. Subsequently, the system displayed whether it was the correct answer as shown in Fig. 5(c). When the subject selected "Next" as shown in Fig. 5(c), the system moved on to the next question. When the subject selected "Result" as shown in Fig. 5(c), or solved all the questions, the system moved on to the results page (Fig. 5(d)), which presented the number of correct answers. When the subject selected "Study again," the menu of the studying items was displayed (Fig. 5(a)). When the subject selected "Study incorrect answers," the study page presented questions that were answered incorrectly (Fig. 5(b)).

4.3 Evaluation criterion

Impression evaluation used the rating scale method, which is a quantitative evaluation of impressions as shown in Fig.6. The rating scale method has 14 questions; it uses a range of values, e.g, "Friendly - Friendless," "Emotional - Intelligent," "Be pleased with me - Be not pleased with me," and "Enjoying studying - Not enjoying studying". The scores of the rating scale method range from 1 to 7. In Fig. 6, the left score is 7, and the right score is 1. In each group, subjects answered this questionnaire at the end of studying. Furthermore, subjects answered a questionnaire on the general comments about impressions. We defined the sum of scores of all questions as the score of good impressions. We conducted paired t-test for impression evaluation.

4.4 Result

Fig. 7 indicates the average impression score in two groups. This graph indicates the score of the Robot-Combination-Group is better than that of the Robot-Facial-Group. The result of paired t-test showed that the good impression score of the Robot-Combination-Group is significantly different from that of the Robot-Facial-Group. These results show that the robot that expresses emotions using both body motions and facial expressions can give learners a better impression than the one that expresses emotions using facial expressions only.

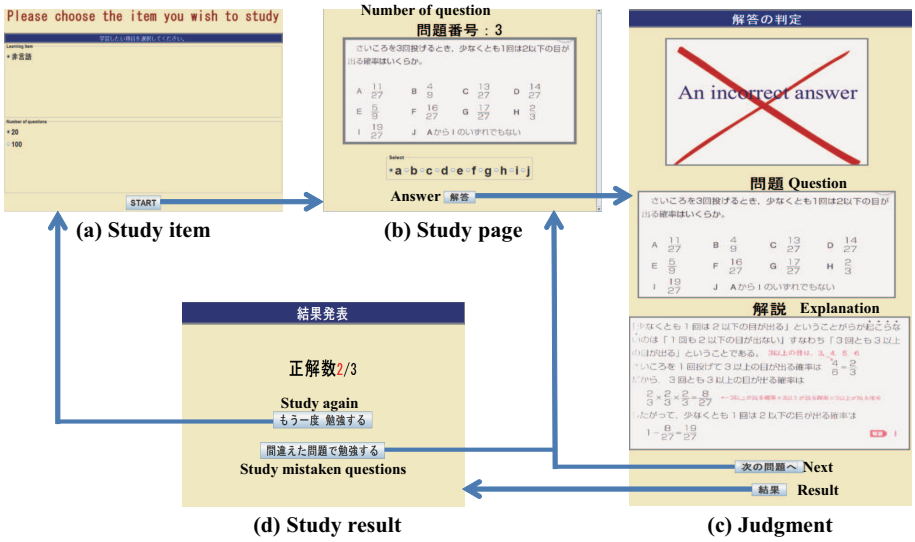


Figure 5: studying system

4.5 Discussion

The result of the experiment indicate that the sympathy expression method that expresses emotions using both body motions and facial expressions can give learners a better impression than a method that expresses emotions using facial expressions only. In the general comments column, somebody stated, "It was difficult for me to look at the robot's facial expressions." This implies that it is possible for some learners to not notice that the facial expression of the robot was changing because they concentrated on the questions of the studying system. Therefore, the impression made by the Robot-Combination-Group was quite different from that made by the Robot-Facial-Group.

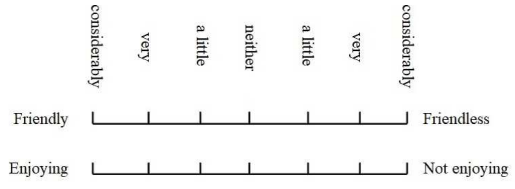


Figure 6: Examples of the rating scale method

5 Experiment 2

5.1 Method

The subjects learned with a robot comprising a sympathy expression method and a studying system. When a subject answers question, the studying system shows a correct or incorrect judgment. Subjects learned by answering the questions which were displayed by the studying system.

20 university students who had never studied with Tabot previously were gathered for this experiment. 20 subjects were divided into two groups of 10 students each; "Robot-Facial-Group" and "Robot-Combination-Group." The subjects in Robot-Facial-Group studied with Tabot, which expresses only facial expressions. The subjects in Robot-Combination-Group studied with Tabot, which expresses a combination of facial expressions and body motions. These subjects were instructed in math by the studying system for 40 minutes, 2 or 3 times a week, for 1 month. Thus, they studied 12 times in a month. During the studying of 1st and 12th time, subjects used studying system on their own to measure their ability before and after studying. These are the pre-test and the post-test scores. From the 2nd time to 11th time, subjects learned using the studying system with the Tabot.

The upper side of Tabot's tablet showed facial expressions. The lower side of Tabot's tablet showed the studying system. Tabot expressed emotions when the studying system gave a correct or an incorrect response.

The robot and studying system used in this experiment is the same as those used in Experiment1.

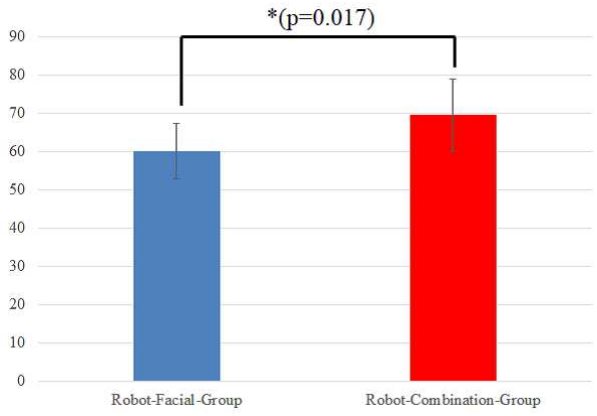


Figure 7: Impression score in experiment1

5.2 Evaluation criterion

Studying effect was calculated based on the improvement score got by difference of the pre-test score and post-test score. Each of these scores was calculated by

100 questions shown by the studying system. Impression evaluation was calculated in rating scale method, that is same as Experiment1. In all groups, subjects answered the questionnaire after the 11th studying. We defined the sum of scores of all questions as the score of good impressions. We conducted a Welch's t-test to improve the studying effect and the impression evaluation. To admit a significant difference based on the fact that p values is 5%, we adjusted the significance level (p = 0.025) using the Bonferroni method.

5.3 Result

5.3.1 Improvement score

Fig. 8 indicates the pre-test average score and the post-test average score in two groups. Fig. 9 indicates the average improvement score in two groups. Fig.9 indicates that the score of the Robot-Facial-Group is higher than the score of the Robot-Combination-Group. We conducted Welch's t-test to compare the studying effects. The results show that there was no significant difference in the improvement score between the two groups.

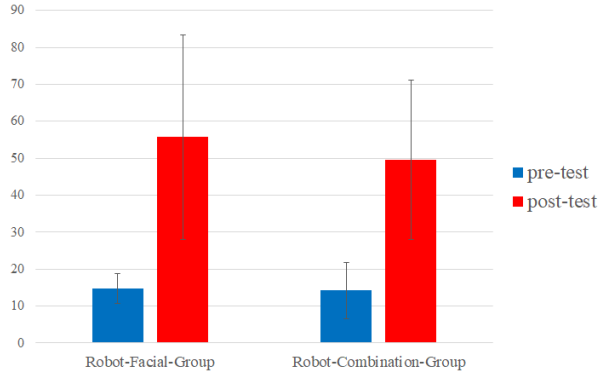


Figure 8: Score of pre- and post-test

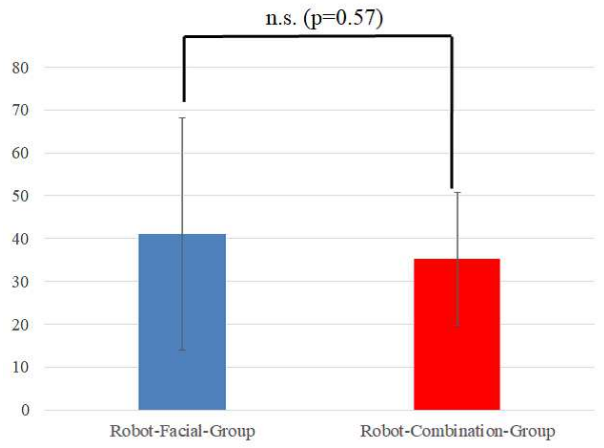


Figure 9: Improvement score

5.3.2 Questionnaire

Fig. 10 shows the average impression score in two groups. This graph shows that the average impression score of the Robot-Combination-Group is higher than that of Robot-Facial-Group. We conducted Welch's t-test to compare the impression scores. The results showed there was no significant difference in the impression score between the two groups.

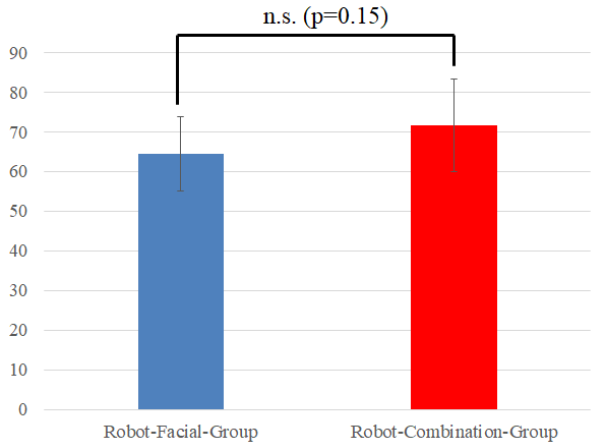


Figure 10: Impression score in Experiment2

5.4 Discussion

The result of studying effect indicates that it is higher in Robot-Facial-Group than in Robot-Combination-Group. Table.1 shows the number of answering questions in Robot-Facial-Group. Table.2 shows the number of answering questions in Robot-Combination-Group. The average number of answering questions is shown in Fig. 11. The answering questions were defined as the number of answered questions in each studying session. The average number of answering questions shows that its number was higher in Robot-Facial-Group than in Robot-Combination-Group. The time to express emotion by body motion is longer than the time to express emotions by facial expression change. Thus, the fact that the robot cannot move to the next problem while expressing emotion produced the difference in the average number of answers between the two groups. Moreover, we determined that the studying time in Robot-Combination-Group was shorter than that in Robot-Facial-Group. We believe that it is possible that the studying effect can be influenced by the number of answering questions.

In contrast, the result of the questionnaire shows that the robot in the Robot-Combination-Group gave a better impression than that in the Robot-Facial-Group. Therefore, the body motion of the robot manifested emotion efficiently as reported in conventional research [7]. Therefore, we believe that learners felt the more sympathy from the robot expressing emotions using body motions than from the robot expressing emotions using facial expressions only.

However, in this experiment, we could not find significant difference in the impression score between the two groups. Thus, we calculated the percentage of the expressed emotions. Fig. 12 shows the percentage of emotions expressed in the Robot-Combination-Group. It shows that only three types of emotion out of 28 were

often expressed by robots in the Robot-Combination-Group. Therefore, learners may feel that robots in this group always expressed the same emotion. Therefore, robots that expressed emotions using body motions and facial expressions did not show any significant difference compared to the robots that expressed emotions using facial expressions only.

Table 1: The number of answering questions in Robot-Facial-Group

No.	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
1	36	29	41	48	52	51	50	59	61	70
2	26	23	18	25	38	41	39	39	38	38
3	20	14	25	24	27	30	36	26	27	38
4	21	24	25	21	19	20	31	27	33	44
5	25	32	29	36	29	38	46	53	70	80
6	24	26	22	21	31	29	31	36	39	41
7	31	31	29	38	38	40	50	72	79	93
8	16	18	20	17	18	14	22	23	14	30
9	38	31	36	36	33	36	51	40	36	45
10	24	33	31	30	40	39	41	55	53	69

Table 2: The number of answering questions in Robot-Combination-Group

No.	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
1	24	32	39	26	32	37	38	47	40	43
2	19	18	20	15	27	28	30	40	45	54
3	25	25	25	29	30	35	40	43	36	37
4	22	18	24	23	32	33	30	38	43	41
5	21	23	27	24	30	26	32	28	36	40
6	18	19	17	27	22	17	23	30	35	26
7	17	17	24	26	20	28	29	29	34	41
8	23	28	26	29	47	38	52	51	63	66
9	17	18	23	30	40	28	38	37	47	46
10	24	23	28	23	38	39	49	59	81	83

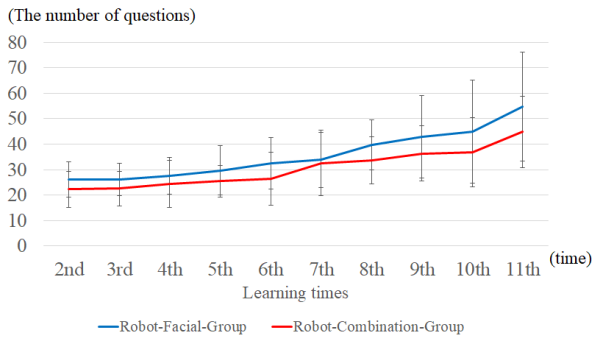


Figure 11: The average number of answering questions

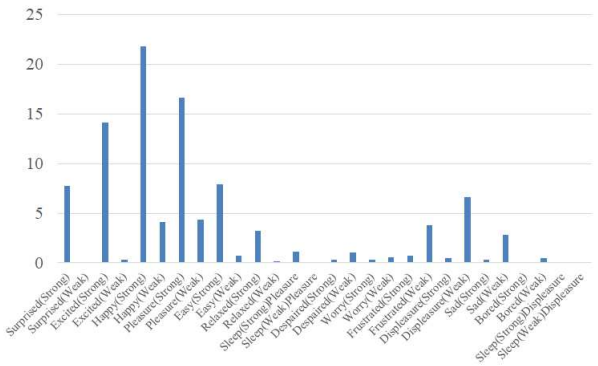


Figure 12: The percentage of expressed emotion

6 Conclusion

This paper examined the effects of sympathy expression method using body motion and facial expression in short-term and long-term experiments. We used "Tabot", a tablet-type robot that has a tablet head. Tabot has a sympathy expression method. In the experiment we performed, we used two types of' Tabot; robots that express emotions using facial expressions only and those that use a combination of body motions and facial expressions to express emotions. The results of Experiment1 suggest that the robot that expressed emotions using body motions and facial expressions prompted more impression to the learners than the robot that expressed emotions using facial expressions only. The results of Experiment2 suggest that robots that express emotions using facial expressions can better prompt learners to improve their studying effect than those that express emotions using the combination of body motions and facial expressions. Moreover, we determined that the robots that express emotions using facial expressions and body motions prompted more impression to the learners than those that express emotions using facial expressions only. However, robot that express emotions using body motions and facial expressions has issues relating to the time of expressing emotions and the kind of emotions they expressed.

In the future, we plan to improve the algorithm for sympathy expression in a way that solves the problem of time of expressing emotions and the kind of emotions expressed by the robot.

Acknowledgment This work was supported by MEXT KAKENHI (Grant-in-Aid for Scientific Research (B), No.16H02889)

References

[1] T. Kanda, T. Hirano, D. Eaton and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," Hum-Comout. Interact., vol.10, no. 1, pp. 61–84, 2004

[2] O.H. Kwon, S.Y. Koo, Y.G. Kim, and D.S. Kwon, "Telepresence robot system for English tutoring," IEEE Workshop on Advanced Robotics and its Social Impacts, pp. 152–155, 2010

[3] S. Koizumi, T. Kanda, and T. Miyashita, "Collaborative studying Experiment with Social Robot (in Japanese)," Journal of the Robotics Society of Japan, vol. .29, no. 10, pp. 902–906, 2011

[4] T. Kanda, T. Hirono, E. Daniel and H. Ishiguro, "Participation of Interactive Humanoid Robots in Human Society: Application to Foreign Language Education," Journal of the Robotics Society of Japan, vol. .22, no. 5, pp. 636–647, 2004

[5] F. Jimenez, M. Kanoh, T. Yoshikawa, T. Furuhashi, "Effects of Collaborative studying with Robots That prompts Constructive Interaction," Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pp. 2983–2988, 2014

[6] F. Jimenez, T. Yoshikawa, T. Furuhashi and M. Kanoh, "Effects of Collaborative studying with Robots using Model of Emotional Expressions," journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol. 28, no. 4, pp. 700–704, 2016

[7] T. Kanda, M. Imai, T. Ono, H. Ishiguro, "Numerical Analysis of Body Movements on Human-Robot Interaction," IPSJ Journal, Japan, Vol.44, No.11, pp.2699-2709,2003

[8] Y. Toyama and L. Ford, "US-Japan Body Talk augmented New version Gestures · Facial expressions · Gesture's dictionary," Sanseido, Japan, 2016

[9] D. Morris and Y. Toyama, "Body talk New style version World gesture dictionary," Sanseido, Japan, 2016

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 91-101 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Systematic Mapping Study of Cloud, Fog, and Edge/Mobile Devices Management, Hierarchy Models and Business Models

Isaac Odun-Ayo[1,*], Rowland Goddy-Worlu[1], Victor Geteloma [1], Emanuel Grant[2]

[1]*Department of Computer and Information Sciences, Covenant University, 112233, Nigeria*

[2]*School of Electrical Engineering & Computer Science, University of North Dakota, 58202, USA.*

ARTICLE INFO

Article history:
Received: 10 January, 2019
Accepted: 22 February, 2019
Online: 15 March, 2019

Keywords:
Cloud computing
Fog computing
Edge computing

ABSTRACT

Cloud computing is an exceptional paradigm, which is facilitating the developments and utilization of resources over the internet. Fog computing operates at the edge of the network saving bandwidth, by not sending all information to the cloud, while edge computing does processing of data at the edge of the cloud. Edge computing reduces the distance data must travel on the network. The unique relationship between cloud, fog and edge computing makes research in these areas mandatory. Deciding on a specific area of research as regards these subjects could be a bulky procedure for a scientist. Therefore, reviews and paper studies for recognizing potential research gaps are required. A systematic mapping study is utilized in giving a summary of the conducted research in a particular study area. The objective of this paper is to conduct systematic mapping studies on cloud, fog, edge/mobile devices management, hierarchy models and business models. The results showed that publications that discussed process in relations to the field of study is 14.04% out of the 114 papers included. Also method contributed 24.56%, model had 42.98% and tool contributed 18.42%. Furthermore, evaluation research in terms of the field of study was 27.5% out of 120 papers included. Also, validation was discussed in 17.5% of the papers, solution was 32.5%, philosophical was 5.83%, experience was 15.83% and opinion was 0.83%. The clearly highlighted gaps ought to inspire more enthusiasm for additional research by both researchers and industry practitioners.

1. Introduction

The Cloud is a parallel and appropriated computing system comprising of interconnected and virtualized computers that are powerfully displayed as a unified computing resource, which is dependent on service level approvals set up through dealings between service providers and users [1]. Cloud computing is basically interconnected computers and applications used to deliver services over the Internet. When systems are connected in this manner, device management also becomes paramount. Cloud computing is presently delivering services at all layers hence, there is everything as a service, but there are three primary services in the cloud which are Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). In SaaS, major cloud service provider (CSP) like Salesforce.com offer custom-built applications for users on the internet, hence the user do not need to bother about installation and license. In PaaS, the user develops and deploys his application through the infrastructure made available by the CSP. In IaaS, the user utilized resources on the CSP infrastructure on a pay as you go basis. The CSPs have massive data centers with state of the art system; hence, the user can start up without the cost of infrastructure. Although the CSPs strive in providing effective and trustworthy services on the cloud, there are also issues of trust [2]. The effectiveness of cloud computing becomes very evident as services improve and expand regularly, due to the fundamental applications running on the cloud [3,4]. In addition, there are four cloud computing models namely the private, public, community and hybrid models. Private clouds provide organizations infrastructures that are hosted on premises or off premises making use of in house staff. Private clouds are considered more secure. Public clouds provide massive elastic, on-demand pool of resources to the users through the CSP

*Corresponding Author: Isaac Odun-Ayo, Contact No. +234 802 88 29456, Email: isaac.odun-ayo@covenantuniversity.edu.ng

infrastructure. Community cloud make services available to institutions that share infrastructure but operate independently. Hybrid clouds provide the potential to host core applications on private cloud and auxiliary applications on public clouds. However, due to the process of virtualization and multitenancy on the cloud, there are concerns about security [5,6].

Fog and edge computers carry out storage and processing at the edge of the cloud. This saves bandwidth and makes usage simple. However, there are technologies that must be used to monitor these applications data and services operating within and at the edge of cloud computing. Edge computing is characterized by proximity to end users, zone geographical distribution, and support for high mobility and value addition for customer IoT services [7]. Mobile edge computing aims at reducing both network latency and resource demand by shifting computing and storage capacities from the Internet cloud to the mobile edge [8]. In [9], the cloud system is described with a 3-layer hierarchy, the cloud computing infrastructure layer, the edge computing or mobile clouds, which are utilized in vertically integrated applications where a part of the processing and analytics happens on the edge device while the cloud is used for coordination and data retrieval [10]. The service measurement index framework discussed in [11] is a business model that provides an understanding of CSP rankings, and has 7 major characteristics namely: accountability, agility, assurance, financial, performance, security and privacy and usability [11].

Systematic mapping studies provides a platform for categorizing research using a scheme and structure indicating the frequency of research in such an area and the results are presented visually on a map. Systematic mapping requires less effort: yet provides a more coarse-grained overview in the area of study [12]. This enables the coverage of the research field to be determined. The systematic process in this work was accomplished using three facets: the topic, contribution and research facets. The topic facet extracts key consideration from the scope of study. The contribution facet deals with method or mode of the work, while the research facet considers the nature of research such as evaluation or opinion. While the goal of this study is to conduct a systematic mapping study, the rest of the paper is organized as follows: Section II examines the related work. Section III examines the materials and methods, Section IV presents the results and discussion and Section V is the conclusion and suggestion for future work.

2. Related Work

In [13], the authors explored the developing stage of a systematic mapping study by identifying the software designs present at the requirement engineering stage of the project, and providing a comprehension of the roles played by these patterns based on basic parameters required in the development process. A protocol was developed for the study with basic steps to enable the replication of their work by the research community, which results in confirming the validation of the research. For this paper, ACM DL, IEEExplore, SCOPUS, and Web of Science digital libraries were utilized, and the guidelines laid down in [12] were adhered to.

In [14], the authors dwelled on describing the protocol for a systematic map, in relation to domain-specific languages (DSL). The authors focused on channeling their work on the advanced apprehension of the DSL domain of research, while emphasizing on trending researches and subsequent expectations, and utilizing the requirements for conducting a systematic review, all within the interval of July 2013 to October 2014.

The authors in [15] analyzed the utilization of concept maps in Computer Science, which resulted in a systematic map focused on gathering and analyzing previous articles on concept maps. This resulted in a thorough examination of concept maps, with emphasis on learning and teaching supports. Also, the searching process involved the utilization of backward snowballing and manual methods, and search strings applied on SCOPUS, ScienceDirect, Compedex, ACM DL, and IEEExplore digital libraries.

In [16], the researchers examined how games related methods are implemented in software engineering, and how these methods assist specified software engineering knowledge domains. Based on publications from 1974 to 2016, the authors utilized 156 primary studies, which were centered on using and evaluating games in software engineering education. The mapping procedure was done in line with the guidelines provided in [12].

In [17], the authors did a mapping of power system model by providing a summary of power system models and its subsequent usage by European organizations in terms of analysis of their modeling features and identification of modeling gaps. 228 surveys were sent out to power experts for data extraction, while 82 questionnaires were eventually completed and utilized for the mapping.

In [18], a mapping study of domain-specific languages was conducted with a primary interest in the type of contribution and research, and the area of focus. By exploring reputable articles from 2006 to 2012, a mapping study was conducted by defining research questions, conducting the search, and extracting the data.

In [19], the authors focused on searching studies on “legal theory” and “legal concepts”. The selected studies were grouped on contributions in terms of language, tool, method, and model. The other steps include identification of the used legal theories in legal core ontologies building process, identification of focus with a clear recommendation on the use of two ontologies, and analyzing each highlighted research for conclusive findings about legal and ontological research.

In [20], the researchers gave a summary of an empirical research in software cloud-based testing in the process of building a classification scheme. The authors examined functional and non-functional testing techniques, their applications, and their peculiarities. Sixty-nine (69) primary studies out of seventy-five (75) research articles were utilized for a rigorous statistical analysis, and an eventual quantitative result. Most articles utilized a singular experiment approach for evaluating their expected result.

In [21], the authors presented a comprehensive review of knowledge management in organization with a focus on the potential role of Information Technology (IT) in the process. They further discussed several important issues bordering on the process of managing knowledge and the role of IT in propelling and assisting these processes. Also, emphasis was made on the need to support in creating, storing, transferring and applying knowledge in organizations.

In [22], the authors discussed the usefulness and limitations of systematic literature review in information system and social sciences. They are of the opinion that the general stand that systematic literature review provide a holistic and superior approach to literature review is not only questionable but also unacceptable. In their argument and justification of this, they concluded that caution and restrains should be exercised when choosing systematic literature review as it could undermine critical engagement with literature and the scholarly nature of academic work.

In [23], the authors discussed the lessons learned from utilizing a systematic literature review process on software engineering domain. The authors summarized the review process, highlighted the reviews of other authors and finally extracted and discussed findings as regards applying this practice to software engineering domain.

In [24], the researchers opined that research reviews must pay close attention to rigorous methodology that is required of primary researcher. The author further conceptualized the research review as a scientific enquiry involving five stages that parallel those primary research, which includes: formulating the problem, collecting data, evaluating the data points, analyzing and interpreting the data, and lastly, presenting the results. All the functions; sources of variance and the other potential treats to validity with each stage are discussed.

In [25], the authors provided useful insights to researchers for carrying out literature review. They suggested synthesizing trends and patterns while preparing to write literature review, among which includes: the purpose and voice before beginning to write must be considered, then consider how to reassemble the notes as well as create a topical outline that traces the argument in the literature review. All of these provides the guidelines for developing a thorough and coherent literature review.

In [26], the researchers assessed the effectiveness of a systematic literature review, which recommends evidence-based software engineering methods for aggregating evidence. The authors utilized manually searched Journals (10) and conference proceedings (4). Of the twenty (20) significant articles considered, eight (8) articles focused on research trends rather than evaluating the technique utilized, and seven (7) systematic literature reviews focused on estimating the cost. The quality of systematic literature reviews was fair with only three scoring less than 2 out of 4.

In [27], the authors have an opinion that the evaluation of the approaches of systematic mapping, as carried out by researchers, is needed in identifying the improved guidelines that are centered on the experiences from existing systematic maps and literature reviews. In the affirmative, the authors carried out a systematic mapping study and considered a few practices of the systematic review guidelines. They discovered that amongst the extensive number of studies carried out, various guidelines were utilized and integrated, leading to multiple methods of conducting systematic mapping studies.

In [28], the authors stressed the importance of literature review in scientific enquiry and the need to avoid standing on the shoulders of dwarf literature search, which remains a point of concern for enhancing a productive literature review. The authors further discussed the difficulties of literature search in the ever-increasing dynamic setting of data frameworks, and proffer solutions on how these difficulties can be dealt with.

In [29], the authors presented the systematic mapping study of high-performance computing and the cloud which based their work on the concepts of [27]. This study provided six classes of studies in the areas of architecture, virtualization, application, optimization, design and implementation, and performances in relation to the focus of study. The selected studies were applied on the contribution facet such tool, method, and model. In addition, the selected studies were used on the research facet which dealt with evaluation, validation and solution research.

In [30], the researchers examined the systematic mapping study of cloud resources management and scalability in brokering, scheduling, capacity planning and elasticity. The classification scheme in relation to resource management, discussed scalability, brokering, scheduling, capacity planning, elasticity, and resource sharing. The selected studies were applied on the contribution facet such metric, tool, method, and model. In addition, the selected studies were applied to the types of research.

3. Materials and Methods

The systematic mapping study is a useful tool for both researchers and practitioners alike. A mapping study provides a visual representation of scope of work in a particular discipline. This study was carried out utilizing the guidelines on systematic mapping studies, as seen in [12]. Systematic mapping study involves replicating a process for deducing and interpreting accessible materials, in line with a research objective [31]. This important process (Figure 1) involves defining the research question and outlining the scope of the review. Next is screening the articles to select the important ones and key wording the abstract of the articles, with the goal of creating a classification scheme. Finally, the last step involves extracting the data, which ultimately leads to creating the systematic map. At every stage of the process there was an outcome which was refined for better overall output of the systematic map.

3.1. Definition of Research Questions

The goal of a systematic map is to provide more insight into the quantity and type of work being conducted in a particular discipline. It may also be necessary to know the places where such work were published. These issues help in determining the appropriate research questions to apply to the study in this paper.

The research questions are:

RQ 1: What areas are addressed and what number of articles are covered in the highlighted areas of the research topic?

RQ 2: What sets of articles are published in the highlighted areas and what are their evaluations and originality?

3.2. Conduct of Research for Primary Studies

For the purpose of getting papers for this study, four (4) major digital libraries based on their impact factor of conferences and journal publication were utilized. Table 1 contains the digital libraries and their corresponding uniform resource locator (URL).

Table 1: Electronic databases used for the systematic mapping study

Electronic Databases	URL
ACM	http://dl.acm.org/
IEEE	http://ieeexplore.ieee.org/xplore
SCIENCE DIRECT	http://www.sciencedirect.com/
SPRINGER	http;//www.springerlink.com/

The search string was designed based on population, outcome and intervention of the papers. The keywords used in the research was gotten from all parts of the structure of this study's title; the search string is as follows:

(TITLE (cloud) AND KEY (cloud) AND (TITLE-ABS-KEY) (fog) OR TITLE-ABS-KEY (edge and mobile) OR TITLE-ABS-KEY ("devices management") AND (KEY (Models) OR TITLE (models)

The searches were performed on the electronic database, which involves utilizing the designed search string above on document metadata to guarantee the inclusion of significant articles. For this study, findings from relevant databases relating to cloud computing and computer science were utilized, and a total of 120 papers were relevant to be included out of an underlying pursuit comprising of 1339 papers. This study covered the period 2001 – 2018. The list of primary studies utilized for this work is at the Appendix.

3.3. Screening of Papers for Inclusion and Exclusion

The importance of selection criteria is identifying and incorporating all papers relevant to the study. This was a vital aspect of the study. It was essential to use the inclusion and exclusion criteria to eliminate irrelevant articles as regards to cloud, fog, and edge/models device management. In addition, the criteria was utilized in removing all materials not providing answers to the research question. Some abstract usually mention only one aspect of the focus of study without further details and such papers are also excluded. This study also excluded panel discussions, editorials, prefaces, presentation slides, summaries and tutorials. It was pertinent to consider articles that had the main focus and able to also provide some secondary details. The main focus of this study was devices management and models as it relates to cloud, fog, and computing. Therefore, the inclusion and exclusion process was done using details depicted in Table 2.

Table 2: Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
The abstract explicitly mentions cloud, fog, edge, mobile devices management, hierarchy models and business models. Furthermore, discussions in this area are meant to achieve cloud management.	The paper lies outside the area of cloud computing as it identifies with cloud management. The paper does not contribute to issues of management in the cloud.

3.4. Keywording of Abstracts

Keyword of abstracts is an important process in systematic mapping, used to design the classification scheme for the study and involves the following stages as shown in Figure 1.

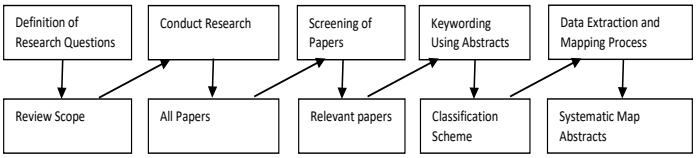


Figure 1: The systematic mapping process [27]

Key wording was useful in decreasing the required time for designing a classification scheme for cloud management and model studies. Also, key wording ensured that only the significant papers were considered in the scheme. The abstracts were studied in order to extract concepts and keywords relating to the focus of this study. Therefore, keywords from different publications were combined together to give satisfactory knowledge about the nature of the contributing research. This was eventually utilized in determining the facts or categories of the study.

A cluster of keywords were finally used in this paper to determine the categories and the eventual systematic map. In this paper, the three categories focused on topics based on all aspect of the title of the study. The second facets discussed the types of contributions in relation to metric, tool, method, process and model [12]. The third facet concentrated on research types such as evaluation and validation.

3.5. Research type Facts with Category and Description

A study of this nature will be inadequate without all the categories outlined. The following is the classification of research methods, which was utilized for the third facet, as listed in [32]:

1. **Validation of research:** The utilized methods are unique but not yet executed, either in form of an application or experiments.
2. **Evaluation research:** The methods have been executed and analyzed with a proof of concept. Outcomes are examined in terms of advantages and disadvantages.
3. **Solution proposal:** The method offers an essential solution to a problem; the benefits and application of such results are also outlined.
4. **Philosophical papers:** The papers discuss techniques that provide new methods of solving a problem in relation to concepts and framework.
5. **Opinion papers:** The research papers do not apply any known methods, but simply expressed the opinion of individuals.
6. **Experience papers:** This paper relates the personal experience of the researcher. It offers insight into how something was done

These categories were considered adequate and sufficient to be used in the classification scheme of this study, as to research facet. The papers included in this study were all examined based on the various research categories.

3.6. Data Extraction and Mapping Studies

Relevant articles were sorted into a classification scheme at the key wording stage. This next step allowed data extraction from the primary studies. The method of extracting data shaped the nature of the classification scheme. The procedure for extracting data was achieved on Microsoft Excel tables. The Excel tables contained the various groups of the classification scheme. The frequency of publications in each category was extracted to different tables. The frequency of publications in each category was combined into a table either contenting the topic/contribution category or the topic/research category. The analysis was focused on presenting the frequencies of articles based on the entries made into the various Excel tables. The essence was to identify which aspect of device management and models as it relates to cloud, fog and edge/mobile computing were given more emphasis in this study This enabled the determination of gaps and made it easy to identify areas for further research.

Based on the results extracted to the Excel tables, bubble plots were subsequently utilized in presenting the frequency of the articles using a map. The map included a two x-y scatter plot with bubbles at the crossing point of the classifications. The size of the bubble coordinates corresponds to the quantity of articles in such combined categories. There were two quadrants based on the three facets under consideration. Each quadrant provided different visualized maps on the basis of the intersection of the topic category with either the contribution or research category. The different facets could be considered simultaneously, providing an overview of the results of the study.

4. Results and Discussion

The analysis of the result is meant to present the frequencies of publication in each category, and to identify which category has been emphasized in past research [12]. The essence of this was the identification of gaps and the possibility of researching further. From the analysis, gaps were identified by using the systematic map, which showed topic areas where there was dearth of study, and indicated the areas that were explored as regards articles. In this systematic study, high-level categories was utilized in assessing the primary studies, which was used to create the systematic map.

4.1. Topic and Contribution Facet

The topics extracted at the classification scheme in this field of study comprise of the following:

1. Architecture
2. Application
3. Implementation
4. Networking
5. Mobile computing
6. Embedded systems

The list of primary studies used for checking the topics against the types of contributions is at Table 3. The systematic map is shown at Figure 2. The contribution facet shows the type of contribution to the field of study. The results showed that publications that discussed process in relations to the field of study was 14.04% out of the 114 papers included. In addition, method contributed 24.56%, model had 42.98% and tool contributed 18.42%. There were no contributions at all from metric.

Models discussion contributed 42.98% of the papers reviewed. In relation to the topics, model had 5.22% in terms of architecture, 12.28% in terms of application, and 3.51% in the area of implementation, 10.53% in the area of networking, 3.51% in the aspect of mobile computing and 7.89% on embedded system. Other aspects of topic and contribution aspect are reflected in Figure 2.

4.2. Type and Research Facet

The list of primary studies used for examining the topics against the types of research is at Table 4. On the x-axis of the right quadrant of Figure 2 are the results of the types of research conducted with respect to the field of study. The result indicated that evaluation research in terms of the field of study was 27.5% out of 120 primary studies. Also, validation was discussed in 17.5% of the papers, solution in 32.5%, philosophical in 5.83%, experience in 15.83% and opinion in 0.83%.

Solution proposal constituted 32.5% of the papers examined in this field of study in relation to the topics. Solution proposal had 1.67% in terms of architecture, 8.33% in terms of applications, 5% in relation to implementation, 11.67% with respect to networking, 3.33% in terms of mobile computing and 2.5% in terms of embedded system. Other aspects of research and topic contribution is as shown in Figure 2.

4.3. Major Findings

From Figure 2,

a. It can be identified from the visual map that there were more publications on implementations and mobile networking that discussed metric with 26%, more publication on applications that examined model with 12.28%, more papers on architecture that discussed method with 6.14% and more work on embedded systems examining process at 6.14%.

b. Similarly, on the left quadrant of Figure 2, there were more articles on mobile computing that discussed evaluation research with 9.17%, more publications on application that examined validation research with 5%, more work on network in terms of solution proposal at 11.67%, more publication on mobile computing in terms of philosophical paper with 3.33%, more materials on application in term of experience papers and only one opinion paper related to embedded system.

c. On the other hand, this study had shown that there were no discussions at all relating to metric in this field of study. There were no papers on architecture in terms of tool. There were no articles on implementation and networking in terms of process. In addition, there was no work on architecture, application, and networking in the area of philosophical papers.

d. Furthermore, publications on application topic as it related to tool, mobile networking and implementations

Table 3: Primary studies for topic and contribution facet

Contribution Facet / Topic	Metric	Tool	Model	Method	Process
Architecture			30, 37, 80, 92, 93, 95,	18, 42, 60, 77, 84, 90, 116,	4, 8,
Application		28,	25, 36, 39, 38, 45, 48, 54, 57, 71, 78, 82, 85, 115, 119,	5, 11, 14, 86, 94,	105, 108,
Implementation		13, 19, 26, 101, 118, 120	50, 55, 87, 100,	2, 43, 46,	
Networking		22, 24, 29, 31, 33,	7, 9, 12, 15, 16, 91, 96, 98, 103, 104,106, 112,	20, 41, 47, 70, 79, 83,	
Mobile computing		3, 6, 10, 23, 109, 110,	27, 32, 34, 35	2, 73, 75, 76, 89, 99	44, 53, 56, 58, 71
Embedded Systems		49, 61, 62,	1, 64, 65, 66, 67, 68, 74, 115, 119	81	17, 21, 51, 52, 97, 113, 114,
Percentage	**0.0%**	**18.42%**	**42.98%**	**24.56%**	**14.04%**

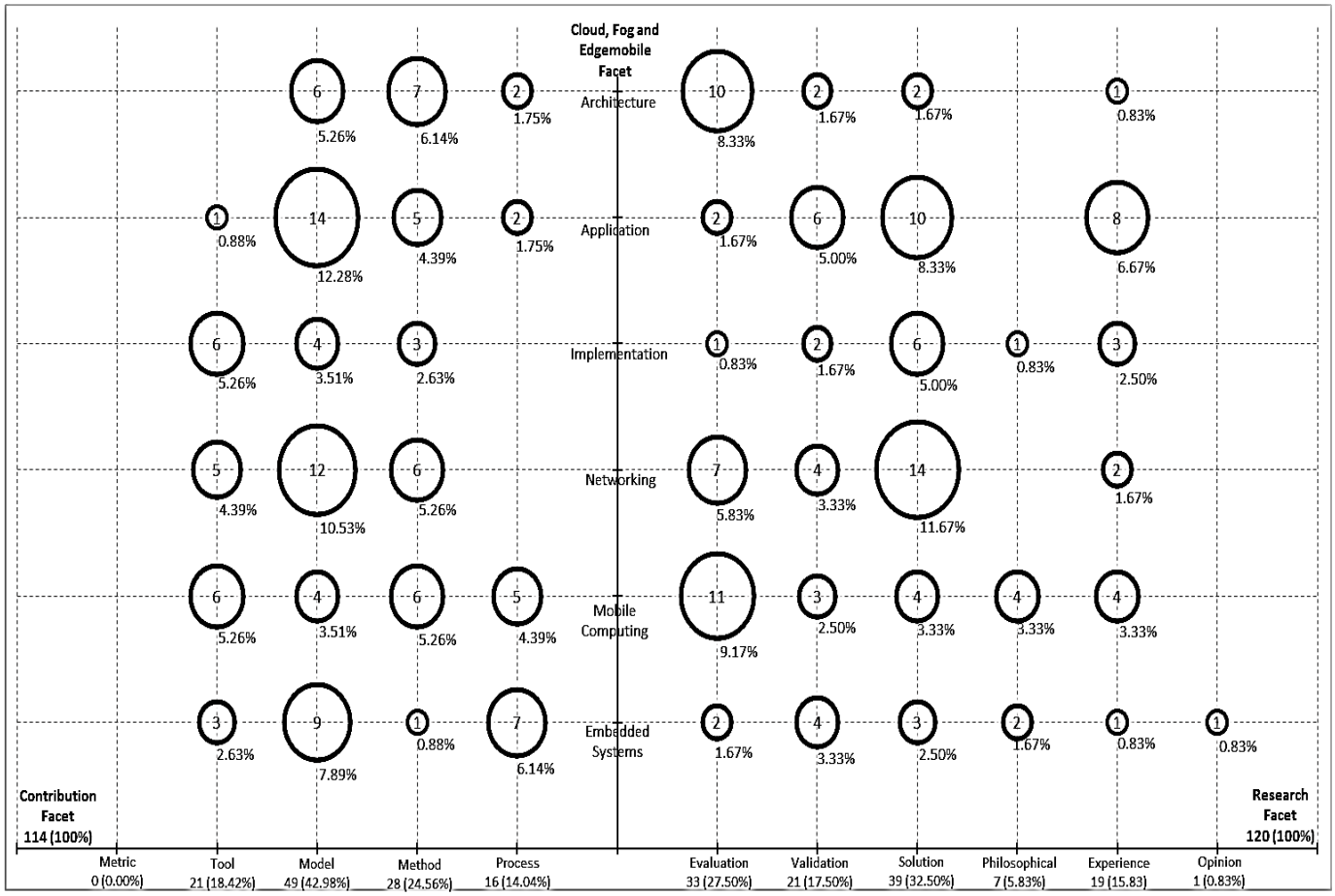


Figure 2: Systematic map of cloud fog, and edge/mobile devices management, hierarchy and business models

Table 4: Primary studies for topic and research facets

Research Facet / Topic	Evaluation	Validation	Solution	Philosophical	Experience	Opinion
Architecture	4, 8, 37, 42, 60, 80, 84, 90, 95, 116	18, 30	92, 93		77	
Application	78, 102	11, 14, 59, 63, 105, 108	5, 36, 39, 38,48, 54, 82, 85, 115, 119		25, 28, 45, 57, 69, 71, 86, 94	
Implementation	2	13, 87	19, 26, 43, 46, 118, 120	101	50, 55, 100	
Networking	7, 20, 22, 24, 29, 31, 33	41, 47, 88, 91	9, 12, 15, 16, 49, 61, 62, 70, 104,106, 112, 96, 98, 103		79, 83	
Mobile computing	3, 6, 10, 23, 34, 35, 75, 99, 111, 113, 114	76, 89, 107	40, 44, 73, 117	53, 56, 58, 72	27, 32, 109, 110	
Embedded Systems	1, 65	17, 21, 51, 52	64, 81, 97	66, 67	74	68
Percentage	**27.50%**	**17.50%**	**32.50%**	**5.83%**	**15.83%**	**0.83%**

in terms of model, embedded systems were the least having, 0.88%, 3.51%, 0.88% and 1.75% respectively.

e. In terms of the topic and research types, implementation topic in relation to evaluation research, implementation and architecture in relation to solution proposals, implementation with respect to philosophical paper, architecture and embedded systems in terms of experience papers had the least publications with 0.83%, 1.67%, 0.83% and 0.83% respectively.

From the foregoing, the visual appeal of a systematic map assists in summarizing and offering researchers with the likelihood of stimulating interest in the results.

The different categories combined together makes the bubble plot more useful. It is imperative to state that the creation of a systematic map in the absence of a successive systematic review is valuable in itself. This is because the systematic map enables the identification of research gaps and shortage of publication in this field of study. This paper provided six classes of studies namely architecture, application, implementation, networking, mobile computing and embedded systems, which could be discussed as regards to either tool, model, method, metric and process or evaluation, validation, solution, philosophical and opinion research. The few aforementioned areas, along these lines, are suggested for additional research, and the rundown of primary studies would likewise help meaning researchers.

5. Conclusion

Cloud computing is providing opportunities in different areas of human life. More areas of cloud research are opening up on a regular basis. This has led to increasing studies and consequent publications in cloud computing. There are volumes of research effort that is useful both to the cloud users and cloud providers alike. The aspect of devices management and models in relation to cloud, fog, edge and mobile computing has continued to receive attention. Despite the large quantity of publications in this field of study, the systematic map has shown that there are shortage of publications in this field of study. The visual representation provided by the systematic map enhances decision on areas of further research. This systematic mapping study has been able to identify some areas where there were less emphasis on the topic, based on the categories used in the scheme. This paper has therefore contributed to knowledge by highlighting the different aspects of the study with gaps, and recommending them for further studies. Also, further research could be carried out to validate this study or resolve contradictory issues. In summary, this study created a systematic map that could be advantageous to the cloud community, and will assist researchers in uncovering the critical gaps that were not previously explored.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation, and Discovery (CUCRID).

References

[1] R. Buyya, J. Broberg, and A. Goscinski, “Cloud computing principles and paradigms,” John Wiley and Son, 2011, pp. 4-10.

[2] I. Odun-Ayo, S. Misra, N. Omoregbe, E. Onibere, Y. Bulama, and R. Damasevičius, “Cloud-based security driven human resource management system,” Frontiers in Artificial Intelligence and Applications. 295, pp. 96 – 106, 2017. doi:10.3233/978-1-61499-773-3-96

[3] I. Odun-Ayo, M. Ananya, F. Agono, and R. Goddy-Worlu, “Cloud computing architecture: A critical analysis,” In IEEE Proceedings of the 2018 18th International Conference on Computational Science and Its Applications (ICCSA 2018), pp. 1-7 doi: 10.1109/ICCSA.2018.8439638

[4] I. Odun-Ayo, B. Odede, and R. Ahuja, “Cloud applications management-Issues and developments,” Lecture Notes in Computer Science (including

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol.10963, pp. 683-694, 2018, Berlin, Germany: Springer.
[5] I. Odun-Ayo, S. Misra, O. Abayomi-Alli, and O. Ajayi, "Cloud multi-tenancy: Issues and developments," UCC '17 Companion, Companion Proceedings of the10th International Conference on Utility and Cloud Computing. pp. 209 – 214, 2017.
[6] I. Odun-Ayo, N. Omoregbe, M. Odusami, and O. Ajayi, "Cloud ownership and reliability - Issues and developments," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 10658, pp. 231-240, 2017, Berlin, Germany: Springer.
[7] S. K. Datta and C. Bonnet, "An Edge computing Architecture integrating virtual IOT Devices," EURECO Sophia Antipolis, France.
[8] T. M. Beck and M. Maier, "Mobile Edge computing challenges for future Virtual Networking Embedding Algorithms," IARIA, 2014.
[9] J. Wenke and M. Jiangbo, "A Reference Model of cloud operating and open source software implementation mapping," 18th IEEE international Work Shop on enables Technologies: Infrastructures for collaborations Enterprises, 2009.
[10] P. Varshney and Y. Simmhan, "Demonstrating Fog computing: Characterizing Architectures, Applications and Abstractions," arxiv: 1702.06331v1 [cs.Dc] 21 Feb. 2017.
[11] G. Baranwal and D. P. Vidyarthi, "A cloud service selection model using improved ranked voting method", Concurring and Computation: Practice and Experience," Wileyonhinelibrary.com, 2008.
[12] K. Petersen, R. Feldt, S. Mujtaba and M. Matisson, "Systematic Mapping Studies In Software Engineering", EASE'08 Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering, Italy — June 26 - 27, 2008, , pp. 68-77
[13] J.L. Barros-Justo, A.L. Cravero-Leal, F. B. Benitti, and R. Capilla-Sevilla, "Systematic mapping protocol: the impact of using software patterns during requirements engineering activities in real-world settings," Cornell University Library, 2017. arXiv:1701.05747v1 [cs.SE].
[14] T. Kosar, S. Bohra, and M. A. Mernik, "Protocol of a systematic mapping study for domain-specific languages," Journal of Information and Software Technology 21(C). pp. 77-91, 2016.
[15] V. Santos, E. F. Souza, K. R. Felizardo, and N. L. Vijaykumar, "Analyzing the Use of Concept Maps in Computer Science: A Systematic Mapping Study," Informatics in Education, 16(2) pp. 257–288, 2017. doi:10.15388/infedu.2017.13
[16] M. Souza, L. Veado, R. T. Moreira, E. Figueiredo, and H. Costa, "A systematic mapping study on game-related methods for software engineering education," Information and Software Technology 95, pp. 201-218, 2018.
[17] C. R. Fernandez-Blanco, F. Careri, K. Kavvadias, I. Hidalgo Gonzalez, A. Zucker, and E. Peteves, "Systematic mapping of power system models: Expert survey," EUR 28875 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-76462-2, doi:10.2760/422399, JRC109123.
[18] M. Mernik, "Domain-specific languages: A systematic mapping study," International Conference on Current Trends in Theory and Practice of Informatics, Lecture Notes in Computer Science, Vol. 10139, pp. 464-472, 2017, Berlin, Germany: Springer.
[19] C. Griffo, J. P. A. Almeida, and G. Guizzardi, "A systematic mapping of the literature on legal core ontologies," In Brazilian Conference on Ontologies, ONTOBRAS 15, 2015, CEUR Workshop Proceedings.
[20] A. Ahmad, P. Brereton, and P. Andras, "A systematic mapping study of empirical studies on software cloud testing methods," In IEEE International Conference on Software Quality, Reliability and Security Companion, pp. 555-562, 2017.
[21] M. Alavi, and D. E. Leidner, "Knowledge management and knowledge management systems: Conceptual foundations and research issues," MIS Quarterly, pp. 107-136, 2001.
[22] S. K. Boell, and D. Cecez-Kecmanovic, "On being 'Systematic' in literature reviews," In Formulating Research Methods for Information Systems, Palgrave Macmillan, London, pp. 48-78, 2015.
[23] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," Journal of Systems and Software, vol. 80, no. 4, pp. 571-583, 2007.
[24] H. M. Cooper, "Scientific guidelines for conducting integrative research reviews," Review of educational research, vol. 52, no.2, pp. 291-302, 1982.
[25] J. L. Galvan, and M. C. Galvan, "Writing literature reviews: A guide for students of the social and behavioral sciences," Routledge, 2017.
[26] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," Information and Software Technology, vol. 51, no. 1, pp. 7-15, 2009.
[27] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," Information and Software Technology, vol. 64, pp. 1-18, 2015.
[28] J. Vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," CAIS, vol. 37, no. 9, 2015
[29] I. Odun-Ayo, R. Goddy-Worlu, O. Ajayi and E. Grant, "A Systematic Mapping Study of High Performance Computing and the Cloud". ARPN Journal of Engineering and Applied Sciences, vol. 13, Issue 24, pp. 9886-9700, 1 Dec 2018.
[30] I. Odun-Ayo, O. Ajayi, R. Goddy-Worlu and J. Yahaya, "A Systematic Mapping Study of Cloud Resources Management and Scalability in Brokering, Scheduling, Capacity Planning and Elasticity". Asian Journal of Scientific Research, vol. 12, pp. 151-166, 2019.
[31] A. C. Muhammed and A. B. Muhammed, "A systematic Mapping study of software architectures for cloud based systems," Software System section, IT University of Copenhagen, 2014.
[32] R. Wieringa, N. A. Maiden, N. R. Mead and C. Rolland, "Requirement engineering paper classification and evaluation criteria. A proposal and a discussion," Requirement Engineering, vol. 11, pp. 102-107, 2006.

Appendix: List of primary studies

[1] Al-Jaroodi, J., Mohamed, N., Jawhar, I., Mahmoud, S. CoTWare: A Cloud of Things Middleware (2017) Proceedings - IEEE 37th International Conference on Distributed Computing Systems Workshops, ICDCSW 2017, art. no. 7979819, pp. 214-219.
[2] Al-Joboury, I.M., Al-Hemiary, E.H. F2CDM: Internet of things for healthcare network based fog-to-cloud and data-in-motion using MQTT protocol (2017) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10542 LNCS, pp. 368-379.
[3] Alonso-Monsalve, S., García-Carballeira, F., Calderón, A. A heterogeneous mobile cloud computing model for hybrid clouds (2018) Future Generation Computer Systems, . Article in Press.
[4] Alsaffar, A.A., Pham, H.P., Hong, C.-S., Huh, E.-N., Aazam, M. An Architecture of IoT Service Delegation and Resource Allocation Based on Collaboration between Fog and Cloud Computing (2016) Mobile Information Systems, 2016, art. no. 6123234, .
[5] Anantula, P.R., Someswar, G.M. Protecting data in cloud environment from intruders using fog computing mechanism and control flows (2018) Proceedings of the 2nd International Conference on Communication and Electronics Systems, ICCES 2017, 2018-January, pp. 176-182.
[6] Arfat, Y., Aqib, M., Mehmood, R., Albeshri, A., Katib, I., Albogami, N., Alzahrani, A. Enabling Smarter Societies through Mobile Big Data Fogs and Clouds (2017) Procedia Computer Science, 109, pp. 1128-1133.
[7] Arun, S.H., Sharma, S.K., Chaurasia, S., Kumar, R. Detection and analysis of fog/low cloud using Ceilometer and INSAT-3D satellite data over Delhi Earth Station, New Delhi (2017) 38th Asian Conference on Remote Sensing - Space Applications: Touching Human Lives, ACRS 2017, 2017-October, .
[8] Bhatia, M., Sood, S.K. Exploring Temporal Analytics in Fog-Cloud Architecture for Smart Office HealthCare (2018) Mobile Networks and Applications, pp. 1-19. Article in Press.
[9] Bruschi, R., Davoli, F., Lago, P., Lombardo, A., Lombardo, C., Rametta, C., Schembra, G. An SDN/NFV platform for personal cloud services (2017) IEEE Transactions on Network and Service Management, 14 (4), pp. 1143-1156.
[10] Bruschi, R., Lago, P., Lombardo, C. In-Network Programmability for Next-generation Personal Cloud Service Support (INPUT) (2016) Procedia Computer Science, 97, pp. 114-117.
[11] Byrne, J., Svorobej, S., Giannoutakis, K.M., Tzovaras, D., Byrne, P.J., Östberg, P.-O., Gourinovitch, A., Lynn, T. A review of cloud computing simulation platforms & related environments (2017) CLOSER 2017 - Proceedings of the 7th International Conference on Cloud Computing and Services Science, pp. 651-663.
[12] Chaudhary, D., Bhushan, K., Gupta, B.B. Survey on DDoS attacks and defense mechanisms in cloud and fog computing (2018) International Journal of E-Services and Mobile Applications, 10 (3), pp. 61-83.
[13] Chaudhary, R., Kumar, N., Zeadally, S. Network Service Chaining in Fog and Cloud Computing for the 5G Environment: Data Management and

Security Challenges (2017) IEEE Communications Magazine, 55 (11), art. no. 8114560, pp. 114-122.
[14] Chen, F., Li, H., Liu, J., Li, B., Xu, K., Hu, Y. Migrating big video data to cloud: a peer-assisted approach for VoD (2017) Peer-to-Peer Networking and Applications, pp. 1-15. Article in Press.
[15] Deepali, Bhushan, K. DDoS attack defense framework for cloud using fog computing (2018) RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings, 2018-January, pp. 534-538.
[16] Deng, R., Lu, R., Lai, C., Luan, T.H., Liang, H. Optimal Workload Allocation in Fog-Cloud Computing Toward Balanced Delay and Power Consumption (2016) IEEE Internet of Things Journal, 3 (6), art. no. 7467406, pp. 1171-1181.
[17] Desirena Lopez, H.J., Siller, M., Huerta, I. Internet of vehicles: Cloud and fog computing approaches (2017) Proceedings - 2017 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2017, 2017-January, pp. 211-216.
[18] Dixit, A., Yadav, A.K., Kumar, S. An efficient architecture and algorithm for server provisioning in Cloud computing using clustering approach (2017) Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016, art. no. 7894532, pp. 260-266.
[19] Elkhatib, Y., Porter, B., Ribeiro, H.B., Zhani, M.F., Qadir, J., Riviere, E. On using micro-clouds to deliver the fog (2017) IEEE Internet Computing, 21 (2), art. no. 7867723, pp. 8-15.
[20] El-Latif, A.A.A., Abd-El-Atty, B., Hossain, M.S., Elmougy, S., Ghoneim, A. Secure quantum steganography protocol for fog cloud internet of things (2018) IEEE Access, 6, pp. 10332-10340.
[21] Elmisery, A.M., Rho, S., Aborizka, M. A new computing environment for collective privacy protection from constrained healthcare devices to IoT cloud services (2017) Cluster Computing, pp. 1-28. Article in Press.
[22] Elmisery, A.M., Sertovic, M. Privacy enhanced cloud-based recommendation service for implicit discovery of relevant support groups in healthcare social networks (2017) International Journal of Grid and High Performance Computing, 9 (1), pp. 75-91.
[23] Elmroth, E., Leitner, P., Schulte, S., Venugopal, S. Connecting Fog and Cloud Computing (2017) IEEE Cloud Computing, 4 (2), art. no. 7912269, pp. 22-25.
[24] Escamilla-Ambrosio, P.J., Rodríguez-Mota, A., Aguirre-Anaya, E., Acosta-Bermejo, R., Salinas-Rosales, M. Distributing computing in the internet of things: Cloud, fog and edge computing overview (2018) Studies in Computational Intelligence, 731, pp. 87-115.
[25] Farris, I., Militano, L., Nitti, M., Atzori, L., Iera, A. MIFaaS: A Mobile-IoT-Federation-as-a-Service Model for dynamic cooperation of IoT Cloud Providers (2017) Future Generation Computer Systems, 70, pp. 126-137.
[26] Farris, I., Girau, R., Militano, L., Nitti, M., Atzori, L., Iera, A., Morabito, G. Social Virtual Objects in the Edge Cloud (2015) IEEE Cloud Computing, 2 (6), art. no. 7397053, pp. 20-28.
[27] Farris, I., Militano, L., Nitti, M., Atzori, L., Iera, A. Federated edge-assisted mobile clouds for service provisioning in heterogeneous IoT environments (2015) IEEE World Forum on Internet of Things, WF-IoT 2015 - Proceedings, art. no. 7389120, pp. 591-596.
[28] Fernandez, E.B., Yoshioka, N., Washizaki, H., Syed, M.H. Modeling and security in cloud ecosystems (2016) Future Internet, 8 (2), art. no. 13, .
[29] Goleva, R., Stainov, R., Kletnikov, N., Achkoski, J., Mirtchev, S., Ganchev, I., Savov, A. Performance Analysis of End-to-End Sensor-to-Cloud Personal Living Platform (2017) Procedia Computer Science, 113, pp. 615-620.
[30] Gupta, R., Garg, R. Mobile Applications Modelling and Security Handling in Cloud-Centric Internet of Things (2015) Proceedings - 2015 2nd IEEE International Conference on Advances in Computing and Communication Engineering, ICACCE 2015, art. no. 7306695, pp. 285-290.
[31] Hajibaba, M., Gorgin, S. A review on modern distributed computing paradigms: Cloud computing, jungle computing and fog computing (2014) Journal of Computing and Information Technology, 22 (2), pp. 69-84.
[32] Hajji, W., Tso, F.P. Understanding the performance of low power raspberry pi cloud for big data (2016) Electronics (Switzerland), 5 (2), art. no. 29, .
[33] Hani, Q.B., Ditcher, J.P. Stand-Out Segmentation Access Control for Cloud Outsourced Data (2017) Proceedings - 2017 IEEE 1st International Conference on Edge Computing, EDGE 2017, art. no. 8029278, pp. 210-215.
[34] Hasan, R., Hossain, M., Khan, R. Aura: An incentive-driven ad-hoc IoT cloud framework for proximal mobile computation offloading (2017) Future Generation Computer Systems, . Article in Press.
[35] Hong, H.-J. From cloud computing to fog computing: Unleash the power of edge and end devices (2017) Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom, 2017-December, pp. 331-334.
[36] Huang, C.-Y., Xu, K. Reliable realtime streaming in vehicular cloud-fog computing networks (2016) 2016 IEEE/CIC International Conference on Communications in China, ICCC 2016, art. no. 7636838, .
[37] Hung, S.-C., Hsu, H., Lien, S.-Y., Chen, K.-C. Architecture harmonization between cloud radio access networks and fog networks (2015) IEEE Access, 3, art. no. 07359093, pp. 3019-3034.
[38] Jalali, F., Hinton, K., Ayre, R., Alpcan, T., Tucker, R.S. Fog computing may help to save energy in cloud computing (2016) IEEE Journal on Selected Areas in Communications, 34 (5), art. no. 7439752, pp. 1728-1739.
[39] Jayalath, C., Stephen, J.J., Eugster, P. Atmosphere: A universal cross-cloud communication infrastructure (2013) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8275 LNCS, pp. 163-182.
[40] Kanapram, D., Lamanna, G., Repetto, M. Exploring the trade-off between performance and energy consumption in cloud infrastructures (2017) 2017 2nd International Conference on Fog and Mobile Edge Computing, FMEC 2017, art. no. 7946418, pp. 121-126.
[41] Khalid, A., Shahbaz, M. Adaptive deadline-aware scheme (ADAS) for data migration between cloud and fog layers (2018) KSII Transactions on Internet and Information Systems, 12 (3), pp. 1002-1015.
[42] Khalid, A., Shahbaz, M. Service architecture models for fog computing: A remedy for latency issues in data access from clouds (2017) KSII Transactions on Internet and Information Systems, 11 (5), pp. 2310-2345.
[43] Kim, O.T.T., Dang Tri, N., Nguyen, V.D., Tran, N.H., Hong, C.S. A shared parking model in vehicular network using fog and cloud environment (2015) 17th Asia-Pacific Network Operations and Management Symposium: Managing a Very Connected World, APNOMS 2015, art. no. 7275447, pp. 321-326.
[44] Kirkham, T., Armstrong, D., Djemame, K., Jiang, M. Risk driven Smart Home resource management using cloud services (2014) Future Generation Computer Systems, 38, pp. 13-22.
[45] Kirkman, S.S., Newman, R. Bridging the Cloud Trust Gap: Using ORCON Policy to Manage Consumer Trust between Different Clouds (2017) Proceedings - 2017 IEEE 1st International Conference on Edge Computing, EDGE 2017, art. no. 8029260, pp. 82-89.
[46] Kitanov, S., Janevski, T. Energy efficiency of 5G mobile networks in hybrid Fog and Cloud Computing environment (2016) CEUR Workshop Proceedings, 1746, pp. 41-46.
[47] Krishnan, Y.N., Bhagwat, C.N., Utpat, A.P. Fog computing - Network based cloud computing (2015) 2nd International Conference on Electronics and Communication Systems, ICECS 2015, art. no. 7124902, pp. 250-251.
[48] Kumar, R., Rajalakshmi, S. Mobile cloud computing: Standard approach to protecting and securing of mobile cloud ecosystems Computer Sciences and Applications, CSA 2013, art. no. 6835687, pp. 663-669.
[49] Liang, Y., Wang, T., Bhuiyan, M.Z.A., Liu, A. Research on coupling reliability problem in sensor-cloud system (2017) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10658 LNCS, pp. 468-478.
[50] Lin, Y., Shen, H. CloudFog: Leveraging Fog to Extend Cloud Gaming for Thin-Client MMOG with High Quality of Service (2017) IEEE Transactions on Parallel and Distributed Systems, 28 (2), art. no. 7465785, pp. 431-445.
[51] Lin, Y., Shen, H. Cloud fog: Towards high quality of experience in cloud gaming (2015) Proceedings of the International Conference on Parallel Processing, 2015-December, art. no. 7349605, pp. 500-509.
[52] Lin, Y., Shen, H. Leveraging Fog to Extend Cloud Gaming for Thin-Client MMOG with High Quality of Experience (2015) Proceedings - International Conference on Distributed Computing Systems, 2015-July, art. no. 7164961, pp. 734-735.
[53] Lobo, F.L., Lima, M., Oliveira, H., Harrington, J., El-Khatib, K. SoLVE: A localization system framework for VANets using the cloud and fog computing (2017) DIVANet 2017 - Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications, Co-located with MSWiM 2017, pp. 17-22.
[54] Lubamba, C., Bagula, A. Cyber-healthcare cloud computing interoperability using the HL7-CDA standard (2017) Proceedings - IEEE Symposium on Computers and Communications, art. no. 8024513, pp. 105-110.
[55] Macías-Macías, M., García-Orellana, C.J., González-Velasco, H., Gallardo-Caballero, R. Independent component analysis for cloud screening of meteosat images (2003) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2687, pp. 551-558.

[56] Macías-Macías, M., García-Orellana, C.J., González-Velasco, H.M., Gallardo-Caballero, R., Serrano-Pérez, A. A comparison of PCA and GA selected features for cloud field classification (2002) Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 2527, pp. 42-49.

[57] Mahmoud, M.M.E., Rodrigues, J.J.P.C., Saleem, K., Al-Muhtadi, J., Kumar, N., Korotaev, V. Towards energy-aware fog-enabled cloud of things for healthcare (2018) Computers and Electrical Engineering, 67, pp. 58-69.

[58] Makitalo, N., Ometov, A., Kannisto, J., Andreev, S., Koucheryavy, Y., Mikkonen, T. Safe and Secure Execution at the Network Edge: A Framework for Coordinating Cloud, Fog, and Edge (2018) IEEE Software, . Article in Press.

[59] Manasrah, A.M., Aldomi, A., Gupta, B.B. An optimized service broker routing policy based on differential evolution algorithm in fog/cloud environment (2017) Cluster Computing, pp. 1-15. Article in Press.

[60] Masip-Bruin, X., Marin-Tordera, E., Jukan, A., Ren, G.-J. Managing resources continuity from the edge to the cloud: Architecture and performance (2018) Future Generation Computer Systems, 79, pp. 777-785.

[61] Masip-Bruin, X., Marin-Tordera, E., Gomez, A., Barbosa, V., Alonso, A. Will it be cloud or will it be fog? F2C, A novel flagship computing paradigm for highly demanding services (2017) FTC 2016 - Proceedings of Future Technologies Conference, art. no. 7821744, pp. 1129-1136.

[62] Meng, X., Wang, W., Zhang, Z. Delay-Constrained Hybrid Computation Offloading with Cloud and Fog Computing (2017) IEEE Access, 5, art. no. 8023966, pp. 21355-21367.

[63] Mo, Y., Peng, M., Xiang, H., Sun, Y., Ji, X. Resource allocation in cloud radio access networks with device-to-device communications (2017) IEEE Access, 5, art. no. 7855644, pp. 1250-1262.

[64] Mohamed, N., Al-Jaroodi, J., Jawhar, I., Lazarova-Molnar, S., Mahmoud, S. SmartCityWare: A service-oriented middleware for cloud and fog enabled smart city services (2017) IEEE Access, 5, art. no. 7990214, pp. 17576-17588.

[65] Mohamed, N., Lazarova-Molnar, S., Jawhar, I., Al-Jaroodi, J. Towards Service-Oriented Middleware for Fog and Cloud Integrated Cyber Physical Systems (2017) Proceedings - IEEE 37th International Conference on Distributed Computing Systems Workshops, ICDCSW 2017, art. no. 7979797, pp. 67-74.

[66] Mohamed, N., Lazarova-Molnar, S., Al-Jaroodi, J. Cloud of things: Optimizing smart city services (2017) 2017 7th International Conference on Modeling, Simulation, and Applied Optimization, ICMSAO 2017, art. no. 7934875, .

[67] Mohan, N., Zhou, P., Govindaraj, K., Kangasharju, J. Managing data in computational edge clouds (2017) MECOMM 2017 - Proceedings of the 2017 Workshop on Mobile Edge Communications, Part of SIGCOMM 2017, pp. 19-24.

[68] Mohan, N., Kangasharju, J. Edge-Fog cloud: A distributed cloud for Internet of Things computations (2017) 2016 Cloudification of the Internet of Things, CIoT 2016, art. no. 7872914, .

[69] Mueller, H., Gogouvitis, S.V., Seitz, A., Bruegge, B. Seamless computing for industrial systems spanning cloud and edge (2017) Proceedings - 2017 International Conference on High Performance Computing and Simulation, HPCS 2017, art. no. 8035079, pp. 209-216.

[70] Mueller, H., Gogouvitis, S.V., Haitof, H., Seitz, A., Bruegge, B. Poster abstract: Continuous computing from cloud to edge (2016) Proceedings - 1st IEEE/ACM Symposium on Edge Computing, SEC 2016, art. no. 7774685, pp. 97-98.

[71] Munir, A., Kansakar, P., Khan, S.U. IFCIoT: Integrated Fog Cloud IoT: A novel architectural paradigm for the future Internet of Things (2017) IEEE Consumer Electronics Magazine, 6 (3), art. no. 7948854, pp. 74-82.

[72] Nan, Y., Li, W., Bao, W., Delicato, F.C., Pires, P.F., Zomaya, A.Y. A dynamic tradeoff data processing framework for delay-sensitive applications in Cloud of Things systems (2018) Journal of Parallel and Distributed Computing, 112, pp. 53-66.

[73] Nan, Y., Li, W., Bao, W., Delicato, F.C., Pires, P.F., Dou, Y., Zomaya, A.Y. Adaptive Energy-Aware Computation Offloading for Cloud of Things Systems (2017) IEEE Access, 5, art. no. 8081765, pp. 23947-23957.

[74] Okafor, K.C., Ugwoke, F.N., Obayi, A.A. Evaluation of virtualized osmotic cloud network using discrete event Branch-And-Bound heuristics (2018) 2017 IEEE 3rd International Conference on Electro-Technology for National Development, NIGERCON 2017, 2018-January, pp. 425-437.

[75] Osanaiye, O., Chen, S., Yan, Z., Lu, R., Choo, K.-K.R., Dlodlo, M. From Cloud to Fog Computing: A Review and a Conceptual Live VM Migration Framework (2017) IEEE Access, 5, art. no. 7896564, pp. 8284-8300.

[76] Ostberg, P.-O., Byrne, J., Casari, P., Eardley, P., Anta, A.F., Forsman, J., Kennedy, J., Le Duc, T., Marino, M.N., Loomba, R., Pena, M.A.L., Veiga, J.L., Lynn, T., Mancuso, V., Svorobej, S., Torneus, A., Wesner, S., Willis, P., Domaschka, J. Reliable capacity provisioning for distributed cloud/edge/fog computing applications (2017) EuCNC 2017 - European Conference on Networks and Communications, art. no. 7980667, .

[77] Pahl, C., Helmer, S., Miori, L., Sanin, J., Lee, B. A container-based edge cloud PaaS architecture based on raspberry Pi clusters (2016) Proceedings - 2016 4th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2016, art. no. 7592710, pp. 117-124.

[78] Pham, X.-Q., Man, N.D., Tri, N.D.T., Thai, N.Q., Huh, E.-N. A cost- and performance-effective approach for task scheduling based on collaboration between cloud and fog computing (2017) International Journal of Distributed Sensor Networks, 13 (11), .

[79] Pradilla, J.V., Palau, C.E. Micro Virtual Machines (MicroVMs) for Cloud-assisted Cyber-Physical Systems (CPS) (2016) Internet of Things: Principles and Paradigms, pp. 125-142.

[80] Ramirez, W., Masip-Bruin, X., Marin-Tordera, E., Souza, V.B.C., Jukan, A., Ren, G.-J., Gonzalez de Dios, O. Evaluating the benefits of combined and continuous Fog-to-Cloud architectures (2017) Computer Communications, 113, pp. 43-52.

[81] Renner, T., Kliem, A., Kao, O. The Device Cloud - Applying Cloud Computing Concepts to the Internet of Things (2014) Proceedings - 2014 IEEE International Conference on Ubiquitous Intelligence and Computing, 2014 IEEE International Conference on Autonomic and Trusted Computing, 2014 IEEE International Conference on Scalable Computing and Communications and Associated Symposia/Workshops, UIC-ATC-ScalCom 2014, art. no. 7306981, pp. 396-401.

[82] Seo, D.B., Jeon, Y.-B., Lee, S.-H., Lee, K.-H. Cloud computing for ubiquitous computing on M2M and IoT environment mobile application (2016) Cluster Computing, 19 (2), pp. 1001-1013.

[83] Shah-Mansouri, H., Wong, V.W.S. Hierarchical Fog-Cloud Computing for IoT Systems: A Computation Offloading Game (2018) IEEE Internet of Things Journal, . Article in Press.

[84] Sharma, P.K., Chen, M.-Y., Park, J.H. A Software Defined Fog Node Based Distributed Blockchain Cloud Architecture for IoT (2018) IEEE Access, 6, art. no. 8053750, pp. 115-124.

[85] Sharma, S.K., Wang, X. Live Data Analytics with Collaborative Edge and Cloud Processing in Wireless IoT Networks (2017) IEEE Access, 5, art. no. 7882669, pp. 4621-4635.

[86] Shekhar, S., Gokhale, A. Dynamic resource management across cloud-edge resources for performance-sensitive applications (2017) Proceedings - 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2017, art. no. 7973763, pp. 707-710.

[87] Shekhar, S., Gokhale, A. Poster abstract: Enabling IoT applications via dynamic cloud-edge resource management (2017) Proceedings - 2017 IEEE/ACM 2nd International Conference on Internet-of-Things Design and Implementation, IoTDI 2017 (part of CPS Week), pp. 331-332.

[88] Shen, S., Huang, L., Zhou, H., Yu, S., Fan, E., Cao, Q. Multistage Signaling Game-Based Optimal Detection Strategies for Suppressing Malware Diffusion in Fog-Cloud-Based IoT Networks (2018) IEEE Internet of Things Journal, 5 (2), pp. 1043-1054.

[89] Shirazi, S.N., Gouglidis, A., Farshad, A., Hutchison, D. The extended cloud: Review and analysis of mobile edge computing and fog from a security and resilience perspective (2017) IEEE Journal on Selected Areas in Communications, 35 (11), art. no. 8060526, pp. 2586-2595.

[90] Shroff, G.Enterprise cloud computing: Technology, architecture, applications (2010) Enterprise Cloud Computing: Technology, Architecture, Applications, 9780521760959, pp. 1-273.

[91] Simonet, A., Lebre, A., Orgerie, A.-C. Deploying distributed cloud infrastructures: Who and at what cost? (2016) Proceedings - 2016 IEEE International Conference on Cloud Engineering Workshops, IC2EW 2016, art. no. 7527837, pp. 178-183.

[92] Sinaeepourfard, A., Garcia, J., Masip-Bruin, X., Marin-Tordera, E. A Novel Architecture for Efficient Fog to Cloud Data Management in Smart Cities (2017) Proceedings - International Conference on Distributed Computing Systems, art. no. 7980258, pp. 2622-2623.

[93] Singh, H., Bonev, B., Petkov, P., Patil, S. Cloud attenuation model at millimeter frequency bands (2018) 2017 International Conference on Infocom Technologies and Unmanned Systems: Trends and Future Directions, ICTUS 2017, 2018-January, pp. 175-180.

[94] Siravenha, A.C., Sousa, D., Bispo, A., Pelaes, E. Evaluating inpainting methods to the satellite images clouds and shadows removing (2011) Communications in Computer and Information Science, 260 CCIS, pp. 56-65.

[95] Souza, V.B., Masip-Bruin, X., Marín-Tordera, E., Sànchez-López, S., Garcia, J., Ren, G.J., Jukan, A., Juan Ferrer, A. Towards a proper service

placement in combined Fog-to-Cloud (F2C) architectures (2018) Future Generation Computer Systems, 87, pp. 1-15.
[96] Souza, V.B., Masip-Bruin, X., Marin-Tordera, E., Ramirez, W., Sanchez-Lopez, S. Proactive vs reactive failure recovery assessment in combined Fog-to-Cloud (F2C) systems (2017) IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD, 2017-June, art. no. 8031528,
[97] Souza, V.B.C., Ramirez, W., Masip-Bruin, X., Marin-Tordera, E., Ren, G., Tashakor, G. Handling service allocation in combined Fog-cloud scenarios (2016) 2016 IEEE International Conference on Communications, ICC 2016, art. no. 7511465, .
[98] Srinivas, K.K., Ramana, T.V. Severe cause of cloud attenuation and rain attenuation on space communication link at millimetre band and differentiation between rain attenuation and cloud attenuation (2017) Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017, art. no. 7976797, pp. 255-259.
[99] Stankova, E.N., Korkhov, V.V., Kulabukhova, N.V., Vasilenko, A.Y., Holod, I.I. Computational environment for numerical modeling of the results of cloud seeding (2016) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9788, pp. 454-462.
[100] Syed, M.H., Fernandez, E.B. Cloud ecosystems support for internet of things and DevOps using patterns (2016) Proceedings - 2016 IEEE 1st International Conference on Internet-of-Things Design and Implementation, IoTDI 2016, art. no. 7471379, pp. 301-304.
[101] Taherkordi, A., Eliassen, F. Poster abstract: Data-centric IoT services provisioning in Fog-Cloud computing systems (2017) Proceedings - 2017 IEEE/ACM 2nd International Conference on Internet-of-Things Design and Implementation, IoTDI 2017 (part of CPS Week), pp. 317-318.
[102] Tang, C., Wei, X., Xiao, S., Chen, W., Fang, W., Zhang, W., Hao, M. A Mobile Cloud Based Scheduling Strategy for Industrial Internet of Things (2018) IEEE Access, 6, pp. 7262-7275.
[103] Tao, M., Ota, K., Dong, M. Foud: Integrating Fog and Cloud for 5G-Enabled V2G Networks (2017) IEEE Network, 31 (2), art. no. 7884943, pp. 8-13.
[104] Teerapittayanon, S., McDanel, B., Kung, H.T. Distributed Deep Neural Networks over the Cloud, the Edge and End Devices (2017) Proceedings - International Conference on Distributed Computing Systems, art. no. 7979979, pp. 328-339.
[105] Tinini, R.I., Reis, L.C.M., Batista, D.M., Figueiredo, G.B., Tornatore, M., Mukherjee, B. Optimal Placement of Virtualized BBU Processing in Hybrid Cloud-Fog RAN over TWDM-PON (2018) 2017 IEEE Global Communications Conference, GLOBECOM 2017 - Proceedings, 2018-January, pp. 1-6.
[106] Truong, H.-L., Narendra, N.C., Lin, K.-J. Notes on ensembles of IoT, network functions and clouds for service-oriented computing and applications (2018) Service Oriented Computing and Applications, 12 (1), .
[107] Tseng, F.-H., Cho, H.-H., Chang, K.-D., Li, J.-C., Shih, T.K. Application-oriented offloading in heterogeneous networks for mobile cloud computing (2018) Enterprise Information Systems, 12 (4), pp. 398-413.
[108] Van Kempen, A., Crivat, T., Trubert, B., Roy, D., Pierre, G. MEC-ConPaaS: An Experimental Single-Board Based Mobile Edge Cloud (2017) Proceedings - 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2017, art. no. 7944867, pp. 17-24.
[109] Varghese, B., Buyya, R. Next generation cloud computing: New trends and research directions (2018) Future Generation Computer Systems, 79, pp. 849-861.
[110] Wang, H., Wang, Z., Domingo-Ferrer, J. Anonymous and secure aggregation scheme in fog-based public cloud computing (2018) Future Generation Computer Systems, 78, pp. 712-719.
[111] Wang, J., Yang, Y., Zhang, G., Yu, S. Climatic trend of cloud amount related to the aerosol characteristics in Beijing during (2010) Acta Meteorologica Sinica, 24 (6), pp. 762-775.
[112] Wang, L., Xie, W., Li, L., Pei, J. A thin cloud and fog removal method for remote sensing multi-spectral images (2013) Shenzhen Daxue Xuebao (Ligong Ban)/Journal of Shenzhen University Science and Engineering, 30 (6), pp. 592-597.
[113] Wang, S., Urgaonkar, R., He, T., Chan, K., Zafer, M., Leung, K.K. Dynamic Service Placement for Mobile Micro-Clouds with Predicted Future Costs (2017) IEEE Transactions on Parallel and Distributed Systems, 28 (4), pp. 1002-1016.
[114] Wang, T., Li, Y., Chen, Y., Tian, H., Cai, Y., Jia, W., Wang, B. Fog-based evaluation approach for trustworthy communication in sensor-cloud system (2017) IEEE Communications Letters, 21 (11), art. no. 8010796, pp. 2532-2535.
[115] Wang, T., Zeng, J., Bhuiyan, M.Z.A., Tian, H., Cai, Y., Chen, Y., Zhong, B. Trajectory Privacy Preservation Based on a Fog Structure for Cloud Location Services (2017) IEEE Access, 5, art. no. 7918561, pp. 7692-7701.
[116] 2016) IEEE Workshop on Local and Metropolitan Area Networks, 2016-August, art. no. 7548853, .
[117] Yannuzzi, M., Milito, R., Serral-Gracia, R., Montero, D., Nemirovsky, M. Key ingredients in an IoT recipe: Fog Computing, Cloud computing, and more Fog Computing(2014) 2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD 2014, art. no. 7033259, pp. 325-329.
[118] Zeng, J., Wang, T., Jia, W., Peng, S., Wang, G. A Survey on Sensor-Cloud (2017) Jisuanji Yanjiu yu Fazhan/Computer Research and Development, 54 (5), pp. 925-939.
[119] Zeng, J., Wang, T., Lai, Y., Liang, J., Chen, H. Data Delivery from WSNs to Cloud Based on a Fog Structure (2017) Proceedings - 2016 International Conference on Advanced Cloud and Big Data, CBD 2016, art. no. 7815194, pp. 104-109.
[120] Zhao, W., Liu, J., Hara, T. Optimal Replica Distribution in Edge-Node-Assisted Cloud-P2P Platforms for Real-Time Streaming (2018) IEEE Transactions on Vehicular Technology, . Article in Press.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 94-105 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

An Extension of Throughput Drop Estimation Model for Three-Link Concurrent Communications under Partially Overlapping Channels and Channel Bonding in IEEE 802.11n WLAN

Kwenga Ismael Munene[1], Nobuo Funabiki[*,1], Md. Manowarul Islam[1], Minoru Kuribayashi[1], Md. Selim Al Mamun[2], Wen-Chung Kao[3]

[1]*Department of Electrical and Communication Engineering, Okayama University, Okayama, Japan.*

[2]*Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Bangladesh.*

[3]*Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan.*

ARTICLEINFO

Article history:
Received: 22 April, 2019
Accepted: 07 June, 2019
Online: 15 July, 2019

Keywords:
WLAN
Access point
IEEE 802.11n
Partially overlapping channel
Throughput estimation model
Three links

ABSTRACT

The IEEE 802.11n wireless local-area network (WLAN) has been extensively deployed around the world due to the flexibility, lower cost, and the high-speed data transmission capability at 2.4 GHz ISM band. However, in the WLAN deployment, one key challenge is to optimize the channel assignment of access-points (APs) under the small number of partially overlapping channels (POCs) to reduce radio interference, particularly for the channel bonding. In POCs, the frequency spectrums of adjacent channels are partially overlapped with one another, which will result to low throughput for concurrently communicating links using them. The accurate throughput estimation of a link is critical in the optimal WLAN deployment. Previously, we studied the throughput drop estimation model using the receiving signal strength (RSS) from the interfered link for two concurrently communicating links under POCs. In this paper, based on measurement results, we have extended this model for three concurrently communicating links. The accuracy of this model extension is verified by comparing the estimated results with the measured ones.

1 Introduction

Nowadays, the *IEEE 802.11 wireless local-area network (WLAN)* has been used across various sectors around the world, due to the unlicensed frequency bands in wireless medium and availability of low-cost devices [1]–[3]. It is suggested that WLAN has become a crucial business and the common asset [4]. For the access to the Internet using WLAN, a host is connected to an *access point (AP)* wirelessly, where the AP is connected to the Internet with wires. Therefore, a host can move randomly in the network area with the current association. As the data transmission speed increases due to technological advancements, WLAN has proved popular even for static settings of personal computers in offices and homes [5].

The IEEE 802.11 WLAN can function in two unlicensed frequency bands [6, 7]. One is the 2.4 GHz *Industrial, Scientific, and Medical (ISM)* band, and the other is the 5 GHz *Unlicensed National Information Infrastructure (U-NII)* band.

For either band, the IEEE 802.11 standards define the small number of channels for use. Each channel has 22 MHz width, conventionally called a 20 MHz channel. The frequency gap between two adjacent channels is merely *5MHz*. Thus, the spectrums of adjacent channels are partially overlapped with one another, called the *partially overlapping channels (POCs)*. Figure 1 shows the 2.4 GHz spectrum with 40 MHz channels to illustrate the POCs in IEEE 802.11n.

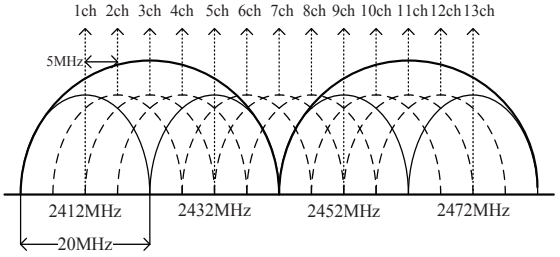


Figure 1: 40MHz channels at 2.4 GHz band.

While the IEEE 802.11 WLANs referred to in this paper use the unlicensed frequency bands, *the Industrial, Sci-*

*Corresponding Author: Nobuo Funabiki, Dep. of Elect. and Comm. Eng., Okayama Univ., Okayama, Japan, Email: funabiki@okayama-u.ac.jp

entific, and Medical (ISM) 2.4 GHz band and *the Unlicensed National Information Infrastructure (U-NII) 5 GHz band*, the 802.11y is licensed to use [8]. The unlicensed IEEE 802.11ad, which represents modifications of IEEE 802.11 physical layer (PHY) and medium access control layer (MAC) [9], provides the very high speed, which is equivalent to the fiber optic. However, its short range limitation coupled with the expensive hardware makes the global adoption of IEEE 802.11ad to be low [10].

Similarly, the IEEE 802.11af is advantageous due to its long-range transmission and low power consumption. However, it requires expensive band-specific hardware, which is not readily available in the global market where it is primarily used in US and Canada. In addition, since it utilizes the frequencies for unused TV channels, they may not be available everywhere [11].

The proposed model in this paper targets WLANs on the IEEE 802.11n unlicensed frequency band. Since this protocol has been widely adopted, the related devices are much easily and cheaply available, compared to those of emerging standards of the IEEE 802.11ad and IEEE 802.11af.

At the 2.4 GHz band, 13, 20MHz POCs are possible, which indicates that the number of *orthogonal channels (OCs)* is four at most. Since orthogonal channels do not overlap with one another, APs assigned OCs are not interfered for the medium access, if any pair of nearby APs do not operate on the same OC.

Currently, the *IEEE 802.11n* is most used in WLAN due to the high-speed data transmission capability at the 2.4 GHz band using the *channel bonding* that combines two neighboring 20 MHz channels together to form one 40 MHz channel as shown in Figure 1, [12]–[14]. Then, the number of OCs is reduced to at most two, and the channel assignment using OCs to the APs without interference cannot be avoided. [15] has shown that channel bonding improves performance, although it may increase frequency competition with adjacent LANs.

To overcome this limitation of the OC assignment in WLAN, it has been reported that the POCs will be beneficial to enhance the performance by fully utilizing the available spectrum [16, 17]. Then, the *throughput estimation model* under POCs is essential to identify the proper POC assignment to the APs while evaluating the performance accurately. Specifically, in a dense WLAN, the model must consider the *throughput drop* caused by the interference that each AP suffers from the neighboring APs on the adjacent POCs.

Previously, we have studied the *throughput estimation model* for a single communicating link, which consists of the *log-distance path loss model* and the *sigmoid function* [18]. Then, we move to estimate the *throughput drop* caused by the interference for two concurrently communicating links under POCs [19]. It is examined according to the *receiving signal strength (RSS)* from the interfering link and the *channel distance (chD)* between the two links.

The proper use of POCs should be considered for the full utilization of the limited frequencies, since there are limited orthogonal channels available in the IEEE 802.11n protocol with the channel bonding. Besides, the estimated throughput should be directly used as the metric for selecting POCs by estimating the network performance, since the interference does not exhibit the exact performance of the network, which has been observed in this paper.

Our model considers the channel bonding and POCs, to directly estimate the throughput from the channel distance and the RSS of the interfering AP. From the estimated throughput, the POC assignment can be optimized under the channel bonding.

Besides, in our previous studies [20, 21], we considered the indoor network environment inside a building for the target field, since WLANs are usually deployed there. A lot of environmental factors, such as wall attenuations and multipath effects, can affect the throughput performance. To accurately estimate the throughput, we have developed the throughput estimation model that consider the environmental factors.

In this research work, we extend the *throughput drop estimation model* for three concurrently communicating links under POCs. Firstly, we conduct extensive measurements of *receiving signal strength (RSS)* and throughputs under three links. Then, we extend the model based on the measurement results.

To confirm the effectiveness of our proposal, we compare the estimated throughputs by the model with the measured ones, and through simulations. It has been demonstrated that the proposal can estimate the throughput for three interfered links under POCs with the considerably high degree of accuracy.

The rest of this paper is organized as follows: Section 2 describes related works. Section 3 discusses the implementations of the auto-channel selection in vendor AP devices and the significance of the throughput estimation model. Section 4 reviews our previous studies. Section 5 presents experiments with three concurrently communicating links under POCs. Section 6 proposes the model extension for them. Section 7 evaluates the proposal with experiments and simulations. Section 8 introduces the application to the channel assignment. Section 9 concludes this paper with future directions.

2 Related Works

In this section, we discuss specific works related to our proposal. A significant amount of research works have addressed the problem of interferences in WLAN to improve the throughput performance by applying partially overlapping channels.

In [22], Mishra et al. revealed that the orthogonal channel assignment to APs in WLAN is inefficient in a network field, if a substantial number of access points are deployed there. In such a case, the number of APs in the network field is higher compared to that of orthogonal channels, hence any AP could exist in the interference ranges of other APs as indicated in [23].

In [24], Mishra et al. also noted that partially overlapping channels have previously been avoided due to their interferences. However, through their model and measurements, it is demonstrated that the careful use of some POCs does not only improve the spectrum utilization but also the throughput performance.

In [25], Feng et al. proposed POCs as a means of reducing interference from the traditional use of OCs. Through

testbed experiments, they reported that POCs offers better flexibility in wireless frequency allocation, and can increase overall network performance. Similarly, in [26], Zhang et al. proposed the use of POCs for interference management which outperforms the OCs.

In [27], Zhao et al. observed that with increased frequency reuse which has been reported in POCs, the network capacity can be scaled up through activating more APs in the network field as the expected distance from the users to their connected APs becomes smaller. However, it is reported that since POCs are partially overlapped with each other, one AP may suffer interference from multiple other APs. They conclude by proposing a need to have a maximum and a minimum bound for the number of APs based on the size of the network field and the number of POCs available.

In [28], Mukherjee et al. explored the AP selection, channel assignment, and the host association. They reported that the three aspects are critical in maximizing the throughput in WLAN, where the appropriate combination of both non-overlapping channels and POCs can further improve the overall throughput of the network.

In [29], Tewari et al. emphasized that multiple overlapping transmissions cause low network performance due to high interference from the limited non-overlapping channels. They have proposed a combined power control and POC assignment, where the former reduces the AP's transmission and the interference range while the latter improves spatial reuse.

In [30], Zhao et al. demonstrated that the effect of interference on the network performance depends on the channel separation and the degree of frequency overlap among the interfered APs. In particular, it is noted that two interferes cause higher performance deterioration than a single interfere, but less than twice the interference caused by a single interfere or the summation of interference from the two interferes.

Furthermore, in [31], Su et al. measured the interference among the APs when POCs are used from the perspective of the physical characteristics of the communication. They reported a clear distinction between in-range and hidden terminal interferences with POCs, where the latter has worse throughput performance due to high packet loss.

In [28], POC interference is evaluated by considering the interference factor defined in [32]. Where more than one APs interferes, [27] [28] and [29] considers summation of interfering signal powers at the target node. However, since the MAC protocol lowers the data transmission rate of the target AP depending on the level of interference with individual interfere, calculating the summation may fail to identify a real value of interference.

At present, there is no research work that has proposed the throughput drop estimation model for multiple interfering APs under POCs that considers the channel distance and the interfering RSS.

3 Auto-channel Selection Implementations and Model Significance

In this section, we survey implementations of the auto-channel selection in vendor AP devices, and discuss the significance of the throughput estimation model under it.

The auto-channel selection is the dynamic adjustment mechanism of the assigned channel to the AP, in order to avoid radio interferences from other WiFi devices. Each AP vendor has its own implementation approach.

In [33], Cisco describes their implementation of the auto-channel selection, where the AP operating in the 2.4 GHz band with 11 POCs can be set only to one of the three orthogonal channels of 1, 6, and 11 under the non-bonded channels with the *(20 MHz)* width. It means that up to three APs can be assigned the channels. The assignment for other POCs must be done manually.

However, when the channel bonding with the *(40 MHz)* width is applied, which is commonly used in IEEE 802.11n WLANs to improve the throughput performance, the AP can have only one choice of channel 3, since all the other channels are partially overlapping with it. It means that the auto-channel selection cannot work for the channel bonding.

On the other hand, our proposal can assign any POC to the AP with the channel bonding by considering the throughput drop from the other up to two interfered links.

Besides, from our experimental observations, the orthogonal channels can interfere with each other when the APs are closely located together, which would further affect the auto-channel selection in the Cisco Meraki APs.

Furthermore, Cisco's approach scans for a possible channel change after every 15 minutes, in addition to the three necessary situations: 1) a new AP joins the network, 2) the network administrator manually runs the auto-channel), and 3) the currently assigned channel fails to work. This would lead to the performance overhead, since the services of an AP must be stopped when channel change is applied. To avoid the frequent service stops, our approach changes channels of APs when an AP joins or leaves the network.

In [34], Fujitsu reports the auto-channel selection such that the total interference is minimized, where the details of the implementation are not given. However, the interference index is not a good metric to evaluate the throughput performance, because the interference and the throughput performance are not proportional as our paper reported. Instead, the throughput drop of each link from the other interfered links should be evaluated to more accurately estimate the network performance under a lot of interfered links.

In [35], Buffalo implements the auto-channel selection that tries to find the non-interfered channel for a new AP. By scanning all the channels, it selects a non-interfered channel if it exists. However, they also reported that the channel selection performs poorly, if no non-interfered channels exist. In real network environments, as demonstrated in our paper, several APs coexist together where any channel is interfered by an existing AP. On the other hand, our proposal is applicable in such real cases.

Similarly, in [36], Vodafone limits the auto-channel selection only to the non-interfered channels or the least interfered channel. In [37], Google WiFi works on the similar principle by selecting an orthogonal channel.

It should be pointed out that the proposed throughput estimation model can be incorporated into an auto-channel selection mechanism to improve the resulting performance of WLAN. Our model considers the channel bonding, POCs, and the direct throughput estimation. Therefore, we believe

that this model can improve the auto-channel selection. In future works, we will compare the performance by our model with existing auto-channel selections.

4 Review of Throughput Estimation Model

In this section, we review our previous works.

4.1 Motivation of Empirical Model

In [5], Reis et al. indicates that most of physical protocol explorations with respect to interference may adopt simple abstract models with multiple assumptions, including that the signal propagation obeys a simple function of the distance, the radio coverage area forms a circle, and the interference range is twice of the transmission range. Unfortunately, experimental data using a real WLAN have shown that all of these models appear to be largely inaccurate [38, 39].

In contrast to a physical model, an empirical model is based on observations on actual network environments. Thus, the empirical model for interference in WLAN is expected to be more descriptive and accurate compared to a physical model. Therefore, we have developed the empirical throughput drop estimation model for multiple-link concurrent communications based on experimental results.

4.2 Definitions of Three Distances

The *channel distance*, the *physical distance*, and the *link distance* are defined to illustrate the throughput estimation model under partially overlapping channels.

4.2.1 Channel Distance

The *channel distance* (*chD*) of the two links is defined as the least channel difference between the channels of these links. For instance, if the two links are activated with the same channel, then *chD* is 0, and they will be entirely overlapped. When one link is assigned *channel 3* and another link is *channel 5*, *chD* is 2, in this case, these channels are overlapped by 50% for 20MHz, and by 75% for 40MHz. The largest *chD* is 12 for 20MHz and 8 for 40MHz, in which any frequency overlapping does not exist theoretically.

4.2.2 Physical Distance

The *physical distance* (*phD*) is defined as the Euclidean distance between the two links. As it represents the separation distance among the interfering links, the farther the two links are placed, that is, the higher the physical distance is, the interference will become lesser. By increasing the physical distance between the links, the signal interference between them fades due to the path loss and the absorption by obstacles on the path.

4.2.3 Link Distance

The Euclidean distance between the sender and the receiver of the link is defined as the *link distance*. The longer link distance generally reduces RSS at the receiver and degrades the throughput.

4.3 Throughput Estimation Model

Generally, the throughput of a link will be affected by a variety of factors like the modulation and coding scheme (MCS), the transmission power, the transmission distance, and the channel interference [13, 40], which have made it hard for the theoretical calculation of the throughput. In [18], we proposed the throughput estimation model for the single link communication based on empirical results. This model first estimates the *receiving signal strength (RSS)* at the host using the *log-distance path loss model* [41]. Next, it estimates the throughput from the RSS using the *sigmoid function*.

4.3.1 Received Signal Strength Estimation

Firstly, the Euclidean distance *d (m)* between each link (AP/host pair) is calculated by:

$$d = \sqrt{(AP_x - H_x)^2 + (AP_y - H_y)^2} \tag{1}$$

Here, AP_x, AP_y and H_x, H_y denotes the x and y coordinates for the AP and the host respectively.

Then, *RSS* at the host from the AP is estimated by:

$$P_d = P_1 - 10\alpha \log_{10} d - \sum_k n_k W_k \tag{2}$$

Here, P_d denotes *RSS* (dBm) at the host, P_1 does *RSS* at the $1m$ distance from the AP where no obstacle exists, α does the path loss exponent, *d (m)* does the Euclidean distance calculated between the AP and the host, n_k does the number of *type_k* obstacles found on the path between the AP and the host, and W_k represents the signal attenuation factor (dB) for *type_k* obstacle.

4.3.2 Throughput Estimation

From P_d, the throughput is estimated by;

$$tp_{ij} = \frac{a}{1 + e^{-(\frac{(120+P_d)-b}{c})}} \tag{3}$$

where tp_{ij} denotes the estimated throughput ($Mbps$) and, a, b, and c are constant coefficients.

4.4 Throughput Drop Estimation Model for Two-Link Concurrent Communications

In [19], we proposed the *throughput drop estimation model* to consider the frequency interference between concurrently communicating adjacent two links using interfering channels. The *throughput drop* of the target link can be estimated by the receiving signal strength at the receiver, called the *interfering RSS*, (RSS^i), from the interfered link, and the channel distance *chD* between the two links:

$$tpD(RSS^i, chD) = p(chD) \times \ln(q(chD) + RSS^i) + r(chD). \tag{4}$$

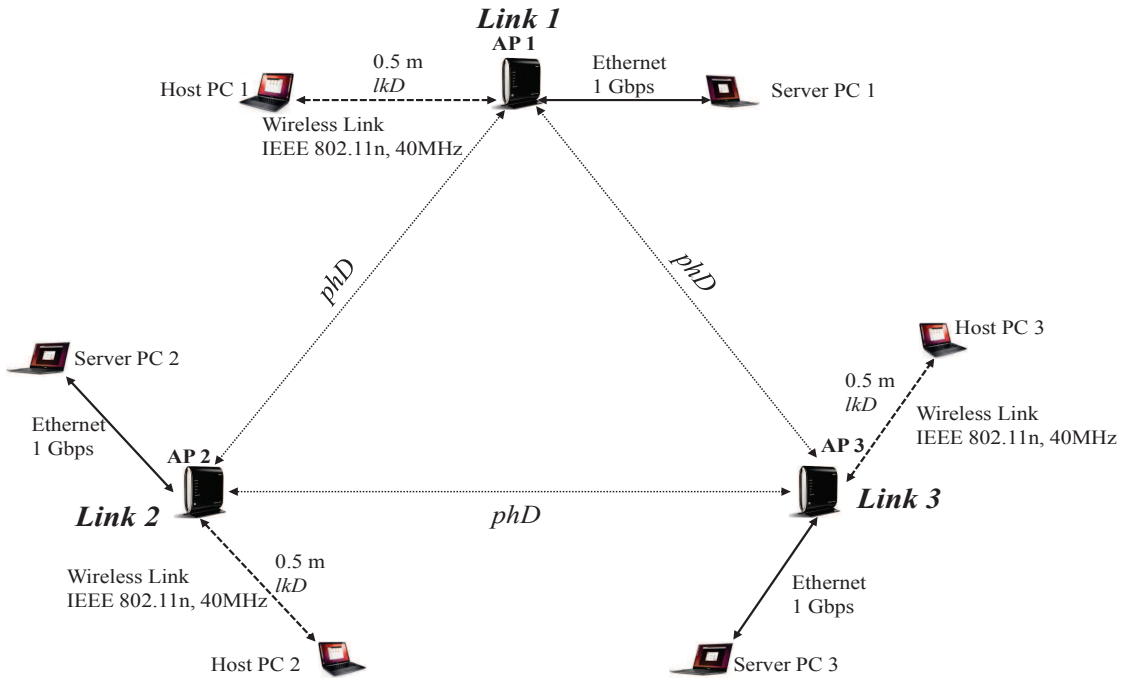


Figure 2: Measurement setup.

where $tpD(RSS^i, chD)$ represents the estimated throughput drop ($Mbps$), and $p(chD)$, $q(chD)$, and $r(chD)$ represent constants determined by the channel distance (chD).

The values of the three constants p, q, and r in the model were obtained from the throughput drop measurement results for every channel distance. Table 1 shows them.

Table 1: Constants for each channel distance.

channel distance	p	q	r
0	27	88.17	-20
1	27	87.36	-20
2	27	89.00	-22
3	25	94.50	-22
4	33	92.00	-56
5	34	92.00	-57
6	45	91.00	-98
7	45	88.00	-100
8	40	75.50	-80

Then, the throughput estimation model under POCs for two interfered links is modified by:

$$tp^i_{ij} = tp_{ij} - tpD(RSS^i_j, chD^i_j) \quad (5)$$

where tp^i_{ij} represents the throughput of the target link from node i to node j under the interference from the interfered link, tp_{ij} does the throughput of this link estimated by the original model, RSS^i_j does RSS from the interfered link at node j, and chD^i_j does the channel distance between the two links.

5 Experiments in Three-Link Concurrent Communications

In this section, we present the experiment results to examine the extension of the throughput drop estimation model for three concurrently communicating links.

5.1 Experiment Setup

Table 2 indicates the necessary devices and software used in the experiments. It is noticed that if different devices are used, the parameter values of the model will vary accordingly.

Table 2: Measurement devices and software.

Access Point (*all links*)	
maker/ type	NEC WG2600HP
protocol	IEEE 802.11n
operating band	2.4 GHz
frequency width	40 MHz
host PC (*all links*)	
maker/ type	Toshiba dynabook R731/B
operating system	Ubuntu 14.04 LTS
processor	Intel Core i5-2520M 2.54 Ghz
network adapter	Intel HM65 Express chipset
server PC ($link_1$)	
maker/ type	Toshiba dynabook R731/B
processor	Intel Core i5-2520M 2.54 Ghz
operating system	Ubuntu 14.04 LTS
network adapter	Intel HM65 Express chipset
server PC ($link_2$, $link_3$)	
maker/ type	Fujitsu lifebook S761/C
processor	Intel Core i5-2520M 2.5GHz
operating system	Ubuntu 14.04 LTS
network adapter	Mobile Intel QM67 Express

Figure 2 shows the topology of three links ($link_1$, $link_2$ and $link_3$) in experiments. The server PC connects to the AP through the Gigabit Ethernet, and the host PC is connected to the AP by the 11n wireless link. The host is 0.5m from its connected AP, and each AP is 5m from the other AP. The maximum transmission power is selected for any AP with the equal antenna gain [16]. The *minimax AP setup optimization approach* in [18] is applied to maximize the throughput for each link. The *iperf 2.05* [42] is used to generate the TCP traffic between the server and the host.

5.2 Experiment Fields

The experiments were carried out in indoor fields on the 2nd floor of *Graduate School of Natural Science and Technology Building* and the 3rd floor of *Engineering Building #2* in Okayama University. Each field consists of several rooms, walls, and floors, which can affect the throughput through the multi-path effect. RSS^i and the throughputs are measured while increasing the physical distance, *phD*, between the two APs and their channel distance, *chD*. By using *Homedale* [43], the frequencies of the links under measurements and the external interfering links are recorded. Figure 3 shows the site while Figure 4 reveals the frequency utilization for measurements, where several channels are highly utilized by non-target APs.

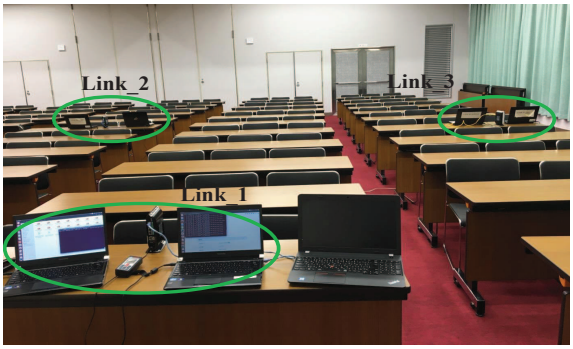
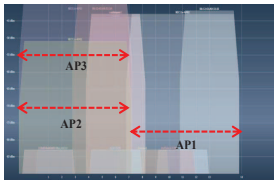
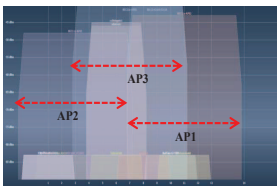


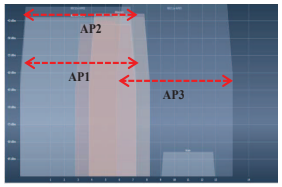
Figure 3: Experiment in Graduate School Building.



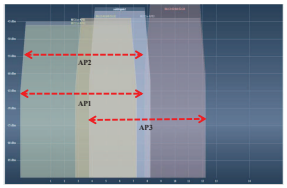
(a) AP1= ch11, AP2 = ch3, AP3 = ch3



(b) AP1= ch11, AP2 = ch3, AP3 = ch7



(c) AP1= ch3, AP2 = ch3, AP3 = ch8



(d) AP1= ch3, AP2 = ch3, AP3 = ch10

Figure 4: Frequency utilization in experiments.

5.3 Experimental Results

For AP_1 (and AP_2 for three links), the bonded channel 3 is always assigned. For AP_2 (AP_3 for three links), the assigned channel is changed from 3 to 11 one by one, so that the channel distance *chD* increases from 0 to 8. The throughput was measured at the same time for all the links. The experiments were conducted on weekends and at night on weekdays to reduce the interference from other Wi-Fi devices.

Figures 5 and 6 show measured individual and average throughput under concurrently communicating two links and three links, respectively. The individual throughput fluctuates, because the contention among the links is not well resolved by the current *carrier sense mechanism* [44], which may cause the unfair channel occupation among them. Thus, in this paper, we use the average throughput among them for the single link throughput.

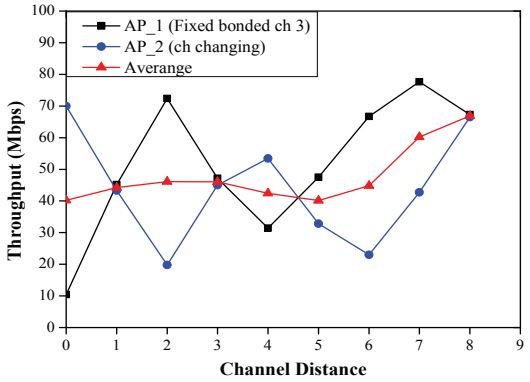


Figure 5: Throughput results for two links.

Figure 6 demonstrates that when the three APs are assigned the same channel, one AP takes the entire medium, which makes the others have the lower throughput. As *chD* increases, the throughput of AP_3 will enhance gradually, due to the reduced channel interference.

In this experimental setup, the measured maximum throughput of one link is about 140$Mbps$ under no interference. Then, it drops significantly to about 40$Mbps$ under two links, when the channel distance is smaller than seven. Furthermore, it drops to about 20$Mbps$ under three links.

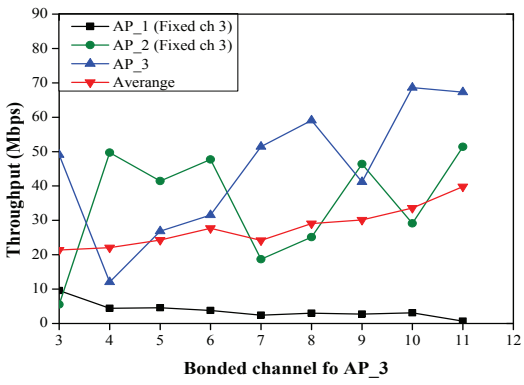


Figure 6: Throughput results for three links.

5.4 Observations from Experiment Results

According to the evaluation, it is observed that for the target link, the interference from the first interfering link causes the larger throughput drop (from 140$Mbps$ to 40$Mbps$), and the interference from the second causes the smaller drop (from 40$Mbps$ to 20$Mbps$). That is to say, when the target link is interfered by the first link, the rate adaptation mechanism lowers the transmission rate by adopting the robust MCS. Then, the second interfering link further lowers the rate by adopting the more robust MCS [30][31]. Here, the rate can be lowered exponentially in MCS.

6 Model Extension for Three-Link Concurrent Communications

In this section, we present the extension of the throughput drop estimation model for three concurrently communicating links based on experimental results.

6.1 Idea of Throughput Drop Estimation under Three Links

As observed before, the throughput drop from multiple interfering links can be estimated one by one through calculating each drop using the model in Section 4.4 in descending order of the interference. That is, the drop from the link with the largest interference is first estimated, assuming the original maximum throughput. Then, the drop from the link with the second largest interference will be estimated, assuming the maximum throughput has been reduced by the first drop.

6.2 Throughput Drop Estimation Procedure

The throughput of the target link $link_{ij}$ under three-link concurrent communications is evaluated by the following procedure:

1. Estimate the throughput of each of the three concurrently communicating links using the original model in Section 4.3.
2. Estimate the throughput drop tpD from each of the two interfering links using Eq. (4) in Section 4.4.
3. Sort them in the descending order: let tpD^{1st} and tpD^{2nd}.
4. For the largest interfering link, adjust tpD^{1st} by the maximum speed of the AP of the target link, because different APs have different throughput performances:

$$tpD^{1st}_{adj} = tpD^{1st} \times \frac{tpM^{AP}}{140} \quad (6)$$

where tpD^{1st}_{adj} represents the adjusted throughput drop by the largest interfering link, and tpM^{AP} does the maximum throughput for the AP of the target link.

Then, the throughput tp^{1st}_{ij} of the target link after considering the drop by the first link interference is estimated by:

$$tp^{1st}_{ij} = tp_{ij} - tpD^{1st}_{adj} \quad (7)$$

5. For the second largest interfering link, adjust the tpD^{2nd} by;

$$tpD^{2nd}_{adj} = tpD^{2nd} \times \frac{tpM^{AP} - tpD^{1st}_{adj}}{140} \quad (8)$$

Then, the throughput tp^{2nd}_{ij} of the target link after considering the drop by the second link interference can be estimated by;

$$tp^{2nd}_{ij} = tp^{1st}_{ij} - tpD^{2nd}_{adj} \quad (9)$$

In [31], Su et al. observed that the throughput drop by the accumulated interference from two interfering links is greater than that from a single interferer, but less than the sum of the drops from the individual interference. The second interfering link will cause a smaller drop than the first one. This observation also supports our throughput drop estimation model for concurrently communicating three links.

7 Evaluations of Model Extension

In this section, we evaluate the throughput drop estimation model for three concurrently communicating links under POCs and the channel bonding through experiments and simulations.

7.1 Experiment Results in One Room

First, the devices and one-room field in Sections 5.1 and 5.2 are used in experiments.

7.1.1 Channel Assignments

In experiments, the bonded channel 3 is always assigned to AP_1. Then, the either channel of 3, 7, and 11 is assigned to AP_2. To AP_3, the assigned channel is moved from 3 to 11 one by one so that the channel distance is changed. The throughputs are measured at the same time for all the links.

7.1.2 Throughput Results

Figures 7, 8, and 9 show the measured average throughput among the three APs when the channel 3, 7, and 11 is assigned AP_2, respectively.

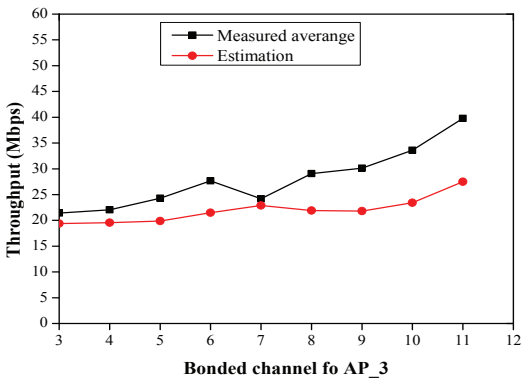


Figure 7: Throughput results in one room for AP_1:ch3, AP_2:ch3.

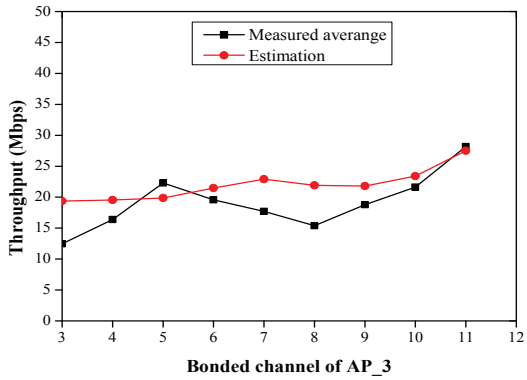


Figure 8: Throughput results in one room for AP_1:ch3, AP_2:ch7.

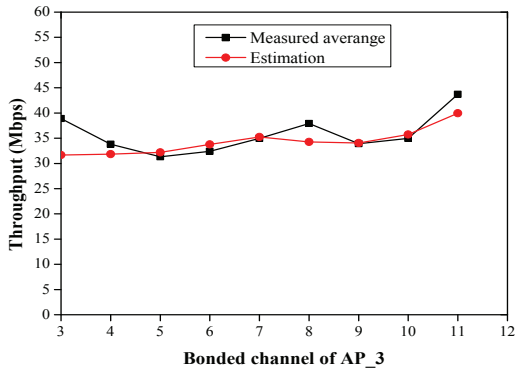


Figure 9: Throughput results in one room for AP_1:ch3, AP_2:ch11.

When the three results are compared, Figure 8 indicates the lowest throughput among them, because AP_2 is interfered with both AP_1 and AP_3. Figure 9 presents the highest throughput in general, because AP_1 and AP_2 are not interfered. On the other hand, Figure 7 shows that as the channel distance increases, the throughput will raise as the reduction of the interference between (AP_1 and AP_2) with AP_3.

Then, the estimated throughput is calculated by the proposed throughput drop estimation model, and is compared with the measured throughput. Figures 7- 9 demonstrate that these throughput are similar in any case. Thus, the accuracy of the proposed model for three concurrently communicating links is confirmed.

7.2 Experiment Results in Three Rooms

Next, the experiments are conducted using three rooms on the 3rd floor of *Engineering Building #2* to examine the accuracy of the model.

7.2.1 Experiment Field

The physical distance between the APs is changed by locating each AP in a different room from the previous experiments, as revealed in Figure 10. By locating each AP in a separate room, the overall interference between them is reduced. The link distance between the AP and its associated host is set 0.5m or 4m. The channel assignments in Section 7.1.1 are adopted.

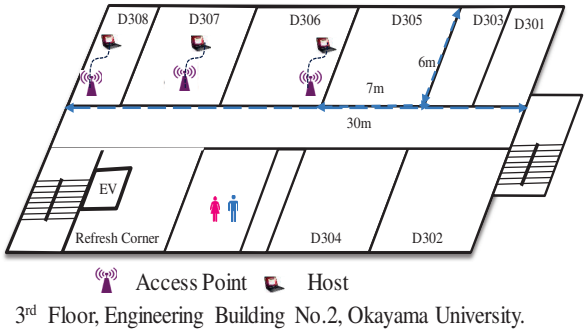


Figure 10: Three room topology.

7.2.2 Throughput Results with 0.5m Link Distance

Figures 11 - 13 show the measured and the estimated throughput results when the link distance is 0.5m. These throughput turns out to be similar at any channel distance, which confirms the accuracy of the proposed model. At the same time, they are higher than the ones in Figures 7- 9, since they are less interfered here.

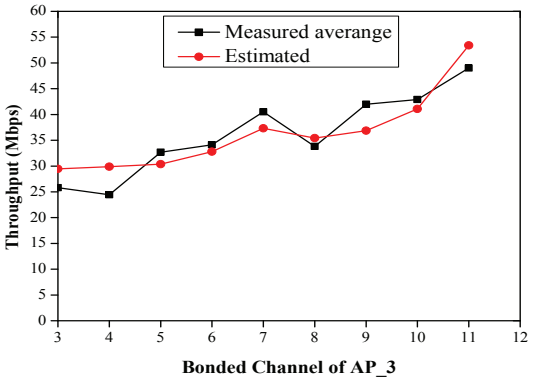


Figure 11: Throughput results in three rooms with 0.5m link distance for AP_1:ch3, AP_2:ch3.

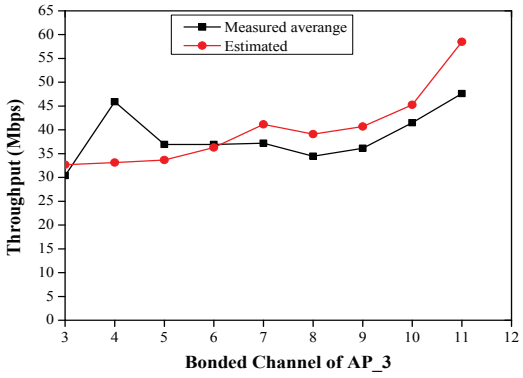


Figure 12: Throughput results in three rooms with 0.5m link distance for AP_1:ch3, AP_2:ch7.

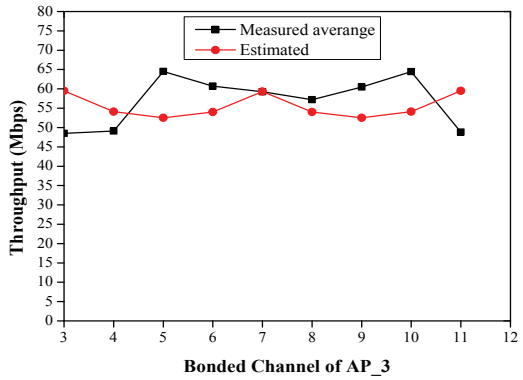


Figure 13: Throughput results in three rooms with 0.5m link distance for AP_1:ch3, AP_2:ch11.

7.2.3 Throughput Results with 4m Link Distance

Figures 14- 16 show the measured and the estimated results when the link distance is 4m. In Figures 15 and 16, they are similar at any channel distance. Thus, the accuracy of the model is verified in the larger link distance as well. However, in Figure 14, the measured throughput is lower than the estimated one at each channel distance. This is because the interference from non-target APs in the field is stronger around the channel 3, as shown in Figure 4. It is a must to conduct experiments in environments with less interference, which will be included in future works.

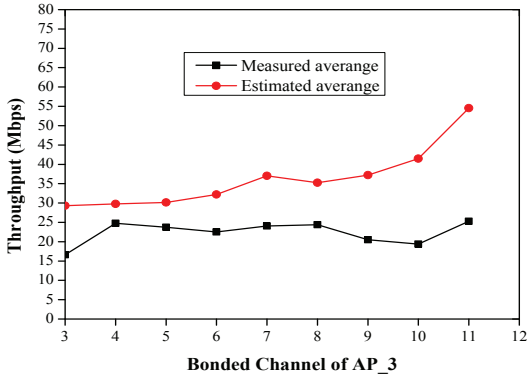


Figure 14: Throughput results in three rooms with 4m link distance for AP_1:ch3, AP_2:ch3.

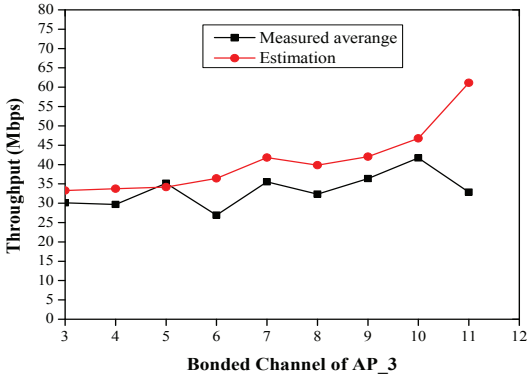


Figure 15: Throughput results in three rooms with 4m link distance for AP_1:ch3, AP_2:ch7.

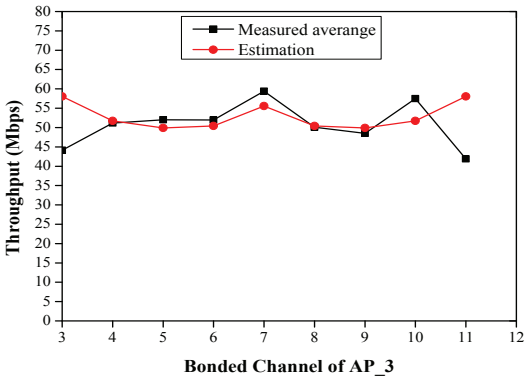


Figure 16: Throughput results in three rooms with 4m link distance for AP_1:ch3, AP_2:ch11.

8 Application to Channel Assignment

In this section, we discuss the application of the proposed throughput drop estimation model to the POC assignment to the APs.

8.1 Modification of Channel Assignment Phase

The channel assignment phase of the *active AP configuration algorithm* in [45] is modified to assign POC using the proposed model. Specifically, the formulations for this phase are revised from the previous one in [45] as follows:

8.1.1 Modification of Input

The number of partially overlapping channels is adopted in place of the number of orthogonal channels for C.

8.1.2 Modification of Output

The partially overlapping channel assigned to every active AP is adopted rather than the orthogonal channel assigned to every active AP.

8.1.3 Modification of Objective

The total interfered communication time E_3 is modified to consider partially overlapping channels by:

$$E_3 = \sum_{i=1}^{N} [IT_i^i] \tag{10}$$

where IT_i^i denotes the interfered communication time under partially overlapping channels for AP_i.

Under POCs, the link speed drop by the interfered links needs to be examined in the throughput estimation model. Therefore, IT_i^i can be simply given by:

$$IT_i^i = \sum_{j \in AH_i} \frac{1}{tp_{ij}^{2nd}} \tag{11}$$

where AH_i denotes the set of hosts associated with AP_i.

Our current throughput estimation model for partially overlapping channels only examines the interference between three links. In this paper, we assume that the target link is interfered by the two strongest interfering links, while the other interfering links may have negligible effects. This can be supported by the results in Section 5.4, where the highest interfering link causes the large drop and the second link causes the far small drop.

8.2 Evaluations by Simulations

First, we evaluate the performance of the POC assignment via simulations.

8.2.1 Simulation Platform

The *WIMNET simulator* [46] is adopted for simulations. Table 3 sums up the parameters for simulations [45].

Table 3: Simulation Parameters in WIMNET Simulator.

parameter	**value**
packet size	2360 bytes
max. transmission rate	150 Mbit/s
propagation model	log-distance path loss model
rate adaptation model	sigmoid function
carrier sense threshold	−85 dBm
transmission power	19 dBm
collision threshold	10
RTS/CTS	yes

8.2.2 Results for Random Topology

To evaluate the performance in various network environments for WLAN, first, the *random topology* is considered. As shown in Figure 17, in this topology we consider a network field of size $800m \times 200m$, where two rooms are located, each of length $400m$. Eight APs and 25 hosts are allocated randomly.

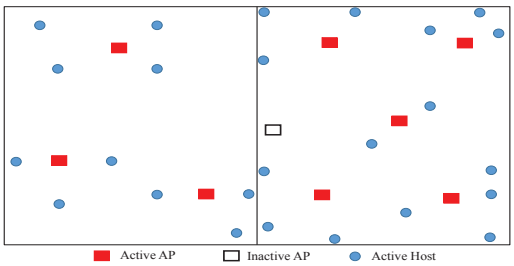


Figure 17: Random topology for channel assignment.

.

Then, the minimum host throughput and the overall throughput are compared between the POC assignment and the conventional OC assignment through simulations. Two channels (3, 11) are used for the OC assignment all the time. On the other hand, three channels (3, 7, 11), six channels (3, 5, 7, 8, 9, 11), and nine channels (3, 4, 5, 6, 7, 8, 9, 10, 11) are used for the POC assignment. Table 4 shows the results.

Table 4: Throughput results for *random topology*.

channel assignment	OC	POC		
# of channels	2	3	6	9
min. host throu. (Mbps)	7.12	9.16	8.42	8.63
overall throu. (Mbps)	174.56	196.74	197.04	198.46

8.2.3 Results for Regular Topology

Next, the *regular topology* in the third floor of Engineering Building #2 at Okayama University is considered. The room size is either $7m \times 6m$ or $3.5m \times 6m$. Eight APs and 55 hosts are regularly allocated, as signified in Figure 18.

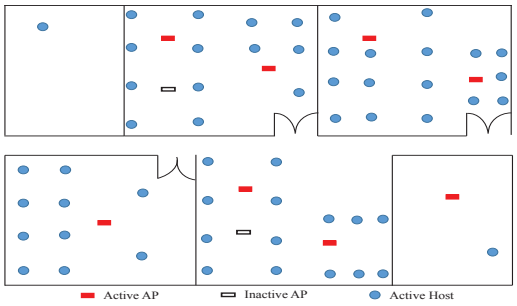


Figure 18: Regular topology.

The same two, three, six, and nine channels as for *random topology* are considered. Table 5 shows the minimum host throughput and the overall throughput for them.

It is noted that in both topologies, as the number of POCs is increased, the overall throughput will enhance by reducing the interference while maintaining the minimum host throughput.

Table 5: Throughput results for *regular topology*.

channel assignment	OC	POC		
# of channels	2	3	6	9
min. host throu. (Mbps)	2.68	3.11	3.06	3.15
overall throu. (Mbps)	147.76	170.88	168.76	173.49

8.2.4 Results for Regular Topology with Large APs

Finally, in Figure 19 we evaluate the model by simulations in a new instance by increasing the number of APs from 8 to 10 in the regular topology. The number of hosts remains as 55.

The evaluation in a real environment using the testbed is more important to confirm the effectiveness of the model. Thus, it will be in our future studies.

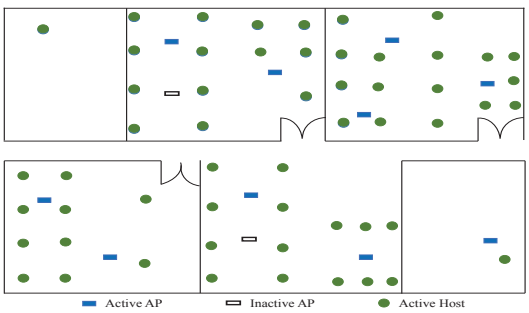


Figure 19: Regular topology for 10 APs.

Table 6 shows the results. As the number of APs increases, the total performance also increases, because more hosts can be located nearer to the APs, which increase the throughputs of them.

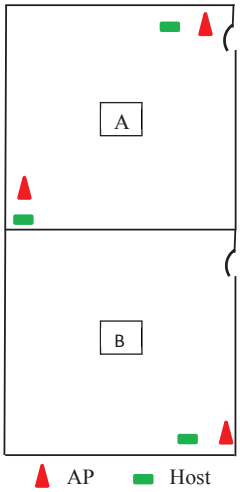
Table 6: Throughput results for *regular topology on 10 APs*.

channel assignment	OC	POC		
# of channels	2	3	6	9
min. host throu. (Mbps)	3.97	4.66	4.49	4.55
overall throu. (Mbps)	218.24	250.72	247.29	251.24

8.3 Evaluations by Experiments

Lastly, the throughput of the POC assignment is evaluated through experiments using the *two-rooms topology* in Figure 20. Two channels 3 and 11 are used for the OC assignment, and three channels 3, 7, and 11 are for the POC.

Table 7 shows the simulation and measurement results. This table indicates the following: 1) the POC assignment improves the overall throughput, and 2) the estimated throughput is well coincident with the measured one. The accuracy of the throughput estimation model and the effectiveness of the POC assignment are confirmed.



(a) Two-rooms topology

Figure 20: Two-rooms topology.

Table 7: Throughput results for *Two-room topology*.

channel assignment	OC	POC
# of channels	2	3
measurement (Mbps)	146.36	158.2
simulation (Mbps)	143.5	157.31

9 Conclusion

In this paper, we presented the *throughput drop estimation model* extension for concurrently communicating three IEEE 802.11n links under *partially overlapping channels (POCs)* and the *channel bonding*. Also, we proposed the model application to the POC assignment to the access-points in WLAN. Through extensive experiments and simulations, we confirmed the accuracy of the model and the effectiveness of the POC assignment. In future, we will upgrade this model for four or more interfering links. Then, we will evaluate it in a variety of network scenarios, such as for dense WiFi environments.

Acknowledgments

This work is partially supported by JSPS KAKENHI (16K00127).

References

[1] M. Balazinska, and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in Proc. Int. Conf. Mob. Syst., pp, 303 - 316, 2003.

[2] M. Elkhodr, S. Shahrestani, and H. Cheung, "Emerging wireless technologies in the Internet of things: A comparative study," Int. J. Wirel. Mob. Netw. (IJWMN) vol. 8, no. 5, pp 67-82, Oct. 2016.

[3] A. B. Makhlouf and M. Hamdi, "Design and experimentation of rate adaptation for IEEE 802.11n WLANs," IEEE Trans. Wirel. Comm., vol. 12, no. 2, pp. 908916, Aug. 2011.

[4] S. Murthy, A. Goswami, and A. Sen, "Interference-aware multicasting in wireless mesh networks," in Proc. Networking, pp. 299-310, 2007.

[5] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Measurement-based models of delivery and interference in static wireless networks," in Proc. Conf. Appl., Tech., Arch. Proto. Comput. Comm., pp. 51-62, 2006.

[6] Supplement to IEEE standard for information technology telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band, IEEE STD 802.11a-1999, 1999.

[7] IEEE Std 802.11-2012, IEEE standard for information technology-telecommunications and information exchange between systems - local and metropolitan area network - specific requirements - Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE STD 802.11-2007, pp. 1184, 2007.

[8] IEEE Standards Association, `https://standards.ieee.org/standard/802_11y-2008.html`, Accessed 28 May, 2019.

[9] IEEE Standards Association, `https://standards.ieee.org/standard/802_11ad-2012.html`, Accessed 29 May, 2019.

[10] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "IEEE 802.11ad: Directional 60 GHz Communication for Multi-Gigabit-per-Second Wi-Fi," IEEE Comm. Magazine, vol. 52, no. 12, pp 132-141, Dec. 2014.

[11] S. Banerji, "Upcoming Standards in Wireless Local Area Networks," Wirel. Mob. Tech., vol. 1, no. 1, pp 6-11, 2013.

[12] "IEEE 802.11n - standard for wireless LAN medium access control (MAC) and physical layer (PHY): enhancements for high throughput," IEEE, Oct. 2009.

[13] T. D. Chiueh, P. Y. Tsai, and I. W. Lai, "Baseband receiver design for wireless MIMO-OFDM communications," 2nd ed., Wiley-IEEE Press, 2012.

[14] L. Deeky, E. Garcia-Villegas, E. Belding, S. J. Lee, and K. Almeroth, "Intelligent channel bonding in 802.11n WLANs," IEEE Trans. Mob. Comput., vol. 13, no. 6, pp. 1242-1255, 2014.

[15] B. Boris, C. Alessandro, Z. Alessandro, B. Jaume, "On the interactions between multiple overlapping WLANs using channel bonding," IEEE Trans. Vehi. Tech., vol. 65, no. 2, Feb. 2015.

[16] S. Bokhari and V. Zaruba, "i-POCA: interference-aware partially overlapping channel assignment in 802.11-based meshes," in Proc. IEEE WoWMoM, 2013.

[17] K. Zhou1, X. Jia, Y. Chang, and X. Tang, "Partially overlapping channel assignment for WLANs using SINR interference model," IInt. J. Comm. Syst, vol. 27, no. 11, pp. 3082-3095, March 2014.

[18] K. S. Lwin, N. Funabiki, C. Taniguchi, K. K. Zaw, M. S. A. Mamun, M. Kuribayashi, and W.-C. Kao, "A minimax approach for access point setup optimization in IEEE 802.11n wireless networks," Int. J. Netw. Comput., vol. 7, no. 2, pp. 187-207, July 2017.

[19] I. M. Kwenga, N. Funabiki, M. Kuribayashi, R. W. Sudibyo, and W.-C. Kao "A throughput estimation model under two-link concurrent communications with partially overlapping channels and its application to channel assignment in IEEE 802.11n WLAN", Int. J. Space-Base. Situ. Comput., vol. 8, no. 3, pp. 123-137, 2018.

[20] M. E. Islam, N. Funabiki, and T. Nakanishi, "An access-point aggregation approach for energy-saving wireless local area networks", IEICE Trans. Commun., vol. E96-B, no.12, pp. 2986-2997, Dec. 2013.

[21] M. E. Islam, N. Funabiki, and T. Nakanishi, "Extensions of access-point aggregation algorithm for large-scale wireless local area networks", Int. J. Netw. Comput., vol.5, no.1, pp. 200-222, Jan. 2015.

[22] A. Mishra, E. Rozner, S. Banerjee, and W. Arbaugh, "Exploiting partially overlapping channels in wireless networks: turning a peril into an advantage," in Proc. ACM Conf. Inter. Meas., pp. 311-316, 2005.

[23] M. Elwekeil, M. Alghoniemy, M. El-Khamy, H. Furukawa, and O. Muta, "Optimal channel assignment for IEEE 802.11 multi-cell WLANs," in Proc. Signal Proc. Conf., pp. 694-698, 2012.

[24] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful", in Proc. Joint Int. Conf. Meas. Model. Comp. Sys., pp. 63-74, 2006.

[25] Z. Feng and Y. Yang, "How much improvement can we get from partially overlapped channels?," in Proc. IEEE WCNC, pp. 2957-2962, April 2008.

[26] Z. Zhang, L. Song, Z. Han, and W. Saad, "Coalitional games with overlapping coalitions for interference management in small cell networks", IEEE Trans. Wireless Comm., vol. 13, no. 5, pp. 2659-2669, May, 2014.

[27] W. Zhao, H. Nishiyama, Z. Fadlullah, N. Kato, and K. Hamaguchi, "DAPA: capacity optimization in wireless networks through a combined design of density of access points and partially overlapped channel allocation", IEEE Trans. Vehi. Tech., vol. 65, no. 5, pp. 3715-3722, May 2016.

[28] S. Mukherjee and S. C. Ghosh, "Throughput improvement using partially overlapping channels in WLAN with heterogeneous clients", Wired/Wireless Inter. Comm., pp. 335-347, 2016.

[29] B. P. Tewari, and S. C. Ghosh, "Combined power control and partially overlapping channel assignment for interference mitigation in dense WLAN", in Proc. IEEE Int. Conf. Adv. Info. Netw. Appli., pp. 646-653, March 2017.

[30] G. Zhao, Q. Wang, C. Xu, and S. Yu, "Analyzing and modelling the interference impact on energy efficiency of WLANs", in Proc. IEEE Int. Conf. Comm., May 2018.

[31] Y. Su, Y. Wang, Y. Zhang, Y. Liu, and J. Yuan, "Partially overlapped channel interference measurement implementation and analysis", in Proc. IEEE Conf. Comp. Comm. Work., pp. 760-765, April 2016.

[32] H. A. Mohammad, H. Xiaoyan, A. Farhana, "Multiple radio channel assignment utilizing partially overlapped channels", in Proc. IEEE GLOBECOM conf., pp. 4737-4743, 2009.

[33] Cisco Meraki, https://documentation.meraki.com/MR/Radio_Settings/Auto_Channel, Accessed 30 May, 2019.

[34] Fujitsu corporation https://www.fujitsu.com/global/about/resources/news/press-releases2015/0909-01.html, Accessed 30 May 2019.

[35] Buffalo corporation, http://buffalo.jp/download/manual/html/air1200/router/whrg300n/chapter118.html, Accessed 30 May 2019.

[36] Vodafone corporation, http://www.vodafone.co.uk/cs/groups/public/documents/contentdocuments/vfcon090503.pdf, Accessed 30 May 2019.

[37] Google WiFi, https://support.google.com/wifi/thread/412017?hl=en Accessed 30 May 2019.

[38] C. Newport, D. Kotz, Y. Yuan, R. S. Gray, J. Liu, and C. Elliott, "Experimental evaluation of wireless simulation assumption," J. Simu., vol. 83, no. 9, pp. 643-661, Sept. 2007.

[39] J. Padhye, S. Agarwal, V. N. Padmanabhan, L. Qiu, A. Rao, and B. Zill, "Estimation of link-interference in static multi-hop wireless networks," in Proc. ACM IMC, pp. 28, Oct. 2005.

[40] M. S. A. Mamun, N. Funabiki, M. E. Islam, M. Kuribayashi and I-W. Lai, "An active access-point configuration algorithm for elastic wireless local-area network system using heterogeneous devices," Int. J. Netw. Comput., vol. 6, no. 2, pp. 395-419, 2016.

[41] D. B. Faria, "Modeling signal attenuation in IEEE 802.11 wireless LANs," Tech. Report, TRKP06-0118, Stanford Univ., July 2005.

[42] iPerf - TCP, UDP and SCTP network bandwidth measurement tool, https://iperf.fr/, Accessed 29 Dec. 2017.

[43] Homedale: Wi-Fi / WLAN Monitor (2017, Dec.), http://www.the-sz.com/products/homedale/.

[44] L. B. Jiang and S. C. Liew, Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks, IEEE Trans. Mob. Comput., vol. 7, no. 1, pp. 34-49, Jan. 2008.

[45] M. S. A. Mamun, N. Funabiki, K. S. Lwin, M. E. Islam, and W.-C. Kao, "A channel assignment extension of active access-point configuration algorithm for elastic WLAN system and its implementation using Raspberry Pi," Int. J. Netw. Comput., vol. 7, no. 2, pp. 248-270, July 2017.

[46] N. Funabiki, ed., "Wireless mesh networks," InTech-Open Access Pub., Jan. 2011. http://www.intechopen.com/books/wireless-mesh-networks, Accessed 20 Jan., 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 97-103 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Toward a Smart Campus Using IoT: Framework for Safety and Security System on a University Campus

Alghamdi Abdullah[*,1], Mohammed Thanoon[2], Anwar Alsulami[3]

[1]*Information Systems Department, Najran University, Najran, Saudi Arabia*

[2]*Electrical and Computer Eng. Dept., Tennessee State University, Nashville, USA*

[3]*Department of Mechanical Eng., Jazan University, Jazan, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 17 May, 2019
Accepted: 09 July, 2019
Online: 03 September, 2019

Keywords :
IoT
Smart Campus
Safety and Security

ABSTRACT

Recently, there is an enormous research on the smart campus concept due to the revolution of the IoT technologies. The motivation of this paper is to: reinforce the safety on campus, reduce the cost, and take one step forward toward a University smart campus. In this paper, we are not only proposing a framework that would act as an instantaneous responder, but we also provide a glimpse of the evolving research on smart campus. In addition, we explore the challenges, and highlight the future work regarding this on-the-spot responder system.

1. Introduction

The smart campus concept has been the main focus of many researchers recently due to the valuable insights gained toward developing smart cities. The university campus theoretically is a small city where it delivers variant services to variant users. There are several factors that attract the investigators to study the smart campus including: delivering high quality services, protecting the environment, and saving the cost. The internet of things is a fundamental part of the smart campus, and it is inescapable getting it invoked.

The internet of things is a communication paradigm that gained its tenacity from its capability of connecting variety of everyday life objects to the internet. These objects include, but not limited to, sensors, robots, security locks, alarms, drones, appliances, smart grid systems, office equipment and so on. Even though IoT is in its early stages, there are many applications and standardization that has been developed in many domains including: home automation, smart grids, water and waste management, traffic control, smart vehicles, healthcare assistance, and industrial automation. Moreover, the realization of the IoT network is facing two main challenges. First, technical challenges because of IoT novelty and heterogeneous nature. Second, business challenges due to the lack of complete and approved business model that would encourage investments.

In this paper, we are proposing a safety and security framework based on the IoT which would enhance the safety on a university campus. The work in this paper is not only an extension of the investigation originally presented at the 2016 International Conference on Future Internet of Things and Cloud (FiCloud) [1], but we also propose a new system, and identify the limitations and research opportunity in this field.

The safety on a university campus is a growing concern among the campus community across the U.S. and across so many other countries due to the horrific crimes and mass shooting committed, lately. Regardless of whether the campus is rural or urban, large or small, significant number of people on campus seems to be reluctant to walk out at night alone [2]. According to Clery Center for security on campus [3], the crimes committed by students on campus is about 80% of the overall crime number. For example, Tennessee State University (TSU) is considered about an average in crime rates [4]; However, there were recent cases which saddened TSU community such as the shooting in 2015 that killed one, and injured three students [5]. Figure 1 shows statistics of the crimes committed on post high school education campuses in the United States over the period of 2015 to 2017 [6]. Our proposed system will not only enhance the safety, but also it would reduce the cost, make efficient use of resources, and

*Alghamdi Abdullah, Assistant Professor, IS Dept., NU, Najran, Saudi Arabia.
Email: aaalghamdi@nu.edu.sa

become one step closer toward a smart campus by utilizing the recent advances in technology.

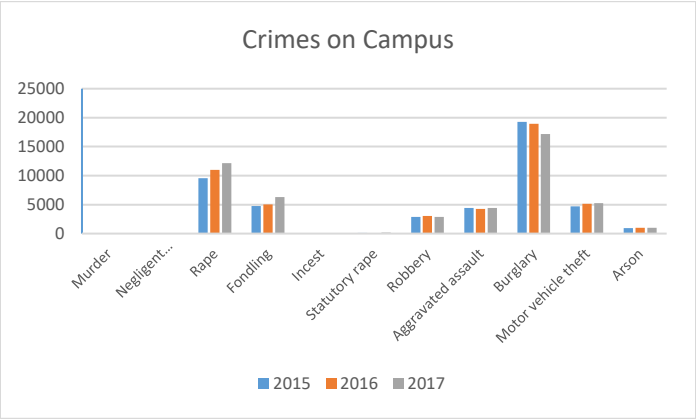


Figure 1 Statistics of The Crimes Committed on more than 11000 Campuses [6].

2. Smart Campus and Its Applications

The smart campus market springs from being a diverse environment regarding the services conveyed and the recipients of these services. The impact of conveyed services is not limited to the academic aspect, but also the social, financial, and environmental aspects. We can categorize the current research in smart campus into four main areas: intelligent buildings, campus smart grid, learning environment, and other applications. Figure 2 shows the context of a smart campus, its impacts, and many examples of its applications. On the rest of this section, we will present the research being done toward achieving a smart campus.

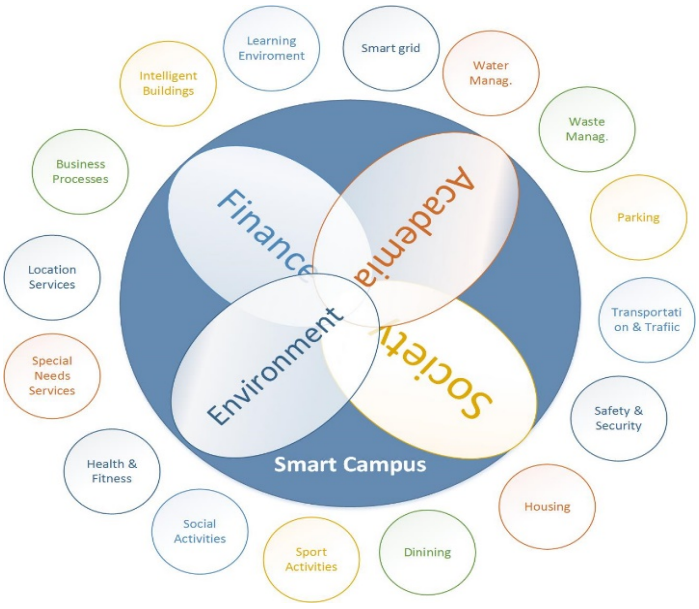


Figure 2 Impacts and Applications of Smart Campus Concept [1].

2.1. The Intelligent Buildings

Given that there is no de facto guide which would lead the people in the construction industry on how to build an intelligent building, a novel approach would be the plan to achieve the stakeholders' goals [7]. The planners need to focus on two main aspects. First, Information and Communication Technology (ICT) infrastructure where at least one fast backend network is needed for both business and building systems. Servers and variant sensors are essential part of the automation process of buildings. Sensors routine includes, detecting carbone dioxide levels in buildings, monitoring building stress, adjusting temperature, turning off the services in empty spaces, and measuring humidity and pollution levels [8,9]. Second, reporting and collaboration between various department is crucial to improve business processes and workflow, and it can be done by sharing information and reports, and enhancing communication [7]. The reports include building occupancy, staff attendance, patterns of utility usage, real-time warnings, energy usage, and so on. Knowing that the productivity gets increased by boosting the comfort of the community on campus, the stakeholders would make that as one of their goals [10].

2.2. Campus Smart Grid

Smart grid "is an opportunity to use new information and communication technologies (ICTs) to revolutionize the electrical power system" [11]. There are many universities adopting smart grids such as Oregon State University, which integrated Synchrophasor technology to have better monitoring and alarming system; however, the optimized results are not realistic due to the fact that this technology is relatively new and needs further studies [12]. In case of failures with SCADA, Santos *et al.* [13] suggests an automation process of power restoration using Open Platform Communications (OPC) protocol, and to characterize detection, isolation, and self-healing. Examining demand patterns, implementing a self-organized mapping (SOM), which is favored load profiling algorithm, and analyzing the current grid to provide a methodology and recommendations for smart microgrid in a university campus are discussed in [14,15]. Web-based systems on a cloud platform are introduced in: [16] for the analysis of energy consumption and behavior patterns, [17] for the purpose of controlling the demand response, and [18] for forecasting of future demand using machine learning (ML) models. Bracco *et al.* [19] propose a quantifying method for the usage of primary energy, predict CO2 emissions cut, and assess the cost reduction at the University of Genoa campus. Finally, IoT made it possible to control and monitor the energy consumption by users, identify energy draining devices, and suggests actions to optimize their behavior.

2.3. The Learning Environment

Given that the primary ambition of e-learning systems is to provide an articulated learning environment for learners based on their aspirations, knowledge, and talent, Wang *et al.* [20] characterize the learning process in an intelligent environment as a recursive process of four stages: learning, assessment, interaction, and analysis. Several papers published in this area including [21] which proposes students profiling based on their behavior over on-campus social network to provide a context-based personalized learning experience to achieve a ubiquitous learning (uLearning); however, additional study in terms of

learners' behavior and content design would exaggerate the benefit of such system. Microsoft and MIT started a research collaboration to achieve an intelligent campus, so called (MIT iCampus) [22]. This collaboration resulted in developing several intelligent systems including: class communicator system (CCS) to overcome communication issues between instructors and students, and class learning partner (CLP), and to provide an exercises and instant feedback during the class. Finally, given that mobile phones, tablets, laptops, and other wearable devices are among the most prominent devices that have influenced this universe [23], and utilizing them in education is inevitable.

2.4. Other Applications

Even though they do not get enough attention, there are plenty of other applications where IoT can be utilized to dramatically change campus to a smart one such as: waste and water management, parking, voting, safety and security, and so on. Regarding waste management, many papers published recently including [24-26] which generally suggest planting sensors at the bins, and waste trucks that collect real-time data for analysis, and a system would use the collected data to suggest better cleaning schedule and a superior and cost-effective route for waste contractors.

One the other hand, Mudumbe and Abu-Mahfouz [27] present a user-centric smart water management system which represent consumption in visual graphs to increase awareness. Gabrielli *et al*. [28] describe a sustainable prototype for smart metering devices and the associated network. Despite the feasibility of the project, the focus of their experiment is to build a sustainable metering device. Moreover, the authors [29,30] demonstrate: a parking guidance system for the parking building, and an electronic voting system that uses RFID technology for the authentication process, respectively.

Keeping in mind that most of recent work is focused on energy, learning environment, and intelligent buildings, the safety and security did not attract the attention of researchers; however, the intrinsic need of: better safety, instantaneous response to emergency, and cost effective solution on campuses actuated our team to investigate this area.

3. Safety and Security System

In this section, we will describe our system in details. At first, we will show the overall architecture of the system followed by the technique that will be used to deploy the sensors. Considering four different types of undesirable incidents: gun possession, assault, burglary and firing arms, we will show how to detect such suspicious activity. After that, we will show how to detect and track the suspect until the backup team arrives to the scene.

3.1. System Architecture

The safety and security system work with a variety of devices including: cameras, microphone sensors, glass break sensors, Raspberry Pi boards (RPis), a drone, and a server connected to the safety and security Operation Control Center (OCC) on campus. Figure 3 shows the general architecture of the system. To avoid overwhelming one server by making it process all the data collected by sensors across the campus, we would use the RPis to accomplish a distributed data processing architecture. Taking advantage of its cheap cost, the RPis would be used to process images, voices and other collected data by attached sensors. In case of detecting suspicious activity, the RPi would notify the server which will identify the location of the RPi using table for mapping the RPi's ID and its predefined location. The server would trigger the alert, identify the fastest route to the incident location, and instruct the drone to observe the situation, and wait for further actions.

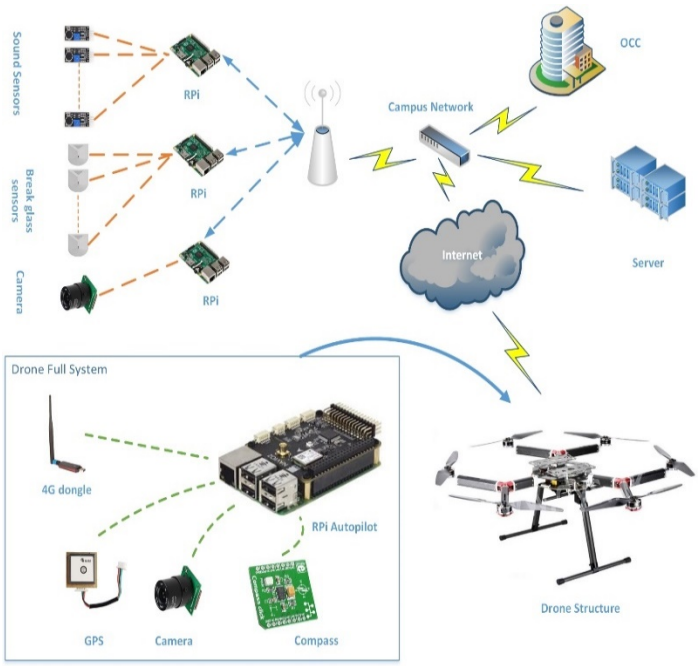


Figure 3 System Architecture

A sophisticated and robust drone need to be used as an instantaneous responder to observe the situation, and send live updates to the OCC for assessment and evaluation. The drone's autopilot would be attached to: a 4G dongle for internet connection, a GPS, a camera, and a compass.

3.2. Sensors Deployment

Since we have different types of sensors and our system depend on them, we need to deploy them carefully. The deployment technique is planned attentively to achieve two goals: to be cost effective regarding the number of sensors, and to avoid degrading the effectiveness of the system by deploying less sensors than required. The cameras would be deployed at parking lots, sidewalks, and buildings' entrances. The glass break sensors need to be deployed inside the buildings. Because one sensor covers around 25 feet, we would deploy one sensor per room wherever applicable. Finally, we decided to choose a microphone that covers 20 to 30 feet in a spherical area such as QSPMIC microphone by Q- See.

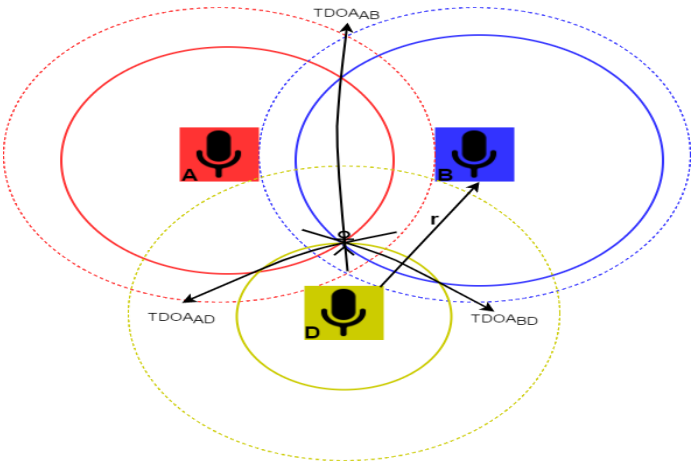


Figure 4 Sound Sensors Deployment

Figure 4 shows the way of planting the sound sensors where every sensor is planted at the end of the previous sensor coverage. Thus, the horizontally and vertically distance between sensors is equal to the radius of the coverage area (r). The sound source location can be determined using the Time Difference of Arrival (TDOA) or Multilateration (MLAT) method [31]. TDOA is an algorithm used to determine the location of the sound source by calculating the differences in arrival time of the sound at spatially separated sensors. The time difference of arrival of sound at different sensors must be known and processed at the central location. When the possible distance of the source from each sensor is drawn, it creates a hyperbolic curve. The location of the sound source lies in the intersection of all the hyperbolas from different sensors. This method works quite accurately for relatively static sources. If dynamic sources to be tagged, modified TDOA can be used [32].

3.3. Suspicious Activity Detection

The proposed system considers four different undesirable incidents to prove its effectiveness which are: gun possession, assault, burglary, and firing arms. The following paragraphs will study these scenarios in details.

Gun possession would be caught using the cameras distributed all over the campus parking lots, entrances, and sidewalks. The Raspberry Pies (RPis) attached to cameras would be able to detect guns automatically using Convolutional Neural Networks (CNNs) [33]. CNN is the state of the art technology in computer vision and it has been proven as a reliable architecture in object detection, giving its high accuracy with sufficient training. CNN is a special case of Neural Networks that apply the convolution operation to an input image for the purpose of feature extraction; however, the usage of the convolution operation demands high computational time, which is a challenge in our environment, knowing that we are using limited, low-energy devices. Fortunately, Iandola *et al.* [34] were able to achieve a smaller CNN architecture that performs the same accuracy as the AlexNet-level on ImageNet with fifty times fewer parameters, and they named it SqueezeNet. As a result, the SqueezeNet can be used on almost any controller or embedded computers. On RPis, it takes SqueezeNet approximately 2 seconds to process an image, which is enough time to: detect a person with a gun on a sidewalk, issue an alarm on the building, automatically close the inner doors of buildings, and notify the safety and security OCC on campus. Using a significant number of images of different types of guns, the SqueezeNet would be trained. The trained model would be deployed on the RPis wherever the cameras are applicable. After the detection of the gun, the RPis would alert the server. Using the predefined mapping between RPis and locations, the server would be able to instantaneously deduce the location of the gunman, plan the fastest route for drone, and instruct the drone to find and tracking the gunman until a backup team arrives to the scene.

There are some cases where the gun would not be detected by the distributed cameras. Therefore, we introduce another method for detection which is by detecting the sounds of firing arms. Usually, firing the arms and assault crimes are accompanied by sounds such as the sounds of gunshots and yelling for help. In case of any of these undesirable incidents happened, the microphones attached to RPis would catch the accompanied signals, and pass it to the neural network to identify if there is any suspicious activity. Since the CNN is proven to be effective in speech recognition [35], likewise, the SqueezeNet architecture would be used here too. Gunshots and yelling for help spectrograms would be used to train the network, and the pre-trained model would be deployed on the RPis. Similar to the previous scenario when the detection happens, the RPis would notify the server. Then, the server would identify the location of the incident and instruct the drone to observe the situation.

For burglary detections, we are using a glass break sensors which are connected to RPis. By listening to acoustical frequencies generated when a glass gets broken, the sensor would alert the RPi over the wireless which would lead to issuing an alarm to the server. Because the RPi mapped the sensor with its location on the building, the server would send the drone to that location to give real-time updates of the broken window. This would be important in case the burglar tried to run away because the drone can always detect and track people on the scene. Meanwhile, the OCC room should have sent a backup team to that room to investigate the situation.

3.4. Target Detection and Tracking

As soon as the drone arrives into the scene, it has to detect and track the suspect. To add more capability, our algorithm would allow two methods of detection: manually and automatically based on the situation. Figure 5 shows the general flowchart of target detection and tracking process. Upon the arrival of drone, it would start: capturing videos frames, notifying the OCC with updates, and analyzing the captured frames for detection of the suspect. Knowing that CNN requires high computation demand, Histogram Oriented Gradient (HOG) [36] would be used to detect upright persons in the scene. Because we need to reduce the computation overhead of the RPi on drone, we would analyze only five frames out of the twenty-five frames per

second that the camera can detect. If there is more than one upright person on the scene, the system would allow the OCC to decide which one to track. Otherwise, the drone will track the detected target. Regardless of the method used to detect the target, the drone will keep the target on its camera frame center, while the OCC will get real-time updates of the location which would be shown on an interactive map. Because we don't want to overwhelm the drone's connection to the internet, the OCC will get few captured frames only.

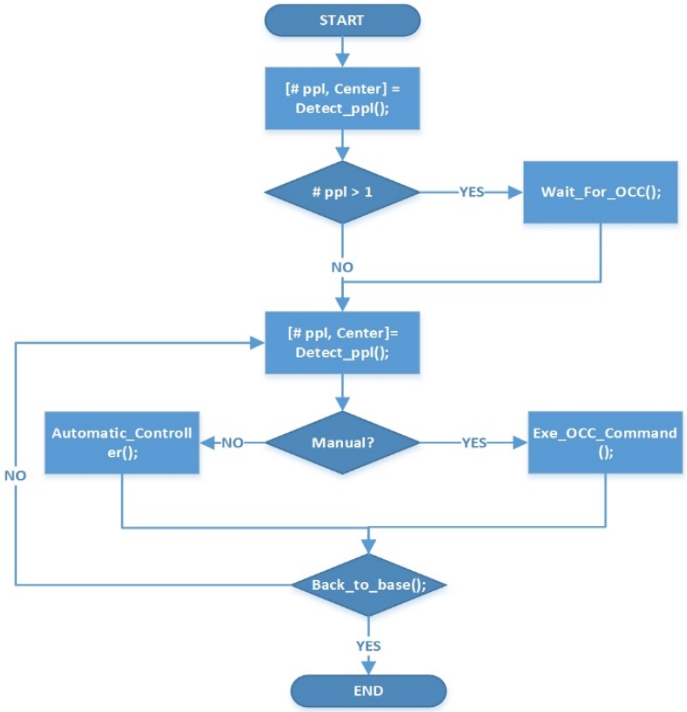


Figure 5 Target Detection and Tracking Flowchart

For the tracking, our system would allow the manual tracking of the suspect where the OCC officer can remotely control the drone without the need for frames analysis; however, this would require a high speed internet because the OCC needs a live streaming to be able to remotely control the drone. In case of automatic tracking, the optical flow method and its estimation using the Lucas-Kanade method will be used to track the suspect at the first frame only of every second [37,38]. The method works as the following: consider a pixel *I(x,y,t)* at the first frame, and it moves by distance *dx, dy,* in the frame that is compared to which is taken at *dt*. Because we are tracking the same person on real time, we assume the intensity does not change.

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (1)$$

by taking Tylor series approximation, and dividing by *dt*, we get:

$$f_x u + f_y v + f_t = 0 \quad (2)$$

$$f_x = \frac{\partial f}{\partial x}; \quad f_y = \frac{\partial f}{\partial y} \quad (3)$$

$$u = \frac{dx}{dt}; \; v = \frac{dy}{dt} \quad (4)$$

Then, using the Lucas-Kanade method, the final solution will be:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i fx_i^2 & \sum_i fx_i fy_i \\ \sum_i fx_i fy_i & \sum_i fy_i^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i fx_i ft_i \\ -\sum_i fy_i ft_i \end{bmatrix} \quad (5)$$

Then, the drone will move to the desired direction to make u, and v in its frame center. This method will be repeated once every second. Because we are capturing from the drone, the capture has a wide area where the suspect is in the center of the frame. It is less likely that the suspect will be able to go beyond the area of the drone's camera in one second. So, we are processing one frame per second only, and to avoid overwhelming the RPi with computation demands. Figure 6 illustrates the detection and tracking process, and the movement of the drone.

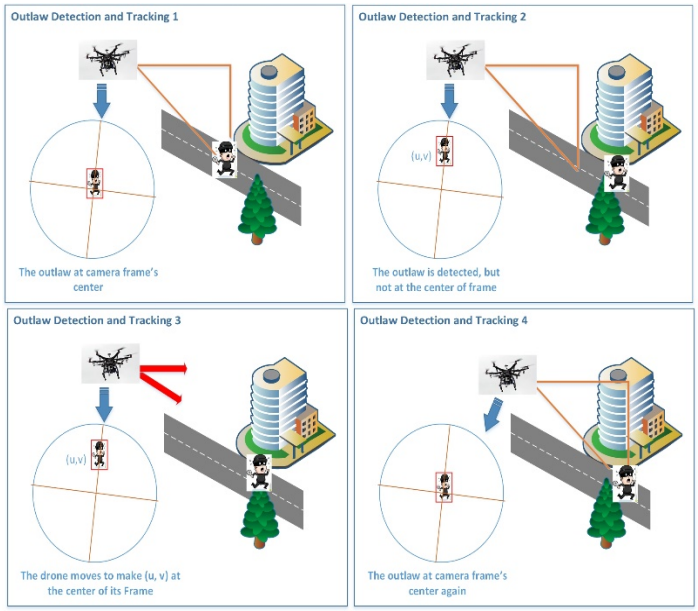


Figure 6 Target Detection and Tracking

4. Challenges and Future Work

Basically, the proposed safety and security framework inherited the obstacles that are facing the smart campus market which we can summarize in three areas: technical, financial, and political.

The technical barriers can be observed from the following perspectives: security, privacy, and configuration [1]. Being able to planet a massive number of low-energy and wirelessly connected devices on campus and meeting the security requirements is a problematic task due to the heterogeneous and intensive communication environment. Giving that we are utilizing various types of sensors, the privacy of people on campus is an issue where we need to be absolute that nobody can misuse the system to invade the people's privacy. Adopting the safety and security system on campus would result in using hundreds if not thousands of sensors which can be an enormous burden to configure them manually. Certainly, new ways of the automated configuration of IoT devices need to be thoroughly investigated.

In general, the smart campus market is facing financial difficulties giving that the limited resources of universities, regardless of immature experiences here and there. Even though the proposed system would save the university tons of many regarding the number of security officers on the long term, the initial investment is an impediment to the implementation of this system.

Regarding political hindrances, this system would need a waiver of some restrictions for unmanned aerial vehicles (UAV) by the Federal Aviation Administration (FAA), in the United States or such agencies in other countries, such as: the maximum altitude of 400 feet above ground level, visual line of sight, daylight only operations and some others [39]. Also, the system may face forceful opposition from people on campus and nearby neighborhoods regarding the invasion of their privacy. Even though it is almost inconceivable, the opposition of anti-tech security officers is a possible hindrance that needs to be considered.

After finalizing the implementation, our work opens the doors for researchers to improve the capability of this system to achieve optimum performance. The detection of outlaws during night-time and tracking them is a possible area of improvement. Moreover, the profiling of crimes respecting the hotspots, timings, and frequent suspicious activities using data collected by sensors would provide copious information about the crimes. The refined information would lead to crimes' prediction and prevention, eventually. Reducing the time between incident and arrival of drone is another area of improvement. Arming the drone with stun gun shocks could be considered for improvement. In addition, finding ways to avoid outlaws targeting the drone such as working on a bulletproof drone structure and trying to balance that with the speed of the drone is another area of improvement.

5. Conclusion

After presenting the recent work in smart campus technologies using IoT, we proposed a safety and security framework tailored to fit any university campus that would like to take advantage of recent advances in technology. The proposed framework would act as an instantaneous responder to incidents that could happen on campus. The proposed system is a good tool that would not only detect the incidents, but also it can track the outlaws even outside the campus until a backup team arrive to the scene. Also, we discussed the challenges of the system from three different perspectives: technical, and financial, and political. Finally, we highlighted the future work for this system to achieve its optimum performance.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A. Alghamdi and S. Shetty, "Survey Toward a Smart Campus Using the Internet of Things," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, 2016, pp. 235–239.

[2] A. L. Ball, "Students Fear Venturing Out Alone at Night on Campus.," *The New York Times*, 20-Jul-2012.

[3] "Clery Center For Security On Campus." [Online]. Available: http://clerycenter.org/. [Accessed: 25-Dec-2018].

[4] College Factual, "Tennessee State University Crime," *College Factual*, 20-Feb-2013. [Online]. Available: http://www.collegefactual.com/colleges/tennessee-state-university/student-life/crime/. [Accessed: 25-Dec-2018].

[5] N. N. Alund, "Two men in custody in 2015 TSU fatal campus shooting," *The Tennessean*. [Online]. Available: http://www.tennessean.com/story/news/crime/2016/08/23/two-people-custody-2015-tsu-campus-shooting/89211076/. [Accessed: 25-Dec-2018].

[6] U.S. Department of Education, "Campus Safety and Security." [Online]. Available: http://ope.ed.gov/campussafety/#/. [Accessed: 25-Dec-2018].

[7] S. Hipwell, "Developing smart campuses #x2014; A working model," in *2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, 2014, pp. 1–6.

[8] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[9] J. P. Lynch, "A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring," *The Shock and Vibration Digest*, vol. 38, no. 2, pp. 91–128, Mar. 2006.

[10] W. Kastner, G. Neugschwandtner, S. Soucek, and H. M. Newmann, "Communication Systems for Building Automation and Control," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1178–1203, Jun. 2005.

[11] A. B. M. S. Ali, Ed., *Smart Grids Opportunities, Developments, and Trends*. London: Springer London, 2013.

[12] C. Huo, J. Song, K. Wagner, G. Harold, and E. Cotilla-Sanchez, "Integrating synchrophasor technology with the Oregon State University campus smart grid project," in *2014 IEEE Conference on Technologies for Sustainability (SusTech)*, 2014, pp. 125–129.

[13] L. A. Santos *et al.*, "Towards a smart grid to the university campus of the federal university of Cear??," in *2015 IEEE 13th Brazilian Power Electronics Conference and 1st Southern Power Electronics Conference (COBEP/SPEC)*, 2015, pp. 1–6.

[14] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Analysis of the electricity demand patterns of a building in a university Campus," in *2013 12th International Conference on Environment and Electrical Engineering (EEEIC)*, 2013, pp. 382–387.

[15] R. M. González, T. A. J. van Goch, M. F. Aslam, A. Blanch, and P. F. Ribeiro, "Microgrid design considerations for a smart-energy university campus," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, 2014, pp. 1–6.

[16] V. Nikolopoulos, G. Mpardis, I. Giannoukos, I. Lykourentzou, and V. Loumos, "Web-based decision-support system methodology for smart provision of adaptive digital energy services over cloud technologies," *IET Software*, vol. 5, no. 5, pp. 454–465, Oct. 2011.

[17] H. Kim, Y. J. Kim, K. Yang, and M. Thottan, "Cloud-based demand response for smart grid: Architecture and distributed algorithms," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 398–403.

[18] Y. Simmhan *et al.*, "Cloud-Based Software Platform for Big Data Analytics in Smart Grids," *Computing in Science Engineering*, vol. 15, no. 4, pp. 38–47, Jul. 2013.

[19] S. Bracco, F. Delfino, F. Pampararo, M. Robba, and M. Rossi, "Economic and environmental performances quantification of the university of Genoa Smart Polygeneration Microgrid," in *Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International*, 2012, pp. 593–598.

[20] M. Wang and J. W. P. Ng, "Intelligent Mobile Cloud Education: Smart Anytime-Anywhere Learning for the Next Generation Campus Environment," in *2012 8th International Conference on Intelligent Environments (IE)*, 2012, pp. 149–156.

[21] Y. Atif and S. Mathew, "A Social Web of Things Approach to a Smart Campus Model," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 2013, pp. 349–354.

[22] "iCampus," *iCampus*. [Online]. Available: http://icampus.mit.edu/. [Accessed: 05-Oct-2018].

[23] V. G. Cerf, "Prospects for the Internet of Things," *XRDS*, vol. 22, no. 2, pp. 28–31, Dec. 2015.

[24] T. Anagnostopoulos, A. Zaslavsky, and A. Medvedev, "Robust waste collection exploiting cost efficiency of IoT potentiality in Smart Cities," in

2015 International Conference on Recent Advances in Internet of Things (RIoT), 2015, pp. 1–6.

[25] F. Folianto, Y. S. Low, and W. L. Yeow, "Smartbin: Smart waste management system," in *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2015, pp. 1–2.

[26] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov, "Top – k Query Based Dynamic Scheduling for IoT-enabled Smart City Waste Collection," in *2015 16th IEEE International Conference on Mobile Data Management (MDM)*, 2015, vol. 2, pp. 50–55.

[27] M. J. Mudumbe and A. M. Abu-Mahfouz, "Smart water meter system for user-centric consumption measurement," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015, pp. 993–998.

[28] L. Gabrielli, M. Pizzichini, S. Spinsante, S. Squartini, and R. Gavazzi, "Smart water grids for smart cities: A sustainable prototype demonstrator," in *2014 European Conference on Networks and Communications (EuCNC)*, 2014, pp. 1–5.

[29] W. Atmadja, J. Yosafat, R. A. Setiawan, and I. I. Irendy, "Parking guidance system based on real time operating system," in *2014 International Conference on Industrial Automation, Information and Communications Technology (IAICT)*, 2014, pp. 5–8.

[30] A. Saad, M. I. M. Roseli, and M. S. Zullkeply, "A Smart e-Voting System Using RFID Authentication Method for a Campus Electoral," in *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*, New York, NY, USA, 2014, pp. 31:1–31:7.

[31] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03), 2003, vol. 6, p. VI-553-6 vol.6.

[32] K. C. Ho and W. Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," IEEE Transactions on Signal Processing, vol. 52, no. 9, pp. 2453–2463, Sep. 2004.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," presented at the Advances in Neural Information Processing Systems.

[34] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv:1602.07360 [cs]*, Feb. 2016.

[35] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[36] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, Washington, DC, USA, 2005, pp. 886–893.

[37] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, Feb. 2004.

[38] OpenCV, "OpenCV: Optical Flow," *Open Source Computer Vision*. [Online]. Available: http://docs.opencv.org/trunk/d7/d8b/tutorial_py_lucas_kanade.html. [Accessed: 25-Dec-2016].

[39] Federal Aviation Administration, "Summary of Small Unmanned Aircraft Rule (Part 107)," FAA, Washington, DC 20591, Jun. 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 100-105 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Multimedia Application Development With Islamic Critical Reflection Through 3-2-1 Technique for Novice Teacher Internship Program

Evi Fatimatur Rusydiyah[1,*], Rakhmawati Rakhmawati[1], Eni Purwati[1], Moh. Hafiyusholeh[2], Ahmad Hanif Asyhar[2]

[1]*UIN Sunan Ampel Surabaya, Faculty of Education and Teacher Training, Indonesia*

[2]*UIN Sunan Ampel Surabaya, Mathematics Department, Indonesia*

ARTICLE INFO

Article history:
Received: 20 April, 2019
Accepted: 30 April, 2019
Online: 24 May, 2019

Keywords:
Macromedia Flash
Islamic Critical Reflection
Research and Development

ABSTRACT

The use of technology in the learning process has become the most significant point in helping teachers reaching the objectives of the learning. This study focuses on the use of multimedia by novice teachers in comprehending the internship program's guidance. The Islamic Critical Reflection model is a new and innovative model in the field of education which development is a result of integration between Islamic value and science. This study aims to answer problems related to the internship supervision which occurs among novice teachers. The multimedia application in the form of Flash Macromedia is developed using Islamic Critical Reflection model for internship program of novice teachers at the Tarbiyah and Teacher Training Faculty UIN Sunan Ampel Surabaya. The research design of this study is Research and Development. It includes the process of collecting information, designing a product, validating the design, improving the design, piloting the product, re-revising the product, trying out the product, revising further and creating products for large numbers. The results of this study indicate that the products produced by Macromedia Flash application integrated with Islamic Critical Reflection Model for Tarbiyah and Teacher Training Faculty of UIN Sunan Ampel Surabaya's internship program are feasible to use. The results are from the experts' analysis and users' trials. Experts involved in the validation of this product are Islamic Education specialists, application experts, educational technology experts, and curriculum expert.

1. Introduction

UIN Sunan Ampel Surabaya has a vision and mission that is hoped to be reached through the milestones between the period of 2019-2045. The milestones are described in the UIN Sunan Ampel Surabaya's Development Master Plan 2019-2045 [1], [2]. One of the biggest duty of UIN Sunan Ampel Surabaya and also of other higher education institutions managed by the Ministry of Religion is to integrate knowledge and religion as the differentiator between the institutions with those managed by the Ministry of Education.

Higher education nowadays has more challenging responsibility in preparing its students to be ready for the 4.0 industrial revolution. This preparation can be managed by an internship program for novice teachers. It is because this module is an integration between theory and practice done at schools. This program aims to prepare students to be professional teachers.

* Evi Fatimatur Rusydiyah, Jl. A. Yani 117 Surabaya, +62 8121635409 & evifatimatur@uinsby.ac.id

Within this context, UIN Sunan Ampel has done many adaptation programs. One of the programs is designed to train novice teachers in mastering teaching skill that is intact and integrated. Therefore, after graduating, novice teachers who have taken the internship module are hoped to be ready for teaching [3], having a positive identity as a teacher, and acquiring more knowledge in the field of teaching [3]. It is possible because the internship program has several elements that are involved: supervisor, mentor teacher, and novice teacher.

The challenge is that the internship program that is running at the moment is considered to be not effective and not efficient yet in its guidance system. The problem that is emerged is the unstructured guidance program within each element. The evaluation and reflection in the internship program have only been partially conducted without a thorough involvement of various weakness points. The result is, there is no comprehensive information toward the development of novice teachers during the internship program. This condition then resulted in less competent novice teachers. Whereas, this internship program is designed to

be the prime program the faculty can offer, and also designed to be the educational core program.

To resolve the problem, Tarbiyah and Teacher Training faculty strives to develop a supervision system through critical reflection approach in the internship program. The technique implemented is the 3-2-1 technique with Islamic values as the base of its integration. Therefore, the process will facilitate teaching and learning, and it would be central to the teacher's development. The implementation of the 3-2-1 technique is done within the discussion process between the supervisor, mentor teacher and novice teacher. The three elements provide three positive inputs of what was found during the internship. After that, two negative inputs are presented and followed by a solution [4]. The Islamic Critical Reflection approach is used during supervision. It is based on Surah Al-Hujurat verse 6-8.

In this research, the concept is made in the form of the application using Flash Macromedia. Included in it, is the first supervision until the eighth. In other words, this research aims to develop a multimedia application with Islamic Critical Reflection for the Novice Teacher Internship Program.

In this research, multimedia as a tutorial application in facilitating teacher mentors, lecturers, and students in carrying out practical activities in field experience, so that they can understand the steps of supervision they will do for each. For teacher mentors, they will understand what they must provide in each supervision. For lecturers, they can understand what must be done when they are supervised. Likewise, for students, they will understand what they should get every time the lecturer supervises. Therefore this multimedia is not used for the learning process in the classroom, but this multimedia is used as a tutorial application in making supervising steps clear.

With the application, it is hoped that the novice teacher's professionalism is improving in the way that novice teachers can critically reflect and integrate it with Islamic values and pedagogical knowledge. "Critical reflection involves concern about the broader issue of teaching and learning*"* [5] is the notion that Murray and Kujundiz hold. They explained that [6] "critical reflection involves a process of analyzing, reconsidering, and questioning experiences within a broad context of issues such as analyzing assumption about teaching, raising awareness about the context of teaching. They imagine alternative ways of thinking".

The concept is visualized in the form of flash Macromedia. The application is developed through the Technology Acceptance Model (TAM). TAM model modification is done by Venkantesh [7] by adding trust variable with the title: Trust enhances Technology Acceptable Model which evaluating the relationship between TAM variable and trust. TAM other modification is Trust and Risk in Technology Acceptance Model (TRITAM) that uses trust and risk variable named TAM variable [8]. This model is one of the models developed from the theory of using a system of technology. It is considered to be the most affected in explaining personal acceptance towards the benefit and information technology used as a system [9].

One research that is related to the novice teacher program is Toman et al., [10]. It develops reflective thinking method and determines the novice teacher reflective thinking level. In 2007, Mathew discussed how mentor teacher creates opportunities for the novice teachers to developed reflective practice during the internship at school [11]. Gonen [12] in the research that was aimed to evaluate the experience that caused some changes after reflection is done. Moreover, it was also aimed to figure out what experience gained by the novice teacher after receiving guidance from the mentor teacher [12].

The importance of this research is for both the mentor teacher and the novice teacher understand the supervision process thoroughly, especially in implementing the 3-2-1 reflective principal as a part of novice teacher professional development. Novice teacher will be more confident with the guidance process that used Islamic Critical Reflection. Mentor teacher will also work professionally in guiding novice teacher.

2. Design

This research is aimed to develop an application in the form of Macromedia flash to be used in the internship program. This multimedia application is used to make an understanding of the duties of lecturers, mentor teacher,, and novice teacher when conducting internship program as a tutorial in supervising. Therefore this application is not used in class as media teaching-learning but is used in the supervision process. According to Borg and Gall [13], this research is categorized as Research and Development (R&D) research.

Steps that are generated from the process are understood to be the R&D cycle. They are consisted of studying research findings, developing the product, testing the product, and revising it to improve its weaknesses.

The R & D design used in this study is as follows.

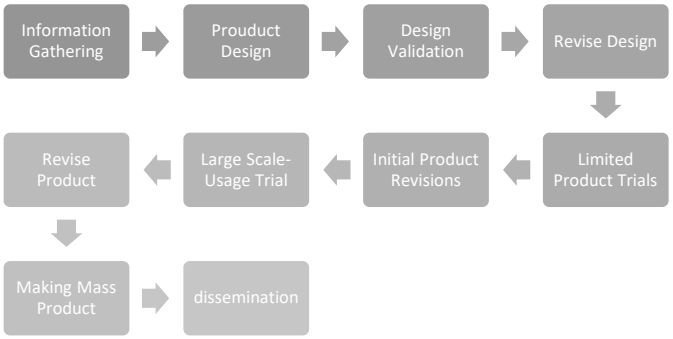


Figure 1. R & D flow

The R & D steps as referred to in the figure are information gathering, product design, design validation, revise design, limited product trials, initial product revisions, large-scale usage trials, revise product, making mass product, and dissemination [13]. These steps are a guide for researchers to conduct this research.

The test subjects in this R & D are stakeholders involved in implementing the practice of the field experience. The

stakeholders are 50 partner schools of the Tarbiyah Faculty UIN Sunan Ampel Surabaya which are occupied by field experience, 50 supervisors, and 150 mentor teachers.

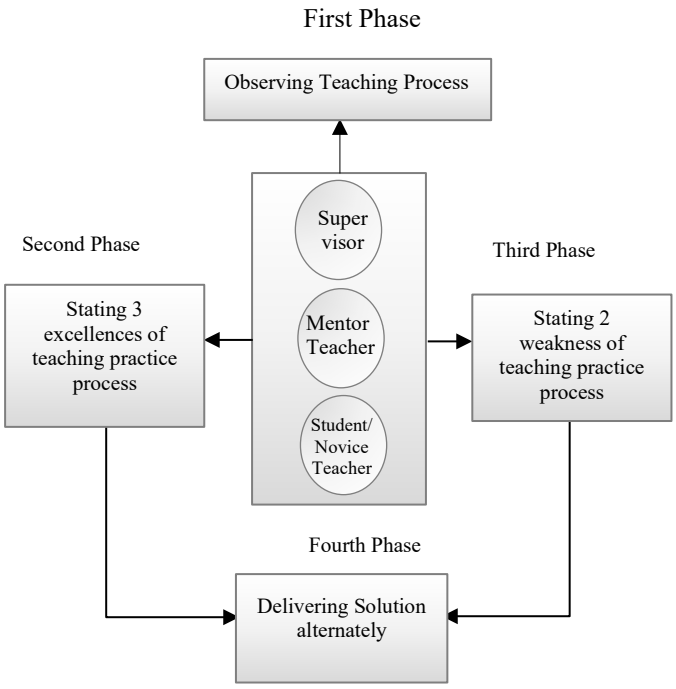


Figure 2: Islamic Critical Reflection Model

The product produced is based on the Islamic Critical Reflection model. This is a distinction as UIN Sunan Ampel integrated its learning process to Islamic principles and research [14], [10], [12]. The steps in Islamic Critical Reflection are observation, advantages, deficiency, and solution. The scheme can be seen in Figure 2.

The instrument which is used to measure the product is the Technology Acceptance Model (TAM). It is to explain personal acceptance towards the benefit and the use of information technology as a system [9]. After the application is made, the validity test was conducted. The experts to validate were ones who are expert in Islamic Education, Technology Education, and Curriculum. Operational suitability of the product was conducted in ten different primary schools and secondary school in Surabaya and Sidoarjo, East Java. The questionnaire was the main instrument for this stage.

3. Finding and Discussion

As in the R & D design above some of the findings in this study are as follows:

3.1. Gather Information

Information about the implementation of internship program in schools was conducted on Sunday, May 13, 2018 through a Focus Group Discussion (FGD) which was attended by 47 good tutors who were in Ibtidaiyah Madrasah, Madrasah Tsanawiyah / Middle School, Islamic Senior High Schools/ Vocational Schools in the region Surabaya and Sidoarjo and 41 Supervisors at the Sunan Ampel State Islamic University (UINSA). The FGD results show the need for understanding the concept of how the guidance model can improve the quality of UINSA internship's students who are qualified, the need for easy access that can be reached by students, civil servant teachers and Field Supervisors about guidelines for implementing effective and efficient internship program.

3.2. Stage of Product Design

On June 5, a Forum Group Discussion (FGD) was held which was attended by experts, namely educational technology experts, multimedia experts, Islamic education experts, and curriculum design experts from development products including a model of student assistance with Pamong Teachers and Field Advisors with 3-2-1 (triad meeting) which is visualized in Flash Media Media multimedia. On June 10, an FGD was held which was attended by experts again by presenting users. This FGD designed a product that will be formed in the internship application development model using Islamic studies on how to provide critical assistance based on an Islamic approach.

At this stage it was found, the practice of field training needs to be done with a reflective approach in communicating to convey information. This 3-2-1 approach is an implementation of Islamic teachings, that motivating students must be done politely with techniques to promote positive aspects rather than negative aspects. Likewise, the technique 3-2-1 strengthens the process of training students in finding solutions based on an analysis of their strengths and weaknesses [6]. Therefore approach 3-2-1 is an approach that leads to the analysis of the theory of critical reflection.

The 3-2-1 approach is also supported by the pyramid theory of human needs from Malow [15]. Humans in their lives have different needs, those needs from the high to the low. The lowest needs are the need to acknowledge oneself, then the need for compassion, love and love, then the need for security, and the highest needs are physiological needs. The higher the education of a person the higher the level of recognition desired [16]



Figure 3: Preliminary Display of the Pre Service Teacher Interference Model Application through the Islamic Critical Reflection Approach

3.3. Stage of Validation Design

The validators involved in this study were 1) Islamic religious education experts, multimedia experts, learning technology

experts, and curriculum experts. The results of the validator are as follows.

Islamic Education Expert

The application was developed through several steps. The first one is the observation step based on the basic Islamic teaching, which is Surah Al Maidah verse 17 and surah Al-Anam verse 46 about observing the process as one of Allah's power. In observing the novice teacher, mentor teacher could see how a novice teacher is going through a process in becoming a professional teacher and how a novice teacher is going through a process to be better.

The second and third step is to give a positive and negative suggestion. The second and the third step is the feedback from the supervisor, mentor teacher, and novice teacher. They all work together to suggest positive points. The inspiration for this step is Surah Al Baqarah verse 219, Surah Al Haj verse 28, Surah Al Isra' verse 14 and Surah Al Ankabut verse 3. Based on the verses, it can be concluded that as a supervisor, a mentor teacher, and a novice teacher, all the three elements have to deliver the truth, which is: conveying positive points to understand the deeper meaning, to contemplate, to evaluate the positive and the negative sides. Contemplation is designed by giving three positive characters and two negative experience.

The fourth step is the solution stage. It is based on Surah Al Baqarah verse 153, Surah Al Baqarah verse 109, Surah Al Ahzab, verse 70-71. All of the surah mentioned gives a clue that supervisor, mentor teacher and novice teacher have to show positive attitude such as to talk politely.

To ease the user in operating the application, guidance and summary result were provided thoroughly as it is seen in Figure 4.



Rangkuman

Praktik Pengalaman Lapangan II (PPL) di Fakultas Tarbiyah dan Keguruan UIN Sunan Ampel Surabaya adalah suatu rangkaian kegiatan untuk menerapkan berbagai teori kependidikan yang

Figure 4: Application Guidelines And An Application Summary

The Macromedia flash application consists of application guidelines, internship material, supervision stages, stages on supervision using the 3-2-1 technique. The third is to show the positive sides of the novice teacher's teaching. The second is to show the weaknesses of the novice teacher's teaching, and the first one is the solution. The 3-2-1 technique was used in a discussion and is conducted when the supervisor supervises the novice teacher. The 3-2-1 technique will be used not only by the novice teacher but also as a supervisor.

The result of the research is suitable with Maslow concept about human need pyramid. It is said that human need recognition. The recognition needed is toward self-actualization, affection, safety, and psychological needs [15]. Those needs are basic human needs. It also supported by several research: Schulte dan Marthann [17], Medcalf et all[18], Emad [19] dan Sari dan Dwiarti [20]. These research found out that motivation has a significant effect on work and study achievement.

The result of this 3-2-1 reflection technique is hoped to shape the novice teacher's self-confident. The teaching and learning will result in positive, and novice teacher can find their professionalism through reflecting every meeting. Therefore, they will be a professional teacher.

Application Analysis

The application that has been made was measured for its suitability using TAM indicator which consists of Perceived of Usefulness with its sub-indicators as follow: *1) Make Job Easier; 2) Work More 3)* Increase Productivity *4)* Effectiveness 5) Improve job performance *(6)* Useful. The second indicator is Perceived ease of use with its sub-indicator as follow: *1)* Easy to learn *2)* Controllable *3)* Clear and understandable *(4)* Flexible*, 5)* Easy to become skillful *(6)* easy to use [21]. This research also used an application that was developed based on the human-machine interaction [22] indicator by Davis [23] and Hendrawati [24]. It has several indicators, such as *learnability, efficiency, memorability, errors, and satisfaction.*

Through the application, it is gained several positive aspects as follow: the application could give guidance for a supervisor so that the supervisor could use it easily. It could also aid the supervisor's understanding in guiding a novice teacher. It helps the supervisor to improve novice teachers' professional competencies. This application is designed to present steps in supervising internship programs using the 3-2-1 technique effectively. This, in return, beneficial for faculty, supervisors and lecturers, novice teachers, school and stakeholders to develop teachers' professionalism.

Based on the application analysis, it is found out that the application design using buttons that are familiar for users is compatible with these Operating System: Windows, Linux, and Android. This development supports Koral and Emel [25], also Akar and Güzin [26] who suggest that technology acceptance relates to content comprehension which becomes the focus. It is also stated by Konak et al. [27] who stated that technology acceptance affects collaborative work. The same notion explained by Davis [21] and Hendrawati [24] who figured that technology acceptance had affected intrinsic motivation to occur.

As the previous research, information technology system acceptance could be explained within group work. The bigger the group who receive the information technology system, the bigger its practice would change. The group would also tend to use or try something new that could be used or developed in an information technology system

3.4. Education of Technology analysis

Education technology is a knowledge which relates to theory and practice on how to design, develop, implement, organize, and measure the education process to reach the educational aim. One

of its important components is a reflection. Reflection is a stage in an educational process so that the goal is reached. Therefore, reflection is used as one of the stages in this application so that it would affect novice teacher positively.

This condition is the same as several research which stated that every novice teacher has a reflective thinking skill [10]. Novice teacher has contributed positively towards the teaching skill that is developed to think reflectively [11]. Reflective needs to be applied when conducting an internship [12].

The same notion goes to video animation. The message is delivered effectively using video. The research that was conducted by Duygu Sonmez and Meral Hakverdi-Can [28] explained that video and application could affect users' comprehension in teaching reflectively.

The result of this study support the research by Mohd dan Elmagzoub A [29], Komalasari, et al. [30], Djamas, et al. [31], and Richter research's [32]. These researchers found out that learning through application brings a positive effect to the learning result. In this research, it was also found that the mentor teacher understood it more on how to do supervision job during the internship. Likewise the results of research conducted by DeVore et al. [33], Linton [34], Hamdi, et al. [35] Syurigin and Krasnova [36], Lailiyah et al. [37] and Darmawan [38].

3.5. Curriculum Analysis

During the internship program, eight supervisions were conducted. The internship was for two months on holiday between July and September every year [3]. The first supervision is an introduction, observation, school introduction, and sit in with mentor teacher as a complete model. The three activities were conducted in the first week within the supervision activity.

The second supervision is a practical observation which was guidance teaching (with the composition of mentor teacher took up 75% and novice teacher took up 25%) and conference (discussion between mentor teacher, supervisor and novice teacher) on a learning process that has been conducted and lesson plan preparation. Other points that were discussed were: practical guidance teaching (mentor teacher took up 50%, and novice teacher took up 50%), and conference (discussion between mentor teacher, supervisor, and novice teacher) about the learning process that has been conducted and lesson plan preparation. The second supervision was done in the second week.

The third supervision was a practical observation on guidance teaching (mentor teacher took up 25%, and novice teacher took up 75%) and a conference (discussion between mentor teacher, supervisor, and novice teacher) on the learning process that has been conducted and lesson plan preparation. This supervision was conducted in the third week of the internship program.

The fourth supervision is a practical observation on independent teaching and a conference (discussion with a mentor teacher, supervisor and novice teacher) on the learning process that has been conducted and lesson plan preparation. This was done in the fourth week.

The fifth supervision is on the fifth week. The supervision was a practical observation on independent teaching and a conference (discussion between mentor teacher, supervisor, and novice teacher) on the learning process that has been conducted and lesson plan preparation.

The sixth supervision is a practical observation on independent teaching and a conference (discussion between mentor teacher, supervisor, and novice teacher) on the learning process that has been conducted and lesson plan preparation. The seventh supervision is a practical observation on independent teaching and a conference (discussion between mentor teacher, supervisor and novice teacher) on the learning process that has been conducted and lesson plan preparation. This was conducted during the sixth week during the internship program. The eighth supervision was an observation on presentation preparation about the internship result and the finishing of the report. This was done in the sixth week.

The activities above can be visualized in a Macromedia flash application. This application was used in order to aid internship guidance comprehension in the means of attractive visual for users. Not only that, the internship guidance scheme as part of the curriculum becomes more organized and systematic. This is because each supervision has a clear aim and target and therefore, could improve the professionalism of novice teacher [16]. This process could be said as a professionalism improvement process by providing review activity in the reflection during the conference. This is also understood as the implementation of democracy in education [39].

The basis of this development and research is on the analysis of curriculum expert. The result was the internship curriculum design that will influence professionalism achievement of a novice teacher. This is as what Capka [40] has found that the curriculum design could influence student professionalism in the field of accountancy. Another research was by Jasperson [41] which stated that the internship curriculum with reflection principles could improve students' self-development. The same result was concluded by Johnson [42] who found that the internship process contributes to student's carrier development after graduation.

4. Conclusion

Internship model application development for novice teacher using the Islamic Critical Reflection was conducted based on four experts. The Islamic religion education expert stated that this application represents Islamic teaching principles as the basis of educational theory development. Application expert stated that this application meets the easiness and readability aspect of TAM (Technology Acceptance Model). Education technology expert stated that educational aspect through reflective teaching is used for feedback to improve the professionalism of novice teacher. It is found in this application. Curriculum expert stated that the cycle of 2-month supervision in this application is a good way in improving novice teacher's professionalism Conflict of Interest.

Acknowledgment

Our gratitude goes to UIN Sunan Ampel Surabaya for giving a research grant in 2018.

References

[1] *RIP UIN Sunan Ampel Surabaya Tahun 2019-2045,*

[2] "Peraturan Menteri Agama Republik Indonesia Nomor 56 Tahun 2015 Tentang Statuta Universitas Islam Negeri Sunan Ampel Surabaya." p. 6.
[3] Tim Penyusun, Buku Panduan PPL I Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Sunan AMpel. 2017.
[4] A. Fred and J. P. A. M. Kessels, "Linking Theory and Practice : Changing the Pedagogy of Teacher Education."
[5] T. Gale, "Preparing Professionals : student teachers and their supervisors at work Asia-Pacific Journal of Teacher," no. November 2018.
[6] M. Murray & Kujundzic, N, *critical Reflection : A text book for critical thinking. quebec*. Canada: McGill-queens University Press, 2005.
[7] Venkatesh, Morris, and Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Q.*, vol. 27, no. 3, p. 425, 2017.
[8] H. K. Lui and R. Jamieson, "eTransformation TRiTAM : A Model for Integrating Trust and Risk Perceptions in Business-to-Consumer Electronic Commerce," *Methodology*, no. September 2001, pp. 349–364, 2003.
[9] Jogiyanto, *Sistem Informasi Keperilakuan*. Yogyakarta: ANDI Offset, 2008.
[10] T. Ufuk, "Analysis Of Pre-Service Science Teachers ' Views About The Methods Which," no. October, pp. 179–189, 2014.
[11] J. Peechattu, "Reflective practices: a means to teacher development," pp. 1–6, 2017.
[12] S. I. K. Gonen, "A Study on Reflective Reciprocal Peer Coaching for Pre-service Teachers : Change in Reflectivity," vol. 4, no. 7, pp. 221–235, 2016.
[13] J. P. G. Gall Walter R. Borg, Educational research : an introduction / Meredith D. 1942.
[14] M. Harvey, "Editorial - Reflection for Learning in Higher Education," vol. 13, no. 2, 2016.
[15] A. H. Maslow, "Toward a Psychology of Being. A Psychology Classic." Start Publishing LLC, 2012.
[16] E. Rusydiyah, "Rehearsal Model as the Practicum Model in Preparing Professional Pre-service English Teachers," in *International Conference on English Language Teaching (ICONELT 2017)*, 2017.
[17] M. Schulte, "Adult Learning Degree and Career Pathways: Allusions to Maslow's Hierarchy of Needs," *J. Contin. High. Educ.*, vol. 66, no. 1, pp. 62–64, 2018.
[18] N. A. Medcalf, T. J. Hoffman, and C. Boatwright, "Children's dreams viewed through the prism of Maslow's hierarchy of needs," *Early Child Dev. Care*, vol. 183, no. 9, pp. 1324–1338, 2013.
[19] G. Emad, "Self-fulfillment development among the Arab sector in Israel," *J. Educ. Train. Stud.*, vol. 5, no. 5, pp. 81–92, 2017.
[20] E. Sari and R. Dwiarti, "Pendekatan Hierarki Abraham Maslow pada prestasi kerja karyawan PT. Madubaru (PG Madukismo) Yogyakarta," *J. Perilaku dan Strategy. Bisnis*, vol. 6, no. 1 February, pp. 58–77, 2018.
[21] J. Fred D. Davis, "A Technology Acceptance Model For Empirically Testing New End-User Information Systems: Theory And Results," 1985.
[22] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "Extrinsic and Intrinsic Motivation to Use Computers in the Workplace' FRED D. DAVIS~," 1992.
[23] J. Fred D. Davis, "A Technology Acceptance Model For Empirically Testing New End-User Information Systems: Theory And Results," 1985.
[24] T. Hendrawati, "Analisis Penerimaan Sistem Informasi Integrated Library System (Inlis)," *Visi Pustaka*, vol. 15, no. 3, 2013.
[25] E. Koral Gümüsoglu and E. Akay, "Measuring Technology Acceptance Level of Teachers by Using Unified Theory of Acceptance and Use of Technology, Online Submission," *Int. J. Lang. Educ. Teaching. v5*, no. 4, pp. P378-394, 2017.
[26] M. Akar and S. Guzin, "A Structural Model for Relationship between Web Pedagogic Content Knowledge and Technology Acceptance of Preservice Teachers.," *Malaysian Online J. Educ. Technol.*, vol. 7, no. 1, pp. 1–14, 2019.
[27] A. Konak, S. Kulturel-Konak, M. Nasereddin, and M. R. Bartolacci, "Impact of Collaborative Work on Technology Acceptance: A Case Study from Virtual Computing.," *J. Inf. Technol. Educ.*, vol. 16, no. 1, 2017.
[28] D. Sonmez, "Videos as an Instructional Tool in Pre-service Science Teacher Education Meral Hakverdi-Can **," *Eurasian J. Educ. Res.*, vol. 46, no. 46, pp. 141–158, 2012.
[29] M. Babiker and A. Elmagzoub, "For Effective Use of Multimedia in Education, Teachers Must Develop Their Own Educational Multimedia Applications.," *Turkish Online J. Educ. Technol.*, vol. 14, no. 4, pp. 62–68, 2015.
[30] K. Komalasari and D. Saripudin, "Value-Based Interactive Multimedia Development through Integrated Practice for the Formation of Students' Character.," *Turkish Online J. Educ. Technol.*, vol. 16, no. 4, pp. 179–186, 2017.
[31] D. Djamas and V. Tinedi, "Development of Interactive Multimedia Learning Materials for Improving Critical Thinking Skills," *Int. J. Inf. Commun. Technol. Educ.*, vol. 14, no. 4, pp. 66–84, 2018.
[32] J. Richter, K. Scheiter, and A. Eitel, "Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis," *Educ. Res. Rev.*, vol. 17, pp. 19–36, 2016.
[33] S. DeVore, E. Marshman, and C. Singh, "Challenge of engaging all students via self-paced interactive electronic learning tutorials for introductory physics," *Phys. Rev. Phys. Educ. Res.*, vol. 13, no. 1, p. 10127, 2017.
[34] J. N. Linton, "Institutional Factors for Supporting Electronic Learning Communities.," *Online Learn.*, vol. 21, no. 1, pp. 238–256, 2017.
[35] A. Z. Hamdi, A. H. Asyhar, Y. Farida, N. Ulinnuha, D. C. R. Novitasari, and A. Zaenal, "Sentiment Analysis of Regional Head Candidate's Electability from the National Mass Media Perspective Using the Text Mining Algorithm."
[36] V. Y. Shurygin and L. A. Krasnova, "Electronic Learning Courses as a Means to Activate Students' Independent Work in Studying Physics.," *Int. J. Environ. Sci. Educ.*, vol. 11, no. 8, pp. 1743–1751, 2016.
[37] S. Lailiyah, T. Nusantara, C. Sa'Dijah, E. B. Irawan, Kusaeri, and A. H. Asyhar, "Structuring students' analogical reasoning in solving algebra problem," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 296, no. 1, 2018.
[38] D. Darmawan, H. Kartawinata, and W. Astorina, "Development of Web-Based Electronic Learning System (WELS) in Improving the Effectiveness of the Study at Vocational High School" Dharma Nusantara"..," *JCS*, vol. 14, no. 4, pp. 562–573, 2018.
[39] C. M. Reigeluth, Instructional Design theories and models. 1999.
[40] J. Capka and C. Foltin, "Policy Implications of a Proposed Framework to Improve the Accessibility and Effectiveness of Internships in Accounting.," *eJEP eJournal Educ. Policy*, 2017.
[41] J. O. Jasperson, "How to Integrate Student Internships into Legal Studies Research and Curriculum: A Case Study.," *Univers. J. Educ. Res.*, vol. 5, no. 9, pp. 1504–1509, 2017.
[42] K. W. Johnson, "A Case Study Exploration of Internships in Undergraduate Business Education," 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 104-109 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

pH Neutralization Process Control based on Active Disturbance Rejection Control

Mohanad Hamad Eljack Elameen*, Mona Bahaeldin Musa Hago, Wen Tan

School of Control and Computer engineering, North China Electric Power University, 102206, China

ARTICLE INFO

Article history:
Received: 06 June, 2019
Accepted: 16 August, 2019
Online: 03 September, 2019

Keywords:
ADRC
Extended state observer (ESO)
rADRC
Nonlinear process
pH neutralization process

ABSTRACT

This paper considers active disturbance rejection control (ADRC) as a new control approach for pH neutralization process. The pH process is challenging to control due to its sensitivity to disturbance, nonlinear dynamics, and time-varying characteristics. Special attention is paid to the impact of disturbances typically exists in the chemical process. First, the pH neutralization process nonlinear model is presented, then the active disturbance rejection control is described briefly and a pH neutralization process closed-loop control system is designed on its bases, and investigated with simulation. More robust and better disturbance rejection performances are achieved compared to the control technique of feedback linearization (FL).

1. Introduction

This paper is an extension of work presented initially in 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) [1]. Control of pH neutralization in the chemical and biotechnological industries is a typical process [2]. The pH of this effluent must be retained within environmental boundaries, before discharging the effluent streams from wastewater treatment plants. Tight pH control is also critically required in pharmaceutical production. However, owing to the high nonlinearity of the pH process, robust and high performance pH process control is often difficult to obtain [3-5]. The process gain of these processes can differ from a small range of pH values to several orders of magnitude. Thus it can exhibit severe static nonlinear behavior.

So far, many authors have proposed, developed, and implemented different control techniques to overcome this problem. The classical PID controller is the first technique applied to control the pH [6-8]. In the field of industrial applications, the PID controller is considered as the most popular controller types, but it has been unable to satisfy the growing demand for efficiency, speed, and accuracy. Nonlinear control techniques also have been applied [9-21]. Other works involve intelligent control technique [22, 23]. In general, most of the above methods require an adequate model, which could not be easy to obtain, add to that the complexity of the controller design based on the above approaches. Hence, pH neutralization process requires a new approach to control it.

In [24-26], Han introduced the powerful control technique active disturbance rejection control (ADRC) and simplified and identified in [27]. The states and disturbance can be estimated in the ADRC before an error can occur, which can ultimately attain early compensation and disturbance resistance. Since it was introduced, the ADRC method has gained worldwide attention and broad implementation and has become common after its linear form has appeared [28]. In many fields, ADRC has been implemented effectively, benefiting from its full control characteristics as fast system response, large scale adaptation, and small overshoot [29].

In this technique, a unique state observer is used to estimate what is called total disturbance. The uncertainties in the dynamics of the internal process and the disturbances of the output are considered as total disturbances. This observer is called the extended state observer (ESO), and it can track the state-space model's extended state, which can achieve this goal. The plant's state-space model should include only the model estimate. A linear model in a state-space form may be an estimate for a nonlinear system [29].

This work proposes a different control technique based on estimating the disturbances and actively rejecting it for pH neutralization process. All the disturbances typically presented in pH process are lumped together in one signal called total disturbances. The extended state observer tracks these disturbances and the control law cancels it.

The organization of this paper is as follows. The pH process's model is given in section 2. Section 3 briefly presents the technique

* Mohanad Hamad Eljack Elameen, mohanadeljack@gmail.com

of feedback linearization. Section 4 shows the ADRC control system for the pH process. Section 5 gives the simulation results, and section 6 gives a general conclusion.

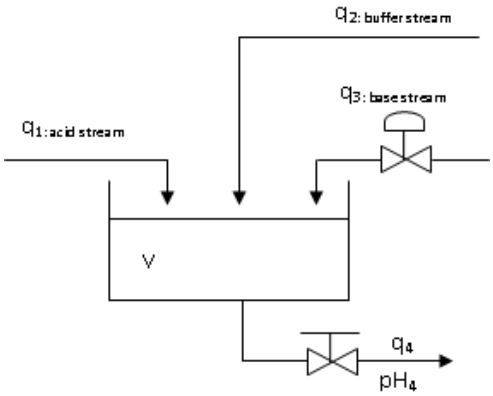
2. Model of pH Process

Figure 1 shows a schematic diagram describing the pH process. The process consists of an acid flow rate (q_1), buffer flow rate (q_2), and the base flow rate (q_3) mixed in the tank. pH_4 represents the output of the process, where q_3 represents the control input. By using the conservation equations of the reaction invariants and equilibrium relations, a dynamic model has been derived. Set tank volume (V), Ideal mixing, constant density, and full ion solubility are considered as assumptions for modeling. The system's chemical reactions are:

$$H_2CO_3 \leftrightarrow HCO_3^- + H^+,$$

$$HCO_3^- \leftrightarrow CO_3^{2-} + H^+,$$

$$H_2O \leftrightarrow OH^- + H^+$$



The reaction's equilibrium constants are

$$K_{a1} = \frac{[HCO_3^-][H^+]}{[H_2CO_3]},$$

$$K_{a2} = \frac{[CO_3^{2-}][H^+]}{[HCO_3^-]},$$

$$K_w = [H^+][OH^-]$$

The chemical balance is formed using the reaction invariant concept [30]. The invariants concentrations are (i=1,2,3,4):

$$W_{ai} = [HCO_3^-]_i + [H^+]_i - 2[CO_3^{2-}]_i - [OH^-]_i,$$

$$W_{bi} = [HCO_3^-]_i + [CO_3^{2-}]_i + [H_2CO_3]_i,$$

where W_{ai} is a charge-related quantity, and W_{bi} is the carbonate ion concentration, and they are independent of the extent of the reaction. The concentration of the hydrogen ion is calculated from:

$$W_{ai} + W_{bi}\frac{\frac{K_{a1}}{[H^+]_i} + \frac{2K_{a1}K_{a2}}{[H^+]_i^2}}{1 + \frac{K_{a1}}{[H^+]_i} + \frac{K_{a1}K_{a2}}{[H^+]_i^2}} + \frac{K_w}{[H^+]_i} - [H^+]_i = 0.$$

The hydrogen ion $[H^+]_4$ with the following relationship calculates the pH of the effluent stream:

$$pH_4 = -log_{10}([H^+]_4)$$

The dynamic model is given by the vessel's material balance:

$$V\frac{dW_{a4}}{dt} = q_1(W_{a1} - W_{a4}) + q_2(W_{a2} - W_{a4}) + q_3(W_{a3} - W_{a4}),$$

$$V\frac{dW_{b4}}{dt} = q_1(W_{b1} - W_{b4}) + q_2(W_{b2} - W_{b4}) + q_3(W_{b3} - W_{b4}).$$

This put the process model in a state-space form as

$$\dot{X} = f(x) + g(x)u + p(x)d \quad (1)$$

where

$$u = q_3, \quad x = [W_{a4}\ W_{b4}]^T, \quad d = [W_{a1}\ W_{b1}]^T,$$

$$f(x) = \begin{bmatrix} -\frac{q_1}{V}x_1 + \frac{q_2}{V}(W_{a2} - x1) \\ -\frac{q_1}{V}x_2 + \frac{q_2}{V}(W_{b2} - x2) \end{bmatrix},$$

$$g(x) = \begin{bmatrix} \frac{1}{V}(W_{a3} - x1) \\ \frac{1}{V}(W_{b3} - x2) \end{bmatrix}, \qquad p = \begin{bmatrix} \frac{q_1}{V} & 0 \\ 0 & \frac{q_1}{V} \end{bmatrix}.$$

Note that Eq. (1) is linear in p, and the unknown parameters are presented in the vector d. The relation between the state variables x and the output y is nonlinear as shown by

$$c(x, y) = 0, \quad (2)$$

where

$$c(x, y) = x_1 + 10^{y-14} - 10^{-y} + x_2 c_{x2}, \quad (3)$$

$$c_{x2} = \frac{1 + 2 \times 10^{y-pk_2}}{1 + 10^{pk_1 - y} + 10^{y-pk_2}},$$

$$pk_1 = -log_{10}k_{a1}, \quad pk_2 = -log_{10}k_{a2}.$$

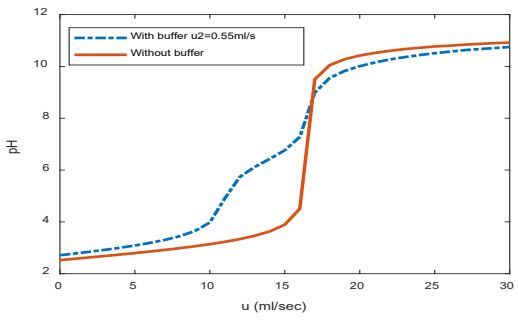


Figure 2: The dynamics of the pH process

Table 1 shows the nominal values of the parameters.

Table 1. Model Parameters Nominal Values

Parameter	Nominal value
V	2900 ml
K_{a1}	4.47×10^{-7}
K_{a2}	5.62×10^{-11}
$[q_1]$	0.003 M HNO3
	5×10^{-5} M H2CO3
$[q_2]$	0.03 M NaHCO3
$[q_3]$	0.003 M NaOH
	5×10^{-5} M NaHCO3
q_1	16.6 ml/s
q_2	0.55 ml/s
q_3	15.8 ml/s
pH_4	7.00
W_{a1}	0.003 M
W_{b1}	5×10^{-5} M
W_{a2}	-0.03 M
W_{b2}	0.03 M
W_{a3}	-3.05×10^{-3} M
W_{b3}	5×10^{-5} M
W_{a4}	-4.50×10^{-4} M
W_{b4}	5.5×10^{-4} M

Figure 2 shows the static dynamics of the pH process, which shows strong nonlinearity.

3. Feedback Linearization Technique

For the severely nonlinear system expressed in (1-3), a nonlinear controller based on feedback linearization (FL) technique is first designed.

Consider the following nonlinear system

$$\dot{x} = f(x) + g(x)u, \quad y = h(x) \tag{4}$$

The objective is to find a linear differential relation between a new input v and the output y. The system has a relative degree r if

$$L_g L_f^i = 0 \qquad 0 \le i \le r - 2 \tag{5}$$

$$L_g L_f^{r-1} h(x) \neq 0$$

where Lf in the direction of f is the Lie derivative. Suppose the system above has a relative degree of r and is linearizable, then the input transformation

$$u = \frac{v - L_f^r h}{L_g L_f^{r-1} h} \tag{6}$$

hence, the relation between y and v is linear as

$$y^{(r)} = v \tag{7}$$

For the pH process, by differentiating (3) using (1) we get

$$\dot{y} = -c_y^{-1}(x,y)c_x(y)[f(x) + g(x)u + p(x)d] \tag{8}$$

where

$$c_x = \frac{\partial c(x,y)}{\partial x} = [c_{x1} \quad c_{x2}],$$

c_{x1}=1,

$$c_{x2} = \frac{1+2\times 10^{y-pk_2}}{1+10^{pk_1-y}+10^{y-pk_2}}.$$

$$c_y = (\ln 10) \times \left\{ x_2 \left(\frac{10^{pk_1-y} + 10^{y-pk_2} + 4 \times 10^{pk_1-pk_2}}{(1 + 10^{pk_1-y} + 10^{y-pk_2})^2} \right) + 10^{y-14} + 10^{-y} \right\},$$

Since $-c_y^{-1}(x,y)c_x(y)g(x) \neq 0$ for all x and y of interest, the relative degree $r = 1$, and the linearization relation is

$$u = \frac{v + [f(x) + p(x)d]c_y^{-1}(x,y)c_x(y)}{-c_y^{-1}(x,y)c_x(y)g(x)}$$

Now, assume v is the output of a PID controller in the form

$$v = k_P e + k_I \int e + \dot{y_{sp}} \tag{9}$$

where $e = y_{sp} - y$ and y_{sp} is the reference signal, k_P is the proportional gain, k_i is the integral gain.

The control structure of the FL controller is shown in Figure 3.

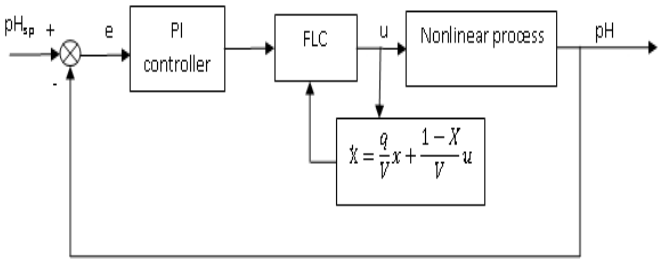


Figure 3: The control structure of the FL controller

4. ADRC Solution for the pH Process

It is shown that the control structure using the feedback linearization method is very complicated. To simplify the control design, ADRC is designed. It is shown that all the nonlinear and time-varying terms under ADRC are components of the total disturbance that need to be estimated and rejected, making the design and structure much more straightforward.

4.1. Conventional linear ADRC

We can rewrite the system described in (1) as

$$\dot{y} = f + bu \tag{10}$$

where f is the total disturbance, and b is the coefficient of the output change rate to the control variable step and varies all through the process. Now, ADRC task is to estimate f in real time and cancel it in the control law, which putting the extended state observer as a choice tool from the modern control theory. Define

the cases as $x_1 = y$ and $x_2 = f$, where the extended state is represented by x_2; then, Eq. (10) rewritten as

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}b\\ 0\end{bmatrix}\dot{f}$$

$$y = x_1 \tag{11}$$

An ESO can be constructed to estimate the states

$$\begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix} + \begin{bmatrix}b\\ 0\end{bmatrix}u + \begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}(y - \hat{y})$$

$$\hat{y} = z_1 \tag{12}$$

where β_1 and β_2 represent the gains of the observer.

The total disturbance $z_2$will be estimated by a well-tuned observer in Eq.(12), and the control law

$$u = \frac{-z_2 + u_0}{b} \tag{13}$$

where u_0 is a virtual control signal, then Eq.(10) reduced to

$$\dot{y} = u_0 + (f - z_2) \approx u_0 \tag{14}$$

which can be easily controlled using a proportional controller.

$$u_0 = k_p(r - z_1) \tag{15}$$

Here k_p is the controller gain and r is the reference signal.

The structure of the ADRC is shown in Figure 4.

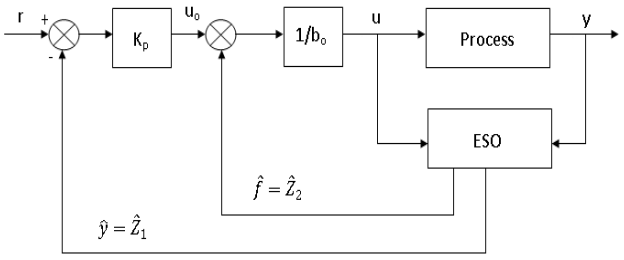


Figure 4: ADRC structure for a first-order process

The gains of the observer are chosen as $\beta_1 = 2\omega_{ob}$ and $\beta_2 = \omega^2_{ob}$ to put the observer poles at $-\omega_{ob}$. In the same way, the controller gains are selected as $k_p = \omega_c$ to put the pole of the control loop at $-\omega_c$. ω_{ob} and ω_c are called the observer bandwidth and controller bandwidth, respectively.

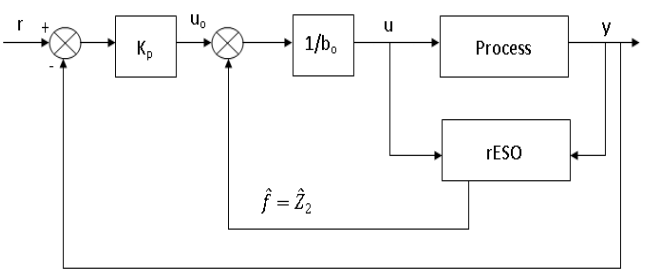


Figure 5: ESO based reduced-order ADRC

4.2. Reduced-order ADRC (rADRC)

In the ADRC, a second-order ESO is used to obtain the extended state compensation estimate; however, redundancy occurs when the state's first component can be directly measured [31]. A reduced-order state observer is employed, as shown in Figure 5.

The state observer of reduced-order for x_2 is

$$\dot{z}_2 = -\omega_{ob}z_2 - \omega_{ob}bu + \omega_{ob}\dot{y} \tag{16}$$

where ω_{ob} (>0) determines the convergent rate of the ESO. To avoid intensifying measurement noises by direct numerical differentiation of signal y in (16), we define a new state as

$$z = z_2 - \omega_{ob}y \tag{17}$$

Therefore (16) may be rewritten as

$$\dot{z} = -\omega_{ob}z - \omega_{ob}{}^2 y - \omega_{ob}bu \tag{18}$$

which combined with

$$z_2 = z + \omega_{ob}y \tag{19}$$

constitutes the reduced-order ESO. The remaining controller design is the same as those above.

5. Simulation Results

A comparison between the performances of the designed FL and ADRC controllers is carried out through simulation. The change in buffer and feed stream flow rate is regarded as a disturbance to the system and shown in Fig. 6. The ADRC parameters are shown in Table 2.

Table 2. ADRC controller parameters

ADRC	$\omega_{ob} = 5$	b = 0.05	$\omega_c = 10$
rADRC	$\omega_{ob} = 0.9$	b = 0.05	$\omega_c = 10$

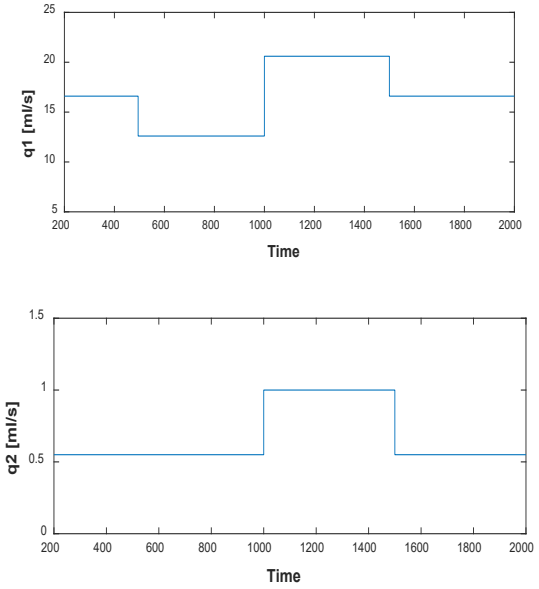


Figure 6: The change of buffer and feed stream flow rate

There are two cases considered in this simulation.

A. Constant Parameters

First, let us look at the case of the constant parameter. The efficiency of both disturbance rejection (Figure 7) and setpoint tracking (Figure 8) is examined to evaluate the effectiveness and robustness of the control system. A comparison to the FL controller carried out to illustrate the advantages of the ADRC as well as the rADRC. It can be observed from Figure 7 the superiority of ADRC and rADRC over the FL controller. We can see from Figure 8 that ADRC tracks the change in setpoint very well, and rADRC tracks it even better, while the FL controller shows some overshoot with longer settling time.

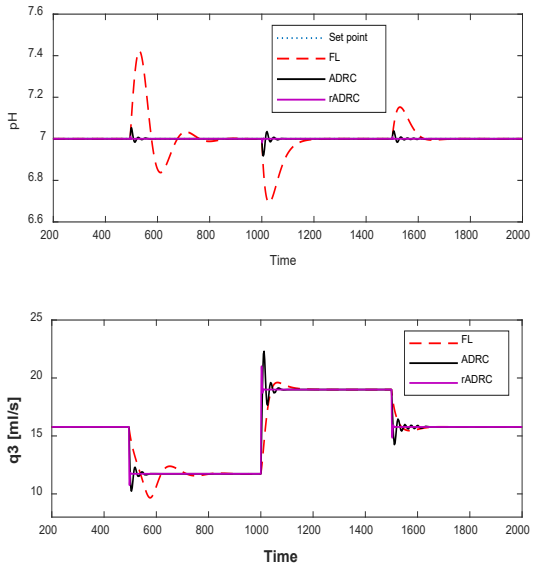


Figure 7: The performance of disturbance rejection

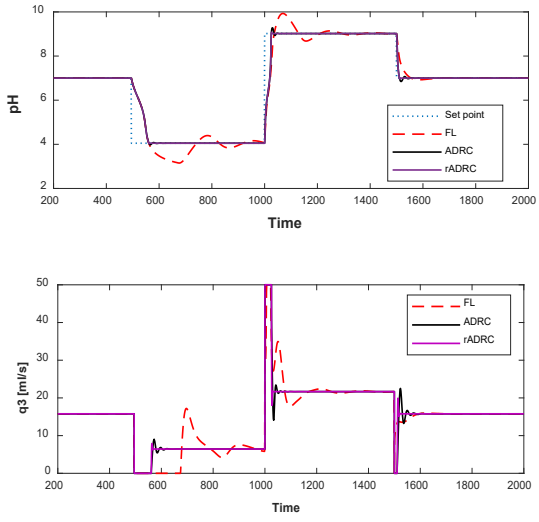


Figure 8: The performance of setpoint tracking

B. Variations in the parameters

The parameters changes are discussed In this case. They are regarded and handled as a system's internal disturbance. Figure 9 shows the changes in the concentrations of the feed stream.

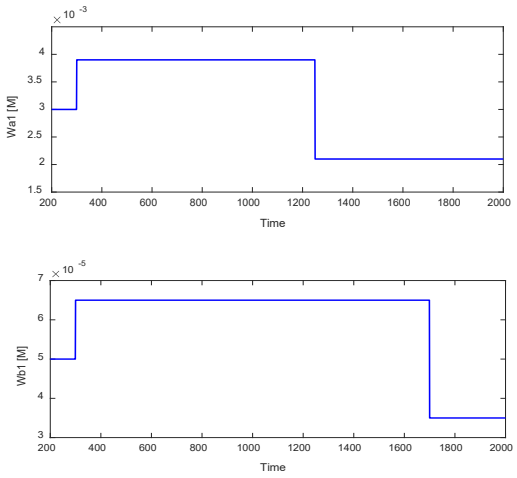


Figure 9: Variations in the concentration of the feed stream

Figure 10 shows the performance of the disturbance rejection, while Figure 11 shows the performance of the setpoint tracking. We can see obviously that ADRC and rADRC are superior to the FL controller.

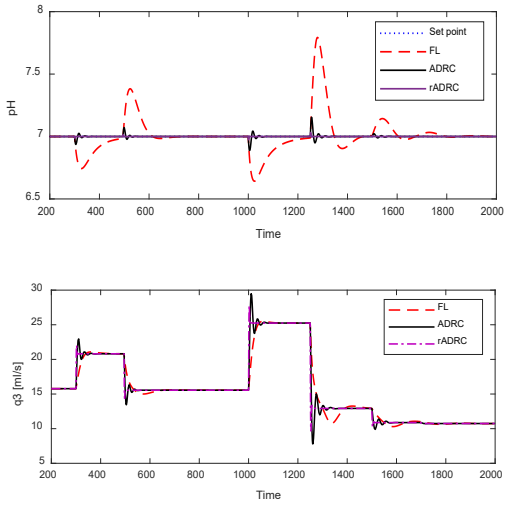


Figure 10: The performance of disturbance rejection with variation in parameters

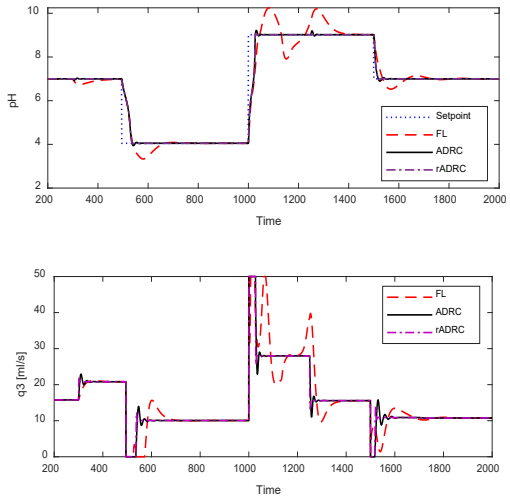


Figure 11: The performance of setpoint tracking with variation in parameters

6. Conclusion

A novel control technique (ADRC) for a pH process is presented in this paper. ADRC is an excellent solution to the process that needs a high performance controller to overcome the modeling errors and disturbances. ADRC's pH simulation studies showed a major enhancement over nonlinear FL control. For more analysis, a reduced-order ADRC was designed and proved that it has the most significant performance in regulating the pH in the existence of the input and external disturbances. The ADRC method is a strong algorithm, particularly in rejection of disturbances. Compared to FL control technique, design and tuning are much more straightforward.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. H. Eljack, M. B. Musa, W. Tan, "pH Neutralization Process Control Using Linear Active Disturbance Rejection Control." In 2018 IEEE 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). 205-210, 2018. https://doi.org/10.1109/iccwamtip.2018.8632617

[2] M. A. Henson, D. E. Seborg, "Adaptive nonlinear control of a pH neutralization process," IEEE transactions on control systems technology, 2(3), 169-182, 1994. https://doi.org/10.1109/87.317975

[3] F. G. Shinskey, pH and PION Confro1 in Process and Waste Streams.New York: Wiley, 1973. https://doi.org/10.1016/0013-9327(74)90073-1

[4] G. K. McMillan, pH Control. Research Triangle Park, NC: Instrument Society of America, 1984.

[5] A. Johnson, "The control of fed-batch fermentation processes-A survey," Automatica, 23(6): 691-705, 1987. https://doi.org/10.1016/0005-1098(87)90026-4

[6] R. A. Aguiar, I. C. Franco, F. Leonardi, & F. Lima. Fractional PID Controller Applied to a Chemical Plant with Level and pH Control. Chemical Product and Process Modeling, **13**(4), 2018. https://doi.org/10.1515/cppm-2018-0004

[7] R. Vilanova, A. Visioli, PID Control in the Third Millennium: Lessons Learned and New Approaches", Springer, London 2012.

[8] R. A. Wright, C. Kravaris, "Nonlinear pH control in a CSTR," American Control Conference Proceeding, 1540-1544, 1989. https://doi.org/10.23919/acc.1989.4790431

[9] T. K. Gustafsson, K. V. Waller, "Dynamic modeling and reaction invariant control of pH," Chem. Eng. Sci. **38**(3): 389-398, 1983. https://doi.org/10.1016/0009-2509(83)80157-2

[10] P. Jutilia, "An application of adaptive pH-control algorithms based on physio-chemical modeling in a chemical waste-water treatment plant," Int. J. Control. **38**(3): 639-655,1983. https://doi.org/10.1080/00207178308933099

[11] G. Pajunen, "Comparison of linear and nonlinear adaptive control of pH-process," in Proc. IEEE Control Systems Magazine. **7**(1): 39-43, 1987. https://doi.org/10.1109/mcs.1987.1105238

[12] J. R. Panish, C. B. Brosilow, "Nonlinear inferential control," AIChE J., **34**(4): 633-644, 1988.

[13] W. C. Li and L. T. Biegler, "Newton-type controllers for constrained nonlinear processes with uncertainty," Ind. Eng. Chem. Res., **29**(8): 1647-1657, 1990. https://doi.org/10.1021/ie00104a012

[14] G. L. Williams, R. R. Rhinehart, J. B. Riggs, "In-line process model-based control of wastewater pH using dual base injection," Ind. Eng. Chem. Res., **29**(7): 1254-1259, 1990. https://doi.org/10.1021/ie00057a029

[15] J. J. E. Slotine, W. Li, Applied Nonlinear Control, 2nd ed., Prentice Hall, Inc., Upper Saddle River, NJ, 1991.

[16] M. H. E. Elameen, M. N. Karsiti, & R. Ibrahim, "Nonlinear model feedback linearization control strategy of a pH neutralization process," In 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS) (pp. 1-5). IEEE, 20014. https://doi.org/10.1109/icias.2014.6869488

[17] S. F. Graebe, M. M. Seron, G. C. Goodwin, "Nonlinear tracking and input disturbance rejection with application to pH control," Journal of Process Control, 6(2-3): 195–202, 1996. https://doi.org/10.1016/0959-1524(96)86054-9

[18] H. Degachi, W. Chagra, & M. Ksouri. "Nonlinear model predictive control for pH neutralization process based on SOMA algorithm." INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, **9**(1), 391-398, 2018. https://doi.org/10.14569/ijacsa.2018.090153

[19] A. W. Hermansson & S. Syafiie. "Model predictive control of pH neutralization processes: a review." Control Engineering Practice, 45, 98-109, 2015. https://doi.org/10.1016/j.conengprac.2015.09.005

[20] J. Chen, Z. Liu, Y. Peng, Q. Guo, "A New Control Method in pH Neutralization Process." Springer, Berlin, 511-519, 2013. https://doi.org/10.1007/978-3-642-38524-7_56

[21] A. Grancharova, J. Kocijan, & T. A. Johansen, "Explicit output-feedback nonlinear predictive control based on black-box models," Engineering Applications of Artificial Intelligence, **24**(2), 388-397, 2011. https://doi.org/10.1016/j.engappai.2010.10.009

[22] M. Heredia-Molinero, J. Sanchez-Prieto, J. Briongos, M. Palancar, "Feedback PID-like fuzzy controller for pH regulatory control near the equivalence point," J. Process Control, 2014, **24**(7): 1023–1037. https://doi.org/10.1016/j.jprocont.2014.05.006

[23] M. H. Dolatabadi, M. N. Monfared, A. Fakharian, "Robust H-infinity control for the pH neutralization process based on fuzzy models." In: Fuzzy Systems (IFSC), 13th Iranian Conference on, IEEE, 1–6, 2013. https://doi.org/10.1109/ifsc.2013.6675681

[24] J. Q. Han, "Auto-disturbances-rejection controller and its Application, Control and Decision," January, **13**(1): 19–23, 1998. (In Chinese).

[25] J. Q. Han, "From PID Technique to Active Disturbances Rejection Control Technique," Control Engineering of China, **9**(3): 13–18, March 2002. https://doi.org/10.1109/tie.2008.2011621

[26] J. Q. Han., "Active Disturbance Rejection Control Technique – the technique for estimating and compensating the uncertainties," National Defense Industry Press, 197-270, 2008.

[27] Z. Gao, "Scaling and parameterization based controller tuning," Proceedings of the American Control Conference, Denver, **6**: 4989-4996, 2003. https://doi.org/10.1109/acc.2003.1242516

[28] Z. Chen, Y. Wang, M. Sun, & Q. Sun, "Convergence and stability analysis of active disturbance rejection control for first-order nonlinear dynamic systems," Transactions of the Institute of Measurement and Control, **41**(7), 2064-2076, 2018. https://doi.org/10.1177/0142331218794812

[29] J. Ping & B. Wang. "Development of User-defined Block Library for Active-Disturbance-Rejection-Control," In 2010 Third International Conference on Information and Computing, Vol. 2, pp. 151-154, 2010. IEEE. https://doi.org/10.1109/icic.2010.132

[30] K. V. Walker, P. M. Makila, "Chemical reaction invariants and variants and their use in reactor modelling, simulation, and control," Ind. Eng. Chem. Process Des. Dev., **20**(1): 1-11, 1981. https://doi.org/10.1021/i200012a001

[31] R. Yang, M. Sun, Z. Chen, "Active disturbance rejection control on the first-order plant." Journal of Systems Engineering and Electronics, **22**(1): 95-102, 2011. https://doi.org/10.3969/j.issn.1004-4132.2011.01.012

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 107-111 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Deep Feature Representation for Face Sketch Recognition

Weiguo Wan[1], Hyo Jong Lee [1, 2, *]

[1]*Division of Computer Science and Engineering, Chonbuk National University, 54896, Korea*

[2]*Center for Advanced Image and Information Technology, Chonbuk National University, 54896, Korea*

ARTICLEINFO

Article history:
Received: 07 January, 2019
Accepted: 28 February, 2019
Online: 15 March, 2019

Keywords:
Face sketch recognition
Face sketch synthesis
VGG-Face network

ABSTRACT

Face sketch recognition aims at matching face sketch images to face photo images. The main challenge lies in modality discrepancy between face photo and sketch images. In this work, we propose a new facial sketch-to-photo recognition approach by adopting VGG-Face deep learning network, with which face images can be represented by compact and highly discriminative feature vectors. Different from existing VGG-Face based methods which directly match face sketches to photos, we firstly transform the gallery photos to sketches for decreasing the modality difference. Experimental results on multiple face photo-sketch datasets indicate the superiority of our method.

1. Introduction

Face sketch to photo recognition is a key branch of face recognition to address the lack of face photos, which has wide application in forensics [1]. If crime happened, and only incomplete information is obtained about the suspect because of the bad quality of monitoring videos, face sketches which drawn by the artists with the description of witnesses are used to identity the possible suspect. The police can shrink the range of suspects by searching the law enforcement face databases or surveillance cameras with the drawn face sketch images [2].

Face sketch to photo recognition approaches can be mainly classified into common space projection approaches and local feature descriptor approaches. Common space projection approaches attempt to transform facial images from various modalities to common subspace where the modality difference is reduced. After that, facial photos and sketches could be directly identified in the common space. Lin et al. [3] translated heterogeneous features to common feature space with common discriminant feature extraction (CDFE). In [4], Lei et al. proposed a subspace learning framework named coupled spectral regression (CSR) for heterogeneous face recognition. Then, they improved the CSR algorithm through learning a mapping from all modalities to all samples [5]. In [6], Sharma et al. employed partial least squares (PLS) algorithm to project heterogeneous facial images into common subspace. Mignon et al. [7] proposed to learn a discriminative latent space by using cross modal metric learning (CMML) method. In [8], Kan et al. suggested a multiple view analysis algorithm, with which the dependencies from intra-view and inter-view can be utilized to obtain a discriminant subspace for heterogeneous face recognition. However, the space projection based methods may lose the underlying information of the source images and lead to degradation of the recognition effect.

Local feature representation approaches aim at extracting robust and modality-invariant face feature representations from the heterogeneous face images. In [9], Klare et al. put forward a face sketch recognition approach with a local feature discriminant analysis (LFDA) framework which taking advantage of multiscale local binary pattern (MLBP) [10] and scale invariant feature transform (SIFT) [11]. In [12], Zhang et al. suggested a coupled information theoretic encoding based local face descriptor for face sketch to photo recognition. And later, the coupled information theoretic encoding algorithm was modified to the random forests by using various sampling methods. Galoogahi et al. [13] put forward a facial feature descriptor named local radon binary pattern (LRBP) for face sketch recognition, which firstly projects the face images into the radon space and then encodes them with the local binary patterns (LBP). In addition, a local face descriptor based on histogram of averaged oriented gradients (HAOG) is suggested by Galoogahi et al. to decrease the modality difference [14]. In [15], Lei et al. put forward a heterogeneous face

*Corresponding Author: Hyo Jong Lee, hlee@chonbuk.ac.kr

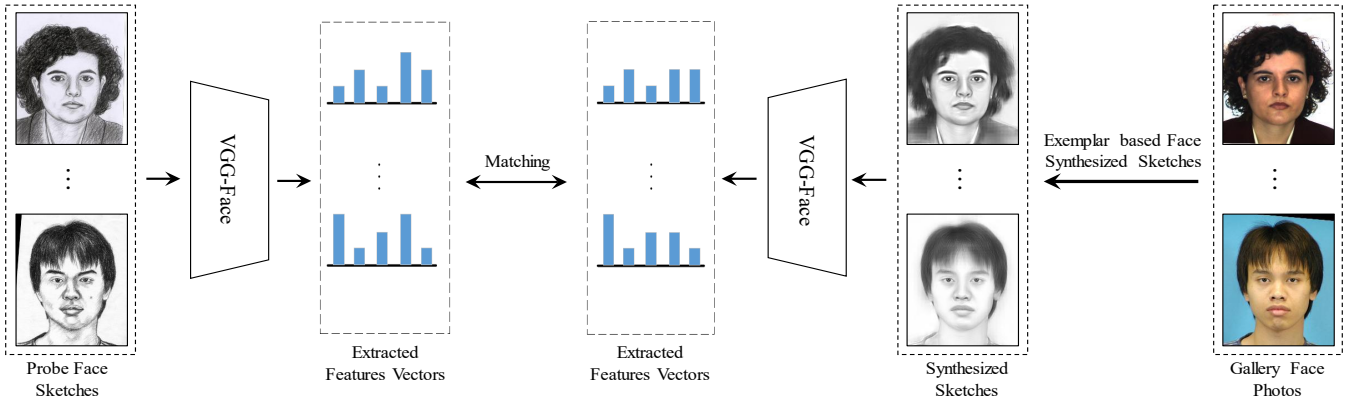


Figure 1. The illustration of the proposed face sketch recognition approach.

recognition method based on discriminant image filter learning. Alex et al. [16] suggested a local difference of Gaussian binary pattern algorithm for heterogeneous face recognition. However, when extracting the face feature representation, the local feature descriptor based approaches lose the sight of the holistic spatial structure of face, which is important for face sketch recognition.

Moreover, because of the huge differences between the face photo and sketch images, it is difficult for traditional homogeneous face recognition approaches to achieve good performance by directly matching the sketches to photos. In this paper, we transform the photo images into sketch images by adopting an exemplar-based sketch generation approach to decrease the modality difference between face photo and sketch images. In [17], Parkhi et al. presented a deep learning network, namely VGG-Face, which is able to map a face images to a compact space where distances can be used for measuring the face similarity. Peng et al. [18] mentioned that it can be adopted for face sketch recognition. Motivated by its efficient performance, we adopt it to extract the discrimination face feature representation in this paper. With the VGG-Face network, the feature vectors which represent the corresponding face images can be obtained. The illustration of the proposed approach is showed in Figure 1.

2. Related Works

2.1. Exemplar-based Sketch Generation

The exemplar-based face sketch generation approach needs a group of training dataset which contains some face photo-sketch pairs. First, each test photo **T** is divided into several image patches with overlapping between adjacent patches. Then, for each test photo patch **t**, several number of closest photo patches are chose from the training photo images which are cut image patches with the same approach. At the same time, we can get *K* corresponding sketch candidate patches with the obtained photo patches. Finally, the sketch patches are generated by combining the *K* candidate sketch patches with the corresponding weighting coefficients. The sketch generation operation can be represented as follow:

$$\min_{w} \| \mathbf{t} - \mathbf{X} \cdot \mathbf{w} \|_2^2, \ \ s.t. \ \mathbf{1}^T \mathbf{w} = 1 \tag{1}$$

where **X is** the *K* selected photo patches from training data. **w** is the weighting vector to combine candidate sketch patches.

Then we can obtain the target sketch patch ***s*** by:

$$\mathbf{s} = \mathbf{Y} \cdot \mathbf{w} \tag{2}$$

where **Y** represents the corresponding sketch patches of the test photo patch **t**.

2.2. VGG-Face Networks

The VGG-Face networks [17] are built on the basis of VGG-Very-Deep-16 convolutional neural network (CNN) architecture. The VGG-Face networks consist of a series of convolutional, pool, and fully-connected layers. The first eight blocks are convolutional layers and the remaining three blocks are fully-connected layers. For each convolutional layers, a ReLU activation layer is followed. The output dimensions of the first two fully-connected layers are 4096 and the output dimension of final fully-connected layer is 2622. The filters of convolutional layers are with size of 3×3 while the pool layers perform subsampling with a factor of two. Figure 2 displays the framework of the VGG-Face model.

In the proposed method, a pretrained VGG-Face model with 37 basic layers is utilized to extract deep face features. The pretrained model is trained on a big face dataset which has 982,803 images from 2622 subjects. The pretrained VGG-Face model only can recognize the people in the training dataset, however, we are able to extract face features from its bottleneck layers for any face images by forwarding the face images through the whole networks. The extracted features are highly discriminative, compact, and interoperable encodings of the input face images. When the face features are obtained from the bottleneck layers of the VGG-Face model, the sketch to photo matching can be performed based on these obtained face features. In this paper, the MatConvNet toolbox, which provides the pretrained implementation of the VGG-Face model, is utilized to extract deep face features.

3. The Proposed Method

In this paper, we presented a new facial sketch-to-photo recognition approach with the VGG-Face network and face sketch generation. The proposed method mainly consists of face image preprocessing, face sketch synthesis, and face sketch recognition three steps, which will be described later.

3.1. Face Image Preprocessing

Face photo and sketch images vary in properties, e.g. varying resolution, pose, and deformation. We firstly normalized all the

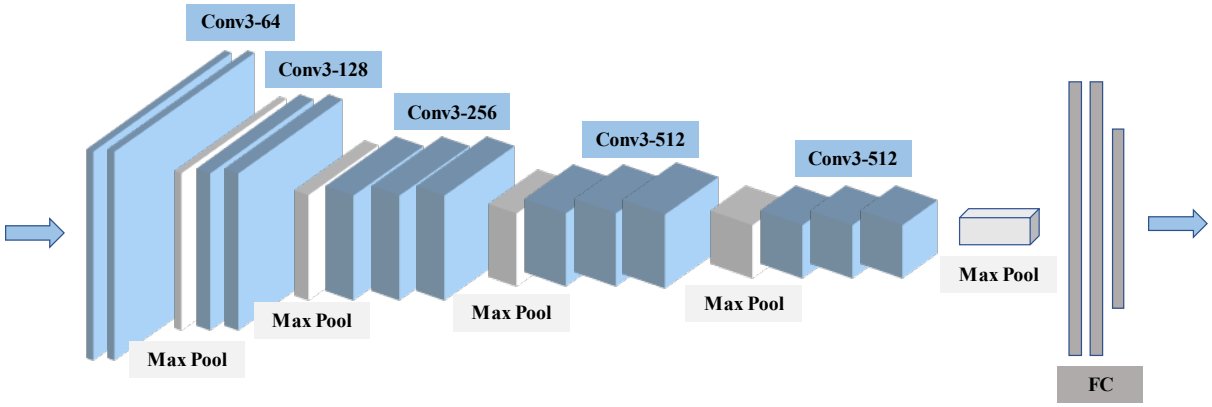


Figure 2. The VGG-Face CNN architecture.

sketch and photo images by fixing the coordinates of eyes centers. The common-used Dlib library was used to detect eyes position, and then translation, rotation, and scaling operations are conducted to align face images based on the detected eyes coordinates. Figure 3 shows some samples of original and preprocessed face photo and sketch images.

3.2. Face Photo to Sketch Synthesis

Assuming there are M geometrically aligned training photo-sketch pairs. We first divide these photo images and sketch images in training data into patches. Then, we reshape every patch to a vector. For each patch position, the searching area is extended c pixel around the patch. Thus, we can obtain $(2c+1)^2$ patches in the searching area for each patch position. We will obtain $(2c+1)^2M$ face photo and sketch patches from the training data, respectively. We employ the random sampling algorithm for selecting K face photo patches $\mathbf{U}^{(i,j)} \in \mathbf{R}^{2p^2 \times K}$ and sketch patches $\mathbf{V}^{(i,j)} \in \mathbf{R}^{2p^2 \times K}$ from training data. (i, j) is the patch position at the m-th row and the n-th column. Each image patch can be reshaped to a d-dimensional column vector.

For each test photo patch $\mathbf{t}_1^{(i,j)}$, we can compute the reconstruction weight by using equation (3):

$$\min_{\mathbf{w}^{(i,j)}} \| \mathbf{t}_1^{(i,j)} - \mathbf{U}^{(i,j)}\mathbf{w}^{(i,j)} \|_2^2 + \lambda \| \mathbf{d}^{(i,j)} \square \mathbf{w}^{(i,j)} \|, s.t. \mathbf{1}^T \mathbf{w}^{(i,j)} = 1 \quad (3)$$

where $\mathbf{d}^{(i,j)} \in \mathsf{R}^{K\times 1}$ represents the L2 distance of the test photo patch $\mathbf{t}_1^{(i,j)}$ and the selected photo patches $\mathbf{U}^{(i,j)}$ from training data, $\mathbf{w}^{(i,j)} \in \mathsf{R}^{K\times 1}$ represents the weighting coefficients of the test photo patch $\mathbf{t}_1^{(i,j)}$.

Equation (3) has the closed-form solution:

$$\begin{aligned} \mathbf{w}'^{(i,j)} &= (C^{i,j} + \lambda \mathrm{diag}(\mathbf{d}^{(i,j)})) \setminus \mathbf{1} \\ \mathbf{w}^{(i,j)} &= \mathbf{w}'^{(i,j)} / \mathbf{1}^T \mathbf{w}'^{(i,j)} \end{aligned} \quad (4)$$

where $\mathbf{1}$ represent a column vector, its elements all are 1. $C^{i,j} = (\mathbf{U}^{(i,j)} - \mathbf{1}\mathbf{t}_1^{(i,j)^T})(\mathbf{U}^{(i,j)} - \mathbf{1}\mathbf{t}_1^{(i,j)^T})^T$ is the covariance matrix, $\mathrm{diag}(\mathbf{d}^{(i,j)})$ is a diagonal matrix which is extended from $\mathbf{d}^{(i,j)}$.

By linearly combining K selected sketch patches of the training data and the weighting coefficients $\mathbf{w}^{(i,j)}$, the result sketch block



Figure 3. Facial photo-sketch examples in CUFS dataset and CUFSF dataset.

$\mathbf{s}^{(i,j)}$ could be generated:

$$\mathbf{s}^{(i,j)} = \mathbf{V}^{(i,j)}\mathbf{w}^{(i,j)} \quad (5)$$

We can generate the final target sketch with overlapping area averaged when get all the target sketch patches,. Fig. 3 shows some generated examples with the exemplar-based face sketch generation. From Figure 3, we can observe that the modality difference between the sketch image and the transformed sketch image smaller than the photo image.

3.3. Face Sketch Recognition with VGG-Face Network

After the photo images are generated to sketch images, the VGG-Face network was adopted to get the facial feature vectors of synthesized sketch images. We then call the synthesized sketch images gallery, and the test face sketch images probe.

We extract the face features by utilizing the VGG-face networks which is provided by the MatConvNet toolbox. The VGG-face architecture is composed of several convolution layers with 3×3 filters, several pooling layers with a factor of 2, and 3 fully-connected layers. In VGG deep networks, the early layers are

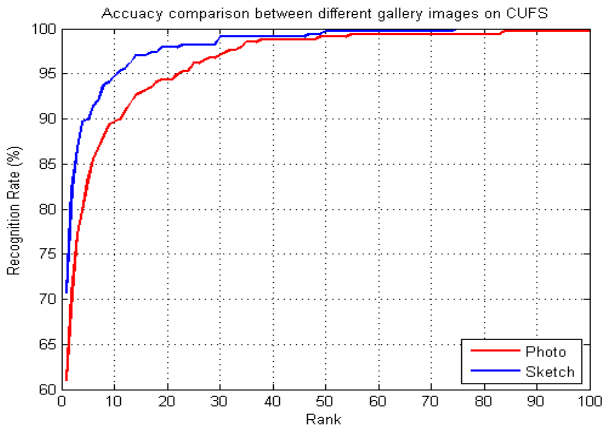


Figure 4. Performance comparison by using different gallery images.

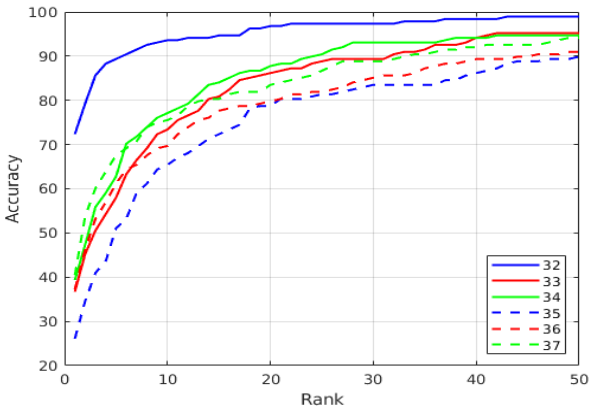


Figure 5. Performance comparison by different VGG-Face layers.

the low-level representation for input images such as edge and corner, the intermediate layers represent mid-level features like parts, and the high-level representations exist in the last layers. In order to find the optimum high-discrimination face features for face sketch-to-photo recognition, we selected activation maps from 32nd layer to 37th layer as the deep feature representation. The performance comparison results, which will be analyzed in the experiment section, indicated that the 32nd layer obtains the highest recognition accuracy. In this case, each face image was represented by a 25,088-dimension deep feature vector.

By using the well-trained VGG-Face network model, we can map all the sketch images in gallery into 25,088-dimension feature vectors. And a same length feature vector can be obtained by using the pre-trained network for each input probe image. We use the squared Euclidean distance of the face feature vectors to calculate the facial similarities. Face images from the different subjects have large distances while facial images of the same person have small distances. Hence, the face sketch recognition can be regarded as a nearest neighbor classification problem. The illustration of the proposed face sketch recognition method is shown in Figure 1.

4. Experiments and Results

4.1. Datasets

Two public available datasets are adopted to assess the recognition effect of the proposed approach: the CUHK dataset (CUFS) and the CUFSF dataset [19]. The CUFS dataset is composed of facial images from 3 datasets: the CUHK student dataset (188 subjects), the AR dataset (123 subjects), and the XM2VTS dataset (295 subjects). The CUFSF dataset has 1194 subjects of FERET dataset. In CUFS and CUFSF datasets, there are one photo image and one sketch image for each subject. All the face photo and sketch images are normalized and cut to with size of 250×200. The first two rows of Figure 3 shows the examples of photo-sketch pairs from these two datasets. The first three columns are face images of CUFS and the final column are face images of CUFSF. The first row are the original face sketches and the second row are the face photos. The final row are the generated sketches.

4.2. Experiments

First. we evaluated the recognition effect between adopting face photo images as gallery and adopting the generated sketch images as gallery. Figure 4 shown the comparison results, from it we can see that using the synthesized sketch images as gallery obtained better recognition performance. The results indicated that modality difference of the face photo and sketch images was narrowed down after synthesizing face photos to sketches.

We extracted last several layers from the VGG-face networks as the deep face features for recognition. To compare the performance by different layers, the experiment on CUHK student sub-dataset was conducted with same distance measurement. The result shown in Figure 5 indicated that the 32nd layer obtains the highest recognition accuracy. Thus, the feature vectors extracted from 32nd layer were set as the face representation in this paper.

Three existing approaches are compared to assess the recognition effect of the proposed approach, namely the HOG based approach, the DCP based approach [20] and the Light-CNN based approach [21]. Figure 6 (a) and (b) shown the comparison results on the CUFS and CUFSF datasets. Table 1 shown the rank-1, rank-5 and rank-10 recognition accuracies on CUFS and CUFSF datasets. From them, it can be seen that the proposed approach achieves the best recognition performance, compared to other methods. For the CUFS dataset, even though it is easy for other approaches to obtain high recognition performance, but our approach is the quickest one achieving 100% accuracy. The CUFSF dataset is more difficult due to the face photo images are taken under various illumination conditions. In addition, the face sketches exist shape deformation compared with the photos in CUFSF dataset. But the proposed approach still achieves the best recognition accuracy. In summary, the proposed face sketch to photo recognition approach is effective and practical.

5. Conclusions

In this paper, we presented a face sketch-to-photo recognition approach based on deep feature representation. The face photos were synthesized to face sketches to decrease the modality gap of the face photo images and sketch images. After that, the VGG-Face network was employed to extract feature representation for each face image. With the extracted face features, the face similarity can be measured directly. Experimental results on the CUFS and CUFSF datasets indicated the superiority of the proposed approach. The proposed approach in this paper is our preliminary research and only the pre-trained VGG-Face network was adopted. In the future, we plan to develop high-performance networks for face photo to sketch recognition.

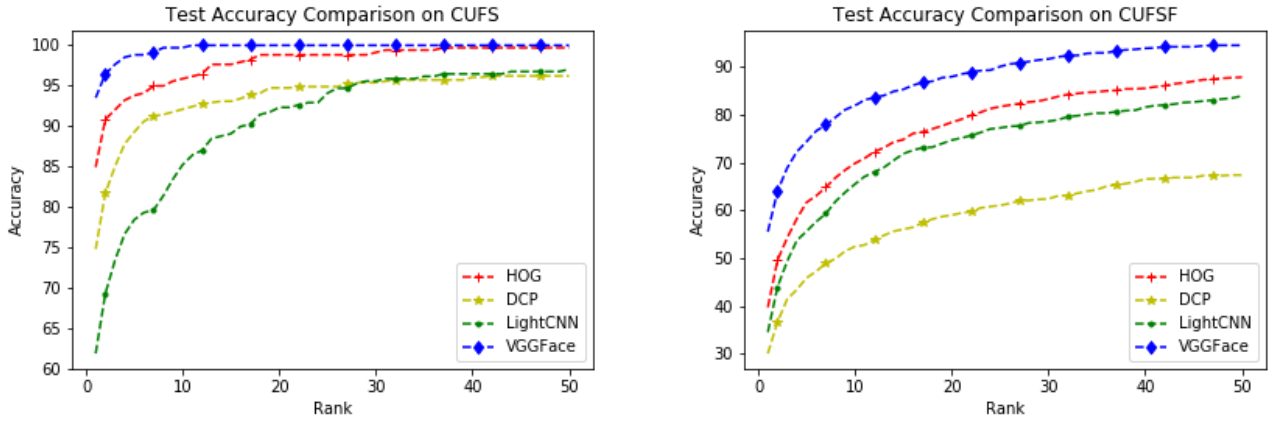


Figure 6. The comparison of recognition accuracies by different approaches on the CUFS and CUFSF datasets.

Table 1. Recognition accuracy (%) by different methods.

Methods		HOG	DCP	LightCNN	Proposed
CUFS	Rank-1	84.91	74.75	61.83	**93.49**
	Rank-5	93.79	89.27	78.40	**98.81**
	Rank-10	95.86	92.08	85.21	**99.70**
CUFSF	Rank-1	39.72	30.08	34.53	**55.51**
	Rank-5	61.65	45.87	55.61	**74.36**
	Rank-10	69.81	52.43	6.546	**81.89**

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by "Research Base Construction Fund Support Program" funded by Chonbuk National University in 2018. This research was also supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2015-0-00378) supervised by the IITP (Institute for Information & communications Technology Promotion). This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (GR 2016R1D1A3B0393 1911).

References

[1] M. Zhang, J. Li, N. Wang, and X. Gao, "Recognition of facial sketch styles," *Neurocomput.*, vol. 149, pp. 1188-1197, 2015.

[2] N. Wang, X. Gao, L. Sun, et al., "Anchored Neighborhood Index for Face Sketch Synthesis," *IEEE Trans. Circuits Syst. Video Technol*, 2017.

[3] D. Lin and X. Tang, "Inter-modality face recognition," *in Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 13–26.

[4] Z. Lei and S. Li, "Coupled spectral regression for matching heterogeneous faces," *IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 1123–1128.

[5] Z. Lei, C. Zhou, D. Yi, A. Jain, and S. Li, "An improved coupled spectral regression for heterogeneous face recognition," *in Proc. Int. Conf. Biomed.*, 2012, pp. 7–12.

[6] A. . Sharma and D. Jacobs, "Bypass synthesis: PLS for face recognition with pose, low-resolution and sketch," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 593–600.

[7] A. Mignon and F. Jurie, "CMML: A new metric learning approach for cross modal matching," *in Proc. Asian Conf. Comput. Vis.*, 2012.

[8] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *in Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 808–821.

[9] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[10] T. Ojala, M. Pietik€ainen, and T. M€aenp€a€a, "Multiresolution grayscale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002

[11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[12] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 513–520.

[13] H. Galoogahi and T. Sim, "Face sketch recognition by local radon binary pattern," *in Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 1837–1840.

[14] H. Galoogahi and T. Sim, "Inter-modality face sketch recognition," *in Proc. IEEE Int. Conf. Multimedia Expo*, 2012, pp. 224–229.

[15] Z. Lei, D. Yi, and S. Li, "Discriminant image filter learning for face recognition with local binary pattern like representation," *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2512–2517.

[16] A. Alex, V. Asari, and A. Mathew, "Local difference of gaussian binary pattern: Robust features for face sketch recognition," *in Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2013, pp. 1211–1216.

[17] O. M. Parkhi, V. Andrea, and Z. Andrew, "Deep face recognition," British Machine Vision Conference, vol. 1, no. 3, 2015.

[18] C. Peng, X. Gao, N. Wang N, et al., "Face recognition from multiple stylistic sketches: Scenarios, datasets, and evaluation," *Pattern Recognition*, 2018, 84: 262-272.

[19] Wang X, Tang X. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(11): 1955-1967.

[20] Ding, C, Choi J, Tao D, Davis L S. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Trans Pattern Anal Mach Intell*, 2016, 38:518-531.

[21] Wu X, He R, Sun Z, Tan T. A light CNN for deep face representation with noisy labels. *IEEE Trans Inf Forensics Security*, 2018, 13:2884-2896.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 110-114 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Forecasting Bio-economic Effects in the Milk Production based on the Potential of Animals for Productivity and Viability

Gennadii Georgievich Cherepanov*, Anatolii Ivanovich Mikhalskii, Zhanna Anatolievna Novosrltseva

Department of Incomplete Data Control, V.A. Trapeznikov Institute of Control Sciences RAS, 117997, Russia

ARTICLE INFO

Article history:
Received: 07 May, 2019
Accepted: 05 August, 2019
Online: 03 September, 2019

Keywords:
Dairy Cattle Breeding
Profitability of Milk Production
Cow's Productivity and Viability
Forecasting Algorithms
Planning in Dairy Industry

ABSTRACT

The most important biological factors, mainly determining the economic efficiency of milk production, are the productivity potential and the level of viability of cows. The aim of the work is to predict the bio-economic effects in a heterogeneous population of dairy cows taking into account the decrease in the length of productive life with increase in milk productivity. In the considered model situation, with an increase in the productivity of cows, the total profit per cow per year increases, while the profitability of production may decline. The developed algorithms are supposed to be used for solving the problems of optimizing the management of the dairy cow herd and planning technological development in animal husbandry.

1. Introduction

Used abbreviations and terms: PL is length of productive life (the number of consecutive lactations of 305 days each); PP is the potential of productivity, VP is the viability potential, CR is constitutive resistance, dry period is the period from the end of lactation to calving; calving period is lactation period + dry period.

Due to the reduction in the total size of dairy cattle population, the rate of decline in milk production in the Russian Federation is not compensated by a positive trend of increasing the average productivity of cows. The use of imported breeding material on a large scale and the use of intensive industrial technologies in milk production are accompanied by a decrease in the level of biodiversity, an increase in "productivity diseases", a decline in reproduction, a deterioration in the quality and safety of the products obtained. Similar problems in the field of dairy cattle husbandry have arisen in recent decades in other countries, since the high productivity of cows is often associated with an increased frequency of culling. [1, 2]. Therefore, to ensure profitability and overall sustainability of milk production, it is necessary to intensify research in the field of biology of the viability and productive longevity of cows [2 - 4]. It is necessary to search for biological traits and technical tools for the collection and transmission of information for early prediction of the health status of lactating cows at the earliest possible stages of postnatal life [2, 3].

Previously, it was shown that for early assessment and prediction of viability in groups of dairy cows, the trait of constitutive resistance can be used as an integral indicator reflecting the effects of age-related decline in protective forces [5], as well as the reciprocal of the rate of cows culling in first lactation. An analysis of empirical data reveals the presence of heterogeneity in terms of survival parameters in populations of various animal and human species. The data from bio-medical studies show that the patterns of mortality in a population considered as one whole may differ significantly from the mortality frequencies in the constituent subpopulations [6, 7]. There are data indicating the heterogeneity of dairy cattle populations on survival parameters [8, 9]. In the context of the stated approach, the main factors determining the lifetime productivity of cows are PP and VP. Age dynamics of milk production in groups of cows with different PL is determined by a specific combination of these factors.

From the viewpoint of planning the development of the dairy industry, it is important to have methodological tools for a comprehensive assessment and forecasting the influence of the most critical factors in the selection, reproduction and housing condition of dairy livestock. In this paper, as a starting point in this direction, we propose a pilot model for predicting bio-economic effects when varying a set of factors, including milk

*Corresponding Author: Gennadii G. Cherepanov, Institute of Animal Physiology, Biochemistry and Nutrition, Borovsk, Kaluga region, 249013, Russia, Email: 8961242110@mail.ru

production potential and the rates of disposal of the herd in a heterogeneous population of dairy cows. When setting the initial data, an empirically fact of a decrease in PL with increasing milk production was taken into account, especially pronounced in highly productive cows.

The aim of this work is the systematization of previously identified empirical patterns and the development of a pilot numerical model for predicting the bio-economic effects in a dairy cows population which is non uniform in terms of PP and VP.

2. Determination of The Numerical Values of Parameters using Empirical Data

2.1. Milk Production Potential

To estimate the potential of milk production in the *i*-th homogeneous group, cumulative 305 d milk yield data are used for consecutive lactations, $y_{mi}(t)$, ranging from 4 ($t = 4$) to the last completed lactation t_{imax}. Age dynamics of 305 d milk yield is described by a three-component multiplicative function

$$y_{mi}=A_i \times\exp(-\exp(-bt))\times D^t{}_t \qquad (1)$$

in which A_i is a potential maximum lactation yield (productivity potential, PP), the exp(-exp(-bt)) function describes an age-dependent increase in the ability to form milk with a gradual reaching a plateau level, the function D^t (constitutive resistance, CR, D<1) describes an age-related decrease in the ability to maintain lactation activity in the *i*-th group. The maximum yield is always less than A by an amount depending on the rate of decline of the CR. At small values of D, the peak of milk yield for lactation occurs earlier and the cows drop out of the herd more quickly.

Earlier, when analyzing empirical data on cows of the Kholmogor and Black-and-White breeds, a linear positive relationship was found between PL, expressed by the number of last lactations t_{imax}, and the value of D_i in groups [5]. Thus, the value of D has the meaning of viability potential (VP), estimated from the age dynamics of 305 d yields in groups. The value of the parameter b in the second component of formula (1) varies within small limits (0.4–0.5) and for proximate calculations it was assumed to be 0.45.

2.2. *Potential of Viability*

The presence of a positive relationship between PL and the "initial" value of the CR in the analysis of the age dynamics of milk yield agrees qualitatively with the identified relationship between PL and the viability index estimated from the dynamics of cows leaving the head (culling by sum of reasons). Culling rate is described by the Gompertz function:

$$y_c(t)=\Delta S(t)/(S(t)\times\Delta t)=B\times\exp(ct) \qquad (2)$$

where $S(t)$ is the cohort size at the moment of time t, $\Delta S\ (t)$ is the reduction of the cohort size for the period of time Δt. For lactating cows, time is usually expressed as the number of the current lactation and $\Delta t = 1$. Parameters B and c can be determined not in cohorts (in groups of animals born during the same period of time), but according to the data of age structure of the herd (cross-sectional study).

The mean PL (T, the average number of lactations in a cohort) can be estimated from empirical data using an analytical expression [10]:

$$T=\exp(B/c)\times\mathrm{Ei}(B/c)/c \qquad (3)$$

where Ei(.) is a special function (integral exponent).

For use in zootechnic practice, we can use the truncated Gompertz distribution, which does not take into account a small number of longevity record holders. Our numerical simulation showed that in the range of values B = 0.04–0.3, c = 0.2–0.8, the maximum PL value is approximately 2 times larger than the average PL i.e. $t_{max}\approx 2\ T$ or $T = t_{max}/2$ [8].

The value of the initial rate of disposal y_{c1} in formula (2) means the difference between the number of cows in the first and second lactation divided by their number in the first lactation. When analyzing production data on the age structure of the dairy herd, in particular, on five breeds of the USA and on eight farms of the Leningrad region, the existence of a positive linear dependence of T on the value of $1/y_{c1}$ was revealed [5]. Therefore, the value of $1/y_{c1}$ has the meaning of predictor of longevity, as measured by the dynamics of the disposal of cows from the population, herd or cohort [8].

The function, inverse to the intensity of disposal, is characterized by a decrease in the viability with age ("resistance to mortality" according to Gompertz). Thus, at t = 1, the "initial" disposal intensity $y_{c1} = B\ e^c$ is determined, and the reciprocal, $1/y_{c1} = B^{-1}\ e^{-c}$, has the meaning of the initial viability level.

3. Forecasting Milk Production Efficiency

3.1. Baseline Data for The Forecast

In the model situations under consideration, there are four groups with consecutive values A = 24, 20, 16, 12 thousand kg (the average for the lifetime milk yield is 12.4; 11.0; 9.35; 7.64 thousand kg); T = 2, 3, 4. 5 (t_{max} = 4, 6, 8, 10 lactations); D = 0.95; 0.91; 0.89, 0.88. An additional list of source data may include a large number of indicators. In this case, this list includes the calving period 13 months and the main cost indicators, including feed costs, the price of sold milk and culled cows (delivery at a meat processing plant, utilization of carcasses of sick cows, etc.).

3.2. Intermediate Calculations and Target Forecast

The initial data for each group are calculated:
- the average lifetime cumulative 305 d milk yield over 10 completed lactations (y_m);
- the number of replacements of the culled cow on the heifer for the total number of 10 lactations ($Nr = 10/T$);
- calendar term corresponding to 10 lactations (years);
- revenue from sold milk over 10 lactations with price of 30 rubles per 1 kg of milk, thousand rubles (milk income, $m.inc = 30 \times ym$);
- revenue from sold culled cow over 10 lactations, thousand rubles (realization income, $real.inc = cow\ price \times Nr$);
- total revenue, thousand rubles (*income, inc = inc.m+ real.inc*);

- the average price of grown or purchased heifers, thousand rubles (heifer cost, *Hc*);
- the cost of growing or purchasing heifers, thousand rubles (reproduction expenses, *repr.exp.* = *Hc* × *Nr*);
- the cost of maintenance and technological spending per cow for the whole period, thousand rubles (technological expenses, *t.exp.* = *12* × *year*);
- feed costs for milk production per 10 lactations, thousand rubles (feed expenses, *f.exp* = *0.6* × *inc.m*);
- the sum of expenses, thousand rubles (*s.exp* = *repr.exp* + *t.exp* + *f.exp*);
- profit, thousand rubles (income over expenses, *IOE* = *inc.);*
- profitability (% rentability, *rent*).

The adopted forecasting scheme assumes a certain method of combining in the same algorithm two heterogeneous data, i.e. the age dynamics of productivity and the longevity of cows. In the literature, there is no description of such an algorithm, and there are also no systematic empirical data necessary for detailed analysis. The simplified calculation method used in this work for ten 305 d lactations at one value of calving interval is as follows. Since the average daily milk yield values are used in groups, to obtain a forecast of the calendar period for 10 lactations per cow place, you can use the above ratio $T = t_{max}/2$. For example, at $T =$ 5 it is necessary to have two periods, i.e. 2 cows with 5 lactations and 4 dry periods (100 months of lactation, 24 months of dry period). At the same time, in the case of culling, a new primiparous cow is transferred to this place, one period of dry period is missed, and the total calendar period for which the cost of one cow per year is calculated for this group is 10.3 years. For the next three groups, these values are 10, 9.9, 9.5.

4. Results and discussion

The results of calculating the profitability of milk production for 4 model groups (subpopulations) with different values of the productivity potential, the average PL, the cost of raising or importing young stock and the main technological costs with an interval of 13 months are given in Table 1 and Figure 1 and Figure 2.

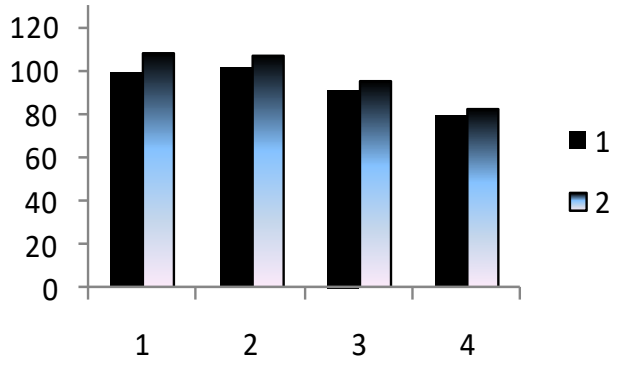


Figure. 1. Total profit per cow per year in the production of milk in four groups of cows with the values of the parameters listed in Table. In groups 1, 2, 3, 4, the average milk yield per life is 12.4; 11.0 9, 35, 7.64 thousand kg.; the calving period is 13 months. The average price of a culled cow sold in option 1 is 10, in option 2 it is 30 thousand rubles.

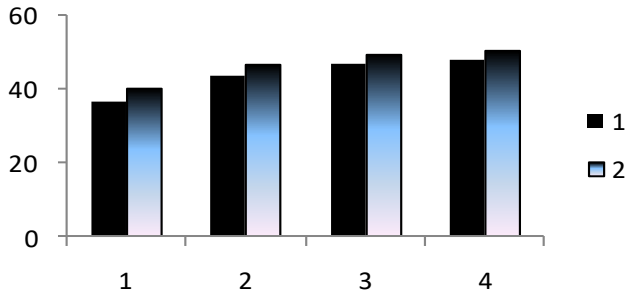


Figure. 2. Profitability of dairy raw materials production (revenue per milk and from sale of culled cows / sum of expenses) in four groups of cows with the values of the parameters listed in Table. In groups 1, 2, 3, 4, the average milk yield per life is 12.4; 11.0; 9.35; 7.64 thousand kg.; the calving period is 13 months. The average price of a culled cow sold in option 1 is 10, in option 2 it is 30 thousand rubles.

In the considered model situation, with an increase in the productivity of cows, the total profit per cow per year increases, while the rentability of production decreases.

To predict bio-economic effects in real conditions (including in terms of long-term planning), it is necessary to take into account the variability of interrelations and a wide range of possible values of parameters. In this work, seven factors were considered, the numerical value of which can vary widely, which is the case for different regions of the Russian Federation. At the same time, four gradations of two factors (level of productivity and viability index) were taken into account with two levels of one additional factor. At this stage of the study, the operability of the formulated concept and the pilot prediction model were tested. With a positive assessment of the results of the first stage, the expansion of the empirical base becomes urgent, i.e. obtaining a sufficient amount of production data while improving the calculation algorithm.

In the extended version of the model, additional indicators may be introduced into the calculations, including the reproduction parameters of the dairy herd. Since, at high levels of milk production, there is a shortage of the young animals number, to preserve the livestock population size the complex methods are developed (separation of bull-sires semen, induction of superovulation, *in vitro* fertilization, transplantation and embryo cloning), which increase the cost of reproduction of the dairy herd. To provide enhanced herds reproduction, the choice of the most appropriate method depends largely on the conditions of the territorial distribution of specific farms, as the initial data in the forecasting can vary greatly for small farms in the regions and for large industrial complexes with intensive technologies located near large urban agglomerations. Therefore, in order to rationalize management decisions, multivariate model calculations are required.

Researchers working in the field of studying the dynamics of human populations have identified the need to identify subpopulations in order to correctly predict demographic processes [11, 12]. In the process of breeding work in animal husbandry, similar problems arise, since an early assessment of the breeding value of parents for productive longevity of offspring is important. For bull-sires, one of the options for such an assessment may be

Table. Baseline data and results of the prediction of the bioeconomic efficiency of milk production in groups that differ in productivity potential (A) and viability (D).

D	*T*	*A*	*ym*	*Nr*	*Hc*	*years*	*repr.*	*t.exp.*	*inc.m*	*s.exp./year*	*inc./year*	*rent.*
0,95	2	24	124	5	80	10,3	400	124	969,2	268	98,9	37
0,91	3	20	110	3,3	65	10	214	120	980,7	231	101,4	44
0,89	4	16	93,5	2,5	50	9,9	125	119	878,2	195	91,2	47
0,88	5	12	76,4	2	35	9,5	70	114	732,8	164	79,2	48

Notes: notations are given in the section 3.2.

the inverse of the rate of disposal in first lactation, measured in a group of daughters [13]. When using the latest genomic, bioinformatics and telecommunication technologies, there are ample opportunities for solving such problems. The use of microelectronic sensors and wireless communication allows to increase the amount of information captured during screening surveys throughout the animal's life, which can dramatically increase the possibilities for monitoring and forecasting multifactorial bio-economic effects.

At this stage, it is important to compare the initial prerequisites of our calculations with known ideas and concepts related to the problems of populations viability. The concept of constitutive resistance as the main determinant of productive animal longevity, used in this work, is in many ways consistent with the concepts of creod according to Waddington [14] and primal health of Odent [15]. There is evidence that metabolic diseases occurring in later life may arise *in utero* as a result of suboptimal intrauterine conditions. The process by which prenatal defects lead to permanent changes in tissue structure and function, and finally to low birthweight, is known as developmental programming; i.e. placental insufficiency invariably affects embryonic development and health in later life [16]. So, the level of viability in heifers is being formed epigenetically in the prenatal and postnatal periods when exposed to internal and external factors. Our data indicate that the value of cow PL depends on the "initial" level of viability indicators (at the beginning of the reproductive period). The existence of this dependence is indicated by empirical data and, as carried out by calculations, a similar relationship follows from the general properties of the Gompertz function, traditionally used in analyzing the survival of populations.

5. Conclusion

The most important biological factors, mainly determining the economic efficiency of milk production in dairy farms, are the productivity potential and the level of cow viability formed to the age of reproductive activity. The length of the productive life of cows largely depends on the initial level of viability indicators. Currently, the average group potential levels of productivity and viability can be determined from indirect information – the data of age dynamics of 305 d milk yields and intensity of cows culling in first lactation.

From the viewpoint of planning the development of the dairy industry, it is important to have methodological tools to predict the influence of the most critical factors in the selection, reproduction and housing condition of dairy livestock. In this paper, we tested the operability of the general approach and the algorithm of numerical simulation using discrete scales and data approximation by the Gompertz function. The results showed that in the considered model situation, with the increase of the average of lifetime milk yield, the total profit per cow per year increases, while the profitability may decline. According to the authors, there is reason to assume that the identified consistent patterns reflect a general trend, at least in a qualitative sense, since empirically established facts and regularities are taken into account in the set of initial data.

After expanding the volume of empirical data, setting parameters and modification of algorithm taking into account the features of real objects, the proposed forecasting model is supposed to be used to optimize the herd management and forecast scientific and technological development in dairy cattle husbandry.

References

[1] K. Adamczyk, J. Makulska, W. Jagusiak and A. Węglarz, "Associations between strain, herd size, age at first calving, culling reason and lifetime performance characteristics in Holstein-Friesian cows" Animal. 11(2): 327-334, 2017. DOI: 10.1017/S1751731116001348.

[2] F. Miglior, A. Fleming, F. Malchiodi, L.F. Brito, P. Martin and C.F. Baes, "A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle" J. Dairy Sci. 100(12): 10251-10271, 2017. DOI: 10.3168/jds.2017-12968.

[3] E. Vasseur, "Animal behavior and well-being symposium: optimizing outcome measures of welfare in dairy cattle assessment" J. Anim. Sci. 95(3):1365-1371, 2017. DOI: 10.2527/jas.2016.0880.

[4] J.H. Britt, R.A. Cushman, C.D. Dechow, H. Dobson, P. Humblot, M.F. Hutjens, G.A. Jones, P.S. Ruegg, I.M. Sheldon and J.S. Stevenson, "Invited review: Learning from the future-A vision for dairy farms and cows in 2067." J. Dairy Sci. 101(5): 3722-3741, 2018. DOI:10.3168/jds.2017-14025

[5] G.G. Cherepanov, "Justification of the concept of the key role of constitutive resistance for the viability and duration of the use of highly productive animals" Problems of Productive Animal Biology, 4: 5-34, 2014 [In Russian]

[6] J.W. Vaupel, J.R. Carey, K. Christensen, T. E. Johnson, A. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y.Zeng, K.G. Manton and J.W. Curtsinger, "Biodemografic trajectories of longevity" Science, 280: 855-860, 1998.

[7] R.V. Ivanov, A.I. Mikhalskii, V.K. Ivanov, S.Yu. Chekin, M.A. Maksyutov and V V. Kashcheev, "On identification of morbidity parameters in a heterogeneous model: The cases of complete and incomplete information" Autom. Remote Control, 78(7): 1329-1340, 2017. DOI: 10.1134/S0005117917070141

[8] G.G. Cherepanov, A.I. Mikhalsky and Zh.A. Novoseltseva, "Estimation of survival parameters for components of a heterogeneous population of productive animals: problem analysis, approximate solution options" Problems of Productive Animal Biology. 4: 83-97, 2017 [In Russian]

[9] G.G. Cherepanov, A.I. Mikhalskii and J.A. Novoseltseva, "Estimation of survival parameters in heterogeneous population" 11th IEEE International Conference on Information and Communication Technologies (AICT). Moscow, Russia, 2017: 433-435. [In Russian]

[10] A.V. Krementsova and N.V. Gorbunova, "The role of environment in the dynamics of the life expectancy distribution". Autom. Remout Control 8: 121-133, 2010.

[11] N. Keyfitz and G. Littman, "Mortality in a heterogenous population" Population Studies, 33: 333-342, 1979.

[12] A.I. Mikhalskii, "Methods of analysis of heterogenic structures in population" Moscow: IPU RAS Publ., 2002. [In Russian].
[13] G.G. Cherepanov, "Prediction of viability of cows: a new look at the old problem" Agricultural Research and Technology. Open Journal (ARTOAJ). 141(5), 2018. DOI: 10.19080 / ARTOAJ.2018.14.555931
[14] C.H. Waddington, "Basic biological conceptions". In: *Na puti k teoreticheskoi biologii. I. Prolegomeny*. Moscow, Mir Publ, P.11-38. 1970. Translated from English: Towards a theoretical biology. I. Prolegomena. IUBS Symposium (Ed. C.H. Waddington), Birmingham: Aldine Publ., 1968.
[15] M. Odent, "Primal Health" London, Century Hutchinson, 1986.
[16] G. Opsomer, M. Van Eetvelde, M. Kamal and A. Van Soom. Epidemiological evidence for metabolic programming in dairy cattle. Reprod. Fertil. 29(1): 52-57, 2016. DOI: 10.1071/RD16410.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 115-122 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Survey on Semantic Similarity Based on Document Clustering

Rowaida Khalil Ibrahim[*,1], Subhi Rafeeq Mohammed Zeebaree[2], Karwan Fahmi Sami Jacksi[1]

[1]*Computer Science Department, University of Zakho, 420011, Zakho, Iraq*

[2]*Department of IT, Akre Technical College, Duhok Polytechnic University, 42004, Akre, Iraq*

ARTICLEINFO

Article history:
Received: 15 June, 2019
Accepted: 16 July, 2019
Online: 03 September, 2019

Keywords:
Clustering
Text Clustering
Semantic Similarity
Ontology
Evaluation Measures

ABSTRACT

Clustering is a branch of data mining which involves grouping similar data in a collection known as cluster. Clustering can be used in many fields, one of the important applications is the intelligent text clustering. Text clustering in traditional algorithms was collecting documents based on keyword matching, this means that the documents were clustered without having any descriptive notions. Hence, non-similar documents were collected in the same cluster. The key solution for this problem is to cluster documents based on semantic similarity, where the documents are clustered based on the meaning and not keywords. In this research, fifty papers which use semantic similarity in different fields have been reviewed, thirteen of them that are using semantic similarity based on document clustering in five recent years have been selected for a deep study. A comprehensive literature review for all the selected papers is stated. A comparison regarding their algorithms, used tools, and evaluation methods is given. Finally, an intensive discussion comparing the works is presented.

1. Introduction

Clustering is a problem of unsupervised learning; the main task is to group a set of objects in a manner that the group of objects in the same collection called the cluster are more similar (in meaning) to each other than those in the other cluster [1][2]. Document clustering or (text grouping) is a block analysis application on text documents, has applications in automatic organization of document, filtering or fast information retrieval and subject extraction [3][4]. The text group divides a set of text documents into different categories so that the documents in the same category group describe the same subject. Text clustering is an essential function in the text mining process [5]. Problem of Description clustering is one of the key issues for the traditional algorithm of document clustering. The objects can cluster in traditional algorithm, but the result of clustering cannot give description concept. It is therefore necessary to find a way to resolve the problem of grouping the document description in such a way to meet the specific needs of document clustering [6]–[8].

For conceptual and meaningful collecting of documents into groups (clusters), it is necessary to exploit the semantic relationships between words and concept in the documents [9]. For this case semantic similarity is needed. The term semantic similarity is a specific measure for several pamphlets/concepts in contrast with the similarity that capable of roughly- calculations with respect to represent grammar (such as the format of the string). For this reason, the mathematical instruments are depended for asset estimation semantic association among language units and ideas or examples, by means of a numerical description obtained by comparing information that supports its meaning or description of its nature [10]–[12].

It is often confused with the term semantic relatedness and semantic similarity. The semantic relationship involves any relationship between two periods, whereas the semantic similarity only includes a relationship “is a” [13]. For example, "car" and “bus" are similar, but is related to “driving" and “road”. By determining topological similarity, Semantic similarity can be estimated using the ontology to determine the distance between concepts/term [14][15].

For semantic similarity purpose between concepts many tools are used, the most common one is the WordNet. The term of WordNet Is the lexicons of the English language database. It’s collecting the words into a group of synonyms named syn-sets, provides general, short info and register a different semantic relationship between these sets of synonyms. Especially WordNet is suitable for similarity procedures, because it regulates nouns and

[*]Rowaida Khalil Ibrahim, Email: rowayda.gravi@gmail.com

verbs in the hierarchy of "is a" relation [16]. Several tools for finding semantic similarity between concepts and words have been proposed in our literature but most of them have been used WordNet. For this paper we reviewed many papers that used semantic similarity but we choose only thirteen systems in these five recent years that using semantic similarity for clustering propose and we give some important information of each of them. Our paper organized as follow: in section 2 it's a literature review of papers, in section 3 we give table of discussion of all that systems we reviewed before, then we conclude all work in section 4.

2. Literature Review

The main framework of the approach addressed by [17] consists of three modules, first one is Page Counts based measures, second Snippets and third is taxonomy-based approach. for page count four co-occurrence standard measures are used namely (WebOverlap, WebJaccard, WebPMI and WebDice), after the page count of two words are calculating then checking the context of those two words locally by using snippets based on probability measure to find out the pattern exist between both words. Also, they have used taxonomy-based measure (LEACOCK and CHODOROW (LCH) [18], WU & PALMER (WUP) [19], and RESNIK [20]) that using WordNet for finding and computing similarity between words (X and Y). then for combination the result using tow class of Support vector Machin (SVM and libSVM). The term of dataset used for evaluation, for training propose using MEN dataset [21] Selecting synonymous word pair at random from dataset and create non-synonymous for each one. However, for testing propose using MC approach [22], and RG [23], WordSimilarity-353 (WS) [24]). And using WEKA [25] for testing and training of the SVM with suitable demonstrating approach LibSVM and the qualified prototypical via determining the division ratio is about 70.51%. The results of accuracy were obtained near by 72%.

As explained in [26], the approach in this paper at first finding similarity between tow documents then extend to between two document sets, after that using clustering and classification to group the documents based on similarity measures .the proposed measure of similarity called (similarity measure for text processing) SMTP and the classification algorithm used are KNN based (SL-KNN) single label classification, KNN based (ML-KNN) multi label classification and for clustering algorithm using K-mean Clustering and (HAC) Hierarchical Agglomerative Clustering. The result checks the SMTP similarity measure's effectiveness that applying in text application (SL-KNN, ML-KNN, K-mean and HAC). Also, Compare SMTP performance with other five metrics, Cosine, Euclidean, IT-Sim, Extended Jaccard (EJ) and Pairwise-adaptive (Pairwise) by using different k value and using evaluating measures like (AC (Accuracy), EN (Entropy)). It shows that the similarity measure usefulness may rely on: 1) the format of feature, e.g., tf-idf or word count; 2) application of domains, e.g., image or text; and 3) classification or clustering algorithm. In addition, the term of datasets which are used for evaluating are three different datasets WebKB download from internet, Reuters-8 [27], and RCV1 [28]. Finally, in each case the result show the proposed measure SMTP have better performance.

In [29], the system approach for clustering the text document is an ensemble approach by using concept from Wikipedia. Two kind of clustering used, First approach is Partitional Clustering (lexical document clustering (LDC)) have three main phases are (Clustering term, finding documents lexical seed and Clustering text documents), in phase1 for the clustering purpose the fuzzy c-means algorithm is used to collecting the columns of matrix the (document-term matrix) hence the Fuzzy c-means collect the document-terms into k term clusters then in phase2 extract the documents representative, that are using as seeds in order for clustering all documents later, finally in phase3 based on those document seeds of each cluster term the centroid of document will be computed. Then the distances among the centroids and documents are used for clustering the documents. Second approach is Ensemble clustering ELSDC (Ensemble Lexical-Semantic Document Clustering) the main phases are (clustering term and topic key term selection, lexical seed documents extraction, find documents seeds, tagged and by using the consensus method collecting the documents). In phases 1 and 2, is same as of (LDC) but in phases 3 and 4, BOC and BOW are used for documents representation. The relevant concepts extract from Wikipedia by wikify module. The Naive Bayes classifier is the component of consensus method, by using documents in the same cluster in both clustering they are trained. The final cluster is generated after the remaining document are classified. However, the feature selection Var-TFIDF method are used for both approaches.

The experimental result has two rounds, in first round find comparison between LDA model and lexical document clustering (LDC) algorithm, the LDA's main idea is that the document can be rate as a distribution probabilistic on the underlying themes seen each subject as the terms of probability distribution. The C++ implementation are used for the LDA model; in this round It has been shown that the LDC can generate results similar to the LDA model on some real text data sets based on the representation of the duration document. In second round of experiments, find comparison between LDC and proposed ensemble algorithm (ELSDC). Show the benefits of incorporating Wikipedia concepts into document clustering. For evaluation proposed using Normalized Mutual Information (NMI) hence It measures the amount of information obtained from the categories given to a group of clusters, and has a maximum value of one clustering process when re-create the layers perfectly and the minimum it is zero. Also, for experiment propose eight datasets are used (20Newsgroups (News-sim3, News-rel3 and 20ng-whole), Reuters-21578, Classic4 is created from SMART data repository, WebKB, SMS Spam Collection and Cade is gathered from the content of Brazilian web pages).

As show in [30], extracting document feature and document vectorization to find similarity between Arabic webpages and showing the semantic annotation. The main steps in proposed approach are, in the first step Extraction of main class features using (Arabic VerbNet). The second one is document vectorization according to semantic feature by two solutions (semantic class probability distribution or Semantic class density) and in third step doing clustering and annotation. the model is work as follow the annotated document is take to be classified and all opinion word will be recognized by using semantic feature extraction and all the word will be aggregate to give semantic annotation to the document. The result gathered tow experiments, at the first one is

showing comparison between clustering only using K-mean (standard) and clustering using K-mean with semantic class density however, k-mean with semantic probability distribution and this is done by using evaluation measures (purity, MICD and DBI) hence the purity show that the clustering with vectorization signifies with high score, MICD tend that proposed model is out perform than standard k-mean and DBI show that proposed model appear to have smaller value than standard is mean that is better performance. The Runtime of standard k-mean is longer than the time consumed for tow solutions k-mean with vectorization. The second experiment showing annotation by using Mean score among cluster with both vectorization solution and Mean ratio among corpus (MRAC) also with two solution of vectorization. However, the datasets used for evaluation propose are collected corpus from the archives of online Arabic newspaper and VerbNet download from internet.

In [31], the approach of semantic clustering the text documents by using lexical chain and WordNet. The proposed approach in this paper are work as follow, at first the polysemy and synonymy are two main problem that effect the representation of text, this problem is solve by using WordNet based on Word sense disambiguation, after that introducing the lexical chain to capture the main theme of text and find the relationship between the word sense, then is show that the method can find a true number of clustering which is useful for finding the number of k that use in K-mean clustering algorithm. Finally, for experimental propose the paper show the comparison between all three (Base, DC, DCS) and with another system like ASG03 [32], LMJ10 [33] and CSF11 [34], hence Base is mean (all nouns) the system use all basic preprocessing techniques, i.e. term set extraction, stop words removing, stemming word and identification of nouns from term set but without performing WSD. However, DC is (disambiguated concepts) corresponding to the Base, but with preforming WSD and finally DCS (disambiguated core semantics) also corresponding to the DC, but adds the core semantics extraction process and is the proposed method.

In addition, in this paper for evaluation propose using Reuters-21578 corpus dataset and the evaluation measure using purity, entropy and F1-measure. Finally, the result show that in all experimental scheme the LMJ10 is the worst and the term of dimensionality reduction the feature account that derived from ASGO3 is lower than the number of base and the reduction of DSC is between 10% and 40% hence that the CSF11 greater than 74%, as well as the cluster quality obtained using the core semantic feature is better than using all nouns and using disambiguated concept (or at least comparable to) however the performance of using the disambiguation concepts is better than using all nouns, This suggests that the proposed disambiguation standard can solve the problem in a highly commendable and volatile manner, improve quality to a certain extent, and that the features of the semantic strings produced by the DCS approach do not only reduce the number of semantic concepts without losing much information, also adequately the main topic of the document that helps the clustering.

As [9], the model of clustering proposed in this paper incorporate Coreference resolution and exploit semantic relationship among the words by tackling polysemy and synonymy problem using WordNet and semantic similarity. The proposed approach consists of five modules are (Coreference Resolution, Preprocessing, Synonymy Identification and Sense disambiguation, Feature Selection and Bisecting k-means) modules. the purpose of those five model are, at first Coreference Resolution is the process of identifying Coreference words that occur in the document, then the propose of preprocessing model to transform the document in more suitable form and this is done in three steps (POS tagging ,Stop word Elimination and stemming),after that Sense Disambiguation and Synonym Identification deal with polysemy and synonymy problem in document and the model use WordNet for this propose, then the feature selection done by weighing the words in document using tf-idf, finally the documents are clustered by using Bisecting K-mean. For experimental propose the paper using four classic Datasets (CACM are 60 Documents, CISI are 44 Documents, CRAN are 44 Documents, and MED consists of 52 Documents) and for evaluating the quality of cluster using purity. The result show comparison between Base and proposed model hence the Base is the model without (the Coreference resolution, sense disambiguation and pruning term), after comparison the term of purity result show that the proposed model is observed that achieve 30% of improvement in clustering purity, that the purity rate of Base configuration is 0.55 hence the proposed model is 0.8.

As [35], Based on style Similarity and structure Clustering Web Pages, the model which used in this paper is DOM (Document Object Model) tree. The approach using two main measures, one of them for structural similarity to find similarity on DOM using TED (Tree Edit Distance) for HTML pages and another one using (Jaccard similarity measure) for CSS which called Stylistic similarity on style sheet information, after that based on those similarity measures using (Near neighbor clustering technique) for clustering propose, however to combine this information using Jarvis and Patrick method (shared nearest neighbors) [36]. It is a respectable feature aimed at aggregating Internet pages built for pairs similarity metrics. For evaluating propose using different threshold value of near neighbor at first using 90% edge for resemblance toward luxury near neighbors' pamphlets and 90% threshold for communal close towards combine clustering then altered the threshold to 95% until found the optimal threshold. The result show that when the threshold was 90% total number cluster was 12, and when it was 95% the total was 24 cluster as well as when it was 85 the total number of clusters was only 5, so after few trails they found that for test dataset the optimal threshold for both resemblances and intersection for communal close parties was near by 85%. As it declares that the result is positive as early evaluation.

As [37], the approach based on Hierarchical agglomerative clustering method (HAC) that is based of conceptual annotation of documents. The key steps of approach are: (1) pairwise calculation of the initial similarity matrix (2) Explanation of newly cluster and updating the matrix of similarity (3) post processing of HAC tree result. The approach chooses to rely on Lin Measure for pairwise similarity and on Base Match Average (BMA) for GroupWise semantic similarity, the proposed method uses GroupWise semantic similarity to compute the pairwise similarities between document for creating the label-base similarities matrix. When two clusters are agglomerative a new cluster are creating then automatically compute the annotation for this new cluster and then the label-based matrix are iteratively update, finally using

postprocessing of the tree hierarchy to take the branch lengths advantage in the tree to only keep the most meaningful cluster. For evaluation propose using (HSC heavy semantic clustering and LSC light semantic clustering) approaches for comparison with baseline (classical HAC with naive cluster annotation) by using Normalized Robinson-Foulds Distance measure hence the lower the distance the higher resemblance. The result after comparison between approaches (HSC, LSC and baseline) the HSC and LSC clearly outperform the baseline but for runtime the baseline was faster. then it shows the comparison of approaches (LSC, Baseline) with postprocessing and without post processing (pp and nopp) show that both are bad without postprocessing, finally it evaluating cluster labeling between our clustering labeling (CL), naive merge annotation (baseline) and expert also, the result show CL is better. And the dataset which used for evaluating propose was derived from WordNet-based disambiguation version of the aforementioned bookmark annotation [38] however, all resources and result of this paper are available at [39].

As [40], the approach for semantic document clustering based on the graph similarity. Graph is mainly used information from WordNet to the degree of semantic similarity between 150,000 of the most common in English terminology. The proposed approach at first Adds all the documents in to the similarity graph, then the distance between a pair of documents is measured by evaluating the paths between them, where a path in the graph can go through several terms that are semantically similar. The idea is to creating a node for each document and connecting this node to the graph, then to find the semantic similarity between the two documents will be measured by calculating the distance between the two nodes, and using three choices for the distance metric: the cosine, linear or logarithmic. Finally, for clustering propose using K-mean algorithm to cluster a set of documents and the algorithm relies on a way for computing the distance between two documents. The term of experimental result, the dataset of documents from Reuters-21578 benchmark and using F-measure for evaluating. The end of evaluating after applying the K-mean algorithm in three different distance metrics (cosine, linear and logarithmic) at first the result show that Using the similarity graph can lead to both higher precision and recall. And with Noting that using the linear or logarithmic similarity metric did not make a difference. The reason is that the two metrics apply different monotonic functions on the average of the sum of the forward and backward paths. Applying these monotonic functions has no effect on the ordering of the distances between nodes and on the clustering result.

In [41] paper the approach named (An Ontology-based and Domain Specific Clustering Methodology for Financial Documents). The main steps in approach are (preprocessing and feature extraction, sense word disambiguation, representation document and clustering). The approach works as follow: at step1 the preprocessing and data extraction is responsible to extract text from document in order to transform word in to more meaningful form and this is done in three steps (Stop Word Removal, Noun Extraction and Lemmatization). In step2 word sense disambiguation the correct sense for the noun are identified and this is done by using (WordNet ontology/database and an information content file) and in this study two disambiguation techniques are used which used different external information to remove ambiguity, used one technique (Brown Content File) which is the default for measuring Resnik and other technical used (file content financial information) that is proposed in this study. In step3 the document representation the features are represented as term frequency (TF or tf) vector to be prepare for clustering. Then, in step4 clustering is the final step which clusters the document vector by using algorithm and in this approach two kind of algorithms are studies (K-mean and sequential Information Bottleneck).

The dataset which used in this study was downloaded from EMMA [42] online repository using 446 documents, EMMA documents are classified under the US jurisdiction and their sectors or purposes and have achieved three types of labeling settings according to the categories provided by EMMA (setting1: purpose at the same time called sector as the class label, setting2: class label as state and setting3: class label as mixing of both state and sector). Then for evaluating propose in this study using purity for external evaluation to evaluate the cluster performance. The result showing the purity rate of both algorithm K-mean and sIB based on three settings also showing comparison as follow: 1) comparison between all term and noun only, Although the names are selected, the number of features is reduced significantly, the purity values of both algorithms do not have a considerable difference; 2) comparison between tf-idf of nouns and without tf-idf, it shows that with tf-idf better purity; 3) comparison between no disambiguation and with disambiguation, both methods of disambiguation have yielded good results, but the finest result produced with one the used financial information content file. Moreover, based on experimental results, the sIB algorithm can be determined as the most appropriate algorithm for aggregation.

As explained in [43], The system called WMDC (Wikipedia matching document classification) several steps are used in this approach. First, select knowledge and concept from Wikipedia. Second, using heuristic selection to pick up related concept. Third, finding similarity between document using combination of (Semantic similarity based on Wikipedia machine and Textual similarity based on keywords matching), and choosing K-mean algorithm for classification in this approach because of its efficiency and accuracy. Also, the evaluation experiment in this paper are divided in to two part. First part, focus on effectiveness of heuristic selection Rules, hence three rules are used to pick up the related concept of word (Rule1: all title, Rule2: all keyword and Rule3: any keyword) for evaluating efficiency of those three rules using selectivity measure and evaluating quality by using relevance. Second part, evaluating the effectiveness of approach.

Using Purity for measuring the accuracy of document classification and for evaluating the efficiency of approach using (vector construction time). The datasets used in this paper are Wikipedia dataset which download directly from internet and published in 2011-10-7, and document dataset from Reuters-21578 divided in to 82 cluster by removing those cluster with documents less than 15 and more than 200 only 30 cluster reminding to enhance the experimental effect. The result in this paper show that after using 4 different thresholds 0,0.5,0.10 and 0.15 for evaluating all rules of heuristic selection given that the smaller rule of selectivity is better efficiency and higher rule of relevance is better quality. then as the overall given that the proposed approach can accurately find out the related concept for a given document. Then the same threshold used for purity and (vector construction time) for evaluating classification accuracy and efficiency it concluded

that the proposed approach can improve the efficiency of language and for the corresponding Wikipedia under the precondition of not compromising the accuracy classification of the document.

As in [44] the system called SEMHYBODC related to (Fuzzy C-Manners). The methodology of clustering works as follow: At first the documents are annotated by "KIM" plugin before clustering. Then the documents are clustered using FPSOFCM (FPSO+FCM) and concept weight is calculated. The term of weight is recalculated through the steps specified in SEMHYBODC algorithm. And the Accuracy of clustering is computed using measures such as cluster purity and compared with various other hybrid approaches. after evaluating by using F-measure and purity the result show that, using swarm intelligence for clustering is not suitable for large dataset because the computational time is more but it give better accuracy in order to solve this problem a hybrid approach combine with algorithms such as PSO and FCM, however the PSO is combine with K-mean it's also leads to optimal solution but the time for computational is still higher, then as finally result the FPSO+FCM it show improvement over another algorithms like K-mean, Fuzzy C-means and Hybrid, It is combined with the ability to search from the globalized and fast algorithm PSO convergence algorithm FCM. It is used as a result of the initial seed FPSO algorithm FCM, which is applied to the purifying with generate concluding outcome. For evaluating propose webpages are collected from internet.

As explain in [45] many methods of clusters are operate based on similarity between documents as well as we explained before in our literature but in this paper the semantic similarity used for clustering the articles. The approach which proposed in this paper are at the first step the semantics of articles based on participating the entities in these articles, build three vectors of representations for each article: One calculates the average vector for all entities, one also with all entities but excluding citations and the other with only citation entities. Once the article vectors are generated, the next step is to identify clusters of articles based on vectorization using (k-mean) Clustering and using the Louvain (Network-based clustering methods) for community detection. For experiments the both clustering methods are using the Astro dataset [46]. At first four solutions of clustering are collected, namely CWTSC5, STS-RG, UMSI0 and ECOOM-BC13 for the pseudo-ground-truth based on adjustF1 that the best K are choosing which gives a highest score of adjustF1 also using adjustedF1 score to evaluate the results of clustering as well from the Louvain method.

The result shows the quality scores of three cluster based on pseudo-ground-truth (no citation, only citation, all entities) and OCLC Louvain then gives the average Adjusted Mutual Information scores (AMI) between current solution and the all four other solutions named, STS-RG, UMSI0, CWTS-C5 and ECOOM-BC13. It indicates that, if citations are used only, the resulting combinations correspond to other solutions of clustering than those if not using citation whose adjusted F1 score is also the lowest. Not surprisingly, the other clustering solutions depend heavily on citation information. So, even if they use different methods of citations, the citation information is still bringing enough agreement between them. Use all entities to represent articles has highest score adjustedF1 and agree with most others. Also, find AMI results among these three solutions and groups based on the Louvain method. Again, the cluster based on citation only agree with the results of Louvain almost as much as using cluster that use all entities. In accordance with these measures, they decided to use all entities as a major choice of characteristics, and to maintain 31 sets as key results for K-Means, which is named OCLC-31.

3. Discussion

The table below show the survey of approaches that using semantic similarity based on clustering. The table give a number of papers each paper has specific method and different tools are used for each of this method as well as provide information of each paper given different options like Measures for similarity and evaluation propose and algorithms for clustering or classification also, provide datasets that used by each paper.

As show in the survey, most of the method using WordNet, the term of WordNet as we mentioned before in introduction is the lexicons of the English language database, as well as in information system the WordNet was used for a number of proposals, for example (information retrieval, word sense disambiguation, text classification automatically, summarization of text automatically, machine translation and also can used for crossword puzzle generation automatically) [47] but one of the WordNet most common application is using to determine the similarity between words as we seen in all that method which are using WordNet the main propose was used for (word sense disambiguation) that deal with polysemy and synonymy problem. However, rather than WordNet there are another tool are used like Wikipedia or Kim Plugging also, there are method based on clustering, similarity graph or even network. Also, most of people using K-mean algorithm for clustering hence it's a must popular clustering algorithm that grouping the documents using a similarity metric that is based on keywords matching, and its useful for huge variables for this reason most of time computationally the k-mean is faster than hierarchical clustering but if keeping (K) small, and another advantage is it can make a tighter cluster [48]. However, many of research using bisecting K-mean because it can deal with large dataset also its very satisfactory quality of clustering with low cost. And another algorithm used as shown in surveys like fuzzy c-mean, HAC, Naïve base classifier, near neighbor, sIB, Louvian method and using PSO with clustering algorithm for optimization propose.

As it can be seen that other measures are used like tf-idf its mean (term frequency-inverse document frequency) it's using as weighting factor the main job is determine how much important the word is to the document in corpus or in group [49], as we seen in survey most of paper used this measure for feature selection. Also, other measures are used like Var-TFIDF, heuristic selection etc. for evaluation and experimental result many types of dataset are used and there are specific measures that using for evaluation purpose most popular one is Purity hence its very simple and primary evaluation measure that using for validation to determine the quality of cluster [50], however there are another one like F-measure, Entropy etc. the main propose are to determine the performance of cluster.

Table 1: Summery of Approaches Using Semantic Similarity Based On Clustering

No.	papers	Date	Method	classification and clustering Algorithm	Evaluation measures	Other Measures	Dataset
1.	L. Sahni, et al. ,[17]	2014	Semantic similarity between words using WordNet	SVM	Training the model by determining a split ratio of about 70.51%.	using 4 co- occurrence measures, probability measure and measures-based approach (LCH, WUP and Resnik)	using MEN dataset for training and 3 Benchmark datasets for testing propose
2.	Y. S. Lin et al. ,[26]	2014	Text classification and clustering	SL-KNN, ML-KNN, k-means and HAC	Accuracy and Entropy in (tf-idf, word count, Z-score)	SMTP proposed measure	Three data sets namely RCV1, WebKB and Reuters-8
3.	S. Nourashrafeddin et al. ,[29]	2014	Document clustering using Wikipedia	fuzzy c-means and Naive Bayes classifier	Normalized Mutual Information (NMI)	Var-TFIDF	using 8 datasets 1-20Newsgroups (News-sim3, News-rel3 and20ng-whole). 2-Reuters-21578. 3- from data repository SMART creating Classic4. 4-WebKaB. 5-SMS Spam Collection 6- Cade from Brazilian web pages content is gathered
4.	H. M. Alghamdi et al ,[30]	2014	Arabic VerbNet	K-Mean	Purity, Mean intra cluster distance (MICD) and Davis Bouldin Index (DBI)	using Mean score among cluster, Mean ratio among corpus (MRAC) for Annotation	Corpus collected from online Arabic newspaper archive
5.	T. Wei et al. ,[31]	2014	Text clustering Using WordNet version 2.0	Bisecting k-mean	Purity, Entropy and F1-measure	Using Lexical chains	Reuters-21578 corpus
6.	S. S. Desai and J. A. Laxminarayana ,[9]	2016	Document clustering Using WordNet	Bisecting K-mean	Purity	Tf-idf for feature selection	using 4 classic Datasets 1-60 Documents from CACM, 2- 44 Documents from CISI, 3-44 Documents from CRAN, 4- 52 Documents from MED
7.	T. Gowda and C. Mattmann ,[35]	2016	Webpage clustering using (DOM)tree	Near Neighbor Clustering technique	using different value of threshold	TED and JS	Dataset from a popular weapons classifieds site
8.	N. Fiorini et al. ,[37]	2016	Semantic clustering using WordNet	HAC	Normalized Robinson-Foulds Distance	lin Measure, BMA, post processing	Derived from WordNet-based disambiguation version
9.	L. Stanchev ,[40]	2016	Document clustering Based on similarity Graph	K-mean	F-measure	cosine, linear, and logarithmic	Reuters-21578 benchmark
10.	C. Kulathunga and D. D. Karunara,[41]	2017	Document Clustering Using WordNet and information content file	K-mean and sIB	Purity	Tf-idf	EMMA
11.	Z. Wu et al. ,[43]	2017	Document clustering Using Wikipedia	K-mean	Selectivity, relevance, purity and vector construction time	Heuristic selection	Wikipedia Dataset published from internet and document dataset from Reuters-21578
12.	J. Avanija et al. ,[44]	2017	Document Clustering using KIM plugging tool	Fuzzy C-mean with PSO and K-mean with PSO	Purity and F-Measure	Tf-idf	Web pages collect from internet
13.	S. Wang and R. Koopman ,[45]	2017	Article clustering based on Network	K-means and Louvain method	F1-measure and adjustedF1	vectoring	Astro dataset

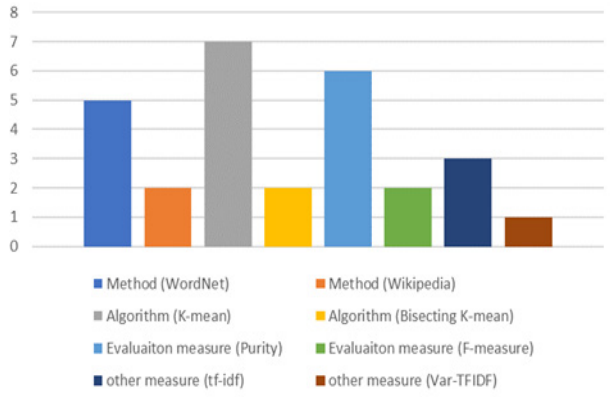


Figure 1: Showing the usage number of main features (Methods, Algorithms and measures).

4. Conclusion

In this paper, after we reviewed all that papers, we conclude that each paper has specific approaches and various tools and measures are available for each approach, and as it can be seen that most popular steps are used in approaches were preprocessing to transform the document in better format and different steps are used in preprocessing like removing stop words, stemming, tokenization etc., word sense disambiguation are used for solving the synonymy and polysemy problems and feature selection another important step that many of approaches was using tf-idf for this case and there was another like Var-TFIDF, heuristic selection etc. however one of the most popular tools used is WordNet the English dataset for meaningful clustering, then clustering done by using clustering algorithms also the most usage one was K-mean algorithm because of its simplicity to use and can give tighter cluster as well as there is another like bisecting K-mean, fuzzy c-mean, hierarchical agglomerative clustering etc. finally, as we see that the main goal of all approaches trying to a chive a better efficiency, accuracy and quality of clustering. after using the experimental measure for evaluating propose like (Purity, F-measure, Entropy etc.) in all case show that the semantic clustering giving a better performance.

References

[1] B. Everitt and Wiley InterScience (Online service), Cluster Analysis. Chichester, West Sussex, U.K: Wiley, 2011.
[2] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and S. R. M. Zeebaree, "Combination of k-means clustering with genetic algorithm: A review," Int. J. Appl. Eng. Res., vol. 12, no. 24, pp. 14238–14245, 2017.
[3] W. L. Chang, K. M. Tay, and C. P. Lim, "A New Evolving Tree-Based Model with Local Re-learning for Document Clustering and Visualization," Neural Process. Lett., vol. 46, no. 2, pp. 379–409, 2017.
[4] K. Jacksi, N. Dimililer, and S. R. M. Zeebaree, "a Survey of Exploratory Search Systems Based on Lod Resources," Turkey. Univ. Utara Malaysia, no. 112, pp. 501–509, 2015.
[5] H. H. Tar and T. Nyunt, "Enhancing Traditional Text Documents Clustering based on Ontology," Int. J. Comput. Appl., vol. 33, no. 10, pp. 38–42, 2011.
[6] Z. Chengzhi and X. Hongjiao, "Clustering description extraction based on statistical machine learning," Proc. - 2008 2nd Int. Symp. Intell. Inf. Technol. Appl. IITA 2008, vol. 2, pp. 22–26, 2008.
[7] K. Jacksi, S. R., and N. Dimililer, "LOD Explorer: Presenting the Web of Data," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 45–51, 2018.
[8] K. Jacksi, N. Dimililer, and S. R., "State of the Art Exploration Systems for Linked Data: A Review," Int. J. Adv. Comput. Sci. Appl., 2016.
[9] S. S. Desai and J. A. Laxminarayana, "WordNet and Semantic similarity-based approach for document clustering," 2016 Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2016, pp. 312–317, 2016.
[10] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, "Semantic Similarity from Natural Language and Ontology Analysis," Synth. Lect. Hum. Lang. Technol., vol. 8, no. 1, pp. 1–254, May 2015.
[11] Y. Feng, E. Bagheri, F. Ensan, and J. Jovanovic, "The state of the art in semantic relatedness: a framework for comparison," Knowl. Eng. Rev., pp. 1–30, Mar. 2017.
[12] F. M. Couto and A. Lamurias, "Semantic Similarity Definition," in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 870–876.
[13] A. Ballatore, M. Bertolotto, and D. C. Wilson, "An evaluative baseline for geo-semantic relatedness and similarity," Geoinformatica, vol. 18:4, pp. 747–767, 2014.
[14] "Semantic similarity." [Online]. Available: https://en.wikipedia.org/wiki/Semantic_similarity. [Accessed: 27-Nov-2018].
[15] A. A.-Z. K. J. A. S. Subhi R. M. Zeebaree, "Designing an Ontology of E-learning system for Duhok Polytechnic University Using Protégé OWL Tool," J. Adv. Res. Dyn. Control Syst., vol. Volume 11, no. 05-Special Issue, pp. 24–37, 2019.
[16] M. A. Hadj Taieb, M. Ben Aouicha, and A. Ben Hamadou, "Ontology-based approach for measuring semantic similarity," Eng. Appl. Artif. Intell., vol. 36, pp. 238–261, 2014.
[17] L. Sahni, A. Sehgal, S. Kochar, F. Ahmad, and T. Ahmad, "A Novel Approach to Find Semantic Similarity Measure between Words," Proc. - 2014 2nd Int. Symp. Comput. Bus. Intell. ISCBI 2014, pp. 89–92, 2015.
[18] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification," WordNet An Electron. Lex. database., no. January 1998, pp. 265–283, 1998.
[19] W. Zhibiao and M. Palmer, "VERB SEMANTICS AND LEXICAL Zhibiao W u," Proc. 32nd Annu. Meet. Assoc. Comput. Linguist., pp. 133–138, 1994.
[20] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy."
[21] K. Sparck Jones and P. Willett, Readings in information retrieval. Morgan Kaufman, 1997.
[22] G. A. Miller and W. G. Charles, "Contextual correlates of semantic similarity," Lang. Cogn. Process., vol. 6, no. 1, pp. 1–28, Jan. 1991.
[23] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," Commun. ACM, vol. 8, no. 10, pp. 627–633, Oct. 1965.
[24] L. Finkelstein et al., "Placing search in context: the concept revisited," ACM Trans. Inf. Syst., vol. 20, no. 1, pp. 116–131, 2002.
[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update."
[26] Y. S. Lin, J. Y. Jiang, and S. J. Lee, "A similarity measure for text classification and clustering," IEEE Trans. Knowl. Data Eng., vol. 26, no. 7, pp. 1575–1590, 2014.
[27] "Datasets for single-label text categorization - Ana Cardoso Cachopo's Homepage." [Online]. Available: http://ana.cachopo.org/datasets-for-single-label-text-categorization. [Accessed: 05-Dec-2018].
[28] D. D. Lewis, Y. M. Yang, T. G. Rose, and F. Li, "RCV1: A new benchmark collection for text categorization research," J. Mach. Learn. Res., vol. 5, pp. 361–397, 2004.
[29] S. Nourashrafeddin, E. Milios, and D. V. Arnold, "An ensemble approach for text document clustering using Wikipedia concepts," Proc. 2014 ACM Symp. Doc. Eng. - DocEng '14, pp. 107–116, 2014.
[30] H. M. Alghamdi, A. Selamat, and N. S. Abdul Karim, "Arabic web pages clustering and annotation using semantic class features," J. King Saud Univ. - Comput. Inf. Sci., vol. 26, no. 4, pp. 388–397, 2014.
[31] T. Wei, Y. Lu, H. Chang, Q. Zhou, and X. Bao, "A semantic approach for text clustering using WordNet and lexical chains," Expert Syst. Appl., vol. 42, no. 4, pp. 2264–2275, 2015.
[32] H. Andreas, S. Staab, and G. Stumme, "Wordnet improves Text Document Clustering," 2003.
[33] L. Jing, M. K. Ng, and J. Z. Huang, "Knowledge-based vector space model for text clustering," Knowl. Inf. Syst., vol. 25, no. 1, pp. 35–55, Oct. 2010.
[34] S. Fodeh, B. Punch, and P.-N. Tan, "On ontology-driven document clustering using core semantic features," Knowl. Inf. Syst., vol. 28, no. 2, pp. 395–421, Aug. 2011.
[35] T. Gowda and C. Mattmann, "Clusteringweb pages based on structure and style similarity," Proc. - 2016 IEEE 17th Int. Conf. Inf. Reuse Integr. IRI 2016, pp. 175–180, 2016.
[36] R. A. Jarvis and E. A. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors," IEEE Trans. Comput., vol. C–22, no. 11, pp. 1025–1034, Nov. 1973.
[37] N. Fiorini, S. Harispe, S. Ranwez, J. Montmain, and V. Ranwez, "Fast and reliable inference of semantic clusters," Knowledge-Based Syst., vol. 111, pp. 133–143, 2016.
[38] R. Bernardi and I. AT4DL 2009 (2009: Trento, Advanced language technologies for digital libraries: international workshops on NLP4DL 2009,

Viareggio, Italy, June 15, 2009 and AT4DL 2009, Trento, Italy, September 8, 2009. Springer, 2011.
[39] "SC - Semantic clustering." [Online]. Available: http://sc.nicolasfiorini.info/. [Accessed: 13-Dec-2018].
[40] L. Stanchev, "Semantic Document Clustering Using a Similarity Graph," Proc. - 2016 IEEE 10th Int. Conf. Semant. Comput. ICSC 2016, pp. 1–8, 2016.
[41] C. Kulathunga and D. D. Karunaratne, "An -ontology-based and domain specific clustering methodology for financial documents," 17th Int. Conf. Adv. ICT Emerg. Reg. ICTer 2017 - Proc., vol. 2018-Janua, pp. 209–216, 2018.
[42] "Municipal Securities Rulemaking Board:EMMA." [Online]. Available: https://emma.msrb.org/. [Accessed: 13-Dec-2018].
[43] Z. Wu et al., "An efficient Wikipedia semantic matching approach to text document classification," Inf. Sci. (Ny)., vol. 393, pp. 15–28, 2017.
[44] J. Avanija, G. Sunitha, and K. R. Madhavi, "Semantic Similarity based Web Document Clustering Using Hybrid Swarm Intelligence and FuzzyC-Means," vol. 7, no. 5, pp. 2007–2012, 2017.
[45] S. Wang and R. Koopman, "Clustering articles based on semantic similarity," Scientometrics, vol. 111, no. 2, pp. 1017–1031, 2017.
[46] J. Gläser, W. Glänzel, and A. Scharnhorst, "Same data—different results? Towards a comparative approach to the identification of thematic structures in science," Scientometrics, vol. 111, no. 2, pp. 981–998, May 2017.
[47] "WordNet Application."
[48] Santini, "Advantages & Disadvantages of k- - - Means and Hierarchical clustering (Unsupervised Learning)," pp. 1–5, 2016.
[49] A. Rajaraman and J. D. Ullman, "Data Mining," in Mining of Massive Datasets, Cambridge: Cambridge University Press, 2011, pp. 1–17.
[50] S. C. Sripada, "Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C Means Clustering," Indian J. Comput. Sci. Eng., vol. 2, no. 3, pp. 343–346, 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 121-127 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Aggrandized Random Forest to Detect the Credit Card Frauds

Jisha. Mulanjur Vadakara*, Dhanasekaran Vimal Kumar

Department of Computer Science, Nehru Arts and Science College, Coimbatore, 641105, Tamilnadu, India

ARTICLE INFO

Article history:
Received: 05 April, 2019
Accepted: 29 June, 2019
Online: 17 July, 2019

Keywords:
Aggrandized random forest
Bagging
Boosting
Credit card
Rose
Random forest
Smote

ABSTRACT

From the collection of supervised machine learning technique, an ensemble procedure is used in Random Forest. In the arena of Data mining, there is an excellent claim for machine learning techniques. Random Forest has tremendous latent of becoming a widespread technique for forthcoming classifiers as its performance has been found analogous with ensemble techniques bagging and boosting. In the present work we have proposed an algorithm, Aggrandized Random Forest to detect fraud from credit card transactions/ATM transactions with high accuracy considering both balanced and imbalanced dataset, comparatively to the defined classification algorithm Random Forest in Data mining.

1. Introduction

The statistical analysis, machine learning, artificial intelligence, database techniques and pattern recognition concepts have magnified the origin of Data mining. Traditional techniques may be inapt due to enormity of data, high dimensionality of data, heterogeneous and dispersed nature of data. In this era, data mining is prevalent in the field of commerce, trade, science, architecture, education and medication fair to remark a few. The area of data mining applications we come across are the mining of genome sequencing, market exchange, clinical experiments and in the transactions of credit card. With the immense of devices gifted of accumulating data, Big data is now used more widely with the collection of data attractively and cheaply.

Frauds existing in tax return, claims of insurances, usage of cell phones, credit card transactions etc. represent substantial problems for corporate and the government but spotting and foiling fraud is not so modest task. Frauds are claimed to be an adaptive crime, so it desires distinct means of intellectual data mining algorithms which are coming raised in the field of investigation to sense and prevent fraud. These are methods that exist in the areas of machine learning, statistical analysis and knowledge discovery databases. These methods in different zones of fraud crimes do offer appropriate and fruitful solutions.

There are mainly two procedures used for fraud detection statistical performances and artificial intelligence. Fraud supervision is a knowledge-intensive activity. The main AI

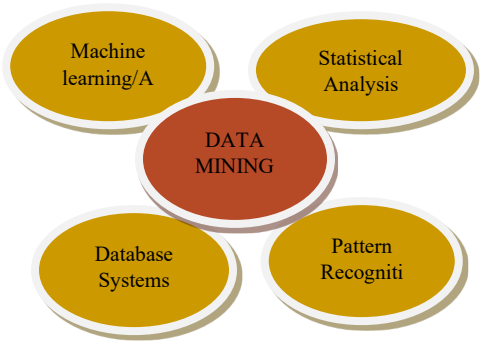


Figure 1. Applications of Data Mining.

Techniques used in the analysis of detecting fraud management include:

- Data mining is the process where facts are classified, clustered and segmented and connotations and

*Jisha.M. V, Nehru Arts and Science College, Thirumalayampalayam, Tamilnadu -641105. Contact No-9745222933 & jisharudhra@gmail.com

instructions in the agreed data are mechanically found that may indicate stimulating patterns, together with those correlated to fraud.

- Expert systems will help to scramble know-how to perceive scam in the way of rules.
- Pattern recognition sense estimated patterns, classes or clusters of mistrustful deeds either routinely or to match assumed inputs.
- Machine learning methods to mechanically find features of frauds.
- Neural networks learn doubtful patterns from samples and used advanced to detect them.

Other techniques such as Bayesian networks, Link analysis, Decision trees and sequence matching are also used for fraud detection.

In field of data mining and knowledge discovery, to handle the problems it is mainly classified into two types supervised learning and un-supervised learning respectively. Supervised learning is further sub-classified depending on the properties of the response variable into classification and regression methods. If the response variable is categorical and discrete, it is classification problems. Otherwise, if the response variable is continuous, it is regression problems. There are different Classification methods for detecting fraud from large data transactions. Selecting Random Forest, Support Vector Machine, Logistic regression and K-Nearest Neighbors [1], our previous work of comparative analysis of these algorithms detected Random Forest to be the best having high accuracy to detect fraud from credit card transactions/ATM card transactions [1].

In this paper, we have given an introduction to Random Forest in section II, description of smote sampling model in section III, about ROSE function in section IV, Theoretical description in section V, proposed algorithm in section VI, followed by Conclusion and References.

2. Introduction to Random Forest

Random Forest can be used to resolve regression as well as classification problems. The dependent attribute or variable is continuous in regression and in classification problems it is categorical. Random Forest is a very influential ensembling machine learning algorithm in which decision tree works as the base classifiers, forming multiple decision trees and then merging the output generated by each of them. Decision tree is a classification model in which it classifies each node with maximum information gain. This search is continued until all the nodes are beat or there is no more information to gain. Decision trees are very modest and informal to know models; however, their predictive power is less, thus called as weak learners [1].

Random Forest works as an ensemble of random trees. It syndicates the yield of multiple random trees and then to originate up with its own output. Random Forest as similar to Decision Tress [1] but will not select all the data points and variables in each of the trees. It arbitrarily assembles data points and variables in each of the tree that it produces and then it is united to arrive at the output at the end. It takes away the prejudice that was encountered in a decision tree model. Also, it improves the predictive power significantly [2].

Random forest stays a tree-based algorithm which take in building several trees (decision trees), were the manipulator generates the number of trees. The elective of all created trees fixes final classifying outcome.

3. Smote Sampling Method

Stereotypically, there are several classification problems that have been anticipated to organize formerly disregarded observations, which will be situated, or recognized set to be tested by application of a classifier is then trained by earlier prearranged observations called training dataset. Numerous typical classifiers are anticipated to deal with this type of snags such as tree-based classifiers, discriminant analysis and logistic regression.

Unbalanced data cause problems in many of the learning algorithms. This is due to irregular quantity of cases that is posed by each class of the problem. Synthetic Minority Over-Sampling Technique algorithm (SMOTE) is used for handling these unbalanced classification problems.

The Smote function can be called as defined, when it discourses the unbalanced problems with the creation of a new smote data set. Alternatively, it returns the ensuing model by routing a classification algorithm on the new data set [3].

Smote (form, data, perc. over = 200, k = 5, perc. under = 200, learner = NULL, ...)

Definition of the arguments-

Form designates the prediction problem. *Data* holds the actual unbalanced dataset. *Perc. Over* is method of over-sampling that finds a figure that initiatives the verdict of how many further cases from the minority class are generated. *Perc. Under* is an under-sampling where among each case made from the minority class, it specifies an amount that initiate the choice of the number of further cases from the majority classes are designated [4,5]. *K* is a quantity signifying the number of adjoining neighbors that is further used to produce the new instances of the minority class. *Learner* is set to Null by default. It acts as a string with a tag of a function that outfits a classification procedure that will be useful in the ensuing smote data set.

The two important constraints that switch the quantity of over-sampling of the minority class and under-sampling of majority class are perc. Over and perc. Under respectively. Perc. Over will approximately a numeral above 100, thereby perc. Over/100 innovative samples of this class are considered for the minority class. If perc. Over value is below 100 then for an arbitrarily selected proportion of the cases, a single case is created and fit into the minority class on the original dataset. The perc. Under constraint panels part of the cases of majority class that erratically selected as the final balanced dataset. With respect to the count of newly engendered minority class cases, the result is obtained [3], [5].

Designed for each existing minority class, ‘n’ new-fangled instances will be formed which is measured by the parameter perc.

Over. The instances will be generated by using the parameter K that holds the count of neighbours [3], [6].

We arrive directly to the classification model from the resulting balanced dataset by using the above-mentioned function [5]. Overall, this function will either return Null value or a new dataset depending on the usage of smote function. If not, the learner parameter will return the classification model [5].

4. Rose: A Package for Binary Imbalanced Dataset

Imbalanced learning signifies the problematic of supervised classification if a class behave erratically over the sample dataset. The Rose, Random Over Sampling Example [7] tackles these problems with its on functions embedded in the package. The class imbalance circumstances are persistent in the multiplicity of applications and fields; this issue had received considerable attention recently [8]. There originate the scarcity of software and procedures clearly meant to handle imbalanced data so to have the non-expert users adopt it. In the R environment, there exist Caret package to validate as well as select classification and regression complications. It highlights the issue of imbalance class as down-sample and up-sample [6][8]. It is worth indicating DMwR (Torgo,2010) which gives an estimation of any classifier, thereby handles imbalanced learning.

The *ROSE* package was motivated to enhance the performance of an imbalanced setting of binary classification providing both standard and more refined tools [7]. Menardi and Torelli at 2014 had developed the smoothed bootstrap-based method. By aiding the phases of model estimation and assessment ROSE [7] relieved the gravity of the effects of an unbalanced distribution of classes.

5. Theoretical Background

5.1. Ensemble Classifiers

An ensemble contains a set of independently trained classifiers whose predictions are combined for categorizing new instances [2].

Bagging [9] and Boosting (Robert E Schapire,2003) are two popular methods for producing ensembles. Bootstrap gathering samples stands for the process called Bagging. As portion of ensemble, if "n" individual classifiers are to be produced from a given original dataset of size "m" then n distinct training set of size m is engendered using sampling. In bagging multiple classifiers created are not dependent on each other. In boosting, from the training dataset each sample is allotted weights. The classifiers generated by boosting is dependent on each other.

5.2 Definition of Random Forest

Random Forest classifier comprise group of tree-structured classifiers $t(x, \Theta_n)$ n=1,2…. , where the Θ_n are self-regulating identical scattered random vectors and every tree casts a unit vote for the most popular class at input x, [2][10].

Once the forest is trained or constructed, each tree give rise to new instance which is recorded as vectors and is joined. The new instance is taken by considering the class having the maximum votes. Nearly one-third of original instances are left, after sampling with replacement of every tree the bootstrap sample set is drawn. The new instances obtained is the out of bag (OOB) data. Out-of-bag error estimation is calculated considering every individual tree's OOB dataset in the forest.

Illustrating Accuracy of Random Forest:

The Generalization error (Pe^*) of Random Forest is Use "(1)",

$$Pe^*=P_{x,y}(mg(X, Y))<0 \quad (1)$$

where mg (X, Y) is margin function. The margin function measures the level to which the average count of votes at the point (x, y) for the exact class surpasses the average votes of any further class. Here x is the predictor vector and y is the classification.

The Margin function is Use "(2)",

$$mg\ (X, Y) =\alpha v_k I(h_k(X)=Y)\ \max j_{=y}\alpha v_k I(h_k(X)=j) \quad (2)$$

Here, I is the indicator function.

Boundary is directly proportional to confidence in the classification.

Benefit of Random Forest is specified in terms of the predictable value of Margin function Use"(3)".

$$B=E_{X,Y}\ (mg\ (X, Y)) \quad (3)$$

A function of mean correlation amongst base classifiers and average strength results in the generalization error of ensemble classifier (Prenger. R, et al,2010). If the mean value of correlation is low an upper bound for generalization error is Use"(4)",

$$Pe^* \leq \rho\ (1- B^2) / B^2 \quad (4)$$

6. Aggrandized Random Forest.

For this work, we have used large dataset of about three lacs of credit card transactions. The classification algorithm is built in the R platform. The proposed algorithm is named Aggrandized Random Forest, which have an increase in its accuracy by 3.06 % for balanced data and an increase of 3.30% for imbalanced data, compared to the previous work as shown in Table.2.

Aggrandized Random Forest is proposed by considering the random forest as the base classifiers with bagging approach and compared with the existing random forest. It is analyzed that Bagged Random Forest gives better results.

Here, we have used the ROSE package of R, to train the imbalanced data set. The functions in ROSE [7,8] have improved the capacity of generalization and reduced the risk conceded by other oversampling methods. As will be expounded, the ROSE technique can be truly considered, which reduces the repetition of instances. In accuracy evaluation step, the overall accuracy may yield misleading results, thus the aggrandized random forest has been evaluated in rapports of class -independent extents such as precision, recall, F-measures etc., as mentioned in the Table.1.

The required library functions in R,

```
library(caret)
library(caTools)
```

library (ROSE)
library (random Forest)

6.1. The Usage of ROSE Function:

ROSE (formula, data, N, subset = options("subset") $subset, seed)

Defining the arguments:

Formula represents the object of class formula where the class label representing the route is allotted with a sequence of routes thru the predictors. The interaction among predictors or their transformations is mentioned by a message. *Data* argument when not specified the variables are taken from environment. It is an optional argument. *N* indicates the anticipated trial proportions of the ensuing dataset. By default, the length of the response variable in formula is considered. *Seed* argument is assigned a single integer value to indicate seeds and preserve the trace of the produced sample [7].

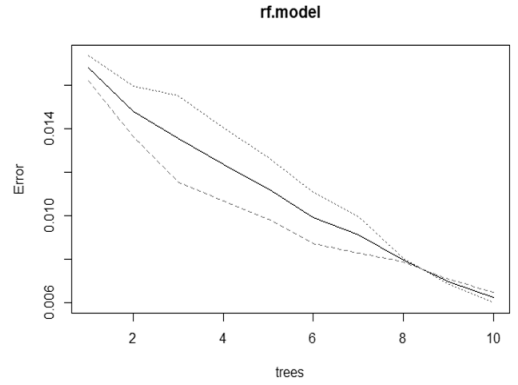


Figure 2(a). Represents the rf. model on Balanced Data

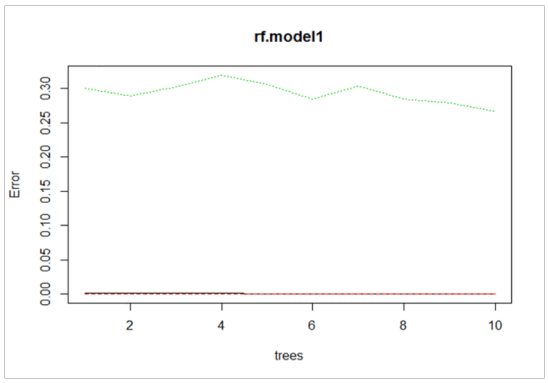


Figure 2(b). Represents the rf. Model on unbalanced data.

6.2. Coding for Imbalanced to Balanced Dataset

###imbalanced classification

creditcard$Class<- as. factor(creditcard$Class)
table (is.na (credit card)) ##-is.na-indicates missing elements.
creditcard$Time
str (credit card)
###
split<-sample. split (creditcard$Class, Split Ratio = 0.7)
train<-subset (credit card, split==T)
test<-subset (credit card, split==F)

###removing unbalanced classification problem and making data balanced
train. rose <- ROSE (Class ~., data = train, seed = 1) $data
table (train. rose$Class)
test.rose <- ROSE(Class ~ ., data = test, seed = 1)$data
table(test.rose$Class)
####Applying random forest on Balanced data
rf.model<-randomForest(Class~., data = train.rose ,ntree=10)
rf.model "As in Figure.2(a)".
####Applying random forest on unbalanced data####
rf.model1<-randomForest(Class~., data = train ,ntree=10)
rf.predict1<-predict(rf.model1,newdata = test)
plot(rf.model1) as plotted "As in Fig.2(b)".

6.3. Confusion Matrix.

After the performance of a classifier, we can engender confusion matrix grounded on the prediction results where each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class.

The table consists of two rows and columns that gives the count of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The minority class is the positive class and the majority class is the negative class "As in Figure.3".

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positives	False Negatives
Actual Negative	False Positives	True Negatives

Figure 3. Confusion matrix representation.

6.3.1 The Confusion Matrix for The Proposed Algorithm.

Confusion matrix to evaluate OOB error rate "As shown in Figure.4".

Actual class	**Predicted class**		**Class error**
	0	1	
0	97821	648	0.006580751
1	639	98186	0.006465975

Figure 4. OOB estimate of error rate: 0.65%.

6.3.2. Confusion matrix for balanced dataset and imbalanced dataset

The "Figure 5 (a)" represents the confusion matrix for the balanced dataset, and "Figure 5 (b)" represents the confusion matrix for imbalanced dataset, where we could get true positive value, false positive, false negative and true negative value to calculate the parameters precision, recall, f-measure, sensitivity and specificity. Thereafter, the analytical measures are calculated accordingly.

Actual class	Predicted class	
	0	1
0	42468	42
1	201	42732

Figure 5(a). For balanced data.

Actual class	Predicted class	
	0	1
0	85284	31
1	11	117

Figure 5(b). For unbalanced data.

6.4. Analytical Measures Used for Comparing the Proposed and Existing.

There are several factors which is used to the detect the best performance of the algorithm. The presentation of the proposed system is evaluated using the analytical measures such as precision, recall, f-measure etc. [1].

- Precision = TP/(TP+FP)
- Recall = TP/TP+FN)
- F-Measure = 2 *(precision * recall) / (precision + recall)
- Sensitivity= TP/FP
- Specificity=TN/FN

Along with the above-mentioned factors, the other evaluated factors are Accuracy, Kappa, prevalence, detection rate, detection prevalence, P -value, Positive Pred. value, Negative Pred. value, Mcnemar's Test P- Value, AUC etc.

Precision measured by the count of true positives divided by the sum of true positives and false positives.

Recall measures the figure of the true positives divided by the sum of true positives and false negatives.

The *F-Measure* indicates the stability between precision and recall values.

Kappa statistics characterize the range to which the data is collected correctly.

Sensitivity refers to the comparison of the count of acceptably recognized as fraud to the amount incorrectly listed as fraud. It is the ratio of true positives to false positives.

Area Under the Curve (AUC) is used to analyze the performance of classification models capable to predict the classes accurately [11,12].

Detection rate is mainly reflected in confusion matrix. This parameter will vary according to the dataset.
Detection rate=TP/(TP+FP+FN+TN).

Specificity indicates the same concept with genuine transactions, or the assessment of true negatives to false negatives.

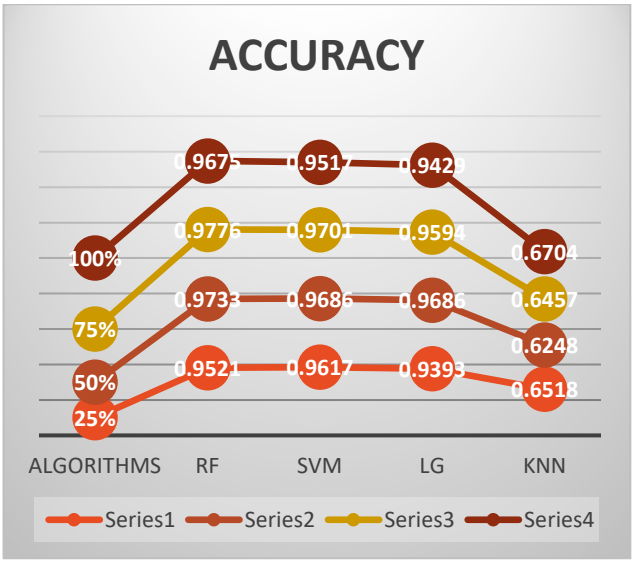


Figure 6. Accuracy of the comparative study of previous work for detecting 100% fraud in the dataset [1].

The previous work was a comparative study of different classification algorithms such as Random Forest [12-15], SVM, K-nearest neighbor and Logistic Regression [1] to detect fraud from credit card transaction. It was found that the Random Forest Algorithm [15] is best in detecting the fraud with an accuracy of 0.9675 "As shown in Figure.6".

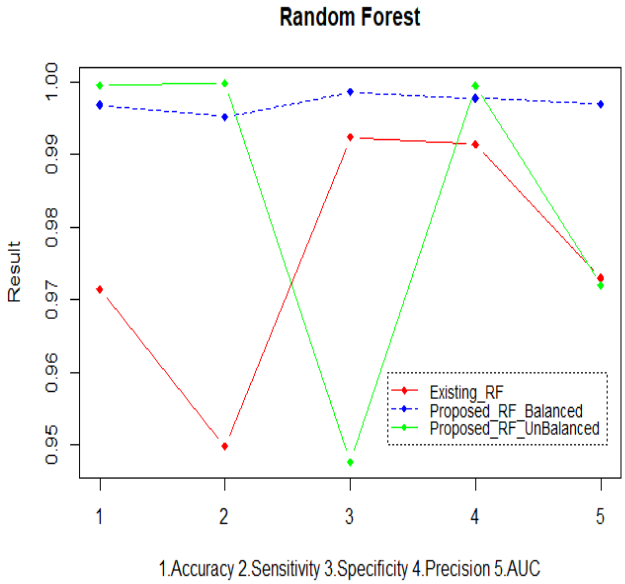


Figure 7. Performance of the proposed Random Forest

Table 1. Comparison of Analytical Measures for the Aggrandized Random Forest Results with Those of the Existing Random Forest.

Factors	Existing Random Forest	For Balanced Dataset	Imbalanced Dataset
Accuracy	0.9675	0.9972	0.9995
95% CI	0.9562, 0.9766	0.9968, 0.9975	0.9993, 0.9996
No Information Rate	0.5182	0.5006	0.9983
P-Value	< 2.2e-16	< 2.2e-16	< 2e-16
Kappa	0.9348	0.9943	0.8476
Mcnemar's Test P-Value	5.806e-07	< 2.2e-16	0.00337
Sensitivity	0.9391	0.9953	0.9999
Specificity	0.9939	0.9990	0.7905
Pos. Predc Value	0.9930	0.9990	0.9996
Neg. Predc Value	0.9461	0.9953	0.9141
Prevalence	0.4818	0.4994	0.9983
Detection Rate	0.4525	0.4970	0.9981
Detection Prevalence	0.4556	0.4975	0.9985
Balanced Accuracy	0.9665	0.9972	0.8952
'Positive' Class	Yes	0	0
AUC	0.970	0.997	0.957
Precision	0.9915	0.9978	0.9995
F-measure	0.9645	0.9965	0.9996

The comparison of the Aggrandized Random Forest with the existing Random Forest "As shown in Figure.7". The analytical measures considered to detect the performance of the new proposed random forest; Aggrandized Random Forest "As shown in the Table 1".

The Analytical measures for the Proposed and the existing is represented by chart diagram "As in Figure.8". The ROC curve which is the receiver operating characteristics of the proposed Random Forest "As shown in Figure 9(a) and Figure 9(b)" for the balanced and imbalanced dataset respectively.

Table 2. Performance of Aggrandized RF.

Dataset	Accuracy	Sensitivity	Specificity	AUC	Precision	Kappa	F-Measure
Existing RF	0.9675	0.9391	0.9939	0.970	0.9915	0.9348	0.9645
Proposed RF Balanced	0.9972	0.9953	0.9990	0.997	0.9978	0.9943	0.9965
Improvement (%)	3.06	5.98	0.51	2.78	0.63	6.36	3.31
Proposed RF Imbalanced	0.9995	0.9999	0.7905	0.957	0.9995	0.8476	0.9996
Improvement (%)	3.30	6.47	-20.46	-1.34	0.80	-9.32	3.63

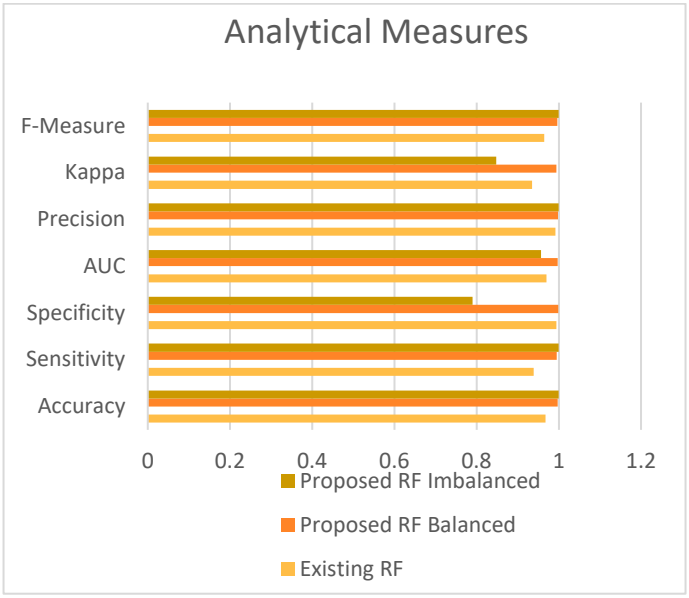


Figure 8. Chart Diagram signifying the comparative recital of Aggrandized Random Forest with Balanced and Imbalanced Dataset to the existing Random Forest Algorithm.

The overall performance of the Aggrandized Random Forest with balanced and imbalanced dataset is evaluated with the existing random Forest, and found to have more accuracy, sensitivity, F-measure, precision, etc. The Table 2 gives the improvement of the proposed algorithm.

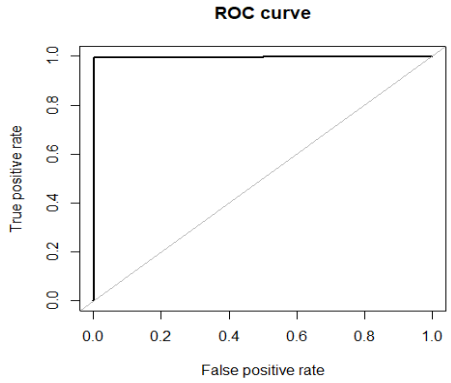


Figure 9(a). ROC of balance dataset.

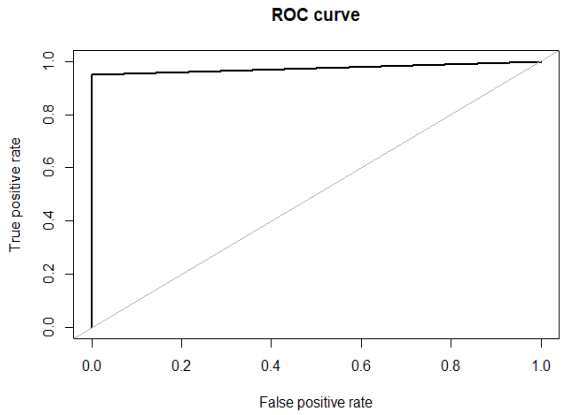


Figure 9(b). ROC of imbalanced dataset.

7. Conclusion

In this paper, our proposed algorithm Aggrandized Random Forest is developed in R platform, having a better accuracy of 0.9972 for balanced and 0.9995 for imbalanced data. The results show that for an imbalanced data, Random Forest outstrips other classification techniques and henceforth stays excessive scope for developing improved Random Forest. Using random forest as a base learner can achieve good outcome in the domain of Imbalanced data.

Acknowledgement

We author thank all the contributors of this journal for considering the article. I would like to thank my guide for giving his support and encouragement in my work. Also, thank the authors of the references giving me the privilege to cite their article and enhance my knowledge. With the responsibility as Ph.D. scholar at Nehru arts and science college, I thank all the teachers and my friends in giving their valuable ideas and support. Wish this article will be beneficial for future scholars.

References

[1] M.V. Jisha., D. Vimal Kumar, "An Efficient Credit Card Fraud Classifier of the four data mining classification algorithms- "A Comparative Analysis."(JETIR)Journal of Emerging Technologies and Innovative Research. Nov.2018, vol 5, issue 11, Pno- 371-378. https://jetir1811656.pdf.

[2] Bernard S Heutte L, Adam S, "Dynamic Random forests, Pattern Recognition Letters", 33 (2012),1580-1586.

[3] Tianxiang Gao, "Hybrid classification approach of Smote and instance selection for imbalanced Datasets". Iowa University, Iowa. Thesis work, 2015.

[4] Vrushali Y, Kulkarni, Pradeep K Sinha. Pruning of Random Forest Classifiers: A survey and future directions."2012 international Conference on Data Science & Engineering (ICDSE), 2012.

[5] R Documentation[online]Http://www.rdocumentation.org.

[6] M.Kuhn. caret: Classification and Regression Training,2014. URL http://CRAN.R-Project.org/package=caret.

[7] N. Lunardon, G. Monardo, and N.Torelli, R package ROSE: Random Over Sampling Examples(version0.0-3). Universität di Trieste and Universität di Padova, Italia ,2013.URL http://CRAN.Rproject.org/web/packages/ROSE/inedx.html.[p79][online].

[8] ROSE:A package for Binary Imbalanced/Balancing data, Https://journal.r-project.org/archieve/2014-008/RJ-2014-008.pdf.

[9] Breiman L, Bagging Predictors, Technical Report No 421, (1994).

[10] Breiman L, Random Forests, Machine Learning, 45,5-32, (2001).

[11] D.Vimal Kumar, M.V. Jisha, "Analytical Measures for Detecting Fraud Using Classification Algorithms". International Journal of Innovative Technology and Creative Engineering (Ijitce), February 2019, vol 9, No.2 636-644. https://ia601504.us.archive.org/19/items/IJITCEFEB19/IJITCE_vol9no202.pdf

[12] Bernard s. Heutte L, Adam S Using Random Forest for Handwritten Digit Recognition, International Conference on Document Analysis and Recognition 1043-1047, (2007).

[13] Bernard S Heutte L Adam S, Towards a Better Understanding of Random Forests Through the Study of Strength and Correlation, ICIC Proceedings of the Intelligent Computing 5th International Conference on Emerging Intelligent Computing Technology and Applications, (2009).

[14] Bernard S Heutte L Adam S, Forests-RK; A new Random Forests Induction Method, Proceedings of 4th International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications – with Aspects of Artificial Intelligence, Springer – Ver lag, (2008).

[15] Bernard S Heutte L Adam S, On the Selection of Decision Trees in Random Forests, Proceedings of International Joint Conference on Neural Networks, atlanta, Georgia, USA, June 14-19,302-307, (2009).

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 127-133 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Automatic Stitching of Medical Images Using Feature Based Approach

Kyi Pyar Win[1,*], Yuttana Kitjaidure[1], Kazuhiko Hamamoto [2]

[1]*King Mongkut's Institute of Technology Ladkrabang, Electronic Department, Faculty of Engineering, 10520, Thailand*

[2]*Tokai University, School of Information and Telecommunication Engineering, 143-0001, Tokyo, Japan*

A R T I C L E I N F O

Article history:
Received: 06 February, 2019
Accepted: 07 March, 2019
Online: 19 March, 2019

Keywords:
Biomedical images
Feature based approach
Image stitching
ORB features
Panorama

A B S T R A C T

Image stitching is a process of creating a panoramic image by combining multiple images that have overlapping regions of the same scene. It is a challenging topic in image processing, multimedia, and medical applications. The proposed system can be applied in medical applications for scoliosis operations and other long limb operations. The problem of the traditional x-ray machine is the narrow view. So, it can't produce the large view of body x-ray image in a single frame. Therefore, this problem is solved by combining two or more x-ray images into a panorama one. This paper proposes a system which automatically stitches the x-ray images and produced a panorama x-ray image. The proposed stitching method is based on feature based approach, ORB (Oriented FAST and Rotated BRIEF). In feature detection stage, Oriented FAST approach is used. In feature description stage, Rotated BRIEF approach is applied. The two important criterias for determining the stitching performance are stitched image quality and processing time. Therefore, the aim of my proposed system is to produce a panorama x-ray image with high resolution and low processing time based on feature extraction approach. We compared our proposed method with three different features detectors. SIFT, SURF, and Harris corner detectors were tested and measured the rate of correct features detection and computation time. Finally, we measured the quality of result images that produced by stitching system of different feature based methods. According to the experimental results, ORB approach can produce high quality panorama image with least processing time.

1. Introduction

Image mosaicing or stitching is creating a panorama image by stitching or mosaicing many images that have overlapping points of the same view. A panorama is a large view of a scenery. It can be obtained by merging many pictures that have overlapping parts of the same scene. Image stitching is very popular in many applications such as construction of large view map, stitching arial images or satellite images and mosaicing radiography images in medical applications. In medical applications, it is needed to view the whole body x-ray images. The large view of x-ray images helps the surgeons to diagnose a disease for example long leg alignment operation, scoliosis patients etc. The traditional x-ray machine cannot produce large view x-ray image. Therefore, our proposed system can be applied in medical applications to view a panorama x-ray image.

* Corresponding Author: Kyi Pyar Win, 1 Soi, Chalong Krung, Ladkrabang, Bangkok, Thailand, +66-6-17430344, kyipyarwin5@gmail.com, 60601169@kmitl.ac.th

Image stitching methods can be divided into two groups. They are direct method and feature based. Direct based method minimizes pixel to pixel mismatching, pixel based approach [1]. Feature based methods find distinctive features and then matched each other between two input images. The robust detectors can extract important characteristics (for example robust in translation, scale, noise and rotation). There has been great progress in medical images stitching system using features based approach. They are SIFT (Scale Invariant Feature Transform) [2], SURF (Speedup Robust Feature detector) [3], HARRIS detector [4] [5], Principal Component Analysis SIFT (PCA-SIFT) [6], and FAST (Features from Accelerated Segment Test). A good feature detector can be choose depending on the nature of the problem. Two main stages in feature based methods are the registration stage and the stitching stage.

The aim of this paper is to develop high quality image stitching system with low processing time. The proposed system contains five main stages. They are preprocessing, feature detection and

description, feature matching, Homography estimation and images stitching. This paper is an extension of work originally presented in 2018 International Conference on Intelligent Informatics and Biomedical Sciences [7].

The rest of this paper comprise of five parts. Part 2 presents the background of research work. Part 3 contains the briefly reviews of three features detection approaches. In Part 4, the proposed image stitching methodology is discussed. In Part 5, the results of proposed system are explained. Finally, Part 6 discusses the conclusion and the future works.

2. Background

Recently there has been done many researches works in image stitching using feature based approaches. Many researchers tried to develop the feature based methods that can detect reliable and repeatable features than traditional methods. But there still a great challenge in image stitching. There are many features detectors that were used in images stitching such as SIFT, SURF, Harris etc. Harris corners fail in image scaling and SIFT is the longest processing time [8].

According to the literatures, there are mainly two methods for image stitching. They are direct based approach and feature based method. Direct methods calculate the camera parameters by minimizing an error function based on the intensity difference of the overlapped region. Transformation matrix is calculated based on intensity difference. Direct based methods give accurate registration, but they fail in image scaling, illumination change and noise. Compared to feature based method, it requires a good initial estimation to achieve transformation matrix which is a major weak point of this method.

Feature based method detects different features, for example edges, points, corners, lines, or other shapes and calculates the relationship between these features. Compared to direct method, it is strong to light change, scale, affine transformation, noise, and orientation of the image. It takes care of locality around detected features to describe that feature as feature descriptor. In feature based method, all main features in an image pair are compared with that of every feature in other image by using the descriptors [9].

There are many feature detector methods exist that are used in medical images stitching system. Feature based approach is suitable for any movement of scene happened in image because of its robustness. This method is very faster, and it has the capability to identify a panorama image by detecting the neighbor correlation between two images. But, feature based techniques rely on accurate recognition of image features.

By using direct based approach, there had been done researched for medical images stitching. Among them, cross correlation was mostly applied because it can be developed using the FFT (Fast Fourier Transform) for stitching two images of the same size. But the weak point is high in costs [10].

There has been great progress in image stitching using feature based approaches. SIFT, SURF, and Harris detectors were mostly used [11]. Gong was presented an image stitching method. In feature extraction, SURF detector was used. The weak point is that it can't detect correct features in image with noise [4] [5]. Singla presented images stitching using x-ray images by combining SIFT and SURF [12]. RANSAC (Random Sample Consensus) algorithm was used to select the correct features points but high in computation complexity. Adel et al. [13] compared many feature based detectors. They showed that ORB was the fastest in processing time compared with other detectors.

Many researchers have been proposed the medical images stitching using feature based approach, but ORB method has not been applied yet. Therefore, this research work is the first work of biomedical images stitching using ORB feature based approach. The proposed system used ORB method in feature detection and description. The correct features points were chosen by using RANSAC algorithm. Homography matrix was estimated by applying the four feature pairs. As a final, the system mosaiced the two x-ray images together using Homography matrix. In blending stage, weight average method was used to eliminate the seam. The proposed system can be used in medical applications to help for surgeons to diagnose the diseases and to save time.

3. Feature Detection and Description

3.1. SIFT

SIFT (Scale Invariant Feature Transform) is an algorithm for local feature extraction and descriptor representation. The SIFT features are robust in image scale and rotation [3]. In SIFT algorithm, there are three major stages. These stages are key point or interest points detection, orientation assignment, and key point descriptor. Key point detection used difference of Gaussian function (DOG) to detect feature points which are invariant to scale and rotation. In orientation assignment stage, one or more orientations are assigned to each key point. In key point descriptor stage, a vector descriptor is developed for each key point.

3.2. SURF

Bay et al. developed the SURF algorithm [1]. SIFT and SURF algorithms process a little different way in detecting features. SIFT creates an image pyramids and then filters each layer with Gaussians by increasing sigma values and taking the difference. SURF makes a stack without down sampling. By using integral images, SURF filters the stack using a box filter approximation of second order Gaussian partial derivatives. Integral images allow the computation of rectangular box filters in near constant time [6].

3.3. Harris

Harris and Stephens presented a corner detector, which is called Harris detector [14]. This detector is widely used to detect feature points and corners. To find the corners in the input image, Harris method takes look at the average intensity which is directional. The intensity change in the small specific area called window around an interested point. The point where the average intensity strongly high changes as compared to the previous one direction is called as a corner point.

4. Methodology

The proposed methodology of x-ray images stitching is shown in Figure 1. There are five main stages in our proposed system. They are preprocessing, feature extraction and description, feature matching, Homography estimation and images blending. Our

proposed system accepted input x-ray images that have overlapped regions and recognized overlapped areas of the images by extracting the ORB features of those images and stitched them automatically to generate the final large view image. In our proposed system, there are two main components, namely overlapped area searching and stitching the images using the searched overlapped area. By using ORB features, the system can find the overlapped areas of the images efficiently. Moreover, the system can stitch the images to produce the seamless high resolution image with low processing time.

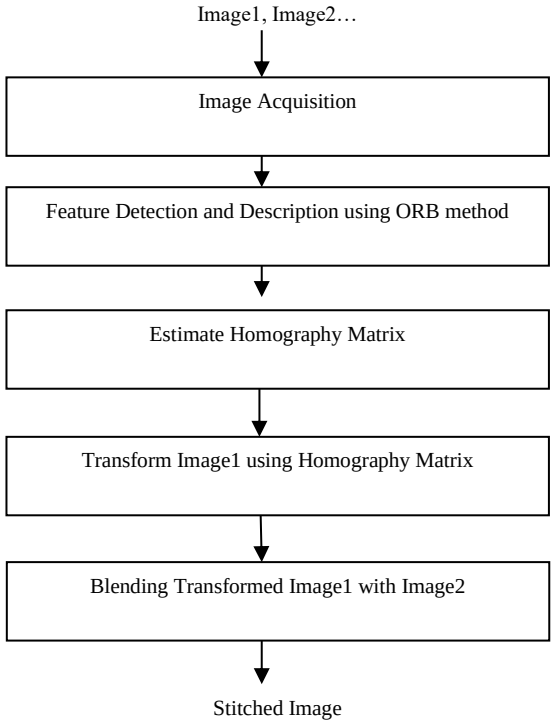


Figure 1. Proposed Images Stitching System Using ORB Feature Based Approach.

4.1. Image Preprocessing

The preprocessing step is an essential step in image processing. First, the original DICOM format is converted into JPEG format to reduce the image size. We used 100 x-ray images with 50 pairs of overlapped images as the input images. It is required about 15-30% of overlapped regions to be stitched. The samples of the input x-ray images are illustrated in Figure 2.

4.2. Features Extraction and Description

Our proposed images stitching system used ORB feature based approach. ORB approach is a combination of Oriented FAST detector and Rotated BRIEF descriptor. Oriented FAST is used to detect feature points and Rotation BRIEF is used to generate descriptors. The speed of FAST and BRIEF are very fast, so ORB has advantages in speed. The good point of this algorithm is fast and having rotational invariance and reducing sensitivity to noise. Figure 3 and Figure 4 show the extracted ORB features of input x-ray images.

4.2.1. Oriented FAST Detector

Both FAST algorithm and BRIEF algorithm do not have scale invariance. FAST detector can detect the features that are translation and rotation invariance, insensitive to noise and high reliability. But the feature points do not have information about direction. Therefore, ORB algorithm uses the Oriented FAST algorithm to detect the feature points which has direction information. Using intensity centroid method, it can calculate the direction of the corner and produce features with direction information.

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y)$$

From Equation 2, we can know the center of gravity or centroid.

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \quad (2)$$

The orientation of the vector from corner to center of gravity can be calculated by using Equation 3.

$$\theta = \arctan\ 2(m_{01}, m_{10}) \quad (3)$$

4.2.2. Rotated BRIEF Descriptor

BRIEF produces binary descriptors by binary coding method. The binary descriptor is simple and storage space is smaller than SIFT and SURF.

$$\tau(\mathrm{p};\mathrm{x},\mathrm{y}) = \begin{cases} 1 \text{ if } & \mathrm{p(x)} < \mathrm{p(y)} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Where p(x) is the pixel intensity at that point. The feature descriptor is defined as follows:

$$f_n(P) = \sum_{1 \le i \le n} 2^{i-1} \tau(P; Xi, Yi) \quad (5)$$

n BRIEF pairs (a1, b1), ……., (an, bn) can be written as a matrix S as follows:

$$S = \begin{bmatrix} a_1 & \cdots & a_n \\ b_1 & \cdots & b_n \end{bmatrix}$$

A rotated version of $\boldsymbol{S_\theta}$ can be computed by using Equation 6.

$$\boldsymbol{S_\theta = R_\theta S} \quad (6)$$

Where, S_θ is the test position for the binary pixel after rotation. θ is the rotation angle, R_θ is the rotation matrix of the main direction.

4.3. Features Matching

After detecting the ORB features, the next stage is features matching. In this step, hamming distance is applied to find the matched features. The results of features matching stage are shown in Figure 5 and Figure 6.

4.4. Compute Homography Matrix

After features matching, the next stage is Homography estimation. The RANSAC (Random Sample Consensus)

algorithm is used to eliminate the wrong matched features and to compute the Homography matrix. The matching relationship of the images are found by using transformed Homography matrix.

4.4 RANSAC (Random Sample Consensus) Algorithm

RANSAC algorithm has three stages. In the first step, it chooses four random feature pairs. Second, it produces the Homography matrix for those selected feature pairs. In the third stage, it calculates an error measure pairs using Homog (8) matrix.

4.4.1. Homography Estimation from Point Correspondences

For two images that have a correspondence between 4 points, homography matrix can be expressed as below:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \cong \begin{bmatrix} a & b & c \\ d & e & f \\ h & i & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

It is required 2 linear equations and 8 unknowns for each correspondence. Therefore, it is needed 8 linear equations for 4 correspondences. Two linear equations can be calculated by Equation 7 and 8.

$$ax_i + by_i + c - x_i'(hx_i + ky_i + 1) = 0 \quad (7)$$

$$dx_i + ey_i + f - y_i'(hx_i + ky_i + 1) = 0$$

In these equations, x_i' and y_i' are coordinates of changed stitched image.



(1) (2)

(3) (4)

Figure 2. The example input images that used in our proposed system.

4.5. Images Blending

The final stage of images stitching is images blending. The aim of this step is to adjust the pixels intensity and to eliminate the visible seams. In this stage, the weighted average method is applied. Figure 7 and Figure 8 show the result images of our proposed system.



Figure 3. The result of features extraction from two knee x-ray images.



Figure 4. The result of features extraction from two leg x-ray images.



Figure 5. The result of features matching stage from the two knee x-ray images.



Figure 6. The result of features matching from the two leg x-ray images.



Figure 7. Final knee stitched image.

5. Results

The proposed system is the automatic images stitching system using medical x-ray images. Our proposed images stitching system is based on feature based approach. In feature detection and description, ORB method was used. The good point of ORB approach is rotation invariant, robust in noise and speed in processing time. We compared our proposed method with three different detectors, SIFT, SURF and Harris. According to the experiments, we can prove that our proposed method is the best in processing time and performance accuracy compared with three features detector methods. We examined the accuracy of our proposed method using 100 x-ray images. Our input samples consist of spine, lower limb and upper limb x-ray images. As the inputs images, our proposed system needs two overlapped x-ray images to be stitched. It is needed about 15-30% of overlapping regions to be stitched. The sample dataset was received from Department of Medicine, Aung Hospital, Lashio in Myanmar. Our proposed system can be used in medical applications such as scoliosis operations. The tested equipment to run for our proposed method is a computer with Core i7 CPU 2GHz and RAM 4GB. Table 1 and 2 describe the comparison of features extraction performance that tested with SIFT, SURF, Harris and ORB. Table 3 shows the comparison of quality measurement of the ORB method and other feature based image stitching methods using MSE and PSNR values. According to this table, images stitching using ORB method has lager PSNR value and smaller MSE value compared with other feature based methods. From experimental results, ORB can detect correct features points with lowest processing time. Therefore, ORB method is suitable for real time applications.



Figure 8. Final leg mosaiced image.

Table 1. The comparison of features extraction performance (SIFT, SURF, Harris and ORB) for Figure 2 (1) and (2).

Methods	Image Pixels	Features 1	Features 2	Time(s)	Matched features
SIFT	1295 x1224 1477 x1297	3295	1043	8.201892	211
SURF	1295 x1224 1477 x1297	459	873	0.682011	145
Harris	1295 x1224 1477 x1297	4321	2639	0.7356092	399
ORB	1295 x1224 1477 x1297	11	24	0.013491	4

Table 2. The comparison of features extraction performance (SIFT, SURF, Harris and ORB) for Figure 2 (3) and (4).

Methods	Image Pixels	Features 1	Features 2	Time(s)	Matched features
SIFT	1521 x1601 1521 x1275	1369	869	10.391036	304
SURF	1521 x1601 1521 x1275	287	172	0.876134	132
Harris	1521 x1601 1521 x1275	3922	2977	1.837210	457
ORB	1521 x1601 1521 x1275	56	43	0.018932	32

Table 3. Quality measurement of the resultant image that produced by ORB method and other feature based image mosaicing methods.

Methods	MSE	PSNR	Average Processing Time(s)
Image Stitching Using ORB	139.23	26.89	0.013491
Image Stitching Using SIFT	146.47	26.51	8.012932
Image Stitching Using SURF	186.75	25.91	0.65411
Image Stitching Using Harris	216.31	23.67	0.75314

6. Conclusion and Future Work

In this paper, we proposed automatic images stitching system using feature based approach. In features detection and matching, ORB features were used. The proposed method was tested with 100 x-ray images with 50 overlapping pairs. The main goal of our proposed system is to develop a high quality images stitching system with low processing time. The contribution of our proposed system is that it is the first work of medical images stitching system using ORB approach. The proposed system was compared with other feature based images stitching algorithms, SIFT, SURF and Harris. The final image quality measured between the ORB algorithms and other feature based methods were computed using MSE and PSNR values. Experimental results show that the ORB algorithm can produce high quality seamless image mosaic of medical images with low processing time.

Our proposed method tends to be extended for 3D image stitching, 2D images stitching with moving objects and using hybrid of feature based with direct based approaches to get better results as a future work.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors highly appreciate the Asian University Network (AUN/SEED-Net) for their financial support to contribute this research. Moreover, we also thank ‘Aung’ Hospital, Lashio, Myanmar, for the x-ray images used in this study.

References

[1] Niu Jing, Yang Fan and Shi Lingyi, "Improved method of automatic image stitching based on SURF," First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech), Jinhua, pp. 1-5, 2013. https://doi.org/ 10.1109/Ubi-HealthTech.2013.6708059

[2] Cansin Yildiz, “An Implementation on Recognizing Panoramas”, Department of computer engineering, Bilkent University, Ankara, Turkey, 2005. https://www.cs.bilkent.edu.tr/

[3] Matthew Brown and David Lowe, “Automatic Panoramic Image Stitching using Invariant Features”, International Journal of Computer Vision, Department of Computer Science, University of British Columbia, Vancouver, Canada, 2007. https://doi.org/10.1007/s11263-006-0002-3

[4] Jun Hui Gong, Jun Hua Zhang, Zhen Zhou An, Wei Wei Zhao and Hui-Min Liu, "An approach for X-ray image mosaicing based on Speeded-Up Robust Features," International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP), Chengdu,pp.432.435,2012.https://doi.org/10.1109/ICWAMTIP.2012.6413530

[5] R. Botan, K. F. Coco and K. S. Komati, "Implementation of an Image Stitching Algorithm to a Low-Cost Digital Microscope,” 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, pp. 285-291, 2017. https://doi.org/ 10.1109/ICIHT.2017.7899142

[6] Luo Juan and Oubong Gwun, “A comparison of SIFT, PCA-SIFT and SURF,” 2009 International Journal of Image Processing (IJIP), pp. 143-152, 2009. https:// http://www.cscjournals.org/

[7] K. PyarWin and Y. Kitjaidure, "Biomedical Images Stitching using ORB Feature Based Approach," 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Bangkok, 2018, pp. 221-225. https://doi.org/ 10.1109/ICIIBMS.2018.8549931

[8] Chakrasali A. and Manjunatha P. “Automatic Image Mosaicing Using Discrete Cosine Transform”, Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking, Lecture Notes in Electrical Engineering, Vol 258, Springer, India, 2013. https://doi.org/10.1007/978-81-322-1524-0_35

[9] Samsudin, S., Adwan, S., Arof, H. et al. J, “Development of Automated Image Stitching System for Radiographic Images”, Journal of Digital Imaging, Volume 26, Issue 2, pp. 361-370, April 2013. https://doi.org/10.1007/s10278-012-9483-5

[10] M. Dawood, C. Cappelle, M. E. El Najjar, M. Khalil and D. Pomorski, "Harris, SIFT and SURF features comparison for vehicle localization based on virtual 3D model and camera," 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), Istanbul, pp. 307-312, 2012. https://doi.org/ 10.1109/IPTA.2012.6469511

[11] Singla, Savita and Reecha Sharma. “Medical Image Stitching Using Hybrid of Sift & Surf Techniques.” International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 8, August 2014. https://www.citeseerx.ist.psu.edu/

[12] Ebtsam Adel, Mohammed Elmogy and Hazem Elbakry “Image Stitching System Based on ORB Feature-Based Technique and Compensation Blending”, International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9, pp. 55-62, 2015. https://www.citeseerx.ist.psu.edu/

[13] S. Adwan and R. Majed, "A development of an augmented registration approach towards Scoliosis’s radiographic images stitching method,". International Conference on Informatics, Health & Technology (ICIHT), Riyadh, pp. 1-6, 2017. https://doi.org/10.1109/ICIHT.2017.7899142

[14] Chia Yen Chen, “Image Stitching Comparisons and New Techniques”, computer science department, University of Auckland, 1998. https://doi.org/10.1007/3-540-48375-6_73

[15] David Lowe, “Distinctive Image Features from Scale-Invariant Key points”, Computer Science Department, University of Columbia, International Journal

of Computer Vision, Vol.60, No.2, 2004. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[16] Yang Fan; He Yan; Deng Zhen Sheng and Yan Ang, "Improvement of automated image stitching system for DR X-ray images", Computer in Biology and Medicine, Vol 71, pp. 108-114, 2016. https://doi.org/10.1016/j.compbiomed.2016.01.026

[17] B. Baheti, U. Baid and S. N. Talbar, "A novel approach for Automatic Image Stitching of spinal cord MRI images using SIFT," International Conference on Pervasive Computing (ICPC), pp. 1-5, June 2015. https://doi.org/ 10.1109/PERVASIVE.2015.7087071

[18] G. Wang, Z. Zhai, B. Xu and Y. Cheng, "A parallel method for aerial image stitching using ORB feature points," IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp. 769-773, 2017. https://doi.org/ 10.1109/ICIS.2017.7960096

[19] A. Pandey and U. C. Pati, "A novel technique for mosaicing of medical images", Annual IEEE India Conference (INDICON), pp. 1-5, June 2014. https://doi.org/ 10.1109/INDICON.2014.7030465

[20] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," International Conference on Computer Vision, Barcelona, pp. 2564-2571, 2011. https://doi.org/ 10.1109/ICCV.2011.6126544

[21] Hassaballah M., Abdelmgeid A.A. and Alshazly H.A. Image Feature Detectors and Descriptors. Studies in Computational Intelligence, SCI vol 630. Springer, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 128-132 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Experimental Evaluation of Transmission between Two XBee Modules Using Radio-over-Fiber Technique

Luis Alejandro González Mondragón[1], Leidy Johana Quintero Rodríguez[1], Ana Gabriela Correa Mena[1], Jorge Rodríguez Asomoza[2], Alejandro García Juárez[3], Ignacio Enrique Zaldívar Huerta*,[1]

[1] *Instituto Nacional de Astrofísica, Óptica y Electrónica, Depto. de Electrónica, Calle Luis Enrique Erro No. 1, Tonantzintla, Puebla, México, 72840*

[2] *Universidad de las Américas, Depto. de Ing. Electrónica, Ex-hacienda Sta. Catarina Mártir, Cholula, Puebla C.P. 72820, México*

[3] *Universidad de Sonora, Depto. de Investigación en Física, Blvd. Luis Encinas y Rosales S/N, Hermosillo, Sonora, 83000, México.*

ARTICLE INFO

Article history:
Received: 05 April, 2019
Accepted: 04 July, 2019
Online: 17 July, 2019

Keywords:
Microwave signals
Radio over fiber
Optical communications
XBee

ABSTRACT

An experimental transmission between two XBee modules using Radio over Fiber technique is demonstrated. Data issued from an XBee module coded on a wireless microwave carrier at 2.44GHz is transmitted through an optical link of 25.24 km. The optical transmission is based on an external modulation scheme over a dispersive channel. Frequency response of the external modulation scheme is determined by simulation and experimentally. In particular, the chromatic dispersion of the optical fiber associated to its length allow the generation of a pass-band filter whose bandwidth is used to code the RF signal. The recovered RF signal exhibits a SNR of 27dB. Obtained results allow proposing this RoF scheme as a good contender for data transmission.

1. Introduction

This paper is an extension of the work originally presented at the 15th International Conference on Electrical Engineering, Computing Science and Automatic Control [1]. Now, the main contribution of this extended work resides in the fact that we are describing the mathematical procedure that allows determining the frequency response of the electro-optical system used for the transmission. Besides, the theoretical behavior of the optical communication system is validated by simulations. This work is situated on the field of Radio-over-Fiber (RoF), a technique that allows joining radio frequency and electro-optic methods. RoF emerges as a promising solution to fulfill the capacity and mobility of access networks providing high data rate, high capacity and a mobility proposal to transmit information [2-4]. Given that the transmission of information is via light through optical fibers, it is evident that its inherent advantages are exploited, such as low losses, immunity to electromagnetic interference and especially the huge bandwidth available. Figure 1 shows a basic RoF scheme composed by a RF-to-Optical stage, an optical link and an Optical-to-RF stage. The RF signal modulates the light (either directly or externally). The modulated light travels across the optical channel. Finally, the light is turned into its electrical form where it can be radiated by an antenna providing wireless coverage in zones of difficult connection such as shopping centers and subterraneous places [5].

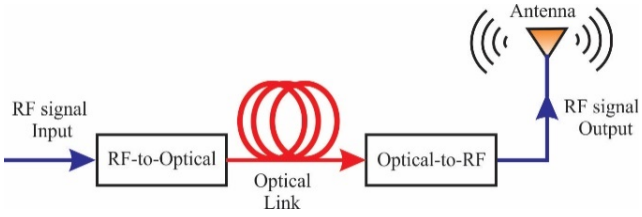


Figure 1. Basic diagram of a RoF system

RoF technology is a candidate that aims to alleviate the bandwidth restrictions for 5G Mobile, delivering wireless data signals via optical fiber channels. [6]. Thus, RoF allows supporting secure, cost-effective coverage as well as excellent capacity mobile-wireless access for oncoming services. Some drawbacks inherent to RoF are due to the distortion and noise derived by the non-linear features of the components employed in the optical network as the optical sources, the chromatic dispersion and photodetectors. For this reason, to mitigate sources of noise and

*Ignacio Enrique Zaldívar Huerta , Luis Enrique Erro #1, Sta. María Tonantzintla, Puebla, México, zaldivar@inaoep.mx

misstatement, intensity-modulation-direct-detection is advisable [7]. In this regard, we propose and experimentally demonstrate a RoF scheme using an optical communication system in the modality of external modulation for the transmission and distribution of a RF-signal. Furthermore, it is demonstrated that the chromatic dispersion (considered as a drawback in this type of systems), is used in an advantageous manner to the generation of notch filters. On the other side, it is known that an XBee is an electronic device whose operation is based on the network transport protocol IEEE 802.15.4, it is able to generate wireless paths to establish communication between devices (http://xbee.cl/que-es-xbee, https://www.digi.com/products/xbee-rf-solutions/xctu-software/xctu). In this sense, XBee modules that use the ZigBee protocol find applications in RoF networks [8, 9]. In this experimental proposal, a wireless signal delivered by an XBee module is used as test signal. A baseband modulation Gaussian Frequency Shift Keying (GFSK) signal at 2.44GHZ is generated by an XBee. This electrical signal modulates an intensity electro-optic modulator. This work is structured as follows. In Section 2, is given the principle of operation of the optical communication system at external modulation and also numerical simulations are presented. In Section 3, experimental set-up of the RoF is described. At the end, a short conclusion is given in Section 4.

2. Principle of Operation

Figure 2 depicts the block diagram the optical communication system in the modality of external modulation used in this work. Basically, it is constituted by a single-mode laser diode as optical source, a Mach Zhender-Intensity Modulator (MZ-IM), a Single Mode-Standard Fiber (SM-SF), as well as a Photodiode (PD).

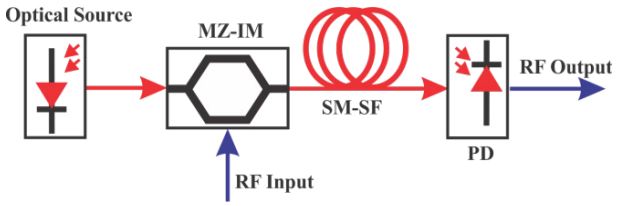


Figure 2. Block diagram of the model used

In the following, a full mathematical analysis explaining the behavior of the model is given. In order to exploit the chromatic dispersion exhibited by the SM-SF, it is necessary to assume that the single-mode laser diode emits at 1550nm. It is demonstrated that the chromatic dispersion (D) and the length (L) of the optical link play an important role that allow the system to act as photonic notch filter.

Mathematically, the optical spectrum of a single-mode laser diode (for example a DFB) can be modeled by a Gaussian envelope centered at an angular frequency ω_0 as [10]

$$S_0(\omega-\omega_0)=\frac{2S_{max}}{\Delta_\omega\sqrt{\pi}}exp\left(\frac{-4(\omega-\omega_0)^2}{\Delta_\omega^2}\right) \quad (1)$$

where S_{max} is the maximum power emission and Δ_ω is the Full Width at Half Maximum of the spectrum that can be expressed in terms of $\Delta\lambda$. Eq. (1) can be simplified, defining $a=4/\Delta_\omega^2$ and $W=\omega-\omega_0$, resulting in

$$S_0(W)=\sqrt{\frac{a}{\pi}}S_{max}exp(-aW^2) \quad (2)$$

The Fourier Transform (FT) of Eq. (2) is the spectral density of the optical source, thus

$$\{S_0(W)\}=\int_0^\infty S_0(W)exp(-j2\pi ZW)dW \quad (3)$$

Defining $Z=v_m\beta_2L$, and expressing the propagation constant $\beta_2=-D\frac{\lambda^2}{2\pi c}$, defining $\omega_m=2\pi v_m$, and $dW=d\omega$, then

$$\mathcal{F}\{S_0(W)\}=\int_{-\infty}^{\infty}\sqrt{\frac{a}{\pi}}S_{max}exp(-aW^2)\,exp(-j2\pi ZW)dW \quad (4)$$

Equation (4) is solved considering that the FT of a Gaussian function $f(t)=exp(-at^2)$ is [11]

$$\mathcal{F}[exp(-at^2)]=\sqrt{\frac{\pi}{a}}exp\left(-\frac{\omega^2}{4a}\right) \quad (5)$$

Rewriting Eq. (4) and considering a, and Z we find

$$\mathcal{F}\{S_0(W)\}=S_{max}exp\left(-\frac{(\pi v_m\beta_2L\Delta_\omega)^2}{4}\right) \quad (6)$$

When $S_0(W)$ is narrow, the frequency response of the system is

$$H(v_m)=cos\left(\frac{\pi v_m^2\lambda^2DL}{c}\right) \quad (7)$$

The squared of the modulation frequency v_m^2 makes the system behave like a non-periodic notch filter.

The notch frequencies f_i, for which Eq. (7) becomes zero, are determined by

$$f_i=\frac{1}{\lambda}\sqrt{\frac{1}{2}\frac{(2i-1)c}{DL}} \quad (8)$$

where λ is the central wavelength in nm, D is in ps/nm·km, c is speed of light in nm/s, and L is the length of the optical link in km.

Equation (7) is numerically evaluated by using MATLAB for the frequency range 0-25GHz, considering: λ=1550nm, c=2.9979x10^{17}nm/seg, L=25.24km, D=15.81ps/nm·km.

Moreover, substituting λ, c, L, and D in Eq. (8), the first and second notch frequency values are f_1=12.50GHz and f_2=21.65GHz, respectively.

Figure 3 shows the result of this numerical evaluation. The result of this evaluation is corroborated by a block level simulation using VPIphotonics software® (http://www.vpiphotonics.com/index.php), and plotted on the same graph. This software is provided with a full library of electro-optics components whose properties can be easily manipulated, therefore, the degree of reliability for the results is high.

From this graph, it is clearly appreciable that the separation between consecutive notches is not constant and decreases as the frequency increases. The bandwidth at -3dB for the low-pass band and for the first lobe (band-pass) are in the order of GHz. The common point between both curves is their tendency. Note that both curves converge to the values of 12.50GHz and 21.65GHz

as stablished by Eq. (8). The justification to use 25km is because the current transmission lengths are designed to meet 20 km, as is determined by an international standard [12].

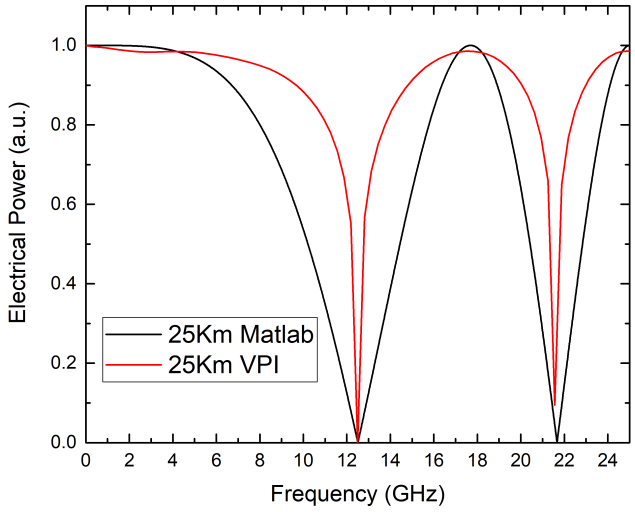


Figure 3. Simulated frequency response using MATLAB and VPI photonics software.

3. Experimental Results

Initially, a test communication between two XBee devices is carried out. Later, a characterization of the optical communication system is done. Finally, the performance of the system composed by a couple of XBee modules over a RoF link is accomplished.

3.1 Test of the XBee devices

XBee is a trademark of radio devices capable of supporting communication protocols as ZigBee, 802.15.4, and Wi-Fi, among others [9]. Figure 4 depicts an XBee network. One module plays the role of Coordinator whereas the other of an End Device.

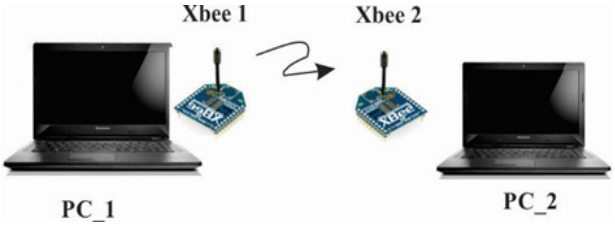


Figure 4. Network generated by using two XBee modules.

The Coordinator handles out addresses and manages the network The End Device is responsible for joining existing networks, transmitting and receiving information. In this work, two XBee modules (XBee-PRO model ZNet 2.5 OEM, indoor range 100m, operating frequency 2.4GHz, transmit power output 18dBm) are used. Both devices are provided with an omnidirectional antenna (Gain 2.15dBi, frequency range from 2.4 to 2.4835GHz) and an USB adapter (Sparkfun XBee Explorer Dongle) that allows the connection via serial communication. The communication between these devices is achieved by using the platform XCTU. Figure 5 shows this interface (https://www.digi.com/products/xbee-rf-solutions/xctu-software/xctu). Data to be transmitted by PC_1 (Coordinator) are coded at the frequency of 2.44GHz and then radiated via an antenna and decoded by PC_2 (End Device).

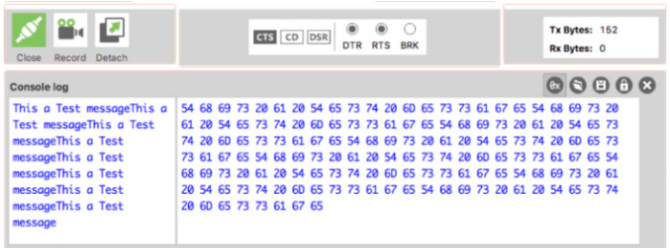


Figure 5. Screenshot showing the interface to control the Coordinator.

Figure 6 shows the spectrum of the emitted radio frequency signal at 2.44GHz measured by an Electrical Signal Analyzer (ESA) exhibiting a SNR value of 55dB. The baseband data is coded within the side lobes.

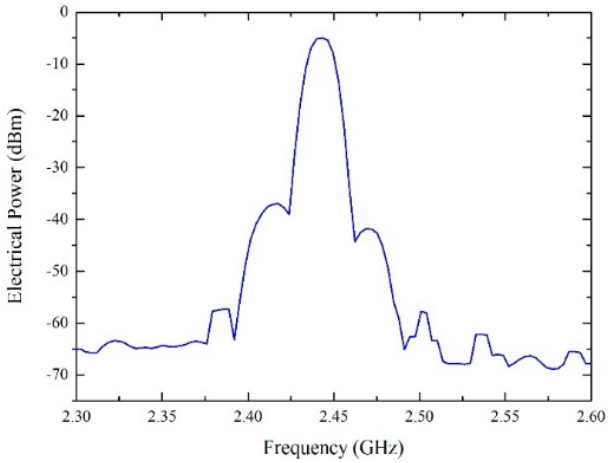


Figure 6. Electrical spectrum corresponding to the Coordinator

3.2 Experimental Frequency Response

Figure 7 illustrates the experimental bench used in this work. It is well known that the use of multimode optical fibers is reserved for short distances, such as in local area networks. Because, we are interested to emulate a FTTH-PON, whose typical length is in the order of 20km [12], thus the use of single-mode optical fiber is mandatory, due to its characteristics of extremely low-losses. Furthermore, most of commercial electro-optical devices (optical sources, optical modulators, photodetectors, etc.) operating at the commercial wavelengths of 1300 and 1550 nm, are supplied with this type of fiber. From the comments above described, SM-SF is used in this experiment. In the next sub-section, a detailed description of the optical communication system is given.

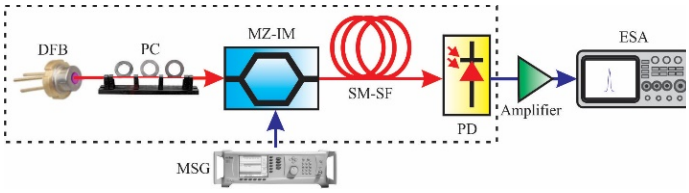


Figure 7. Optical communication system externally modulated

First, the optical spectrum of a DFB (NX8508-55, Side Mode Suppression Ratio SMSR=40dB, linewidth 0.1nm) is recorded by means of an Optical Spectrum Analyzer. According to the datasheet, this optical source has its own internal optical isolator to avoid reflections, therefore a good optical stability is guaranteed. Figure 8 shows the optical spectrum for an operation current of

20mA exhibiting a central wavelength λ_0=1553.5nm. The shape of the optical spectrum corresponds to Eq. (1).

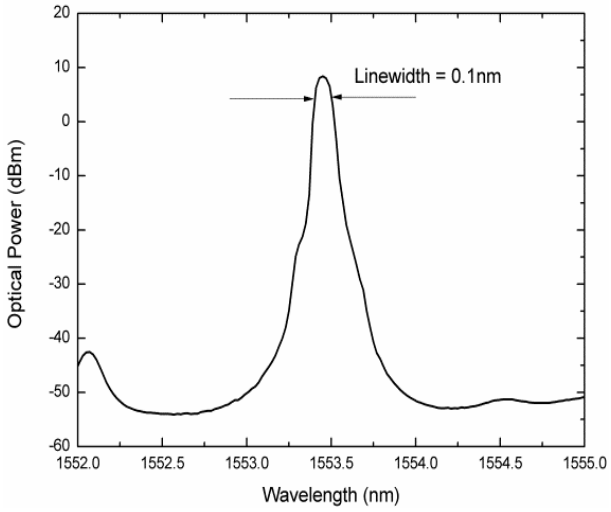


Figure 8. Optical spectrum of the DFB laser source used

Once the characteristics of the optical source are known, the light generated by the DFB is modulated via the electro-optic MZ-IM (insertion loss of 2.7dB, BW-20GHz, operating wavelength 1530-1580 nm, Vπ=7.5 V). The Polarization Controller (PC) device allows optimizing the optical power level at the output of the modulator. The RF signal used to modulate the light is provided by a Microwave Signal Generator (MSG) through a sweep in frequency range of 0.01-13 GHz at a power of 0 dBm. The modulated beam is injected to a reel of SM-SF (D=15.81ps/nm·km @ 1550nm, α=0.22dB/km, L=25.24 km). The light that emerges of the optical fiber is turned to its corresponding electrical signal by the PD (Miteq, BW-13GHz, Responsivity=0.9Amp/Watt). This signal is submitted to a process of amplification (Minicircuits, ZVA-183-S+ Ultra-Wideband Amplifier, 0.7-18GHz, Gain 26dB) and finally visualized in the screen of the ESA. Figure 9 corresponds to the measured frequency response where the simulation curves have been superimposed. Since the frequency emitted by the XBee Module is at 2.44GHz, the low-pass band will be used to code the transmission of the RF signal.

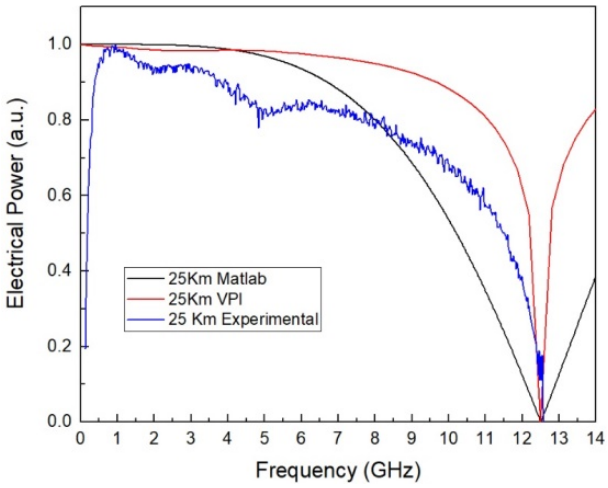


Figure 9. Comparison between simulated and experimental results

From this graph, it is noticeable that the response is restricted by the BW of the PD.

3.3 *Experimental RoF Transmission*

Figure 10 shows the test bench for carry out the transmission of the signal issued from the XBee module. The main difference to the scheme shown in Fig. 7 is that now the MSG is replaced by the coordinator module connected to PC_1 generating the information to be sent (an electrical signal at 2.44GHz at -6dBm) that is emitted from Antenna 1. Antenna 2 receives this signal which is amplified and launched to the RF port of the MZ-IM to modulate the optical signal. The modulated light is injected to the SM-SF. At the end of the link, the modulated light is converted into electric signal by the PD. Subsequently, amplified and separated by a power divider. A percentage of the recovered signal is evaluated by the ESA and the remainder is connected to the Antenna 3 where is emitted. An anechoic chamber is used to guarantee isolation between the signals emitted by antennas 1 and 3. Lastly, antenna 4 captures the signal to be handled by the second XBee module coupled to computer number 2 (PC 2), recovering in this manner the original transmitted message. Figure 11 corresponds to the measured electrical spectrum corresponding to the recovered signal with a SNR of 27dB.

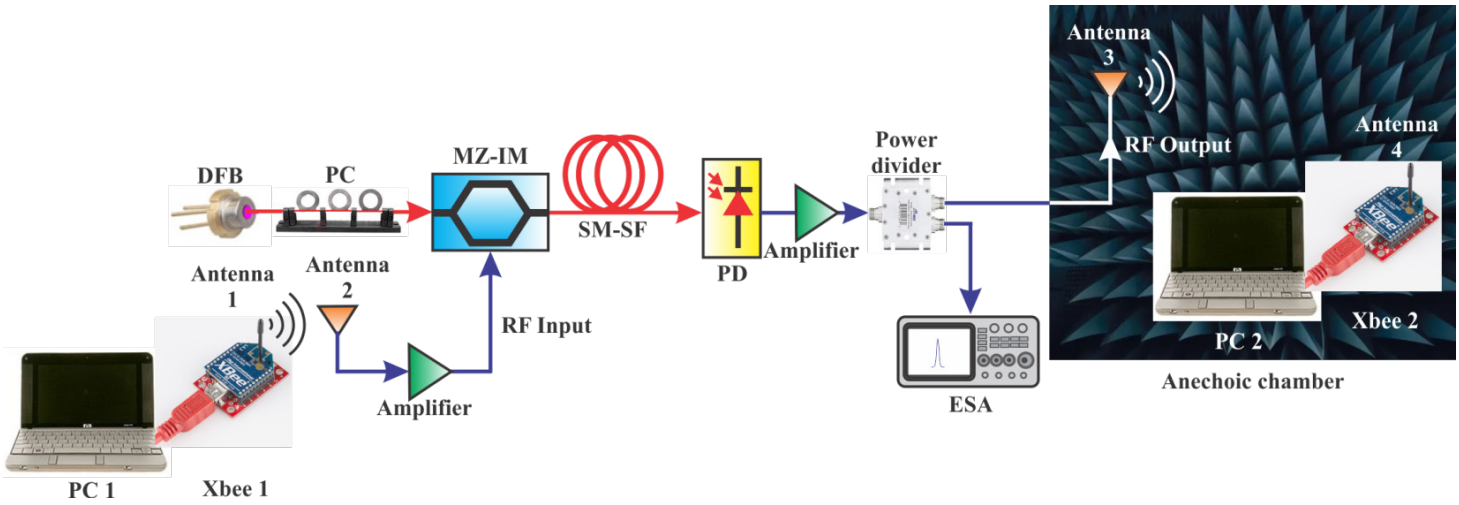


Figure 10. Experimental scheme assembled to test the RoF XBee transmission system

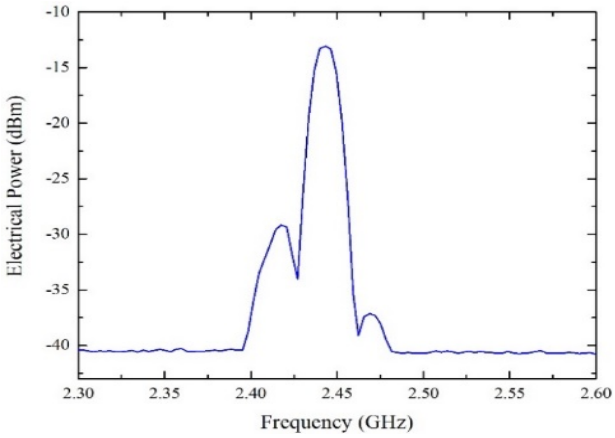


Figure 11. Recovered spectrum corresponding to the output of the system

4. Conclusion

In this work, we have experimentally demonstrated an efficient communication between two XBee modules using the RoF technique. The optical transmission was carried out by using an external modulation scheme. A full mathematical analysis explaining the principle of operation of an optical communication system was given. The relationship that allows determining the frequency response of the optical communication system was evaluated by a numerical simulation achieved in Matlab. Later, a block level simulation of the optical system was carried out by using VPIphotonics software corroborating the previous results. One thing to keep in mind is that the usage of an optical source whose central wavelength is 1310nm associated with SMSF allows the cancellation of the chromatic dispersion effect [12]. However, in this work, we have used a DFB laser whose central wavelength is around 1550nm and SM-SF; thus, the chromatic dispersion value was considerable. This effect, as well as, the length of the optical link has an important role to determine the electrical bandwidth of the system as it was evident in Eq. (7) and Eq. (8). The frequency response was composed by a low-pass band as well as a series of bandpass or notches. In particular, the first lobe was used for the transmission of a wireless signal of 2.44GHz issued by the XBee module. Optical transmission was successfully achieved through an optical link of 25.24Km. The recovered RF signal exhibits a SNR of 27dB. Considering that the electrical bandwidth can be tailored in function of the length of the optical link, it is possible to assure the possibility to accommodate future services distributed by means of RoF schemes using a larger bandwidth. Furthermore, another interesting feature of this RoF system is that, at the end of the link, data can be distributed by antennas to several customers. Direct and external modulation can be used to implement a D-RoF system in our case we have opted for external modulation using a Mach Zehnder modulator (MZM) to modulate a DFB laser in order to generate a simpler configuration [13]. As future work, we propose the use of a VCSEL as optical source in order to benefit of its properties as threshold currents of very few mA and facilitate the coupling with optical fibers [14].

Acknowledgments

L. A. González-Mondragón, L. J. Quintero-Rodríguez, and A. G. Correa-Mena, wish to thank the CONACyT for the student scholarships number 553921, 465594 and 335148, respectively.

References

[1] L. A. González-Mondragón, L. J. Quintero-Rodríguez, A. G. Correa-Mena, J. Rodríguez-Asomoza, A. García Juárez, I. E. Zaldívar-Huerta, “Performance Evaluation of Transmission Between Two Wireless Devices Based on Radio-over-Fiber Technology”, in 2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City , MX, 2018. https://doi.org/10.1109/ICEEE.2018.8533916

[2] A. M. Zin, M. S. Bongsu, S. M. Idrus, and N. Zulkifli, “An overview of Radio-over-Fiber Network Technology”, in 1st International Conference on Photonics 2010 (ICP 2010), Langkawi, Malaysia, 2010. https://doi.org/10.1109/ICP.2010.5604429

[3] A. K. Vyas, N. Agrawal, “Radio over Fiber: Future Technology of Communication”, International Journal of Emerging Trends & Technology in Computer Science, 1(2), 233-237, 2012.

[4] J. Beas, G. Castañon, I. Aldaya, A. Aragón-Zavala and G. Campuzano, “Millimeter-Wave Frequency Radio over Fiber Systems: A Survey”, IEEE Communications Surveys Tutorials, 15(4), 1593-1619, 2013 https://doi.org/10.1109/SURV.2013.013013.00135

[5] S. Kaur, M. Srivastava, K. Singh Bhatia, “Radio over Fiber Technology–A Review,” International Conference of Technology, Management and Social Sciences, Special Issue ICTMS-15, 85-89, 2015.

[6] D. Apostolopoulos, G. Giannoulis, N. Argyris, N. lliadis, K. Kanta, and H. Avramopoulos, “Analog Radio-over Fiber Solutions in Support of 5G”, in 2018 International Conference on Optical Network Design and Modeling (ONDM), Dublin, Ireland ,2018. https://doi.org/10.23919/ONDM.2018.8396143

[7] Pooja, Saroj, Manisha, “Advantages and Limitations of Radio over Fiber System”, International Journal of Computer Science and Mobile Computing, 4(5), 506-511, 2015.

[8] Li Zonglei, Yan Lianshan, Jiang Hengyun, Chen Zhiyu, Guo Yinghui, “Simultaneous transmission of RFID, WIFI and ZigBee over fiber”, in 22nd Wireless and Optical Communication Conference (WOCC), Chongqing, China, 2013. https://doi.org/10.1109/WOCC.2013.6676438

[9] Z. Li, L. Yan, W. Pan, B. Luo, J. Ye, H. Jian, “Simultaneous transmission of multiple wireless services over fiber with reduced network complexities”, IEEE/OSA Journal of Optical Communications and Networking, 6(1), 26-32, 2014. https://doi.org/10.1364/JOCN.6.000026

[10] J. P Goedgebuer, A. Hamel, ,H. Porte, and N. Butterlin, “Analysis of optical crosstalk in coherence multiplexed systems employing a short coherence laser diode with arbitrary power spectrum”, IEEE Journal of Quantum Electronics, 26(7):1217-1226, 1990. https://doi.org/10.1109/3.59661

[11] B. Saleh and M. Teich,. Fundamentals of photonics. Wiley Interscience, 1991.

[12] C. Lee, W. V. Sorin and B. Y. Kim, "Fiber to the Home Using a PON Infrastructure," in Journal of Lightwave Technology, 24(12), 4568-4583, 2006. doi: 10.1109/JLT.2006.885779

[13] M. U. Hadi, H. Jung, S. Ghaffar, P. A. Traverso, G. Tartarini, “Optimized digital radio over fiber system for medium range communication”, Optics Communications, 443, 177-185, 2019. https://doi.org/10.1016/j.optcom.2019.03.037

[14] M. U. Hadi, P. A. Traverso, G. Tartarini, O. Venard, G. Baudoin and J. Polleux, "Digital Predistortion for Linearity Improvement of VCSEL-SSMF-Based Radio-Over-Fiber Links", IEEE Microwave and Wireless Components Letters, 29(2), 155-157, 2019. doi: 10.1109/LMWC.2018.2889004

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 1, 128-134 (2020)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Sensor Based on-the-go Detection of Macro Nutrients for Agricultural Crops

Shraddha Balayya Hiremath*, Raghavendra Shet, Nikita Patil, Nalini Iyer

School of ECE, KLE Technological University, 580031, India

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 02 November, 2019
Online: 22 January, 2020

Keywords:
Agriculture
Macro nutrients
SPAD meter

ABSTRACT

Agriculture, being the vital sector, contributes to India's GDP by 18%. It is necessary to enhance the production in this sector with minimal need of resources to get the good yield of crops. It in turn boosts up the necessity of automation in the field of agriculture. This paper discusses about the experimental study which includes the invasive/non-invasive method to sense the approximate supply of macro nutrients such as Nitrogen, Phosphorous & Potassium to the plants. The nutrient content of leaf is sensed by the usage of optical sensors. The results have been obtained based on the sensor data carried out for the sample test leaf. Thus, the level of the estimated nutrient contents is indicated by three different ranges such as Low (L), Medium (M) and High (H). A mathematical model is developed for nutrient measurement especially for nitrogen and validated with SPAD meters for the measurement of nitrogen content. The other macro nutrients were also estimated, and the entire system is tested at different test conditions with the support of UASD in the field of University of Agricultural Science.

1. Introduction

As there is an increase in population and the per capital income, the demand for crop yield is drastically increasing with the advancement in technological growth. It is necessary to identify the severe problems involved in crop yield and address the challenges [1]. Even though there is much advancement in the field of science and technology still farmers across the country adopt the traditional practices. The availability of the resources in the process of farming is one of the most important aspects in eco-system. In addition to this nutrient are very important resources for the plants that give structure to the leaf. Since the population growth of the nation is getting enhanced, this creates a necessity to increase the food production. Supplying the appropriate amount of fertilizers and water becomes evident in enhancing the plantation yield as a whole. The basic nutrients provided to the plants are Nitrogen (N), Phosphate (P), and Potassium (K) in the ratio of 4:2:1. On the other hand due to the lack of awareness and quick measurement techniques, the farmers oversupply nutrients and water to plants, leading to destruction of the crops. Especially the excessive supply of nitrogen nutrient to the crops causes the chemical to evaporate and create oxides of nitrogen (NOx) mainly nitric oxide and nitrogen dioxide into the atmosphere. The formed NOX contributes to the formation of smog and acid rain, affecting ozone layer resulting to global warming. Hence there is a necessity for efficient use of nitrogen broadly known as Nitrogen Use Efficiency (NUE).

The fertilizers should be supplied in appropriate manner so that good number of crops can be obtained without any defects. The exact amount of fertilizers will be supplied based on the type of crop, its growth status and the quality of fertilizers or soil. The growth of the plant depends upon absorption and consumption of the nutrients. One of the major challenges in agriculture in order to get the optimized plant yield is reducing the amount of water consumption and fertilizers [2]. Different nutrients are available in the market for plantations. They are scattered to the plantation area for better crop output. Nitrogen related chemicals are distributed in abundant volume as they play an active role in managing an edulcorated premises in the field. Also, potassium and phosphorous aids in plant growth. The growth stage of the plant can be managed in a better way by maintaining proper supply of these nutrients. But many times, unnecessary supply or inadequate supplies of nutrients are evident in all field area leading to environmental hazards and wastage of crop yield. In association with the technology, a smart way has to be established to accurately measure the nutrient content and supply concrete chunk of fertilizers.

*Shraddha Balayya Hiremath, Email: shraddha_h@bvb.edu

The conventional methods of Chlorophyll and Nitrogen analysis require the destruction of the leaf sample. It is also time-consuming and expensive process. The above techniques cannot satisfy the requirements of precision farming. Non-destructive optical techniques based on absorbance and reflectance of light by leaves have been proven as alternative time-saving and simple techniques to quantify Chl in a number of agricultural species such as birch (*Betula pendula Roth.*)[3], [4], wheat (*Triticum aestivum L.*)[5] and potato (Solanum tuberosum L.), maize (*Zea mays L.*)[6] and soybeans (*Glycine max L. Merrill*.), sorghum (*Sorghum bicolor L. Moench*.) and pigeonpea (*Cajanus cajan L. Millsp*.). Strong relationship between CCI and nitrogen was also reported in sugar maple (*Acer saccharum Morsh*.) (Van den Berg and Perkins2004) [7], [8], and rice (*Oryza sativa L*).

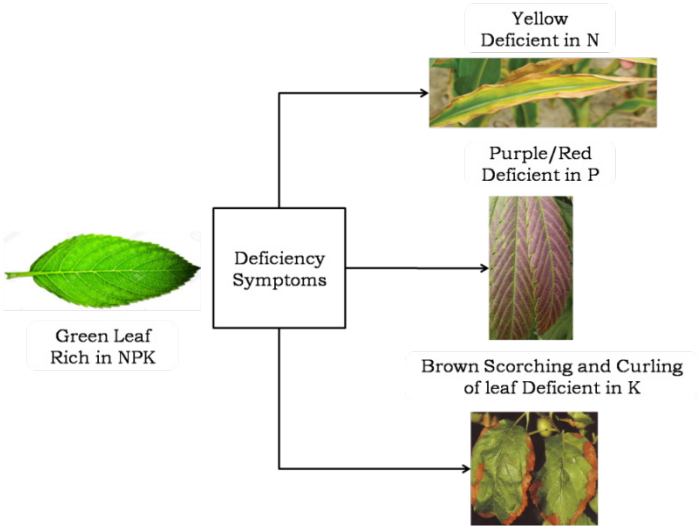


Figure 1: Change of color showcasing nutrient deficiency in leaf

From Figure 1, it is apparent that the deficiency in nutrients results in an outcome in the form visual look of the leaf which proves to be a nutrient deficient plant [9]. The deficiency symptoms in plants completely depend upon the color of the leaf. They are as follows,

- If the color of the leaf results in yellow, it proves to be deficient in N
- If the color of the leaf results in purple/red, it proves to be deficient in P
- If the leaf results in brown scorching and curling of leaf, it proves to be deficient in K

Finally, if the leaf appears to be completely green at its initial and later stages of growth, it results to be rich in all the three macro nutrients.

Methodology carried out to detect the approximate value of nutrients is by the study of spectral reactance of the leaf of the plant. The reactance alters at different stages of the plant growth on the grounds of variation in nutrient content [3]. In particularly the nitrogen is absorbed by the plant through atmosphere and is closely proportional in a linear fashion with chlorophyll. Hence the penetration and reactance of the leaf with light modulates with variation of the color of the leaf from green to yellow. On the other way the different macro nutrients are directly sourced from the soil of the plants.

The techniques present for nutrient and moisture content in plant are:

- Invasive Method
- Non-invasive method

Invasive technique uses destructive method for measurement of nutrients. The invasive methods used for measuring micro and macro nutrients for soil, leaf and irrigation water are:

- Electro-chemical method.
- Ion-selective electrode method.

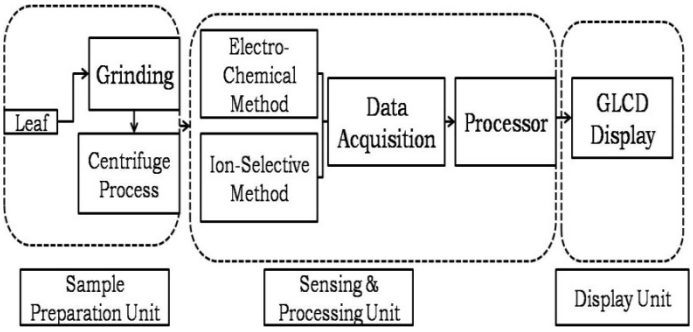


Figure 2: Architecture for Invasive Method

The overall system level block diagram for Invasive technique is depicted in Figure 2. It comprises of the various blocks such as sample preparation unit, sensing & processing unit and display unit. The sample preparation unit consists of sample leaf which is grinded into small pieces and further sent to the centrifuge process where in the sedimentation principle is used to separate the denser substances from less dense ones. The sensing and processing unit inculcates two different methods such as electro-chemical method and ion-selective method to acquire the data from previous unit. The data is further processed and given to the display unit in order to display the required data. The architecture for Non-invasive method is shown in Figure 3. The non-invasive technique uses non-destructive method for measurement of nutrients. This method uses non-contact type of sensors. The optical based sensors with the appropriate wavelength are selected for measuring the nutrients. The output obtained from optical sensor is processed using an embedded system and displayed. Based on the concentration of the nutrients an automatic alarm signal is generated for an inadequate concentration of particular nutrients which in turn controls the switching ON and OFF of the existing drip irrigation system.

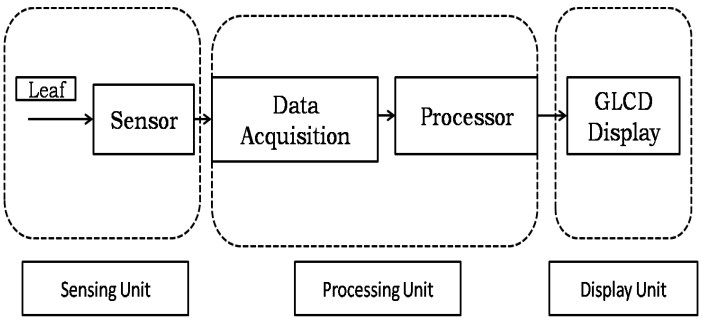


Figure 3: Architecture for Non-invasive Method

In order to quantify the spectral absorption optical sensor was used termed as TCS230.A waveform is generated as a result of the sensor which is later sampled to quantify the changes in nitrogen nutrient in leaf [10]. Different plant families exhibit variations mathematically involving Chlorophyll content Index (CCI) and entire quantity of chlorophyll/nitrogen content [11], [12]. For the product under design a self reliant mathematical model is developed for the species under consideration to a belonged zone. The main feature of the system should include its portability and user friendly behavior.

The organization of the paper is as follows. Section II discusses the implementation details of the system involving nitrogen, phosphorous and potassium detection. Section III discusses the results obtained for each nutrient and lastly section IV concludes the observational study.

2. Proposed Methodology and Implementation

This section describes about the implementation carried out to detect the three essential macro nutrients in plants such as Nitrogen, Phosphorous and Potassium. Two techniques were implemented for stress measurement of different nutrients.

1. Non invasive technique for detection of nitrogen and phosphorous.

2. Invasive technique for detection of potassium nutrient.

Intermediate sub models were built to reach the set goals. Different methods to improve the efficiency by the means of which precision and smart fertigation is obtained is discussed. This part of the paper also confers about different system models that are developed in order to achieve the set goals. It also discusses various methodologies of optimizations to the built models that are adopted to lock the constraints on various criteria's and approach by which appropriate nutrient distribution is carried out.

2.1. Nitrogen Detection

This segment of paper describes the details of nitrogen detection in maize leaf using non-invasive technique by discussing the system architecture proposed and implemented, formulating a mathematical model and optimization. Figure 4 shows the structural view of estimating the content of nitrogen in maize plant. It consists of various sub blocks such as color sensor, Pulse modulation wave processing, ADC block, and comparator and display device. The color sensor was used to sense the color of the leaf. Depending upon the color the amount of nitrogen will be estimated. Constant light source was provided to the leaf. The sensor output will be in digital domain. The color of the leaf changes accordingly with the frequency of the output wave. Visible light and infra red are the two light sources which were considered for experimentation. Different set of readings were obtained depending on the lighting conditions. The digital output obtained from the color sensor is provided from the controller for the signal to be processed. The frequency, wavelength of light reflected from the processed data was calculated with the help of algorithm. The frequency obtained is directly proportional to the reflectance of light that in turn changes with the content of chlorophyll of the leaf. Different frequency values are obtained as per the change in the color of the leaf. The relationship between obtained and reference frequency is gained by using mathematical model.

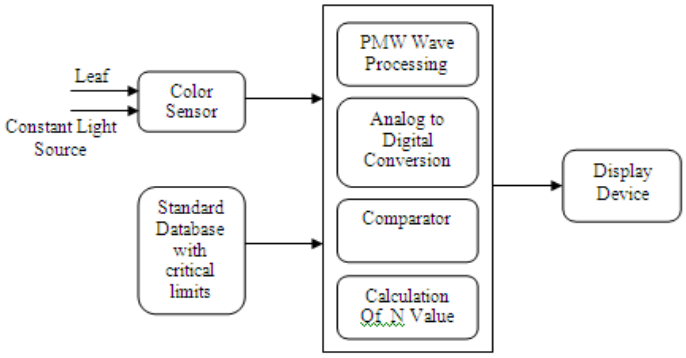


Figure 4: White box of N detecting sub module

The data set for the worst value was aggregated for analytical measurement from UASD College. Based on these values and also the value obtained from the system were compared and the assessed value of nitrogen in percentage of the leaf is unveiled and authenticated.

2.1.1. Mathematical Modeling

There are several authoritative systems in the market which act as chlorophyll meters. SPAD 502 meter is one such device that is made use in this project to act as a reference for spectral assessment of the leaf [13], [14]. Figure 6 provides an insight of the methodology undertaken to arrive at the inter-relationship between the standard device and the device under design.

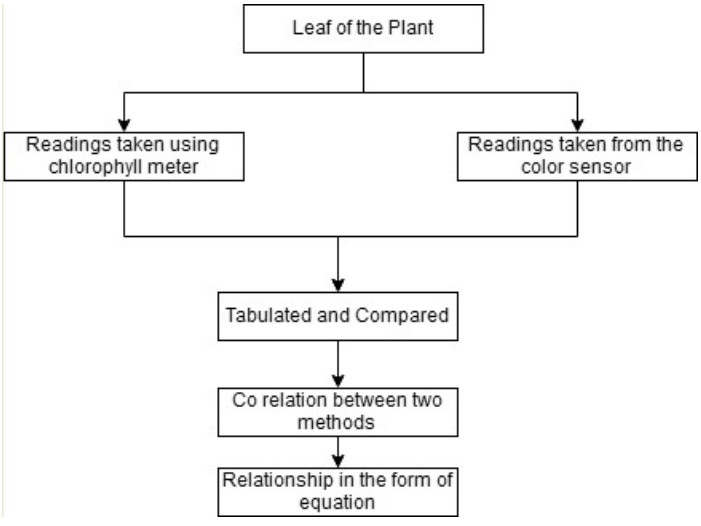


Figure 5: Method to arrive at the relationship between SPAD meter and color sensor

The readings of the leaf using both the devices were taken under different environmental conditions.

Equation (1) describes a linear inter-relation between the data from meter and the color sensor observed by plotting the graph.

$$Sensor = a * SPAD + b \quad (1)$$

Parameter 'a' & 'b' in Equation 1 constitutes to the slope and intercept respectively. This produces a dedicated mathematical model for nitrogen estimation. While creating a data from the reference meter, the readings were taken from all the different areas of the leaf at various locations. The same technique was also carried out while collecting the leaf data using color sensor

The data tabulation of SPAD meter and N_Sensor plot characterizes non- uniformity and was proportional in an inverted manner with each other revealing no patterns. An in-built function in MatLab i.e. fit function was exploited to fit the curve obtained with the help of polynomial of degree. The equation procured after the fitting procedure is as depicted in Equation 2 which shows inverse proportionality relation between the two data sets. Later stage involved modification of Equation 2 in terms of N_Sensor as exemplified in Equation 3.

$$SPAD = (-1.7393)N_sensor 93.83 \quad (2)$$

$$N_sensor = (-0.5748)SPAD + 93.83 \quad (3)$$

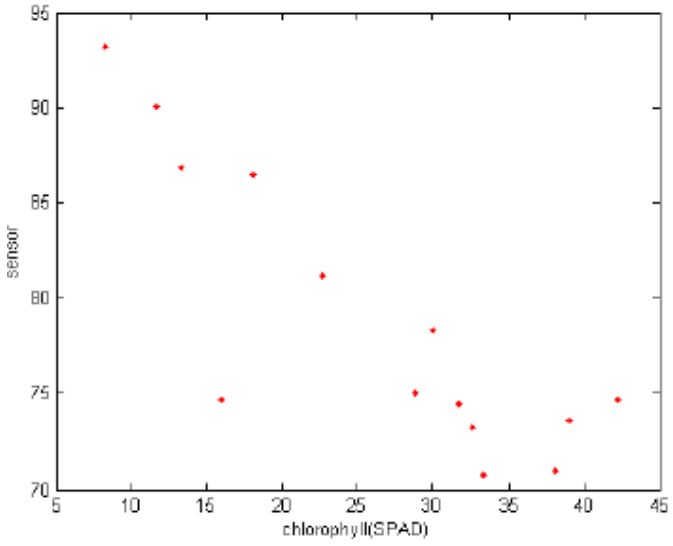


Figure 6: Relationship between chlorophyll content in SPAD and chlorophyll content in sensor

2.1.1. Optimization

Taguchi's technique acts as a reference and narrates a systematic approach for planning and conduction of experiments most effectively when there are numerous parameters related to the end results [15]. Hence the most appropriate data is collected which provides valid and objective conclusions as shown in Figure 7. Figure 8 depicts the inclusion of this approach to obtain different features.

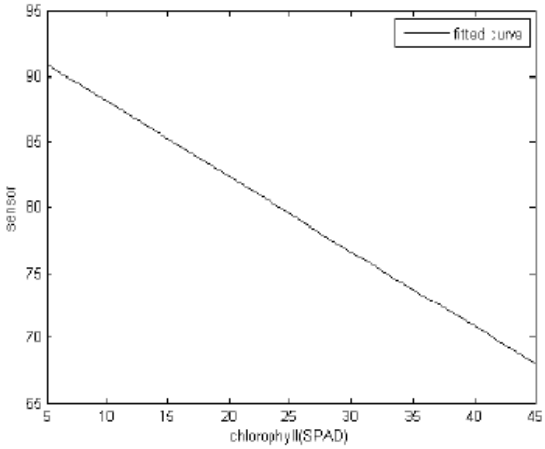


Figure 7: Best fit curve displaying the connection between SPAD device and chlorophyll

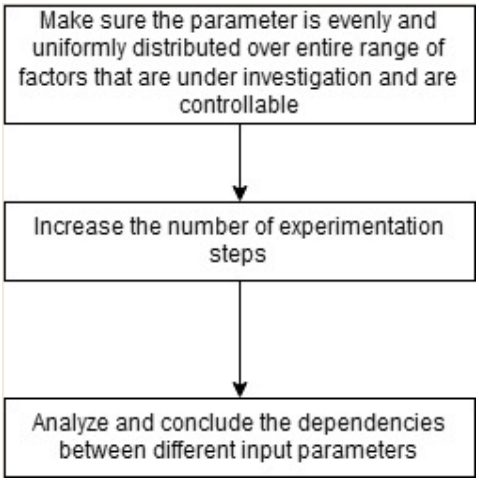
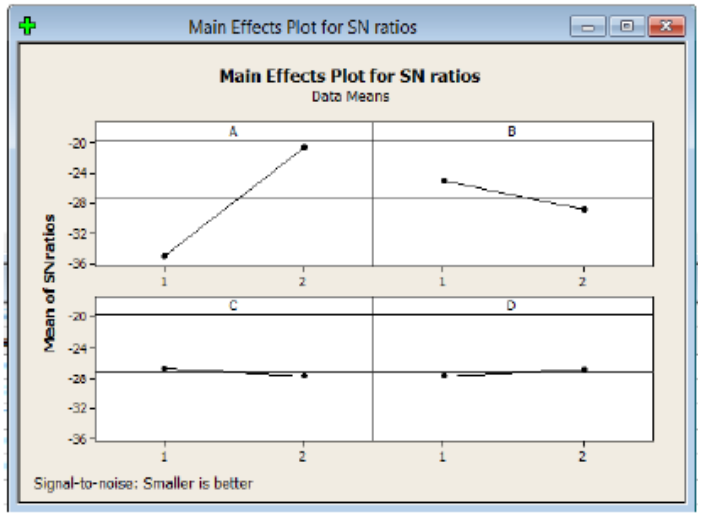
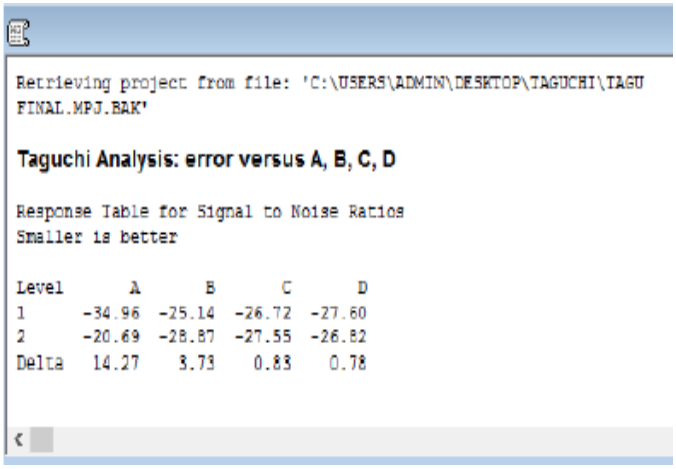


Figure 8: Features list requirement

The signal to noise ratio was plotted and the sensor error was defined as shown in below Figure 9.



In terms of achievement a lower range of the ratio of signal and noise should be present. Hence the gilt edge level of data set should be chosen in that range which has very least level of ratio of noise and signal. After above mentioned experimentation the best result is achieved at 5Volts present at second level with the distance of 1cm and at the lower range of frequency of first level. This also covered a large area at 2nd level as shown in Figure 10. A thorough study was investigated to identify the major design factors affecting the quality of the results using ANOVA technique. The outcome of this investigation led to the conclusion that voltage variation mainly influences the results compared to other design parameters.



2.2. Phosphorous Detection

This is another essential nutrient required in large quantity for the yield to enhance their productivity [16]. As per the observation, this nutrient has absorbance peaks in red regions with the wavelength ranging from 600nm to 700nm. It has no transmission in the near infra-red region. Hence the design in selection of the sensor is enabled such that the emission in light in preferably in the visible region. At the later stage of the sub system an optical filter id designed based on the wavelengths of interest that has to be passed. For this work the wavelength taken into account are beyond 665nm.

To calculate the nutrient amount approximately in the maize, crop a conceptual model was developed as depicted in Figure 11. The instrument used to calculate the phosphorous amount is equipped with series of electro-luminescent diodes that eradicates light with greater intensity in the visible range. The output of the designed system produces the range of phosphorous content in three levels – High, Medium and Low. The specimen of interest is exactly positioned between transmitter and receiver. Later the optical sensor detects the reflectance of those wavelengths.

Table 1: Set Threshold Values on Experimental Basis

Threshold Value	Nature of Leaf
2	No Green
3	Green

Table 2: Readings Obtained for Phosphorous

Readings	Approximate Phosphorous Content
> 3.4	Sufficient
< 3	Deficient

Table 1 discusses about the threshold value set based on the numerous readings that differentiates between green and non-green leaf. Table 2 discusses about the obtained readings which provides an approximate phosphorous content. The advantage of the device under investigation is, it identifies the change in Phosphorous content in an approximate manner well before it is visible by naked eye.

2.3. Potassium Detection

This nutrient is considered as the quality nutrient which is required in large amounts for proper growth as well as reproduction of the plants. The detailed block diagram for detecting the potassium content in coconut leaf is shown in Figure 12.

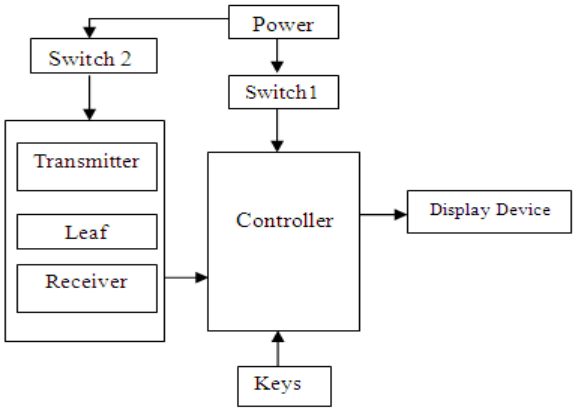


Figure 11: Functional Block Diagram

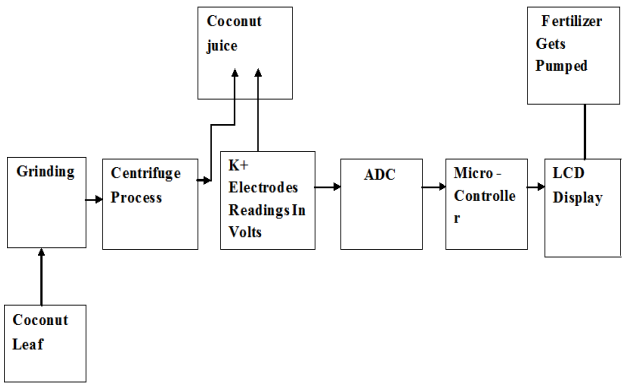


Figure 12: Functional Block Diagram

An experiment was conducted on coconut leaf to estimate the potassium content. One of the invasive techniques, ion selective electrode method was carried out for estimating the nutrient content. In this approach a K^+ electrode is used and the approximate amount of K^+ in a standard unit of PPM is determined.

Table 3: Source- Visible light; Measuring parameter- Nitrogen in maize leaf

Case	Gap between leaf and measuring device in cms	Dark green	Light green	Yellow
Day	2.4	47.05	48.81	53.26
	2.7	45.33	47.07	51.81
	3	43.10	46.01	50.95
Night	2.4	46.75	46.18	48.16
	2.7	44.83	45.98	46.445
	3	43.46	44.48	45.66

Table 4: Source-Infrared; Measuring parameter- Nitrogen in maize leaf

Case	Gap between leaf and measuring device in cms	Dark green	Light green	Yellow
Day	2.4	46.23	54.56	62.05
	2.7	37.27	46.34	52.34
	3	31.07	39.8	43.07
Night	2.4	41.3	45	50.06
	2.7	36.76	39.09	41.89
	3	30.12	32.76	36.23

3. Results and Discussion

This part of the study details about the tabulation of results obtained from the designed system for nutrient management in plants. Initial sub sections discuss the detection of nitrogen and phosphorous content in maize plant using non-invasive technique and the last sub section describes the results obtained for potassium content in coconut leaf using invasive technique. An IoT application was built to store the acquired data onto the cloud for further analysis. The created home page with experimental values is been listed.

3.1. Nitrogen Detection

SPAD meter is used as a reference device to carry out the investigation work. The data was recorded and the average of sensor data was calculated for each segment of the leaf. There is a need to compare a wide range of chlorophyll data which is obtained from the spectroscopy. The readings were considered from different segments of the leaf. The most important observation was that the content of chlorophyll was decreased at the center portion of the maize leaf. As it is apparent that leaf's nitrogen content directly varies with the content of chlorophyll in it. Thus, there is a linear relation between the two. Exploiting this mathematical relation maize leaf nitrogen content is calculated. Tests are conducted during day and night for validating the impact of sunlight on the designed sensor. The sensor is also validated for different sources of light particularly visible and infrared light as depicted in Table 3 and 4

T



Figure 13: Prototype showing high nitrogen in maize leaf

Figure 13 exhibits the proof of concept of the sub-component implemented.

3.2. Phosphorous Detection

For proof of concept the specimen leaf of maize plant was taken into account. A total of 10 leaves were procured on random basis for analysis. The test was conducted on the leaf using two methods for validation. The work carried out mainly concentrates on developing a strong interdependency between nutrient levels of phosphorous and reflectance numbers in two regions i.e. red and NIR range. [17].

Figure 14 & 15 displays the values obtained by the component under design and web page that includes the real time readings.



Figure 14: Home Page

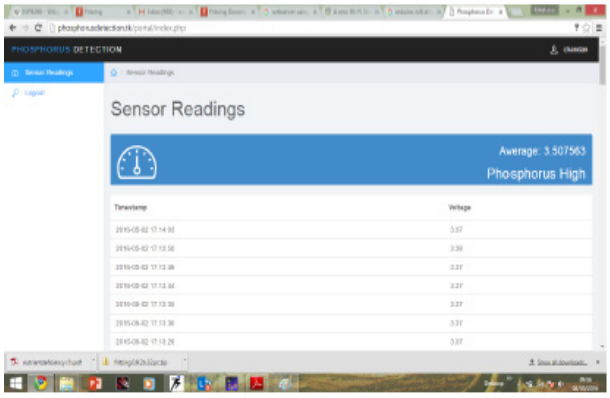


Figure 15: Dash Board

3.3. Potassium Detection

The measuring device initially requires preparation of coconut leaf extract that involves grinding followed by centrifuge process. The electrodes are used to estimate the reading in volts and the data is refined using signal conditioning circuits. The controller processes the conditioned data and displays the results of the nutrient content in terms of High (H), Medium (M) and Low (L). Table 5 displays the results obtained by ion selective method for potassium content in the coconut leaf. There was a linear relationship between the electrode output readings and concentration in PPM. This method of detecting the potassium content was further corroborated by taking the coconut leaf samples for UAS Dharwad which was categorized into less potassium content, medium content and rich in potassium. Using these leaves the results were obtained that stated the potassium content as per its presence in the leaf.

Table 5 : Results Obtained for Potassium Nutrient

Electrode Output in Volts	Concentration in PPM
0.2	4.503
0.4	20.28
0.6	106.204
0.8	478.3
1.0	2154.43
1.2	9703.49

4. Conclusion

Irrigation and fertigation play a pivotal role in better yield of crops. Excessive usage of fertilizers, in particular nitrogen leads to evaporation of this chemical which in turn creates global warming. Hence this impacts the environment. The projected work assesses the quantity of the nutrient present in the leaf approximately in the scale of high, medium and low. Three nutrients are addressed which are treated as macro nutrients. The nitrogen and phosphorous was tested and validated for maize plant where as potassium nutrient was initially experimented from coconut leaf. The entire work is based on the spectral analysis which was conducted in isolation for N and P and by inculcating non-invasive method of stress detection. Compared to conventional procedure, sensor-based reasoning of nutrients contributes to large spectrum of readings. Different optimization techniques were involved and among them Taguchi's approach was used in nitrogen detection.

A server is created to observe the real time data remotely. An invasive technique was also subjected in the proposed work for detecting potassium content in coconut leaf by using ion selective electrode method which was found to be time consuming and involving huge devices for making the required solution.

5. References

[1] B. H. Shraddha, R. M. Shet, P. Nikita and N. C. Iyer, "Nutrient Detection for Maize Plant Using Noninvasive Technique," *2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, Sangamner, 2018, pp. 504-509.doi: 10.1109/ICACCT.2018.8529659

[2] Ghorbani, Amir Mirzaei Mossivand and Abazar Esmali , "Utility of the Normalised Difference Vegetation Index (NDVI) for land/canopy cover mapping in Khalkhal County".Ouri 2012

[3] Bodo M, Schmidhalter U ," Estimating the nitrogen nutrition index using spectral canopy reflectance measurements". Europ J Agron. 29:184-190.,2008.

[4] Betula pendula Roth,"Effects of nitrogen on growth properties and phenology of silver birch" JOURNAL OF FOREST SCIENCE, 58, 2012 (9): 391–399

[5] Guijun Yang; Chunjiang Zhao; Ruiliang Pu; Haikuan Feng; Zhenhai Li; Heli Li; Chenhong Sun; "Leaf nitrogen spectral reflectance model of winter wheat"(Triticum aestivum L).

[6] J.E. McMurtrey ; E.M. Middleton ; L.A. Corp ; P.K.E. Campbell ; L.M. Butcher ; C.S.T. Daughtry "Optical reflectance and fluorescence for detecting nitrogen needs in Zea mays L."

[7] Van Den Berg, A.K.; Perkins, T.D. Evaluation of a portable chlorophyll meter to estimate chlorophyll and nitrogen contents in sugar maple (Acer saccharum Marsh.) leaves. For. Ecol. Manag. 2004, 200, 113–117.

[8] P Nikita, P Jyoti, B Shraddha, P Preeti, S Raghavendra, Nalini C Iyer, " Active Learning in Electronic Measurements and Instrumentation Course through Hands-on " Journal of Engineering Education Transformations.

[9] Ahmad U, Mardison S, Tjahjohutomo R, Nurhasanah A , " Developmentof automatic grading machine prototype for citrus using imageprocessing," Aust J Agric Eng. 1(5):165-169. , 2010.

[10] BH Shraddha, M Subhas, P Nikita, P Preeti, K Bhagyashree, RM Shet, P Jyoti, Nalini C Iyer, "Model based Learning of Linear Integrated Circuit Course: A practical approach" Journal of Engineering Education Transformations.

[11] Anatoly A. Gitelson, Yuri Gritz and Mark N. Merzlyak "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves".

[12] Anand M, Haripriy M, Byju G, "Chlorophyll meter and leaf colour chart to estimate chlorophyll content, leaf color, and yield of cassava.Photosynthetic",. 46: 511-516., 2008.

[13] M. T., Voroney, P. and Parkin, G. 2005, "Predicting nitrogen fertilizer requirements for corn by chlorophyll meter under different N availability conditions". Can. J. Soil Sci. 85: 149159.

[14] Chang SX, Robison DJ ," Nondestructive and rapid estimation of hardwood foliar nitrogen status using the SPAD-502 chlorophyll meter". Forest Ecol Manag. 181: 331-338.,2003.

[15] Preeti Pillai, S Raghavendra, B Shraddha, P Nikita, Nalini Iyer, " Digital Signal Processing: An abstract mathematics to real world experience" Journal of Engineering Education Transformations.

[16] Osborne, S. L., J. S. Schepers, D. D. Francis, and M. R. Schlemmer. 2002. "Detection of Phosphorus and Nitrogen Deficiencies in Corn Using Spectral Radiance Measurements".

[17] Ciganda V, Gitelson A, schepers J," Non-destructive determination of maize leaf and canopy chlorophyll content. J Plant Physiol" 166: 157-167.,2009.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 132-138 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Optimization Method of Wideband Multilayer Meander-Line Polarizer using Semi-Analytical approach and Application to 6-18GHz Polarizer including test with Horn Antenna

Wafa Abdouni-Abdallah*, Muhammad Saeed Khan[1], Athanasios Konstantinidis[1], Anne-Claude Tarot[2], Aziz Ouacha[1]

[1]*Emirates Technology and Innovation Center, Communication and Electronics Department, Khalifa University, Abu Dhabi, UAE*

[2]*Institute of Electronics and Telecommunications of Rennes, Antennas & Microwave Devices Department, University of Rennes 1, 35700 Rennes, France*

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 28 August, 2019
Online: 24 September, 2019

Keywords:
Meander-Line Polarizer
Wideband Polarizer
Multilayer Structure
Transmission-Line Model
Simulated Annealing Algorithm

ABSTRACT

An optimization method using a semi-analytical approach is detailed in this paper to design a wideband multilayer meander-line polarizer. It consists of the transmission line equivalent circuit study to find the best values of the shunt components that allow the maximum transmission and a 90 phase difference between two orthogonal components over the entire bandwidth, followed by a simulated annealing algorithm to optimize the polarizer's dimensions by converging the values of the shunt components found earlier, to some rigorous empirical formulas of the equivalent admittances.
To validate the proposed method, a wideband (6-18 GHz) four layer's meander-line polarizer is optimized and manufactured. The polarizer prototype is stacked to a horn antenna operating at the same frequency band and tested. The experimental results verify that the linear waves of the horn antenna alone were converted to circular ones when the multilayer meander-line polarizer is stacked to it.

1. Introduction

This paper is an extension of work originally presented in the 48th European Microwave Week [1] where a four layer's wideband (6 – 18 GHz) meander-line polarizer has been optimized and simulated with different patch antenna for validation.

As well known, the polarizer is a passive broadband device, used to change the polarization properties of an Electro-Magnetic (EM) wave. Combining antennas with those devices to change the polarization of the wave is very useful. A well-designed polarizer permits to change the antenna polarization and to propagate a different polarization without any significant variations in the pattern performance [2]. The Meander-Line (ML) polarizer is the most popular device used to change the Linear Polarized (LP) waves into the Circular Polarized (CP) ones [3]. It is a multilayer structure made of several printed ML sheets spaced by about one-quarter wavelength apart. The principle of operation is to convert an incident wave into two equal components at ±45º as summarized in [4]. One component passes through a structure equivalent to a broad-band shunt-inductive filter while the other passes through a broad-band shunt-capacitive filter. the other component, by about 45º.

The ML polarizer has been proposed in [4], and improved in [5]. Later, [6-9] gave analytical formulas to compute the phase delay and characterize the polarizer's grating layer. Transmission Line (TL) theory together with Method of Moments (MOM) based algorithms has been used by [7, 10-13] to analyze the ML polarizer.

In this paper, the basic principle of the ML polarizer is described in section 2. Section 3 describes the detail of the optimization procedure which includes the TL equivalent circuit and Simulated Annealing (SA) algorithm. Finally, in section 4, the four layer's wideband (6 – 18 GHz) meander-line polarizer studied previously in [1] is used as a case study. The optimization study is presented, followed by experimental validation using a wideband horn antenna covering the same bandwidth.

2. Meander Line Polarizer: Basic Principle

The ML polarizer is a multilayer structure (Figure 1). Each layer consists of MLs printed on a low-loss dielectric substrate and separated by a spacer respecting one-quarter wavelength apart. The

*Wafa Abdouni-Abdallah, Emirates Technology and Innovation Center, Khalifa University, Main Campus, Abu Dhabi, UAE, +971 690 3881, wafa.abdouni@etic.ac.ae

spacer could be air (vacuum) or any kind of low loss material such as foam or dielectric substrate.

The principle of operation is to convert an incident wave into two equal and perpendicular components at ±45 degrees as summarized in [3]. The ML polarizer's layer is placed in the *xy*-plane in such a way that the meander axes are all oriented at an angle Ψ = 45 degrees with respect to the incident electric field $\boldsymbol{E_{inc}}$ along the *y*-axis (Figure 2).

The incident electric field $\boldsymbol{E_{inc}}$ will be decomposed into two orthogonal components: one parallel and the other one perpendicular to the ML axis, noted respectively $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$. As well known, a CP wave can be obtained by designing a ML polarizer that introduces a differential phase shift of 90^0 between the components $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$ while keeping their magnitudes identical over the entire required frequency bandwidth. Right-hand or left-hand circular polarizations can be easily obtained by switching the orientation angle $\boldsymbol{\Psi}$ from $+45^0$ to -45^0.

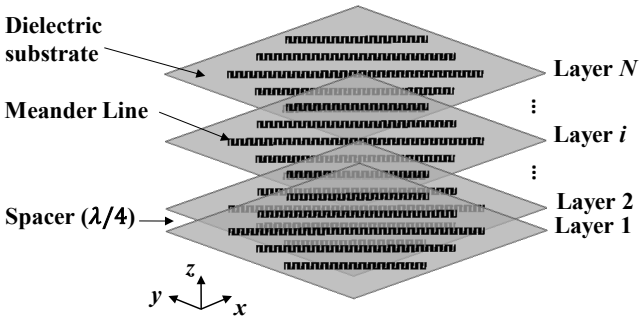


Figure 1: N-Layers ML polarizer design

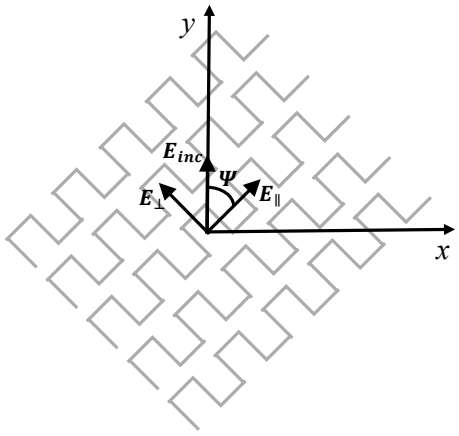


Figure 2: Incident field ($\boldsymbol{E_{inc}}$) exiting the polarizer layers with the parallel $\boldsymbol{E_{\parallel}}$and the perpendicular $\boldsymbol{E_{\perp}}$ components resulted.

The difficulty in the design of a meander-line polarizer is to determine the inductive and capacitive impedances of the grating for the $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$ polarizations with respect to frequency and constitutive dimensions of the meander line.

An optimization method is developed to optimize the ML polarizer's dimensions and described in detail in the below section. It consists of a semi-analytical analysis based on the combination of TL circuit theory and SA algorithm and allows to find the dimensions of the MLs that keep the magnitude of the components $\boldsymbol{E_{\parallel}}$and $\boldsymbol{E_{\perp}}$ identical with a phase differential shift of 90º over the desired frequency bandwidth.

Our method of optimization has the advantage of avoiding the heavy computations and simulations used in [13] to solve the TL circuit and the unit cell of each layer. This implies a large reduction in the computation time.

- The TL equivalent circuit is solved using the optimization tool of ADS. This tool has the advantage of optimizing the shunt elements of all the layers at one simulation and avoid the heavy computation of the transmission matrices of each layer then the scattering transmission parameters and finally optimize the shunt elements using quasi-Newton algorithm as used in [13].
- SA algorithm is used to optimize the physical dimensions of the MLs. This algorithm optimizes the dimensions and minimize the absolute values of the differences between the theoretical admittances introduced by Chu and Lee in [14], and their lumped-element counterparts optimized by solving the TL circuit using the optimization tool of ADS. This method avoids the full-wave analysis of the unit cell and all the steps described before which are used in [13].

3. Meander Line Polarizer: Optimization and validation Procedures

3.1. TL Equivalent Circuit Model

As summarized in [4], the incident electric field $\boldsymbol{E_{inc}}$ will be decomposed into two orthogonal components $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$. One component passes through a structure equivalent to a broad-band shunt-inductive filter while the other passes through a broad-band shunt-capacitive filter. The phase shift through either filter should have almost the same slope, so that if the differential phase shift is 90º at one frequency in the common passband, it remains close to 90º everywhere in the common passband.

The TL equivalent circuit of the N-layers ML polarizer excited by an incident wave along *y*-axis (Figure 1) can be represented as shown in the Figure 3 where η_0 is the characteristic impedance of the vacuum ($\eta_0 = \sqrt{\mu_0/\varepsilon_0}$) with μ_0 and ε_0 the permeability and the permittivity of vacuum, respectively. The parameters h_{d1}, h_{d2}, …, h_{dN} and $\varepsilon_{r_{d1}}$, $\varepsilon_{r_{d2}}$, …, $\varepsilon_{r_{dN}}$ are respectively the thickness and the relative permittivity of the 1st, 2nd, …, *N*th layer. The parameters h_{s1}, h_{s2}, …, h_{sN} and $\varepsilon_{r_{s1}}$, $\varepsilon_{r_{s2}}$, …, $\varepsilon_{r_{sN}}$ are respectively, the thickness and the relative permittivity of the 1st, 2nd, …, *N*th separator. The L_{eq_1}, L_{eq_2}, …, L_{eq_N} and C_{eq_1}, C_{eq_2}, … , C_{eq_N} are respectively the lumped inductance and capacitance equivalents of the 1st, 2nd, …, *N*th periodic strip lines on each layer.

The input parameters of the TL model of ML polarized are: the frequency bandwidth (center frequency f_0), the number of layers and the relative permittivity and thickness of each dielectric substrate and separator. This model must be solved by optimizing the shunt components L_{eq_1}, L_{eq_2}, …, L_{eq_N} and C_{eq_1}, C_{eq_2}, …, C_{eq_N} which verify the maximum transmission coefficients (S_{21} and S_{43}) in term of magnitude (~0dB) while

attaining 90^0 differential phase shift between both components for the CP properties.

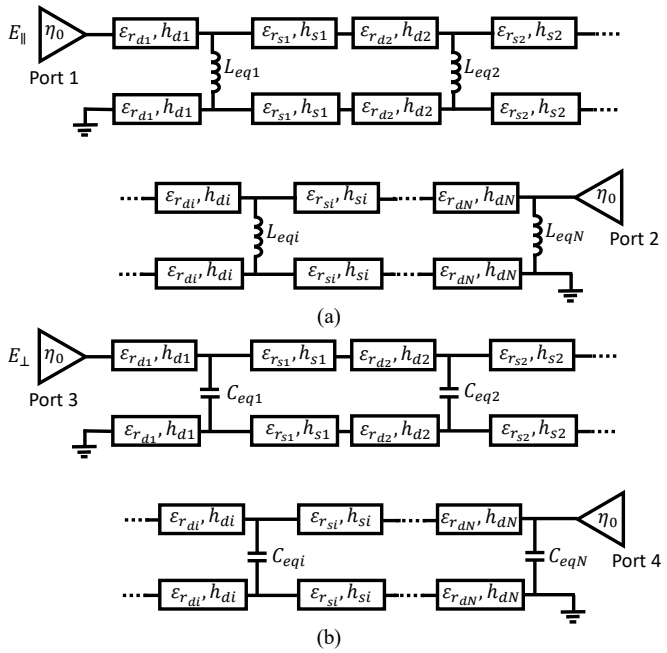


Figure 3: Transmission Line equivalent circuit for N-Layers ML polarizer observed by (a) the parallel wave $\boldsymbol{E}_\parallel$ and (b) the perpendicular wave $\boldsymbol{E}_\perp$.

3.2. Simulated Annealing Algorithm

Simulated Annealing is a mathematical equivalent of the thermodynamic annealing. The energy of the particle in thermodynamic annealing process can be compared with the cost function to be minimized in optimization problem. The particles of the solid can be compared with the independent variables used in the minimization function. The Simulated Annealing Algorithm is detailed in [15].

The MLs dimensions must be optimized to verify that the impedance equivalent of each layer is equal to the one found by solving the TL circuit model.

Chu and Lee in [14], presented rigorous empirical formulas of equivalent admittances for parallel ($Y_\parallel$) polarization $\boldsymbol{E}_\parallel$ and perpendicular ($Y_\perp$) polarization $\boldsymbol{E}_\perp$ of the ML, dependent on the dimensions a, b, h, and w (Figure 4).

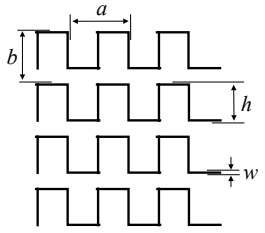


Figure 4: Polarizer's dimensions

The SA algorithm is used to optimize the best values of the ML dimensions of each layer i by making the admittances values ($1/j\omega L_{eq_i}$ and $j\omega C_{eq_i}$), found using the TL circuit model, equal to the ones given by Chu and Lee ($Y_{\parallel_i}(a_i, b_i, h_i, w_i)$ and $Y_{\perp_i}(a_i, b_i, h_i, w_i)$). In other word, for each layer i, the dimensions a_i, b_i, h_i, w_i are optimized to minimize the difference between the admittances ($\delta_i \approx 0$ and $\delta_i' \approx 0$) as shown in the equation (1).

$$\begin{aligned} &\left|Y_{\parallel_i}(a_i, b_i, h_i, w_i) - \frac{1}{j\omega L_{eq_i}}\right| \le \delta_i \\ &\left|Y_{\perp_i}(a_i, b_i, h_i, w_i) - j\omega C_{eq_i}\right| \le \delta_i' \end{aligned} \quad \text{for } i = \{1, \dots, N\}. \quad (1)$$

The optimization procedure is summarized in Figure 5.

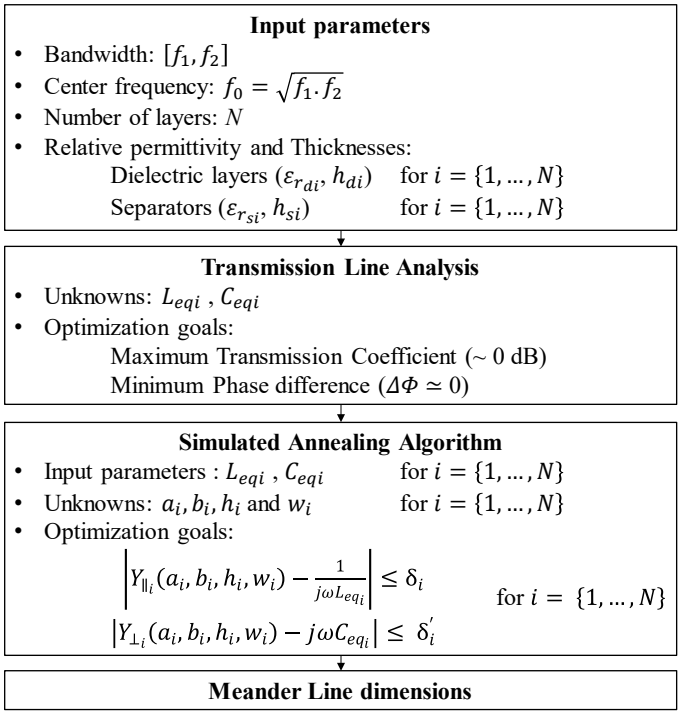


Figure 5: Optimization Procedure Summary

4. Case Study: 6-18GHz Four Layers ML Polarizer

A four layers ML polarizer covering the bandwidth from 6 GHz to 18 GHz is studied in this section.

4.1. Polarizer Optimization

The optimization procedure starts by defining the input parameters which are the frequency bandwidth [6-18GHz], center frequency f_0 = 10.4GHz, number of layers N = 4, relative permittivity and thickness of the dielectric substrates ($\varepsilon_{r_{di}}, h_{di}$) and separators ($\varepsilon_{r_{si}}, h_{si}$) for $i = \{1, \dots, 4\}$.

To simplify the TL equivalent model, the equivalent circuit was studied with one-quarter wavelength apart ($\lambda/4$) air separators and without considering the substrate effect since a very thin substrate ($h_{di} = 0.127mm$) with low permittivity and low loss (Rogers RT/duroid 5880 ($\varepsilon_{r_d} = 2.2, Tan\delta_d = 0.0009$)) is used.

The choice to have the same MLs dimensions (i.e. same L_{eq} and C_{eq}) for each layer was also made to make the system integration simple and flexible. The TL equivalent circuit of the four layers ML polarizer is shown in the Figure 6.

The optimization tool of ADS is used to find the best values of the equivalent inductance and capacitance to verify the maximum transmission coefficients (S_{21} and S_{43}) in term of magnitude (~ 0dB) and 90^0 differential phase shift between them

over the entire bandwidth. The optimum values are: C_{eq}=3 fF and L_{eq} = 0.494 pH.

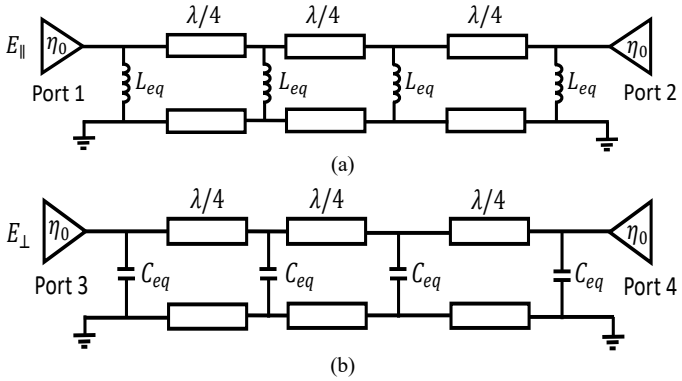
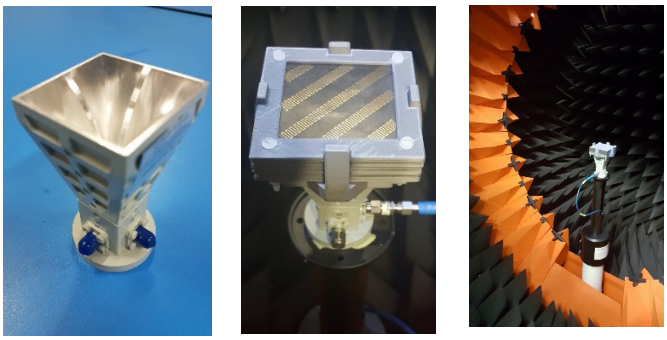


Figure 6: TL circuit model of the four layers Meander Line polarizer observed by the (a) parallel wave ($\boldsymbol{E}_{\parallel}$) and the (b) perpendicular wave ($\boldsymbol{E}_{\perp}$).

The ML dimensions (a, b, h, and w) are optimized using the SA algorithm, as detailed previously, by making the admittances values ($1/j\omega L_{eq}$ and $j\omega C_{eq}$), found using the TL circuit model (Figure 6), as close as possible to the ones given by Chu and Lee (1). The optimum values found are a = 2.17mm, b = 13.08mm, w = 0.446mm and h = 5.151mm.

4.2. Polarizer test with 6-18GHz Horn Antenna

To verify the performance of ML polarizer, a wideband horn antenna (6 to 18 GHz) shown in figure 7(a) is used. The four ML layers of the polarizer were fabricated. A mechanical support was designed and realized using a 3D printer to maintain the quarter wave length distance between each layers and the antenna. The picture of the polarizer attached to the antenna is shown in figure 7(b). Figure 7(c) shows the antenna with the ML polarizer under test in the near field anechoic chamber StarLab (by Microwave Vison Group), available in IETR.



(a) (b) (c)

Figure 7: Picture of (a) horn alone - (b) horn with ML-Polarizer – (c) horn with ML-Polarizer under test

The measurements were performed for the horn antenna alone and with the ML polarizer excited in vertical polarization. The reflection coefficients of the horn antenna alone (without the ML polarizer) and with the ML polarizer are plotted in Figure 8.

Both reflection coefficients are very close, which proves that the ML polarizer doesn't mismatch the antenna except at the around 6GHz, where the reflection coefficient is measured around -6.5dB.

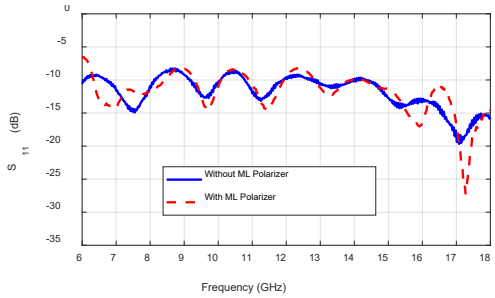


Figure 8: Reflection coefficients of the horn antenna with and without the ML polarizer

The co-polarization and cross-polarization of the Electrical filed (E_θ and E_ϕ) are plotted for different frequencies in the figures 9-15.

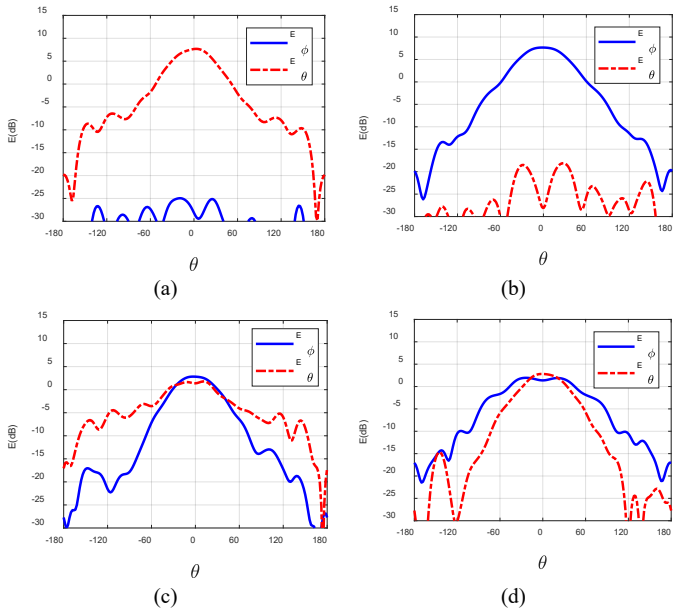


Figure 9: Measured E_θ and E_ϕ at 6 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

In figure 9, the E_θ and E_ϕ components of horn antenna alone and with ML polarizer are plotted at 6 GHz in both planes ($\phi = 0^0$ and $\phi = 90^0$). The results seem to show that the horn antenna initially excited for a vertical polarization (Figure 9 (a) and (b)) is radiating a circular polarized wave when it is stacked with the ML polarizer (Figure 9 (c) and (d)). At the boresight, the difference between the co and cross components of the horn antenna measured alone is more than 28 dB while it reduces to 1dB when the horn antenna is measured with the ML polarizer. The measured Axial Ratio (AR) of the horn antenna with the ML polarizer at this frequency is around 6dB (Figure 16). Therefore, the ML polarizer is converting the linear polarization waves into elliptical polarized ones. The measured insertion losses are around 2.5dB (Figure 17) which is due to the mismatch of the antenna with the ML polarizer at this frequency (Figure 8).

The E_θ and E_ϕ components of the horn antenna alone and with ML polarizer are plotted in figure 10 – 14 for the frequencies from 8GHz to 16GHz. The vertical polarized waves of the horn antenna alone are perfectly converted to circularly polarized ones when it is stacked to the ML polarizer with the maximum AR of

2.9 dB at 8GHz (Figure 16). The insertion losses are less than 1.4 dB with the minimum of 0.24 dB at 14 GHz (Figure 17).

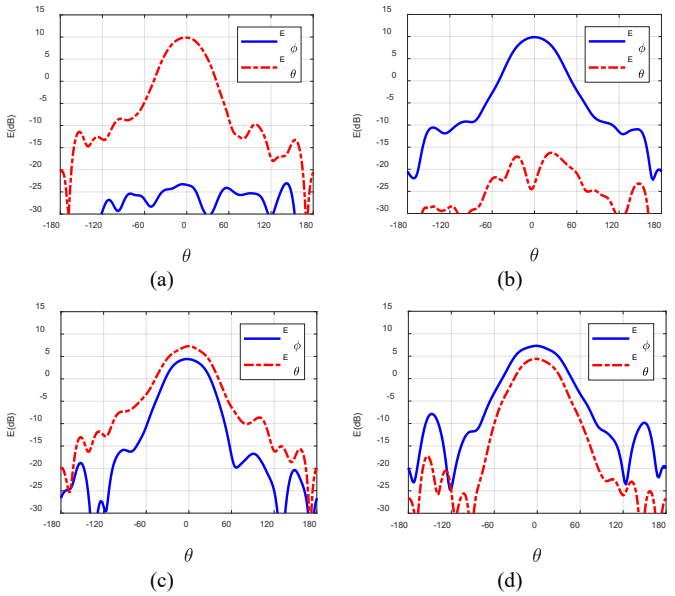


Figure 10: Measured E_θ and E_ϕ at 8 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

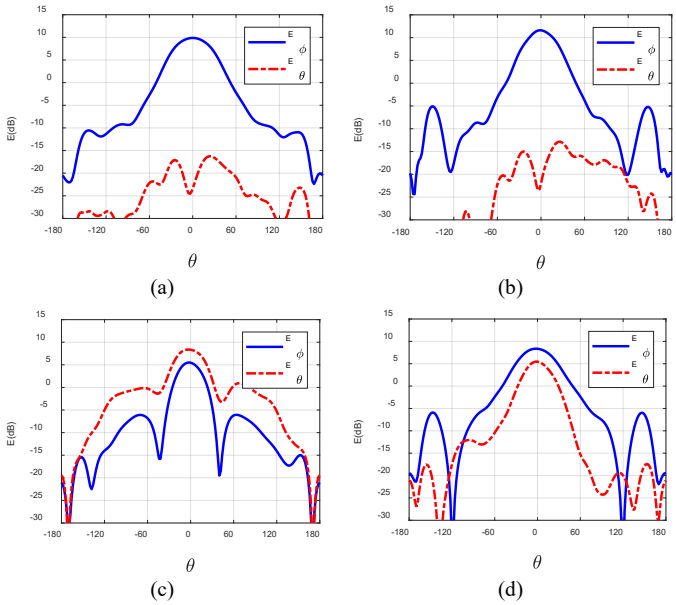


Figure 11: Measured E_θ and E_ϕ at 10 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

The measured results at 18 GHz are plotted in figure 15. At the plane $\phi = 0^0$, some ripples are observed in the pattern of the co-component when the antenna is measured with the ML polarizer. This comportment is not reproduced at the other cut plane ($\phi = 90^0$) where the directive radiation pattern shape is preserved. The measured Axial Ratio (AR) of the horn antenna with the ML polarizer at this frequency is around 5.4 dB (Figure 16). Therefore, the linearly polarized waves of the horn antenna alone are converted to elliptical polarized waves. The measured insertion losses are around 0.6 dB (Figure 17).

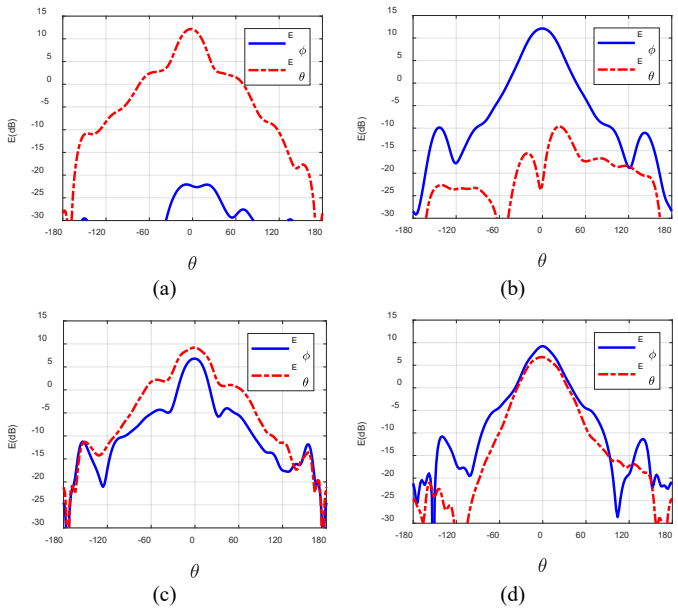


Figure 12: Measured E_θ and E_ϕ at 12 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

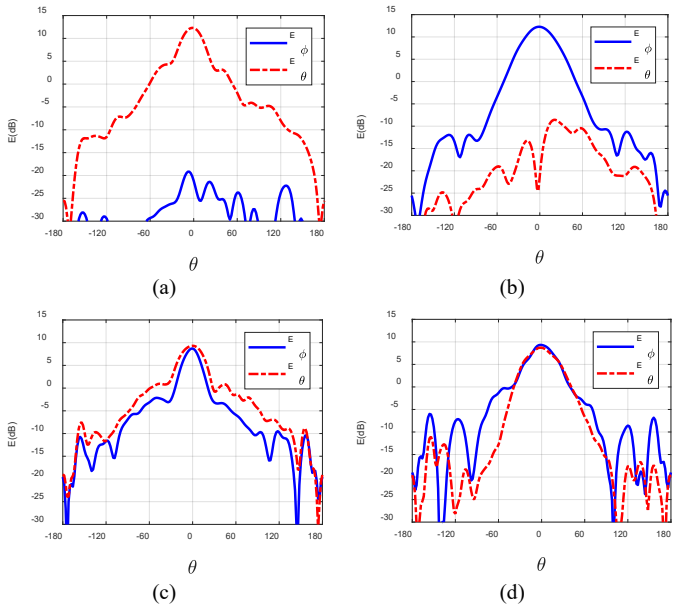


Figure 13: Measured E_θ and E_ϕ at 14 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

The measured values of the AR at boresight are reported in figure 16 versus frequency for the horn antenna with the ML polarizer. The AR is less than 3dB at the frequency band between 8 GHz and 16 GHz. Therefore, the ML polarizer converts the linear polarization of the horn antenna to a circular one. In both ends of the frequency band (6 GHz and 18 GHz), the ML polarizer converts the linear polarization of the horn antenna to an elliptical one.

The total electrical fields (E_{tot}) values at boresight are reported in figure 17 versus frequency for the horn antenna alone and with the ML polarizer. This figure shows clearly that the E_{tot} of the horn

antenna with the ML polarizer is following the same allure as the E_{tot} of the horn antenna alone variation with some insertion losses.

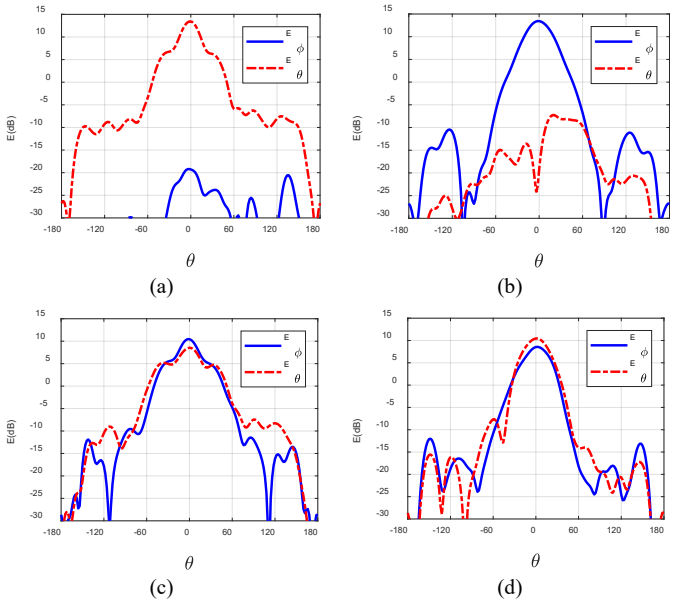


Figure 14: Measured E_θ and E_ϕ at 16 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

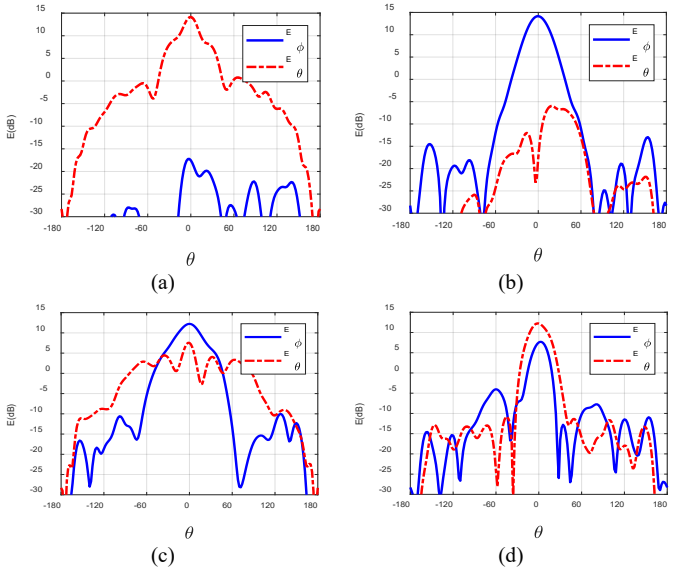


Figure 15: Measured E_θ and E_ϕ at 18 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

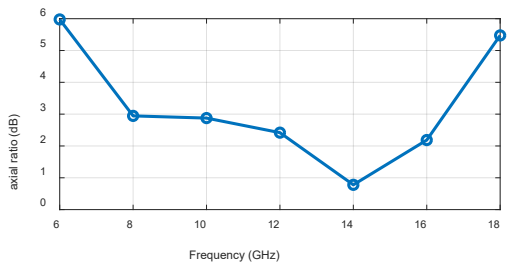


Figure 16: Measured Axial Ratio (***AR***) versus frequency for the horn antenna with the ML polarizer

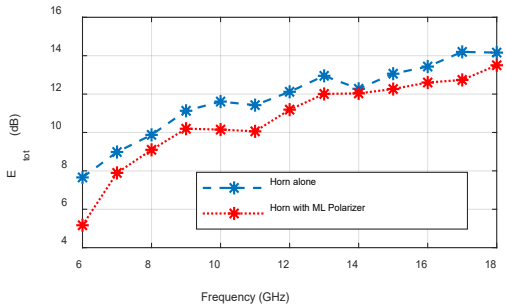


Figure 17: Measured total electrical field (***E_{tot}***) versus frequency for the horn antenna alone and with the ML polarizer

In almost the entire bandwidth, except at 6 GHz, the insertion losses are less than 1.5dB with the minimum of 0.24 dB at 14 GHz. The maximum insertion losses (2.5dB) are measured at 6 GHz which is due to the mismatch of the horn antenna with the ML polarizer.

5. Conclusion

An optimization Method of Wideband Multilayer Meander-Line Polarizer using a Semi-Analytical approach is described in this paper. The originality of this method lies in the optimization of: i) values of the shunt equivalent inductance L_{eq} and capacitance C_{eq}, determined by the ADS software - ii) ML dimensions (a, b, h and W) using the SA algorithm. This method was applied to optimize a four layers wideband (6 – 18 GHz) ML polarizer as study case.

The optimized polarizer was manufactured and a wideband Horn antenna covering the same bandwidth was used for the testing. The test consists of the measurements of the horn antenna vertically polarized without and with the ML polarizer. Good measured results were observed where the linear polarization of the antenna was converted to a purely circular one in almost the entire frequency band (8 GHz – 16 GHz, corresponding to an octave). In both ends of the frequency band (6 GHz and 18 GHz), an elliptical polarization was achieved with a maximum axial ration of around 6dB.

At 6GHz, the axial ratio is around 1dB but the insertion losses are around 2.5dB which is due to the mismatch of the horn antenna with the ML polarizer at this frequency. At 18 GHz, the axial ratio is around 5dB and the insertion losses are around 0.6 dB.

The future work will be devoted to the study of five layers ML polarizer to improve the performances, and especially at both ends of the frequency band.

References

[1] W. Abdouni, M. S. Khan, A. Konstantinidis "Design of a Wideband Multilayer Meander-Line Polarizer (6 – 18 GHz) using a Semi-Analytical Method," in 48th European Microwave Conference (EuMC), Madrid Spain, 2018, https://doi.org/10.23919/EuMC.2018.8541675

[2] R. A. Marino, "Accurate and efficient modeling of meanderline polarizers," Microw. J., 41 (11), 22-34, Nov. 1998

[3] J. F. Zurcher, "A meander-line polarizer covering the full E-band (60-90 GHz)," Microwave Opt. Technol. Lett., 18, 320-323, Aug. 1998. https://doi.org/10.1002/%28SICI%291098-2760%2819980805%2918%3A5<320%3A%3AAID-MOP3>3.0.CO%3B2-A

[4] L. Young, L.A. Robinson, and C.A. Hacking "Meander-Line Polarizer," IEEE Transactions on Antennas and Propagation, 21(3), 376-378, 1973. https://doi.org/10.1109/TAP.1973.1140503
[5] J. J. Epis, "Broadband antenna polarizers," US Patent 3754271, Aug. 1973.
[6] T. L. Blackney, J. R. Burnett, and S. B. Cohn, "A design method for meander-line circular polarizers," 22nd Ann. Antenna Symp., 1-5, Oct. 1972.
[7] C. Terret, J. R. Level, and K. Mahdjoubi, "Susceptance computation of a meander-line polarizer layer," IEEE Transactions on Antennas and Propagation, 21, 1007-1011, Sep. 1984. https://doi.org/10.1109/TAP.1984.1143456
[8] S. W. Lee, and G. Zarrilo, "Simple formulas for transmission through periodic metal grids or plates," Antenna Application Sym., University of Illinois, Urbana, IL, Sponsored by RADC (EEA) Electromagnetic Sciences Division, Hanscom AFB, 1983. https://doi.org/10.1109/TAP.1982.1142923
[9] A. K. Bhattacharyya, and T. J. Chwalek, "Analysis of multilayered meander line polarizer," Int. J. of Microwave and Millimeter-wave Computer-aided Eng., 442-454, 1997. https://doi.org/10.1002/(SICI)1522-6301(199711)7:6%3C442::AID-MMCE6%3E3.0.CO;2-P
[10] M. A. Joyal, and J.-J. Laurin, "Analysis and design of thin circular polarizers based on meander lines," IEEE Transactions on Antennas and Propagation, 60, 3007-3011, Sep. 2012. https://doi.org/10.1109/TAP.2012.2194659
[11] C. C. Chen, "scattering by a two-dimensional periodic array of conducting plates," IEEE Transactions on Antennas and Propagation, 18, 660-665, 1970. https://doi.org/10.1109/TAP.1970.1139760
[12] R. S. Chu, and K. M. Lee, "Radiation impedance of a dipole printed on periodic dielectric slabs protruding over a ground plane in an infinite phased array," IEEE Transactions on Antennas and Propagation, 35, 13-25, 1987. https://doi.org/10.1109/TAP.1987.1143966
[13] M. Letizia, B. Fuchs, C. Zorraquino, J.-F. Zurcher, and J. R. Mosig, "Oblique incidence design of meander-line polarizers for dielectric lens antennas," Progress In Electromagnetics Research B, 45, 309- 335, 2012. https://doi.org/10.2528/PIERB12080807
[14] RUEY-SHI CHU, KUAN-"I LEE, "Analytical Model of a Multilayered Meander-Line Polarizer Plate with Normal and Oblique Plane-Wave Incidence," IEEE Transactions on Antennas and Propagation, 35(6), 652 – 661, 1987. https://doi.org/1987 10.1109/TAP.1987.1144158
[15] E.S. Gopi, "Algorithm Collections for Digital Signal Processing Applications Using Matlab", Mathematics\\Wavelets and signal processing, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 133-140 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Effects of Using Fuzzy Material Handling Inputs in the Genetic Algorithm for Machine Layout

Wanwanut Boongsood*, Chiranuwat Jadram

School of Manufacturing Engineering, Suraneree University of Technology, 30000, Thailand

ARTICLE INFO

Article history:
Received: 28 April, 2019
Accepted: 26 June, 2019
Online:17 July, 2019

Keywords :
Machine layout
Fuzzy numbers
Genetic algorithm

ABSTRACT

This study introduces the implementation of fuzzy set theory to solve machine layout design issues, in order to handle vague information, using a genetic algorithm with tournament selection as the selection operator. The material handling inputs, including frequency and volume of materials that move between machines, were the parameters regarded as fuzzy numbers. The experimental results came from 2 case studies in a manufacturing system. In the first case, examining the difference in shapes of the triangular membership functions of input data, the total distances were reduced from 38.45 m to 29.72 m, a 22.71% reduction in distance. In the second case, examining the uncertainty of fuzzy data by an expert, the total distances were reduced from 103.45 m to 82.45 m, a 20.03% reduction in distance. It was found that given the uncertainty in input data, a shorter total material handling distance might not give a lower cost. The selection operator of tournament selection can compete effectively to converge to near the optimum solution. This can, therefore, be an alternative technique in managing manufacture.

1. Introduction

Facility layout design plays a large part in business competition. Facilities can be machines, workstations, construction sites or departments, depending on the type of business. Costs can be reduced by having an appropriate layout. However, it is a complex problem. One of the essential goals of the design is to minimize total material handling costs, which are affected by many factors, such as types of materials, processing and production flow. Especially in the manufacturing industries, some produce not just one product but a variety of goods.

Moreover, in reality, the factors that affect material handling costs are vague and fuzzy. In a manufacturing system, they are dependent generally on the seasons, product variety and business growth. In order to engage with this ambiguity, there are levels of the vagueness of information, and the decision is usually made under uncertainty [1]. Fuzzy sets theory is a suitable concept to deal with this situation [2-6]. Therefore, this paper extends the research to understand the effects when the values in fuzzy sets are predictable and if they are extremely vague.

In many manufacturing industries, there can be several products manufactured in a factory, and each product can have different routes through machines or workstations. Production efficiency depends highly on this kind of problem in production planning [7]. Production costs can be reduced by 10–30% when the facility layout is more effective [8]. Shorter handling distances of materials' flow between workstations are required for faster transfer times within a factory floor.

The inputs necessary to design a facility layout for an actual production system are usually imprecise. Besides, the issue is considered a complicated situation because of the choice of products, materials, processes and other factors involved with the business. These factors are also time-dependent, such as customer orders and product growth, so the cost of material handling between machines usually changes. Different methods [9-12] have been presented for solving these kinds of problems; however, uncertain and ambiguous data inputs have been paid little attention [13]. Therefore, the fuzzy set theory is applied in this paper.

Many researchers have proposed different ways to solve this complex problem. In choosing the solution, the aim should be to achieve maximum efficiency, and there should be a simple method for finding the answer. Drira, et al. [14] gathered techniques and methods used for solving machine layout problems. Several methods can be used in the analysis of these problems, such as branch and bound, dynamic programming, simulated annealing, tabu search, genetic algorithms, ant colony, intelligence methods.

To find a solution with accurate answers to a complex problem, the costs of solving are very high [15]. Thus, it is better, rather, to have a method that can solve the vague problem obtaining an

*Wanwanut Boongsood, Email: wanwanut@sut.ac.th

acceptable answer with less time and cost than that with high accuracy and high cost. Since mathematical modeling techniques used to solve the facility layout design problem are complicated, challenging, and time-consuming, the issues require a powerful technique [5].

For the problems of machine layout design, the genetic algorithm is one popular method for finding answers, or approximate answers [16]. The Simple Genetic Algorithm (SGA) applies the principle of evolution and natural selection, by which the best individuals are selected to generate answers for the next generation. To calculate the answer by SGA is fundamental to solving the optimization problem. Many studies adopting this method have been released [14].

Conventionally, Genetic Algorithm (GA) includes these steps: generate an initial population, select high potential chromosomes to be parents, crossover and mutate parents' chromosomes to produce offspring chromosomes. Since there are a couple of genetic operators used in GA steps, improving the results fulfilling genuine applications can be achieved by adopting some of these operators (see [17-19] and [20]). The action of selecting high potential chromosomes to be parents called the selection process is the operation that has been broadly studied. Tournament Selection (TOS) and Roulette Wheel Selection (RWS) are general ways for making a selection. It was suggested that TOS is more effective than RWS [21-23]; the consequence is that a reasonable answer can be reached faster.

The research of Vitayasak [24] on fuzzy methods is interesting. The fuzzy approach can help manage uncertainty in practice, which may arise from a lack of information, not having enough data for the application, or the data not yet being available [25]. The experience of experts in reasoning or decision making in the human situation can help analysis and help manage uncertainty. The fuzzy theory is a theory of uncertain boundaries. Currently, it is widely used in the study of problems containing uncertain information [26].

It is important for machine layout design decision making and an efficient way to deal with inaccurate and vague information. Kritwattanakorn, et al. [27] used the RWS method for the selection approach in GA and suggested not to apply mutation for machine layout problems. However, based on the suggestions given by some research [21-23], the TOS method seems to be superior in finding a solution qualitatively (for instance, minimizing the total material handling cost) and obtain answers faster and closer to the optimum. In our paper presented initially to a conference [28], we related to the research of Kritwattanakorn, et al. [27] that not to apply mutation, used TOS for selection, and showed how to deal with the vagueness of information and assist solving problems taken from earlier research works [25, 26]. Our proposed method performed corresponding to the work presented by Jinghui, et al. [21] that using TOS is competitive in terms of solution quality and achieving minimization faster.

However, realistically, there are also levels of the vagueness of information. To understand the effects when the values in fuzzy sets are predictable and if they are extremely vague, in this study, the tournament method was used in the selection process for the genetic algorithms solving the machine layout problem. Material handling information was considered employing fuzzy uncertainty. An original layout and new layouts constructed from three GA based approaches were compared for 2 case studies. The first was the case where fuzzy information was thought predictable, examining different forms of skewed shapes. In the second case, the fuzzy data was obtained from a process expert in a new factory when asked to give the possible values for each variable. Thus the fuzzy sets were utterly formless. Every approach in the study was programmed and run with Visual Basic for Applications in Microsoft Excel for determining the layout, distance, and cost.

2. Methods

The fitness function, which is a particular type of objective function, is required for the GA algorithm to execute optimization. In this paper, the objective (1) is to minimize the total costs of material handling. This objective has been chosen by many researchers [29-31] in determining facility layouts. The function can be written as

$$Minimize\ C = \sum_{i=1}^{M} \sum_{j=1, i \neq j}^{M} f_{ij} c_{ij} d_{ij} \quad (1)$$

$$f_{ij} = \mathrm{F_{ij}} \times \mathrm{V_{ij}} \quad (2)$$

where C represents the total cost of the material handling system.

f_{ij} is the flow of materials in the production system.

c_{ij} is the unit cost of material handling between machines *i* and *j*.

d_{ij} is the rectilinear distance between the centroids of machines *i* and *j* (see for instance Fig. 1, where the distance between machines B to E is equivalent to $\Delta x + \Delta y$).

M is the total number of machines.

F_{ij} is the frequency flow between machines *i* and *j*.

V_{ij} is the volume flow between machines *i* and *j*.

In the calculation, minimize C is the objective, the variables used in consideration in this paper, $\mathrm{C_{ij}}$ and $\mathrm{d_{ij}}$ are considered as crisp numbers, and f_{ij} is information that is uncertain.

In general, $f_{ij} = \mathrm{F_{ij}} \times \mathrm{V_{ij}}$. However, in this study the information was considered both as if it is fuzzy to compare with the conventional idea, and as if it is crisp. Thus, when the frequency and the volume of material flows are considered as fuzzy numbers, they will be represented as $\mathrm{F'_{ij}}$ and $\mathrm{V'_{ij}}$, respectively, and the material handling flow as a fuzzy number is f'_{ij} (3).

Since the total cost, C, is a crisp number, f'_{ij} needs to be converted back to be an exact number. This method is called defuzzification, see (4).

$$f'_{ij} = \mathrm{F'_{ij}} \otimes \mathrm{V'_{ij}} \quad (3)$$

$$\mathrm{Defuzzification}(f'_{ij}) = f_{ij} \quad (4)$$

However, in some cases, the materials frequency and volume flow information are exact numbers, so they need to be converted

to a fuzzy number. This method is called fuzzification (5) when X is a number.

$$Fuzzification(X_{ij}) = X'_{ij} \quad (5)$$

The framework:

- Data input: the data input includes the production routing of each product (P_I, P_{II}, P_{III} ...), number of machines (N), sizes of machines (width, w_n, and length, l_n), shop floor area (W is width and L is length), frequency of material flow per round (F'_{ij}), volume flow per unit (V'_{ij}), unit material handling cost between locations of machines (c_{ij}) and distance between machines (d_{ij}), population size (v) and the number of generations to simulate (G).
- Step 1) Fuzzification: The uncertain flow frequency of materials, f_{ij} which is the function of F_{ij} and V_{ij} , was converted to fuzzy numbers (f'_{ij}).
- Step 2) Fuzzy computing: Fuzzy arithmetic was used.
- Step 3) Defuzzification: In order for the next step following the GA approach, crisp numbers are essential for calculation. The fuzzy number f'_{ij} was required to be transformed into crisp numbers.
- Step 4) Genetic algorithm: The sub-processes in the adapted GA are shown in Fig. 2.
- Stop: The calculation stops when conditions are met as in Fig. 2.

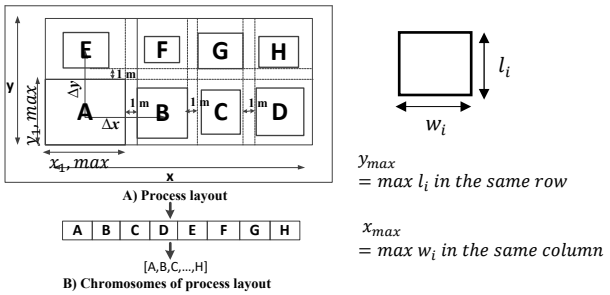


Figure 1: Chromosomes representation.

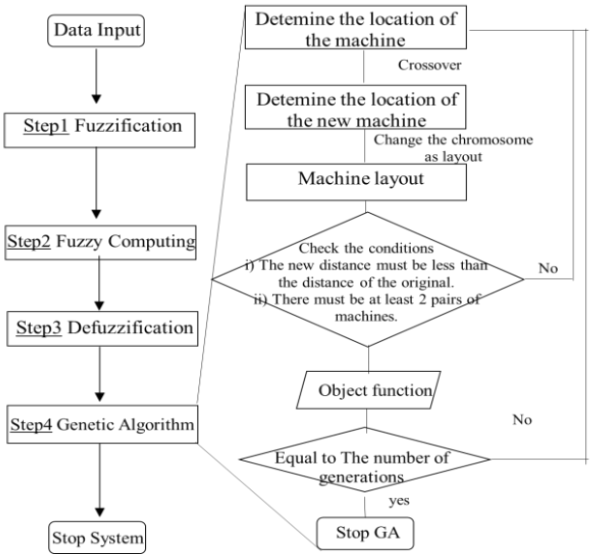


Figure 2: The framework.

2.1. Fuzzy number

Fuzzy numbers are the foundation of fuzzy set theory [2]. They are an extended form of real numbers that do not refer to one single value, but a set of possible crisp numbers. The fuzzy technique is used to represent data with uncertainties.

2.1.1. The shape of membership: Triangular fuzzy numbers (TFNs), graphically represented as Fig. 3, are used in this research. They are suitable because of their computational advantages and commonly used for subjective description [6].

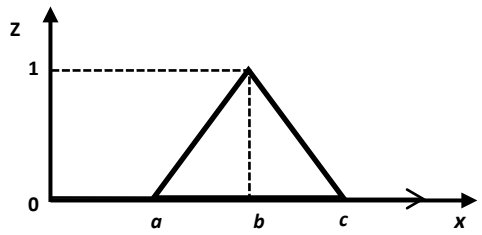


Figure 3: Triangular fuzzy number.

The membership function of a TFN can be explained by:

$\mathbf{Z(x)}$ is the grade of membership,

where $Z(x) > 0$ when $a < x < c$

$Z(x) = 0$ when $x \leq a$ *or* $x \geq c$

$Z(x) = 1$ when $x = b$ *"highest grade of membership at the modal value."*

Therefore,

a = the lowest possible value,

b = the most likely possible value,

c = the highest possible value in the fuzzy set.

2.2.2. Fuzzification: First, it is necessary to convert the input information to fuzzy numbers and calculate them using fuzzy operators. For input data F_{ij} and V_{ij}, the fuzzy sets of F'_{ij} and V'_{ij} are

$$F' = [\mathrm{F_L}, \mathrm{F_M}, \mathrm{F_H}] \quad (6)$$

$$V' = [\mathrm{V_L}, \mathrm{V_M}, \mathrm{V_H}] \quad (7)$$

$\mathrm{F_L}, \mathrm{F_M}$ and $\mathrm{F_H}$ are the lowest, most likely, and highest possible values of the frequency flow and $\mathrm{V_L}, \mathrm{V_M}$ and $\mathrm{V_H}$ are the lowest, most likely and highest possible values of the volume flow, and their grades of membership are between 0 and 1. If, for example, the total material flow frequency of all products is possibly 20 times and the total material flow volume is approximately 5 units each time, the material flow F could be a set of numbers with the most likely value of 18 such as [16, 18, 19] as well as V could be [3, 5, 6].

2.1.3. Fuzzy arithmetic: The operations [32] can be indicated by

Fuzzy addition $\oplus$:

$$\mathrm{F} \oplus \mathrm{V} = [\mathrm{F_L} + \mathrm{V_L} \ , \ \mathrm{F_M} + \mathrm{V_M} \ , \ \mathrm{F_H} + \mathrm{V_H}] \quad (8)$$

Fuzzy multiplication $\otimes$:

$$F \otimes V = [\min(F_L V_L, F_L V_H, F_H V_L, F_H V_H), F_M V_M, \max(F_L V_L, F_L V_H, F_H V_L, F_H V_H)] \quad (9)$$

With the above example values

$$F \otimes V = \begin{bmatrix} \min(16\times 3,\ 16\times 6, 19\times 3,\ 19\times 6), 18\times 5, \\ \max(16\times 3,\ 16\times 6, 19\times 3,\ 19\times 6) \end{bmatrix}$$

$$= \begin{bmatrix} \min(48, 96, 57, 114), 90, \\ \max(48, 96, 57, 114) \end{bmatrix}$$

$$= [48, 90, 114]$$

2.1.4. Defuzzification: Before entering the GA procedure, the fuzzy numbers for the material flow must be changed to crisp numbers using the fuzzy weight method [33].

$$f_{ij} \cong f'_{ij} \quad (10)$$

$$f_{ij} \cong \{[(0.5 \times (a + c)) + b]/2\} \quad (11)$$

therefore, $f_{ij} = \{[(0.5 \times (48 + 114)) + 90]/2\}$

$=85.5$

2.2. Genetic algorithm

The objective function (1) is used in this paper to find the optimum value by applying the genetic algorithm. The suggested GA has been modified from the Simple Genetic Algorithm (SGA) [16]. It begins with an initial set of random solutions for the problem being considered. The set of these alternatives is called the population. The population's individuals are called chromosomes. The objective function evaluates the results from the chromosomes. The chromosomes showing high potential is then carried out to be a part of the new generarion. In the crossover phase, a chromosome is chosen, along with the checking conditions, to produce the offspring. The details follow.

2.2.1. Initialization: GA approach begins with creating an initial population, $\boldsymbol{v}$. In our cases for solving machine layout problems, the genes represent the machines, and the chromosomes are for the layouts. Thus, the initial population can be converted from the "prior to improvement" process layout, see Fig. 1 (B) for an example.

2.2.2. Fitness evaluation: This is the procedure for assessing the chromosome's value. The chromosome that is more likely to be selected for the reproduction process would be selected by taking the proportionate fitness into consideration. The objective function of the fitness function, as presented in (12), can be described by total distance, in order to minimize the total costs of the material handling in the machine design issue.

$$\text{Total distance: } d_{ij} = \sum P_I + P_{II} + \cdots + P_k \quad (12)$$

Every material handling distance of product P is computed from the machine routing when each row or column of the machines is 1 m apart, as shown in Fig. 1. Assuming that there are 3 products required to be manufacture in a plant, as shown in Table 2, thus k=3. The machines have to be arranged on the floor plan to determine the total distance derived for a chromosome. The machines were arranged onto a certain floor area of W × L. The difference between the number of columns and rows of the machines was 1 at most.

2.2.3. Selection: This step decides which chromosomes should be parents in the reproduction process to generate offspring for the next generation.

Tournament selection (TOS): The simple tournament selection method [14] was used in this research. First, it divides the population into groups. Next, in each group, n chromosomes are randomly selected. The next generation of parents shall then be assigned from the chromosomes that make the most efficient solution in their groups. There were 2 groups in a tournament in this work. Each group had 3 chromosomes (n = 3), then the best chromosomes in each group were chosen as the next generation of parents.

2.2.4. Crossover: A new solution, namely offspring, is created by the crossover process. In this work, a crossover point, as shown in Fig. 4, was randomly chosen. After the crossover process, 2 conditions were checked. For condition 1, at least 2 pairs of machines in the chromosome must be operated side-by-side. For condition 2, after simulating the machines onto the shop floor area, and the new distance derived from the offspring must be reduced in comparison to the distance achieved by the layout of the parent.

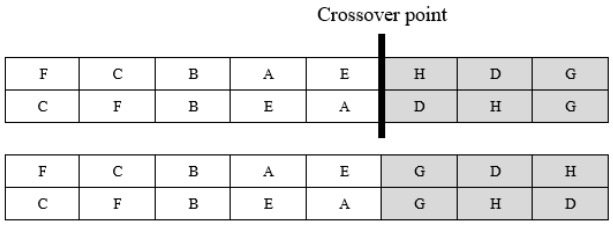


Figure 4: One point crossover process.

2.2.5. Mutation: The mutation process is one in which genes are randomly selected and their positions in the chromosome are swapped to generate additional offspring. It is a traditional step in conventional GA. Kritwattanakorn et al. [27], however, proposed that the mutation should be disregarded in this kind of problem. They have found it produced machines positioned further away while they should be put side by side when considering routing and lastly produced poor outcomes.

2.2.6. Stop: The proposed GA procedure was looped until the given number of generations was met, see Fig. 2.

3. Example

Two layout problems were studied in this research. The details are as follows:

- Case 1: This case uses original information from the manufacture of plywood furniture. There were 3 products and 8 machines in the same shop floor area. Table 2 shows the production routing for each product. Table 3 exhibits the sizes of the machines. The material handling information was initially received as crisp numbers.
- Case 2: This case studied a bus manufacturer. The fuzzy material handling information was given when their

process experts were asked for the values in the fuzzy sets of material flow between machines. There were 12 products and 8 machines. Table 4 shows the production routing of each product, and the machine sizes are shown in Table 5.

Since GA is widely applied for solving machine layout problems, the layouts constructed by different approaches were compared as listed. A summary of GA based approaches is shown in Table 1.

1. Initial layout (unknown approach)
2. SGA [16]
3. Modified GA with RWS and additional conditions (Kritwattanakorn et al. [27])
4. Modified GA with TOS and additional conditions (proposed approach)

Table 1: Summary of GA based approaches.

No	Information	Selection	Crossover	Mutation	Special Conditions
1	Crisp	Unknown (Initial layout)			
2	Crisp	RWS	One point	One point swap	No
3	Crisp	RWS	One point	No	Yes, see 2.2.4
4	Fuzzy	TOS	One point	No	Yes, see 2.2.4

Here the proposed GA approach used Thai Baht (32.50 THB = 1 USD), the number of the initial population, v, was 10, and the number of generations, G, was 100 given to stop the loop. The machine layout design problems were solved using VBA in MS Excel running under Windows 8 OS, on an Intel (R) Core (TM) i7-4500U 1.8GHz Ram 4 GB device.

Table 2: Products and production routing for case 1.

Product	Production Routing
1	A-B-C-D
2	A-E-F-G
3	A-H

Table 3: Sizes of the machines for case 1.

Machine	Dimensions	
	Width (m)	Length (m)
A	2.00	2.00
B	1.35	1.30
C	1.30	1.25
D	1.35	1.30
E	1.22	1.25
F	1.00	1.00
G	1.22	1.25
H	1.20	1.22

4. Results and Discussion

4.1. Case 1: When the level of fuzziness was considered predictable.

In case 1, the costs from layouts exposed by the existing methods including: from the SGA method [16], the adapted GA that applied GA with RWS selection and without mutation [27], and the proposed method in this study using 6 different fuzzy data sets combined with the GA approach (Fig. 5). The case studies used various fuzzy set differences of 0%, 15%, 30%, and 50% fuzziness for the frequency flow and volume flow. The fuzzy set with the difference of 0% represents a crisp number.

Table 4: Products and production routing for case 2.

Product	Production Routing
1	A - F
2	B - C
3	A - F
4	B - D
5	A - G
6	B - H
7	A – G – H
8	B – G – H
9	A – E – F
0	B – E – C
11	A – E – F
12	B – E – D

Table 5: Sizes of the machines for case 2.

Machine	Dimensions	
	Width (m)	Length (m)
A	1.20	1.50
B	1.20	1.50
C	1.70	1.50
D	1.20	1.40
E	1.50	1.20
F	1.30	6.00
G	1.20	1.00
H	3.50	2.00

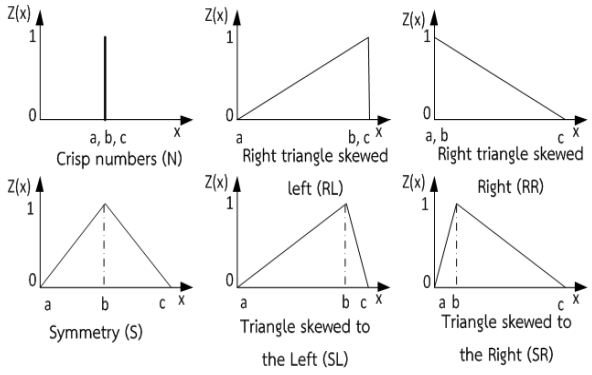


Figure 5: Data sets representing fuzzy numbers.

The results in Fig. 6 show that the total costs vary if the input information is uncertain or fuzzy. If there is the high difference between the lowest and highest possible values, the total costs increase or decrease considerably, such as for the right triangle skewed right (RR), the highest possible value gives a high total cost, but the lowest possible value gives a low total cost. The costs increase or decrease depending on the nature of the skew. If the input information is skewed fully to one side (such as RR and SR), it is likely to make the total costs higher, but if the input is skewed to the minimal side (such as RL and SL), it shows the opposite impact.

Table 6: Results of case 1 with different fuzzy types.

Difference	Fuzzy Type	F', V'	Total Cost (USD)
0% [a]	N	[20, 20, 20], [5, 5, 5]	914.46
15%	S	[18, 20, 22], [4, 5, 6]	923.61
	RL	[17, 20, 20], [4, 5, 5]	841.30
	RR	[20, 20, 23], [5, 5, 6]	1,001.34
	SL	[18, 20, 21], [4, 5, 5]	861.88
	SR	[19, 20, 22], [5, 5, 6]	976.19
30%	S	[17, 20, 23], [4, 5, 6]	955.61
	RL	[14, 20, 20], [3, 5, 5]	781.86
	RR	[20, 20, 26], [5, 5, 7]	1,101.93
	SL	[16, 20, 22], [3, 5, 6]	868.74
	SR	[18, 20, 24], [4, 5, 7]	1,005.91
50%	S	[15, 20, 25], [3, 5, 7]	960.18
	RL	[10, 20, 20], [2, 5, 5]	731.57
	RR	[20, 20, 30], [5, 5, 8]	1,234.52
	SL	[13, 20, 23], [2, 5, 7]	884.74
	SR	[17, 20, 27], [3, 5, 8]	1,067.63

a. Crisp number

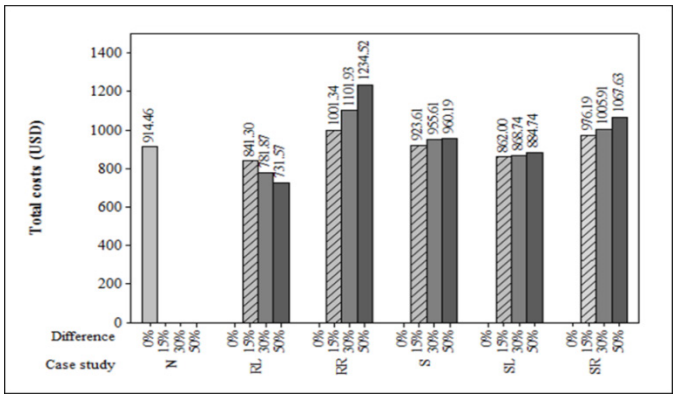


Figure 6: The relationship between various fuzzy data sets and the total cost.

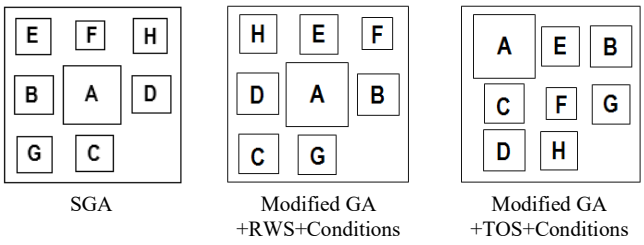
From the findings of this case, it can be seen that the suggested technique shows total costs that increase or decrease depending on the flow of the material handling system in the production. These findings agree with the existing research works [9-12] that information is essential for making a decision on layout planning and design, and fuzzy sets is an appropriate idea in dealing with imprecise information [25, 26].

Table 6 shows the outcomes of the 4 different methods and their layouts for the machines are shown in Fig. 7. Table 6 demonstrates the results of total costs and distances. Before arranging the layout, the total cost initially was 1,179 USD and the total distance was 38.45. The proposed method with a different variation of 15% in fuzzy data input (due to natural associated vagueness) has a total cost of 923.61 USD and a total distance of 29.72 m. It can reduce the total cost by 21.66% and the distance by 22.71%. Both the traditional GA (SGA) method [16] and the Kritwattanakorn et al. method [27] give total costs and distances greater than the proposed method.

Table 7: Comparison of the different approaches in case 1.

Approach	Total cost (USD)	Total Distance (m)
Unknown (initial layout)	1,179.00	38.45
SGA	1,060.92	34.48
Modified GA+RWS+Conditions	978.46	31.80
Modified GA+TOS+Conditions *	923.61	29.72

* Fuzziness at 15 % difference



SGA | Modified GA +RWS+Conditions | Modified GA +TOS+Conditions

Figure 7: Layouts of machines obtained using different methods for case 1.

4.2. Case 2: When the level of fuzziness was extremely vague.

In fact, the shape of fuzzy sets might not be specific. This study used a bus manufacturer as an example. They are a family run business. Machines were laid out based on the experience of the systems expert, not an analysis of manufacturing data. Table 8 shows the frequency and volume of material transport between machines, by letting the expert respond what were the minimum, modal, and maximum values, as for the machine layout design input.

Table 8: Fuzzy data input of case 2.

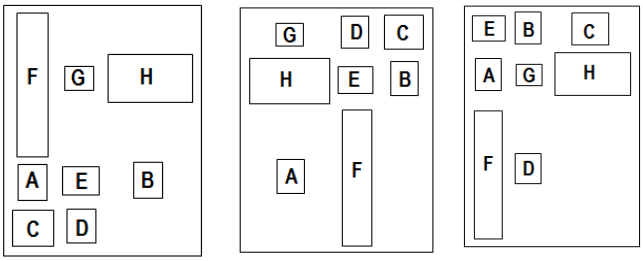
Machine	Frequency flow (F')	Volume flow (V')
M1-M5	[3, 10, 16]	[5, 16, 22]
M1-M6	[6, 8, 14]	[7, 10, 24]
M1-M7	[3, 4, 7]	[8, 12, 18]
M2-M3	[1, 1, 1]	[0, 2, 2]
M2-M4	[3, 6, 8]	[6, 8, 14]
M2-M5	[2, 2, 4]	[7, 8, 13]
M2-M8	[1, 3, 4]	[1, 2, 2]
M5-M3	[3, 6, 8]	[6, 8, 14]
M5-M6	[2, 2, 5]	[6, 11, 15]
M7-M8	[2, 2, 2]	[3, 4, 6]

Table 9 shows the results that the totals cost of material handling can be considerably lowered. The costs generated from the layouts obtained from the method described in this paper were compared competitively with the initial machine layout and 2 existing methods. The machine layouts can be seen in Fig. 8. Table 9 compares the approaches. The total cost and the total distance before arranging being 84.22 USD and 103.10 m, respectively. The proposed method, with fuzzy data input acknowledged from the product experts, has a total cost and distance of 66.40 USD and 82.45 m, correspondingly. The total cost and distance can be reduced by 21.16% and 20.03%.

Even if the total cost obtained from the proposed method was higher than SGA [16] and the method with RWS (Kritwattanakorn et al. method [27]), the total distance was the shortest. This emphasizes the effect of fuzziness of information on decision making. More attention should be paid to uncertainties of data.

Table 9: Comparison of the different approaches in case 2

Approach	Total cost (USD)	Total Distance (m)
Unknown (initial layout)	84.22	103.10
SGA	48.56	92.40
Modified GA+RWS+Conditions	41.24	90.25
Modified GA+TOS+Conditions	66.40	82.45



SGA | Modified GA+RWS+Conditions | Modified GA+TOS+Conditions

Figure 8: Layouts of machines obtained using different methiods for case 2.

5. Conclusion

Acknowledgment

This research was supported by grants funded by the Graduate School of Manufacturing Engineering, Suranaree University of Technology, and the Ministry of Science and Technology, Thailand.

References

[1] L. A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility," Fuzzy Sets and Systems, 100, 9-34, 1999. https://doi.org/10.1016/S0165-0114(99)80004-9

[2] L. A. Zadeh, "Fuzzy Sets," Information and Control, 8, 338-353, 1965. https://doi.org/10.1016/S0019-9958(65)90241-X

[3] İ. Kaya, M. Çolak, and F. Terzi, "A Comprehensive Review of Fuzzy Multi Criteria Decision Making Methodologies for Energy Policy Making," Energy Strategy Reviews, 24, 207-228, 2019. https://doi.org/10.1016/j.esr.2019.03.003

[4] A. L. Guiffrida and R. Nagi, "Fuzzy Set Theory Applications in Production Management Research: A Literature Survey," Journal of Intelligent Manufacturing, 9, 39-56, 1998. 10.1023/a:1008847308326

[5] B. K. Wong and V. S. Lai, "A Survey of the Application of Fuzzy Set Theory in Production and Operations Management: 1998–2009," International Journal of Production Economics, 129, 157-168, 2011. https://doi.org/10.1016/j.ijpe.2010.09.013

[6] S. Balin, "Parallel Machine Scheduling with Fuzzy Processing Times Using a Robust Genetic Algorithm and Simulation," Information Sciences, 181, 3551-3569, 2011. https://doi.org/10.1016/j.ins.2011.04.010

[7] M. Ficko, M. Brezocnik, and J. Balic, "Designing the Layout of Single- and Multiple-Rows Flexible Manufacturing System by Genetic Algorithms," Journal of Materials Processing Technology, 157–158, 150-158, 2004. http://doi.org/10.1016/j.jmatprotec.2004.09.012

[8] J. A. Tompkins, J. A. White, Y. A. Bozer, and J. M. A. Tanchoco, Facilities Planning: Wiley, 2010.

[9] F. M. Defersha and A. Hodiya, "A Mathematical Model and a Parallel Multiple Search Path Simulated Annealing for an Integrated Distributed Layout Design and Machine Cell Formation," Journal of Manufacturing Systems, 43, 195-212, 2017. https://doi.org/10.1016/j.jmsy.2017.04.001

[10] P. B. G. S. N. Murthy, J. Ranganayakulu, K. P. Vidhu, and K. V. Rao, "Heuristic Search Algorithm for the Single-Row Machine Layout in an Automated Manufacturing System," Procedia Technology, 25, 1088-1095, 2016. https://doi.org/10.1016/j.protcy.2016.08.213

[11] Y. Ojaghi, A. Khademi, N. M. Yusof, N. G. Renani, and S. A. H. b. S. Hassan, "Production Layout Optimization for Small and Medium Scale Food Industry," Procedia CIRP, 26, 247-251, 2015. https://doi.org/10.1016/j.procir.2014.07.050

[12] M. A. El-Baz, "A Genetic Algorithm for Facility Layout Problems of Different Manufacturing Environments," Computers & Industrial Engineering, 47, 233-246, 2004. https://doi.org/10.1016/j.cie.2004.07.001

[13] S. Safarzadeh and H. Koosha, "Solving an Extended Multi-Row Facility Layout Problem with Fuzzy Clearances Using Ga," Applied Soft Computing, 61, 819-831, 2017. https://doi.org/10.1016/j.asoc.2017.09.003

[14] A. Drira, H. Pierreval, and S. Hajri-Gabouj, "Facility Layout Problems: A Survey," Annual Reviews in Control, 31, 255-267, 2007. https://doi.org/10.1016/j.arcontrol.2007.04.001

[15] F. Karray, E. Zaneldin, T. Hegazy, A. Shabeeb, and E. Elbeltagi, "Computational Intelligence Tools for Solving the Facilities Layout Planning Problem," in Proceedings of the 2000 American Control Conference. ACC (IEEE Cat. No.00CH36334), 3954-3958, 2000.

[16] J. H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence: MIT press, 1992.

[17] N. M. Razali and J. Geraghty, "Genetic Algorithm Performance with Different Selection Strategies in Solving Tsp," in World congress on engineering 2011, London, U.K., 1134-1139, 2011.

[18] M. Gen, K. Ida, and C. Cheng, "Multirow Machine Layout Problem in Fuzzy Environment Using Genetic Algorithms," Computers & Industrial Engineering, 29, 519-523, 1995. https://doi.org/10.1016/0360-8352(95)00127-M

[19] R. Tavakkoli-Moghaddain and E. Shayan, "Facilities Layout Design by Genetic Algorithms," Computers & Industrial Engineering, 35, 527-530, 1998. https://doi.org/10.1016/S0360-8352(98)00150-8

[20] H. M. Pandey, "Performance Evaluation of Selection Methods of Genetic Algorithm and Network Security Concerns," Procedia Computer Science, 78, 13-18, 2016. https://doi.org/10.1016/j.procs.2016.02.004

[21] Z. Jinghui, H. Xiaomin, Z. Jun, and G. Min, "Comparison of Performance between Different Selection Strategies on Simple Genetic Algorithms," in International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna, Austria, 1115-1121, 2005. 10.1109/cimca.2005.1631619

[22] H. Kılıç and U. Yüzgeç, "Tournament Selection Based Antlion Optimization Algorithm for Solving Quadratic Assignment Problem," Engineering Science and Technology, an International Journal, 22, 673-691, 2019. https://doi.org/10.1016/j.jestch.2018.11.013

[23] H. Kılıç and U. Yüzgeç, "Improved Antlion Optimization Algorithm Via Tournament Selection and Its Application to Parallel Machine Scheduling," Computers & Industrial Engineering, 132, 166-186, 2019. https://doi.org/10.1016/j.cie.2019.04.029

[24] S. Vitayasak, "Facility Layout Problem: A 10-Year Review and Research Perspectives (in Thai)," Naresuan University Engineering Journal, 5, 46-62, 2010.

[25] Z. Güngör and F. Arıkan, "Application of Fuzzy Decision Making in Part-Machine Grouping," International Journal of Production Economics, 63, 181-193, 2000. https://doi.org/10.1016/S0925-5273(99)00010-9

[26] S. K. Deb and B. Bhattacharyya, "Fuzzy Decision Support System for Manufacturing Facilities Layout Planning," Decision Support Systems, 40, 305-314, 2005. https://doi.org/10.1016/j.dss.2003.12.007

[27] S. Kritwattanakorn, K. Chamniprasart, and W. Boongsood, "A Permutation-Based Genetic Algorithm for Solving the Machine Layout of Manufacturing System," in The 10th SEATUC Symposium, Tokyo, Japan, 2016.

[28] C. Jadram and W. Boongsood, "Application of Fuzzy Materials-Handling Inputs with the Genetic Algorithm for Machine Layout," in 2018 7th International Conference on Industrial Technology and Management (ICITM), Oxford, UK, 265-269, 2018. 10.1109/icitm.2018.8333958

[29] R. G. Askin and C. R. Standridge, Modeling and Analysis of Manufacturing Systems: Wiley, 1993.

[30] C. K. H. Lee, "A Review of Applications of Genetic Algorithms in Operations Management," Engineering Applications of Artificial Intelligence, 76, 1-12, 2018. https://doi.org/10.1016/j.engappai.2018.08.011
[31] W. Yi, H.-L. Chi, and S. Wang, "Mathematical Programming Models for Construction Site Layout Problems," Automation in Construction, 85, 241-248, 2018. https://doi.org/10.1016/j.autcon.2017.10.031
[32] S. Gao and Z. Zhang, "Multiplication Operation on Fuzzy Numbers," Journal of Software, 4, 331-338, 2009.
[33] W. Dong and F. S. Wong, "Fuzzy Weighted Averages and Implementation of the Extension Principle," Fuzzy Sets and Systems, 21, 183-199, 1987. https://doi.org/10.1016/0165-0114(87)90163-1

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 139-142 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Development of Teaching Aids Selection Standards in Maker Education

Woochun Jun*

Seoul National University of Education, Department of Computer Education, 06639, Korea

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 28 August, 2019
Online: 24 September, 2019

Keywords:
Maker Education
Selection Standards
Teaching Aids
Physical Computing

ABSTRACT

Due to the recent development of information and communication technology, such as sensor technology, IoT technology, and smart technology, interest in maker education is increasing. Maker education offers students experience-based education that allows them to make their own results, giving them various interests and motivations. Recently, various research results on maker education have been under way. However, there is little research work on the teaching aids selection criteria for maker education. The purpose of this study is to justify the teaching aids selection standards that are developed in the previous research works. In the previous research work, 14 selection standards were developed in a total of eight areas. In this work, through rigorous statistical analysis, justification of the 14 standards are verified. The findings are expected to help a lot in future policy proposals and related research on maker education.

1. Introduction

This paper is an extension of work originally presented in ICTC2018[1].

Various changes are taking place throughout the industry following the recent advent of the 4th industrial revolution era. Information-oriented industries are emerging at the forefront of the industry in the existing product-oriented manufacturing industry structure. The changes are prompting the acceleration of the modern knowledge and information society, calling for, among other things, interest in and understanding of the information-oriented society.

This change in the industrial structure also calls for many changes in public education in schools. In particular, interest in information education such as information collection, analysis, and processing is increasing in schools as the knowledge and information society accelerates. Information education is evolving into a software education form, especially from Information and Communication Technology Education, which started from 2000 in Korea. While the human image sought in ICT education was the core part of a computer-savvy user, the human image sought by the government in 2015 became the producer or manufacturer who could actually make software [2].

Maker education began in the early 2000s with the U.S. as the Do-It-Yourself Movement [3]. The DIY movement aimed to produce and use the necessary items in everyday life and also to share the necessary processes and products with each other. The DIY movement evolved into the Maker Movement, which, combined with the Maker Space, was ultimately born into Maker Education [4].

Meanwhile, according to [5], maker education was defined as follows by relating it to information education (or software education). In [5], maker education is a new type of educational paradigm to foster so-called "maker geeks" who can create things themselves. In other words, what students have imagined autonomously is produced directly using digital devices. It was also defined as process-oriented project education rather than the results that lead to the sharing of knowledge and experience learned in this process.

Maker education is in the spotlight as an experience-type education that can increase students' interest and motivation for learning. It is also receiving attention at the schools in a typical form of convergence education, combined with various subjects, such as science and practical subjects. In addition, various forms of hands-on exhibitions and contests are attracting interest from students and parents, and teachers are given various training opportunities to encourage interest in maker education.

Along with the spread of maker education, academic research is under way on various maker education. However, in order to carry out actual maker education on the schools, there is a lack of standards on how to choose teaching aids, which is the core of the maker education, among other things Generally, teaching aids are

*Corresponding Author: Woochun Jun, Seocho Dong, Seocho Gu, Seoul, Korea, wocjun@snue.ac.kr

tools used in class activity. The typical teaching aids in class are scissors, colored paper, glue, and so on. In maker education, the typical teaching aids includes 3D printers, 3D pens, cutting tool, sewing machine, and so on. Teaching aids are either raw materials or making tools like a machine tool. In maker education, teaching aids are essential for class activity. Thus, choosing right teaching aids in maker education is very important in successful class.

The purpose of this study is to justify teaching aids selection standards that are developed in the previous research works. That is, for those 14 selection standards, rigorous statistical work is done and selection standards are proved to be useful in schools.

The rest of this paper is as follows. First, in Chapter 2, related works are presented. In Chapter 3, the existing 14 selection standards are introduced. In Chapter 4, statistical analysis is presented for selection standards. Finally, in Chapter 5, conclusions and further research issues will be discussed.

2. Related Works

This chapter presents an overall relevant study of maker education. First, the typical characteristics of maker education can be summarized as follows [6].

First, knowledge can be organized through interaction with objects or people through direct-creating activities. In addition, outputs created directly by learners can induce social participation through various sharing, thereby providing a meeting place for various participants to communicate together through common tasks.

Second, it can provide self-directed and cooperative learning. Makers are engaged in maker activities, choosing their own themes from areas of personal interest and social issues. In addition, working on cooperative projects can lead to mutual respect and communication through cooperative learning.

Third, it provides an environment for making activities. In other words, the maker education will require a maker space for direct production activities and a variety of teaching aids and materials. The maker space needs an environment with the tools needed for the maker's activities.

Fourth, maker education requires a position as an inventor and creator as a learner. In other words, students go through the process of choosing their own themes through making activities. Through this process, learners experience creative problem solving.

Fifth, in maker education, a teacher will act as a helper. Maker activities are basically activities that can improve students' creative thinking and problem-solving skills. In addition, since students decide on their own topics, tools, methods, etc., they should provide a free learning environment and a free environment for evaluation regardless of the outcome's success or failure. Therefore, a teacher's role should be a learning guide or facilitator.

Sixth, students can start with personal participation in maker education and continue with social participation. This is because the maker education can share and open information with other learners through personal problem solving, starting from social issues and creating experiences, and to see if personal and social problems can be solved through product sharing.

On the other hand, a recent study by [7] developed and presented a measure of the ability of making as part of the software coding- based maker education program for scientific gifted children. The Matching Capacity Scale means the competency that students who participate in the maker education should have, and the makeup competency measures presented in [7] are as follows. Overall, the making capabilities comprise three core competences: analysis capabilities, design capabilities, and implementation capabilities. Specifically, analysis capability refers to the ability to collect data and analyze problems, and design capability corresponds to the ability to generate and refine ideas for problem resolution, and to come up with solutions through sharing ideas. Finally, implementation capability refers to the ability to select a tool for problem solving, generate results through active participation and evaluate results. The three core competences were divided into more detailed sections to provide specific capabilities.

Table 1. The Original Selection Standards

Number	Standard	Detailed Criteria
1	Safety	-There should be no risk of safety accidents. -Material that can be harmful to health should not be used
2	Subject Compatibility	-It should be available for various educational activities -It should be possible to integrate with other teaching materials and learning materials.
3	Multi-purpose	-The teaching aids must support various functions through the connection of built-in or additional parts.
4	Ease of Manipulation	-(Teacher) It should be easy to use in class by simple training. -(Student) It should be easy to use after simple teaching and learning about how to use.
5	Economics	-It should be easy to buy within budget items of the school. -Parts can be easily procured within reasonable time and price.
6	Universal Design	-It should be designed for the older, people with mental and physical disabilities. -It should provide alternative functions for main functions
7	Gender Equality	-It should be used regardless of gender.
8	Durability	-Failure or breakage should not occur even if students repeatedly use. -It should be able to withstand the minor impact (collision, drop, etc.) that occurs between educational activities.

3. Development of Teaching Aids Selection Standards

3.1 Design Principles

The selection standards for teaching aids in maker education are developed through the following principles [1].

First, it is in line with the general teaching aids selection criteria used in primary and secondary schools. That is because the teaching aids used in maker education are essentially one of the teaching aids.

Second, it is based on the selection criteria for the physical computing education, which is mainly used in recent software education.

Third, the universal design principle is adopted so that it can be used by anyone. In other words, care is taken to make it easier for anyone, regardless of age, gender, age, ordisability.

3.2 The Original Selection Standards

In [1], the original selection standards of teaching aids in maker education are proposed. The following table 1 show the initial standards.

4 Statistical Analysis

4.1 Design of Analysis

Verification of the reliability of the tool to measure the importance of teaching aids selection standards in maker education showed that Cronbach α was 0.77 and was reliable, as shown in Table 2.

Table 2. Reliability of Measurement Tool

Item Number	Cronbach's α
14	0.77

The data gathered in this work were analyzed using the Statistical Package for the Social Science (SPSS) WIN 25.0 program. Cronbach α is used to verify the reliability of the measuring tool. In addition, the average and standard deviation were calculated in order to find out the importance of the selection standards in maker education, and the correlation analysis was conducted to understand the relationship between the sub-areas of selection standards and selection standard in maker education.

4.2 Analysis Results

-Descriptive Statistics

The average and standard deviation of questions asked about the importance of teaching aids selection standards in maker education are as shown in Table 3.

4.3 Sub-area of Descriptive Statistics

According to the results of the review of the average and standard deviation in the sub-areas of teaching aids selection standards in maker education are shown in table 4. The average importance of the selection standards for the teaching aids in maker education was 4.13, and the selection standards for the teaching aids were important.

Among the sub-areas of the standards for selection standards for teaching aids, “safety” 4.84, followed by "ease of manipulation" 4.52, "durability" 4.30, "economics" 4.28,"subject compatibility" 4.05, "gender equality" 4.00 and "multi-purpose" 3.78, with the lowest "universal design."

As above, it is shown that “safety” is the highest in the sub-areas of the teaching aids selection standards in maker education, and that “universal design” is less important than other areas.

Table 3. Results of Descriptive Statistics

Standard	AVG	SD
1. There should be no risk of safety accidents.	4.84	0.37
2. Material that can be harmful to health should not be used.	4.84	0.37
3. It should be available for various educational activities	4.13	0.91
4. It should be possible to integrate with other teaching materials and learning materials.	3.97	0.82
5. The teaching aids must support various functions through the connection of built-in or additional parts.	3.78	0.94
6. (Teacher) It should be easy to use in class by simple training.	4.41	0.80
7. (Student) It should be easy to use after simple teaching and learning about how to use.	4.63	0.61
8. It should be easy to buy within budget items of the school.	4.38	0.66
9. Parts can be easily procured within reasonable time and price.	4.19	0.86
10. It should be designed for the older, people with mental and physical disabilities.	3.06	1.16
11. It should provide alternative functions for main functions	3.00	1.14
12. It should be used regardless of gender.	4.00	1.19
13. Failure or breakage should not occur even if students repeatedly use.	4.28	0.85
14. It should be able to withstand the minor impact (collision, drop, etc.) that occurs between educational activities.	4.31	0.69

Where AVG and SD stand for average and standard deviation, respectively.

4.4 Correlation Analysis

The results of a review of the correlation between the sub-areas of the standards for teaching aids selection standards in maker education and the whole teaching aids selection standards in maker education are shown in Table 5.

Safety (r=.455; p<.01) and subject compatibility (r=.627, p<.001), multi-purpose (r=.548, p<.01), ease of manipulation (r=.391, p<.05), economics (r=.534, p<.01), universal design (r=.725, p<.001), gender equality (r=.824, p<.001), and durability (r=.453, p<.01) showed a statistically significant static correlation with the importance of the standards for teaching aids selection

standards in maker education. Therefore, the higher importance of safety and subject compatibility, multi-purpose, ease of manipulation, economics, universal design, gender equality, and durability of the standards for teaching aids selection standards in maker education, the higher importance of the standards for teaching aids selection standards in maker education.

Table 4. Results of Descriptive Statistics of Sub-area

Evaluation Standard Area	AVG.	SD
Safety	4.84	0.27
Subject Compatibility	4.05	0.73
Multi-purpose	3.78	0.94
Ease of Manipulation	4.52	0.63
Economics	4.28	0.61
Universal Design	3.03	1.07
Gender Equality	4.00	1.19
Durability	4.30	0.69
Overall Selection Standards	4.13	0.43

where AVG and SD stand for average and standard deviation, respectively.

Table 5. Results of Correlation Analysis

Area	Correlation Value
Safety	0.455**(0.009)
Subject Compatibility	0.627***(0.000)
Multi-purpose	0.548**(0.001)
Ease of Manipulation	0.391*(0.027)
Economics	0.534**(0.002)
Universal Design	0.725***(0.000)
Gender Equality	0.824***(0.000)
Durability	0.453**(0.009)

Where * p<.05, ** p<.01, *** p<.001

5 Conclusions and Further Research Issues

Along with the development of various information and communication technologies and smart technologies, maker education is receiving attention at schools due to the increase in the number of available teaching aids at schools. Maker education is an experience-based education in which students directly participate in designing and analyzing the work they want to produce, and finally participate in production, which has the advantage of generating interest and motivation for students. Maker education is also very popular as a project-oriented cooperative study because teachers and students can participate and discuss the entire process of producing works together.

The role of the teaching aids in maker education is very important and the interest and academic performance of students in the whole class, depending on how they choose the teaching aids, also affects safety issues. However, there has been no standardized research or guideline on the standards for teaching aids selection standards in maker education.

The purpose of this study is to justify the teaching aids selection standards that are developed and proposed. In the previous research, 14 selection standards are developed and analyzed statistically. In this research, with more rigorous statistical analysis, the justification of the results is verified.

The immediate further research issues are as follows. At first, it is necessary to present more detailed selection standards depending on tools of maker education such as 3D printer. Also, it is also necessary to present more integrated and extensive selection standards for both maker education and software education. This is because there is no clear limit between maker education and software education.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] W. C. Jun, “A Study on Development of Selection Standards of Maker Education Teaching Aids” in 2018 International Conference on Information and Communication Technology Convergence, Jeju Island, Korea, 2018.

http://10.1109/ICTC.2018.8539704

[2] Korea Ministry of Education, “Software Education Training and Operation Guidelines”, 2015.

[3] E. R. Halverson, and E. M. Sheridan, “The Maker Movement in Education”, Harvard Business Review, 84(4), 495-504, 2014.

https://doi.org/10.17763/haer.84.4.34j1g68140382063

[4] D. Dougherty, ”The Maker Movement, Innovations: Technology, Governance, Globalization“, MIT Press, 7(3), .11-14, 2012.

https://doi.org/10.1162/INOV_a_00135

[5] Seoul Metropolitan Office of Education, “'2018 ~ '2022 Mid and Longterm Maker Education Development Master Plan”, November 1, 2017.

[6] M. J. Kang, “Development of Checklist for Development and Operation of Maker Education Program”, Master Thesis, Kyunghee University, 2018.

[7] J. H. Lee and J. H. Jang, “ Development of Maker Education Program based on Software Coding for the Science Gifted”, Journal of Gifted/Talented Education, 27(3) 331-348, 2017.

http://doi.org//10.9722/JGTE.2017.27.331

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 140-152 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Educational Domain Management Knowledge Content Identification and Knowledge Updating Method, Based on Enterprise Management Information Interactions

Jurij Tekutov[*, 1, 2, 4], Saulius Gudas[3], Vitalijus Denisovas[2, 4], Julija Smirnova[1]

[1]*Klaipeda State University of Applied Sciences, Faculty of Technology, Information Technologies Department, Bijunu St. 10, LT-91223 Klaipeda, Lithuania*

[2]*Klaipeda University, Faculty of Marine Technologies and Natural Sciences, Informatics and Statistics Department, Bijunu St. 17, LT-91225 Klaipeda, Lithuania*

[3]*Vilnius University, Faculty of Mathematics and Informatics, Institute of Data Science and Digital Technologies, LT-04812 Vilnius, Lithuania*

[4]*Lithuania Business University of Applied Sciences, Informatics Department, Turgaus St. 21, LT-91249 Klaipeda, Lithuania*

ARTICLE INFO

Article history:
Received: 29 December, 2018
Accepted: 04 March, 2019
Online: 20 March, 2019

Keywords:
The Detailed Value Chain Model
The Elementary Management Cycle
Educational domain knowledge content

ABSTRACT

The aim of this paper is to develop the educational domain management knowledge content identification and knowledge updating method, based on formal descriptions of enterprise management information interactions. This is very important because the progress of modern science leads to the emergence of new factual or conceptual knowledge. There is a need for proper methods that allow, first, to study and make decisions regarding the relevance of the content, and then take practical steps on the content and possibly of the corresponding structure adaptation. The authors suggest a systematic solution for this problem. The hierarchical Detailed Value Chain Model and Elementary Management Cycle model of educational domain knowledge content identification and updating is formally described, computerized process measures are proposed. The paper provides a method for updating the knowledge of the analyzed domain, referred to as "enterprise domain", based on enterprise modelling in terms of management information interactions. A method, whose formal DVCM and EMC descriptions are provided in BPMN notation, was designed, allowing to develop a two-level (granular) model for describing knowledge of educational domain management information interactions. In implementing this model and its algorithms in technological terms, a subsystem of enterprise knowledge has been created in a knowledge-based CASE system (computerized knowledge-based IS engineering), performing the function of a domain knowledge database.

1. Introduction

This paper is an extension of work originally presented in 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) [1].

Meeting modern, constantly evolving educational enterprise needs requires enterprise integration with information technologies and computerized information systems (IS) that would be able to adapt to organizations' enterprise changes. In the context of this paper enterprise is understood as a educational institution.

Among different IS development approaches one of the most promising is model-driven approach, which aims to utilize formal domain knowledge to perform, adjust and manage and engineering process. One of the earliest stages of IS development life-cycle is enterprise modelling, which becomes especially important in applying model-driven IS development methodology (MDE – Model Driven Engineering) and its methods [2]. The aim of enterprise modelling is to create an enterprise model of an

[*]Jurij Tekutov, Email: jurij.tekutov@yahoo.com

organization, which would help to deal with issues of an analysed domain. The concept of Model Driven Architecture (MDA) defines IS development process as a three-level model (CIM – the Computation Independent Model, PIM – the Platform Independent Model, PSM – the Platform Specific Model) interaction, whose purpose is to directly link enterprise modelling and further stages of IS analysis, design and implementation. Thus, from enterprise modelling professionals' point of view MDA concept [3, 4] should be thought of as an effort to integrate enterprise modelling methods with full life-cycle Computer-Aided Systems Engineering (CASE) technologies characterized by the bottom-up approach, i.e. from engineering methods to enterprise modelling methods for IS engineering needs. Hence, modern IS development methodologies pay increasingly more attention not only (and not as much) to software and IS engineering methods, but rather organizations' enterprise modelling issues, i.e. domain knowledge mapping.

IS engineering methods (IS life-cycle stage models) based on intra-enterprise engineering, the basis of which is subject area (domain) pattern model, are called knowledge-based methods [5]. The purpose of MDE-based computerized IS engineering enterprise model is to systematize and formalize knowledge on a certain analysed domain (subject area). IS engineering methods are emerging that seek to integrate knowledge about operational functions, processes, rules in CASE systems and use them in intellectualizing IS development process. There have also been attempts to apply domain knowledge content creation and updating methods [6] from similar areas, such as e-learning, Business Intelligence systems [7, 8], business rule modelling [9].

The analysis of IS engineering stages (1. traditional IS engineering, 2. computerised software engineering, 3. computerised IS engineering (including enterprise modelling), 4. knowledge-based IS engineering) reveals the key features of the variations in the IS engineering process (methods, knowledge sources, software types). Hence, it enables to define the key characteristics of the knowledge-based IS engineering methodology and the composition of the basic components. In traditional computerized IS engineering a system is created empirically, by first collecting, analysing and specifying consumer needs. Works linked to knowledge-based IS development allow to take software development to a higher level of abstraction, partly automate these processes and use the models in other systems. Business activity modelling, the improvement of knowledge formalization and the application of the conceptual modelling in developing the IS are the key things for the further development of the IS engineering methods. In knowledge-based computerized IS engineering a system is created using an enterprise knowledge base and domain knowledge stored in it. It is believed that the main characteristic of knowledge-based IS engineering is a subsystem of domain knowledge, which is based on an organization's enterprise model and is designed to acquire knowledge on a computerized domain. Such knowledge acquisition subsystem should perform the functions of a storage of domain knowledge, required for the creation of IS project models. The following structure for a subsystem was proposed: enterprise model and enterprise metamodel (formalized enterprise management knowledge structure).

Domain knowledge acquisition subsystem should become the essential component of a knowledge-based CASE measure, intellectualizing the whole process of IS development. The creation of such subsystem is one of the most urgent scientific problems addressed in this paper. The development of knowledge-based IS engineering methods requires an internal approach to business process modelling, which should be based on the internal business process logic – the causal links between business processes, i.e. necessary and sufficient components of activity elements and interactions between them (the activity goals of material processes, information processes and structural information units). Therefore, the above-mentioned diagram shows the managerial information exchange (in proposed modified Value Chain Model).

In our previous works it was proposed that the development of these systems should be performed by focusing on Value Chain Model (VCM) [10], since computerization involves enterprise management functions and processes, as well as interactions between organization's functional areas, which are revealed by the VCM. In computerizing enterprise, a research of informational interactions between enterprise management functions and processes is performed, requiring further decomposition of value chain elements, i.e. further VCM modification is performed [11], a Detailed Value Chain Model (DVCM) is created. Hence, knowledge-based IS development is based on knowledge on enterprise management, which defines fundamental attributes of enterprise management information interactions. In previous work [12] was shown the enterprise management information interaction is specified as an Elementary Management Cycle (EMC), linking enterprise management function and process by a feedback loop. The novelty of the work is the application of the expanded VCM and the EMC to manage operational knowledge; after defining the assessment criteria, such management model allows to objectively assess and constantly improve the knowledge content of the subject area.

This paper elaborates a proposed domain management knowledge modelling method, which is based on formal DVCM and EMC descriptions presented in a BPMN notation. Educational domain has been chosen intentionally, on the basis of methodological terms. In academic field IS development typically starts with enterprise management function analysis, in which enterprise management information interactions are important. The analysed higher education study domain is multivariate, dynamic and requires constant knowledge content updating. Only by actualizing study programme structure and knowledge content in a timely manner is it possible to properly take into account changing needs and requirements from all stakeholders.

This paper is devoted to the actual topic of modern university management. Improving the quality of higher education and the corresponding management approaches are very important. The authors propose to apply the proven methods of enterprise management. This is a very promising decision, because, firstly, such business-oriented models have proven their effectiveness, and then such models are focused on practical solutions.

Previous works. One of the pioneers of analysing study programme development as an IS engineering process were MOCURIS – Modern Curriculum Development in Information Systems at Master Level project promoters, who offered in their works a model study programme design method [13]. They have

introduced the notion of study programme engineering, defined study programme architecture and by analyzing a study programme as a system of courses and modules, they have applied the standard IS engineering process to develop a Master‘s study programme. Using this method, a large IS-related Master's programme specification document was created, which would act as a guideline for developers of certain programmes, courses and modules. However, already in the project fundamental flaws of this method became evident. Such programme development process could only be effectively implemented by the authors, meanwhile, the resultant artefact (textual specifications document) because of its technicality, complexity and size was unsuitable to perform the main function of requirement specification which is to ensure communication of all process participants and coordinate their work. After analysing requirement identification, localization and specification (granularity) issues, it was found that these limitations are inherent in traditional non-automated requirement engineering processes. Therefore, other researchers (including the authors of this paper) have suggested to automate this process by creating a study programme requirement engineering system on the basis of CASE measures [14]. With a direct involvement of the authors, such a system has been created and applied for the development of new study programmes [15, 16]. However, in a fast changing environment timely study programme modernization, quality assurance and knowledge content updating become especially relevant. It all requires effective feedback from all stakeholders. To solve the issues mentioned in this paper, a DVCM has been used, ensuring constant information feedback, which sends descriptive data (attributes) to the enterprise management function and retrieves management decisions to direct enterprise process. The method is implemented by creating an enterprise knowledge database in IS engineering CASE system, which is designed on the basis of a computerized educational domain enterprise model.

The creation of knowledge content specification includes the collection of domain knowledge, its preparation and analysis, verification and validation. In this paper domain knowledge is understood as a dynamic set of knowledge attributes, described according to requirements as a system of facts, principles and theories related to a specific professional activity. In other words, knowledge attribute set is an expression of a domain in current context. Enterprise modelling can be assigned to requirement collection and analysis stage. Then enterprise models are transformed into requirement specification for the system being

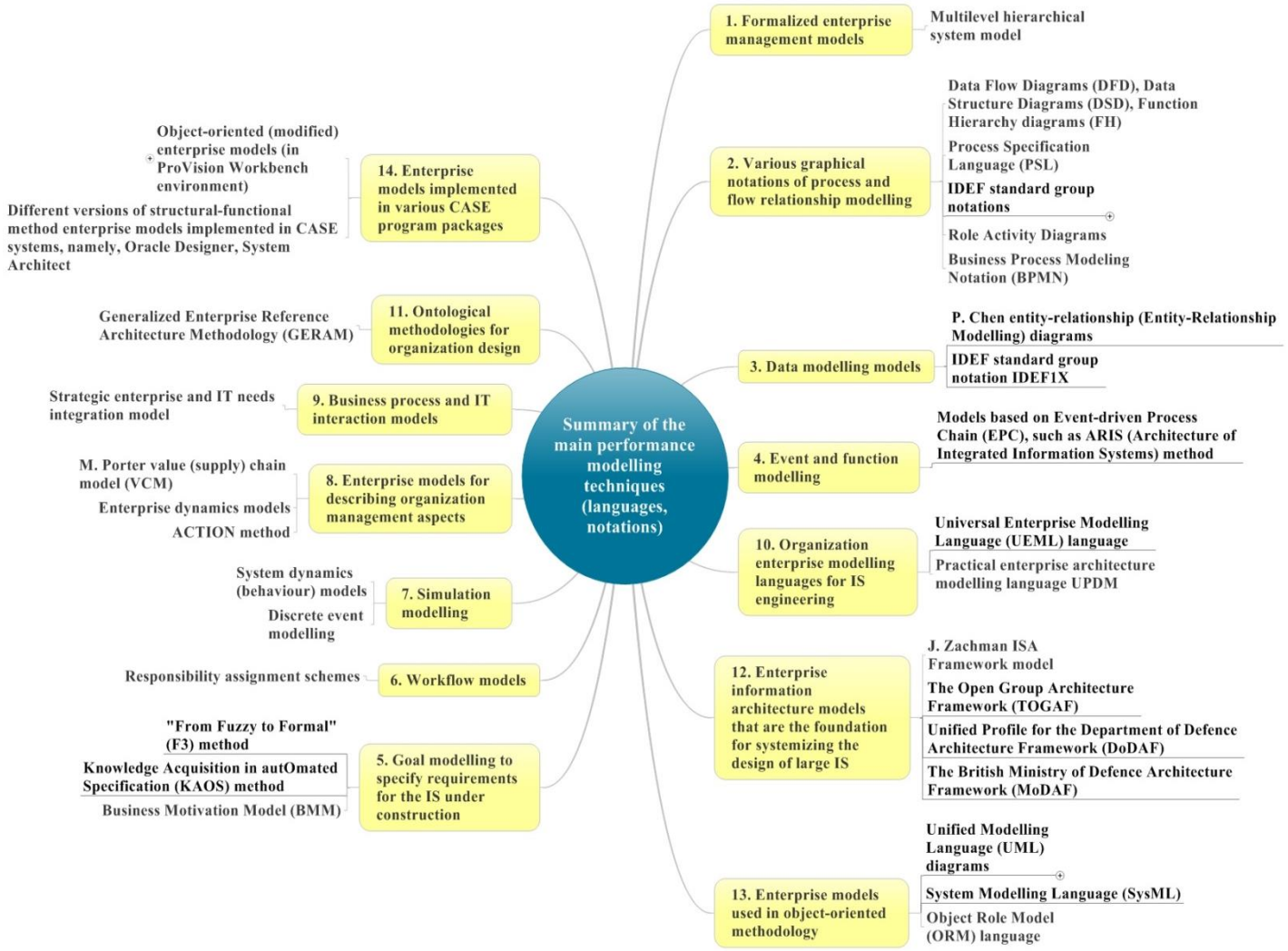


Figure 1. Concept map of enterprise modelling techniques (languages, notations) (using *Mindjet MindManager Pro™* tool)

developed. Systems (for example, domain knowledge content) and separate component specifications are produced during IS engineering process. The most important step in every effective domain knowledge management process is the writing down of the requirements identified during the selection. It helps to ensure accurate communication and management of the requirements during their development. Documents are an understandable form for requirement recording. These documents provide a basis for context identification and requirement collection, yet are not very useful in supplementing and modifying requirement information.

The main aim of this paper is to develop the educational domain management knowledge content identification and knowledge updating method, based on formal descriptions of enterprise management information interactions.

The remainder of the paper is structured as follows: Section 2 reviews domain modelling methods in terms of knowledge content modelling. Section 3 introduces the created educational domain management knowledge modelling method, based on enterprise management information interactions. The application of a Detailed Value Chain and Elementary Management Cycle model for knowledge content description and updating is demonstrated. The created educational domain management knowledge content identification and knowledge updating computerized process model are laid out in Section 4.

2. The Analysis of Current Domain Modelling Methods in Terms of Knowledge Content Modelling

Due to the fact that many new enterprise modelling approaches, methodologies and methods were recently created, a separate research area has occurred, called enterprise modelling.

Therefore, there is a need to systematize the variety of these methods in order to select the most convenient way of modelling management activities. To facilitate this selection a concept map of different enterprise modelling techniques (languages, notations) [17-23] was designed (Figure 1).

Enterprise models are divided into the following categories: business process models, data flow models, data models, business objective models, control flow models, management process models, knowledge management processes [24]. Widely known enterprise modelling methods are structural-functional modelling (Data Flow Diagrams or DFD), business process modelling languages, BPMN (Business Process Modeling) notation, IDEF (Integration of computer aided manufacturing DEFinition) set of standards and enterprise modelling language UEML (Unified Enterprise Modelling Language), and business architecture modelling language UPDM (The Unified Profile for DoDAF/MODAF). Well known enterprise modelling methodologies and standards are CIMOSA, GERA, EPC, DoDAF, MODAF, MDA [23]. However, amongst them one can rarely find those that would model enterprise in terms of management, i.e. performance management information components and their interactions.

When comparing modern enterprise modelling methods (languages, notations), it is becoming clearer that enterprise modelling needs to have capable methodologies because it has to be ensured that all necessary structures can be modelled. For this purpose, the above mentioned concept map can be used, that shows primary and systematized information and provides references to all other required sources of domain modelling methodologies. When analysing known enterprise modelling methods (notations, languages) in terms of domain management, it is becoming clearer that only a few methods can actually evaluate management information interactions that are necessary to manage domain from a theoretical perspective.

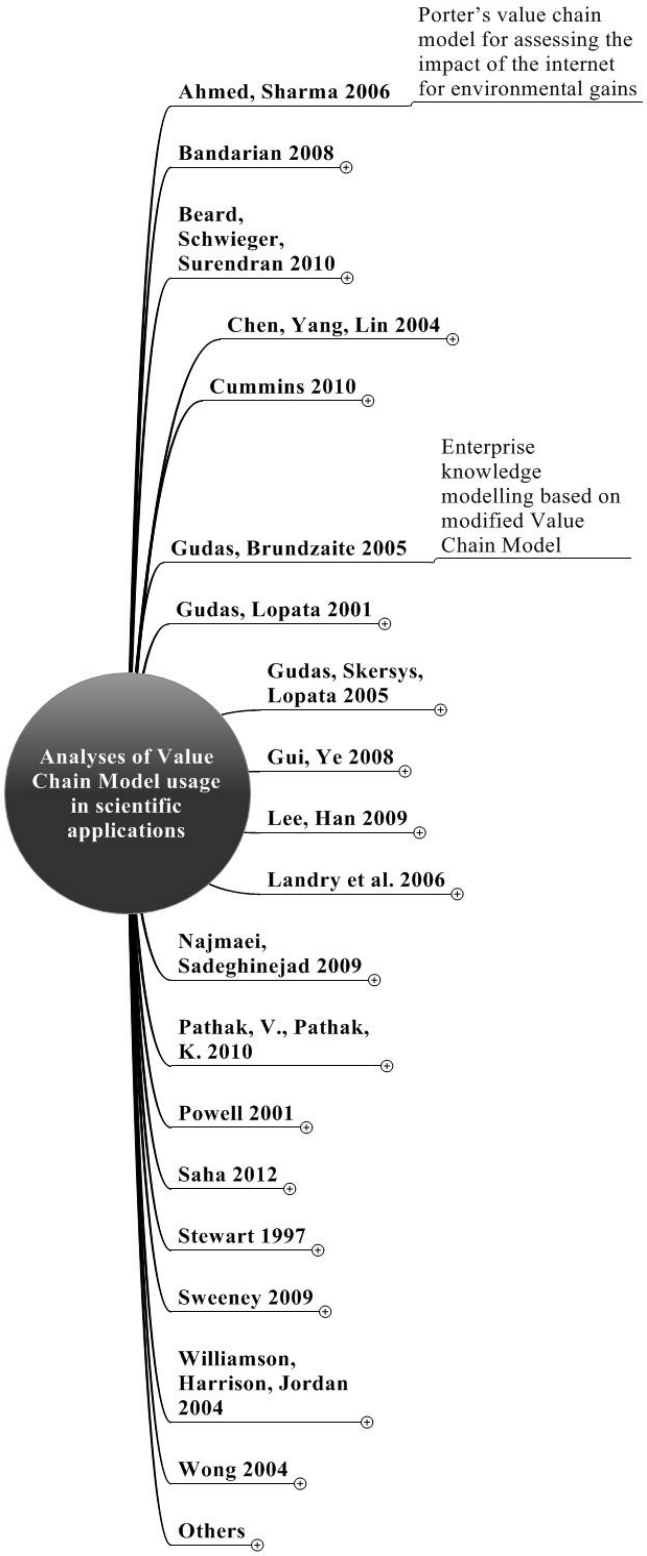


Figure 2. Conceptual map of Value Chain Model in scientific applications

In organizational management practice Value Chain Model (herein referred to as VCM), created by Porter, is popular and has been used in many works (Figure 2) [26-42].

The concept map (Fig. 2) illustrates examples of the application of VCM in scientific research (other examples of the analysis can be accessed by clicking on the plus sign in the concept map in Mindjet MindManager™ environment): Porter's value chain model for assessing the impact of the internet for environmental gains [26]; Enterprise knowledge modelling based on modified Value Chain Model [40]; others.

The structured DVCM is used to identify the information transactions between Management Functions and Enterprise Processes. Detailed Value Chain Model (DVCM) embodies a procedural approach to enterprise consisting of: primary activities (operational processes) and support activities (operational functions). The formal description of DVCM is as follows:

$$DVCM = \{(Fi) x (Pj)\} \quad (1)$$

Where: (F1,…, Fi,…., Fn) – a set of Enterprise Management Functions, (P1,…, Pj,…., Pm) – a set of Enterprise Processes, x – relationship between a set of Management Functions {F} and a set of Enterprise Processes {P} (management information transactions $(Fi \times Pj)$.

The interrelationship between primary and secondary business processes explored in Gudas, Lopata [43] identified a different nature of these 2 enterprise activities: secondary processes possess informational nature and are referred to as enterprise (management) Functions; while primary processes are concrete (non-informational) and are named enterprise (material) Processes. This paper presents more detailed content of Function Fi since it defines a sequence of definite types of interacting information activities directed to control Process Pj.

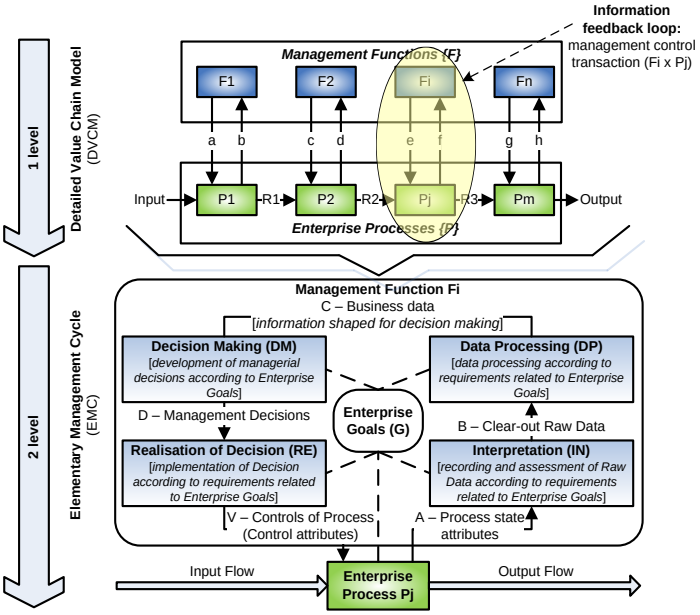


Figure 3. The formalized model of the interaction between the Function Fi and Process Pj (Elementary Management Cycle – EMC) in modified (structured) DVCM

In brief, the concept of EMC is a formalized description of the Enterprise management control as the interaction between the Function and the Process – as two core components of enterprise from the control point of view. On the basis of these findings, the Value Chain Model is modified. The decomposition of the information interactions $(Fi \times Pj)$ in the DVCM between the Management Function Fi and Enterprise Process Pj is defined as Elementary Management Cycle (EMC) on the lower level in Figure 3. The interaction between the core elements a Function and a Process is formally assumed as a Control Process with the Feedback Loop between the Function Fi and the Process Pj. The two levels of granularity: 1) the Detailed VCM (DVCM) and 2) the Elementary Management Cycle (EMC).

A detailed model of the informational interactions between the Function Fi and Process Pj can be defined by:

$$EMC(Fi,Pj) = \begin{pmatrix} Pj(A,G) \rightarrow IN(A,B,G) \rightarrow P(B,C,G) \rightarrow \\ M(C,D,G) \rightarrow RE(D,V,G) \rightarrow Pj(V,G) \end{pmatrix} \quad (2)$$

Where A – state attributes of the process Pj, needed in terms of G; B – systemized (interpreted) primary data, needed for the enterprise management function Fi in terms of G; C – enterprise data, formed by data processing DP, prepared for decision making DM and needed in terms of G; D – goal-congruent management decision formed by the decision making process DM; V – goal-congruent effects of management on the process Pj, formed by the realization process RE.

The expanded DVCM and EMC was printed in the general format [15, 16]. As proposed in works [44, 45], the Elementary Management Cycle (EMC) is the basic construct of Enterprise Management modelling, it refines the components of management (control) cycle as well as content of management information transformations. The mandatory steps (Interpretation – IN, Data Processing – DP, Decision Making – DM, Realisation of Decision – RE) of the EMC are defined as information transferring processes focused on the control of the content of the Management Functions Fi. The Management Function Fi consists of a sequence of definite types of goal-driven information transformation activities (steps of EMC) aimed to control a state of an Enterprise Process Pj. For example, in the structure of an EMC: Pj (A, G), Pj (V, G) – Technological Process (managed object), its input (I) and output (O) are Material Flows. Material Flows are defined by state attributes of a specific Process Pj, which are necessary to perform a specific enterprise management Function Fi in combination with enterprise Goals (G).

Based on this general model, the authors have proposed a specialised model for particular problem domain.

3. Educational Domain Management Knowledge Modelling Method, Based on Enterprise Management Information Interactions

Educational domain management knowledge modelling method, based on enterprise management information interactions, was created in order to identify the discrepancy between the existing knowledge model and the actual domain knowledge model. The resulting method is designed to create a two-level

(granular) knowledge description model for the analyzed educational domain management information interactions, to form current knowledge and domain knowledge models and to perform the analysis of knowledge models. The computerisation of the organisational activities involves an examination of the information interactions between the management functions and the business processes, which requires the decomposition of the value chain elements, i.e. the VCM is modified and a Detailed Value Chain Model (DVCM) is created. Thus, the knowledge-based IS development is performed based on that business management knowledge which defines the key features of business management information interactions. A business management information interaction is specified as an Elementary Management Cycle (EMC), which links the business management function and the business process through a feedback circuit. A two-level subject area management knowledge model was selected. The two levels of knowledge description are:

1. First level of detail – DVCM that identifies information interactions.

2. Second level of detail – a detailed model of each management information interaction by applying EMC.

Such model is believed to be sufficient in all cases of activity knowledge modelling. The decomposition of the organisational system business management functions and the business processes can be continued up to the required level of detail of the business management function.

Paper comprise known approaches to create educational domain content knowledge models updated existing. Model is created on to levels that combines DVCM and EMC model presenting the transformation of the existing and the domain knowledge models on the basis of management information interactions", BPMN notation (Figure 4).

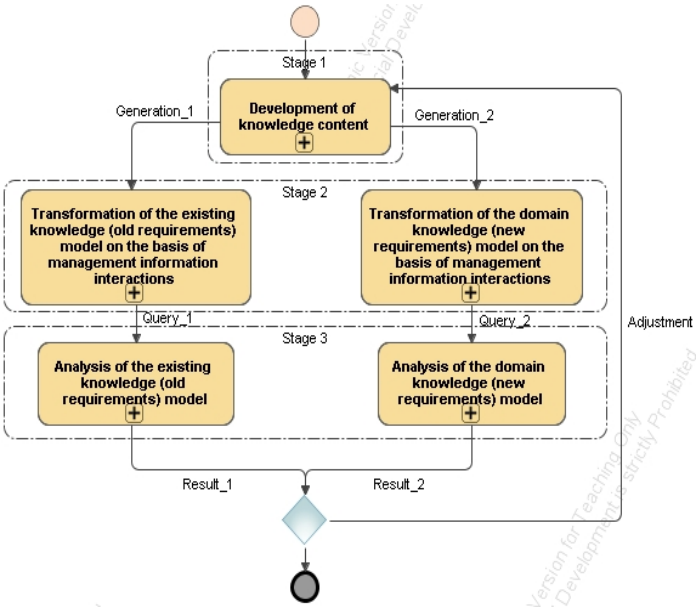


Figure 4. A principal scheme of educational domain management knowledge modelling method (BPMN notation)

Development of knowledge content (Figure 5) involves structural, functional, quality and other requirements.

The method is elaborated in a specific domain – higher education study programme design and updating. In terms of the current domain (for example, programme structure), the smallest structural component is a study module. Module composition is described in terms of attributes. To establish a system hierarchy, modules are combined into study blocks (subject groups). Main blocks are defined in the general requirements for study programmes (component_1 in Fig. 5). Functional (study content) requirements are defined by the purpose, stage and objectives of a study programme that are defined in regulatory documents and derived from other requirement sources (component_2 in Fig. 5).

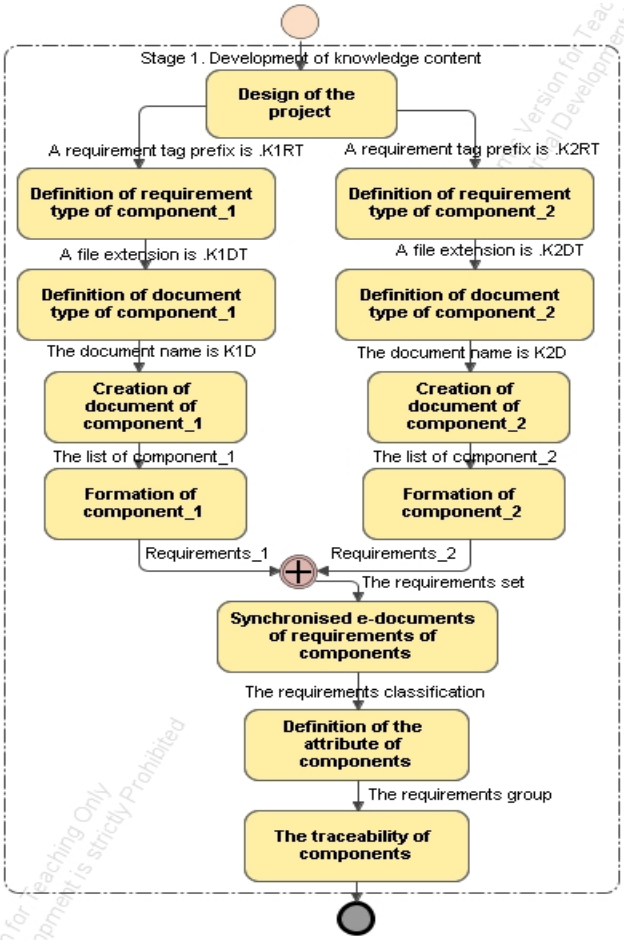


Figure 5. Part of the method (Stage 1 "Development of knowledge content", BPMN notation, using *MagicDraw*^{TM} tool)

Designing a knowledge base on the basis of DVCM and EMC. Management knowledge model based on DVCM and EMC is created in two stages (Figure 6):

1. On the basis of the Detailed Value Chain Model (DVCM) identifying management Functions and their information resources.

2. Each intersection between management Function and Process is modelled in detail as the managed process by creating their Elementary Management Cycle (EMC) models.

1st stage of designing a management knowledge model and a knowledge base. Designing a knowledge base on the basis of the Detailed Value Chain Model (DVCM) requires to evaluate not only DVCM structure, but also all the informational attributes

(data) of interactions between management Functions and Processes:

1. Analyze the general case of the Detailed Value Chain Model (DVCM) which is the formalized DVCM structure.

2. Verify that management Function hierarchy exists.

3. Verify that Process hierarchy exists.

4. Verify that data (information attributes) related to each intersection between a management Function and a Process exists.

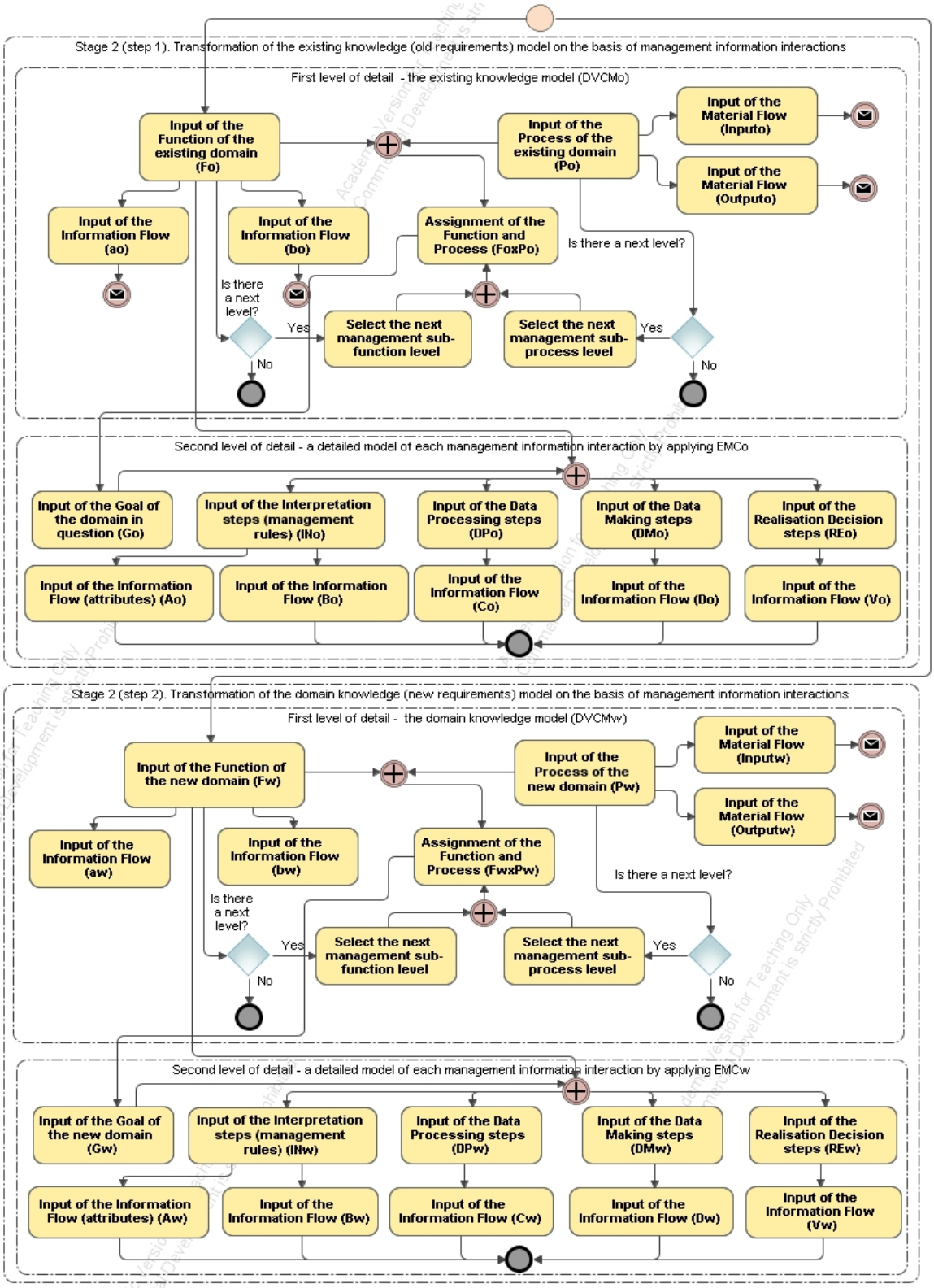


Figure 6. Part of the method (Stage 2 "Transformation of the existing and the domain knowledge models on the basis of management information interactions", BPMN notation)

2nd stage of designing a management knowledge model and a knowledge base. The resulting management knowledge model does not assess the internal structure (informational transformations performed by the enterprise management Function itself) of management Functions, i.e. information interactions between the structural parts (components) of a management Function. Next, the internal structure of each management Function is modelled on the basis of Elementary Management Cycle (EMC) description:

1. Select a specific management Function and Process pair (intersection) identified with DVCM, that will further be modelled as a managed process (formally described EMC).

2. Create an Elementary Management Cycle (EMC) model of the selected management Function and Process intersection by identifying (naming) all EMC components: IN, DA, SP and RE processes, related goals G, information flows between EMC components.

3. Illustrate DVCM and EMC through domain entity class model.

4. Develop a prototype of a knowledge base.

When requirements documents are created and linked together while the requirements themselves are kept in the Knowledge Base (KB), an opportunity appears to automate requirements analysis by using computerized systems (Figure 7). Within the selected CASE tool (IBM Rational RequisitePro™) special visual environments (so called views) for knowledge analysis are selected. It is possible to review knowledge (requirements) presented in different views in parallel by using various matrices or hierarchical structures (trees) where requirements with their attributes and/or traceability links between different requirement types are represented. For example, one of the problems of requirement analysis is the determination of requirements attributes.

When applying the visual environment of Attribute Matrix, it is possible to comfortably revise all the requirements of a particular type and attributes related to them. When analyzing a current domain (for example, study programme), the following actions with the matrix are performed: the establishment and editing of a requirement title, text, attributes and traceability links; the saving of matrix query; matrix printing. Filtering and sorting functions can also be applied to analyze requirements, thus maximizing the informational value of each requirement. By choosing one or several attributes and/or traceability signs (indicators), it is possible to perform requirement filtering and sorting, therefore, selecting the necessary requirements or forming new categories of them. In particular, it is used to design the architecture of a study programme and it's structural components (study blocks and modules).

Modification of a structured Value Chain Model to identify educational domain knowledge content. To reflect these needs the authors have proposed a new problem domain life cycle model, which is based on modified VCM. This specialised model is presented in Figure 8.

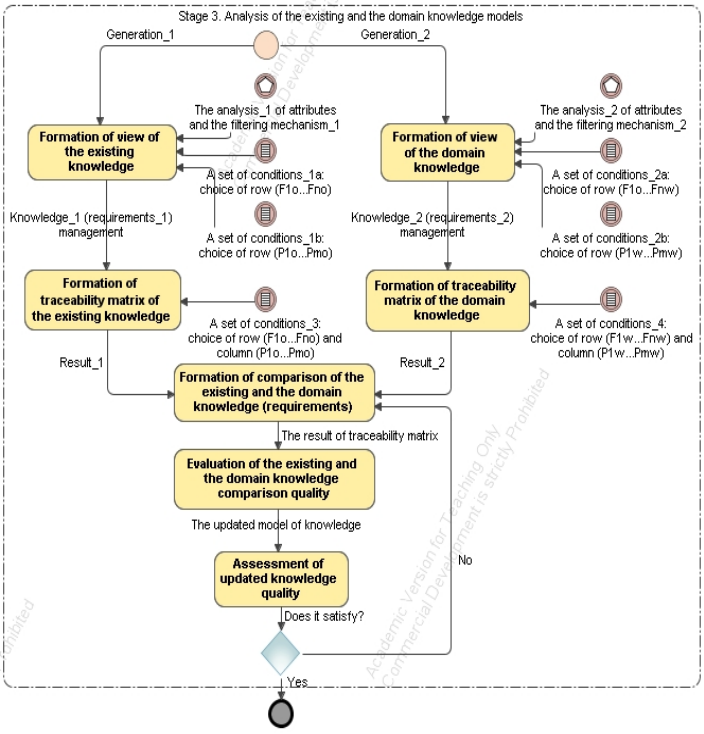


Figure 7. Part of the method (Stage 3 "Analysis of the existing and the domain knowledge models", BPMN notation)

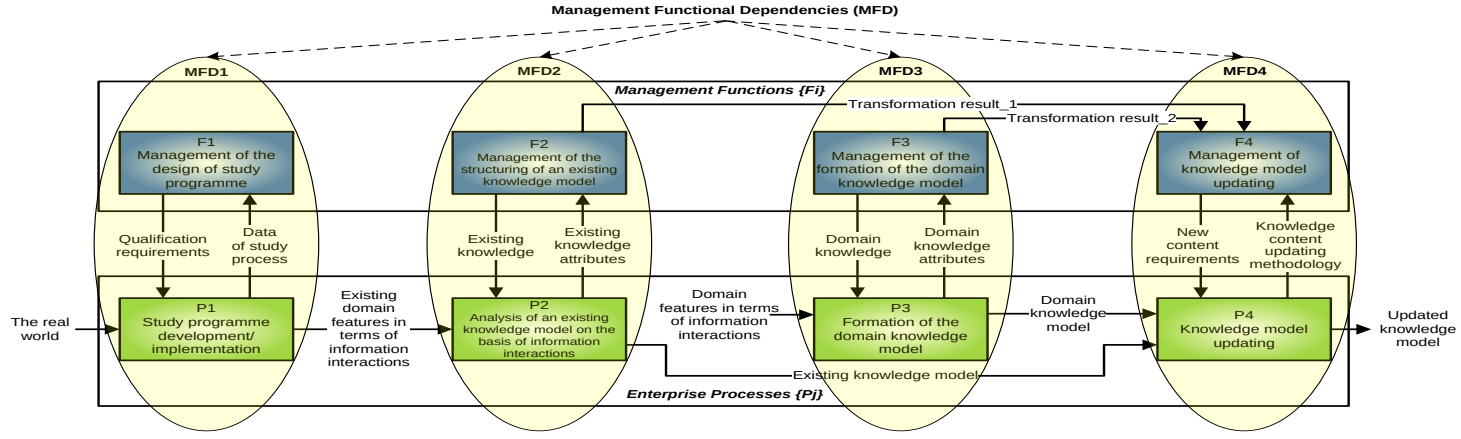


Figure 8. Informational model of educational domain knowledge content identification process

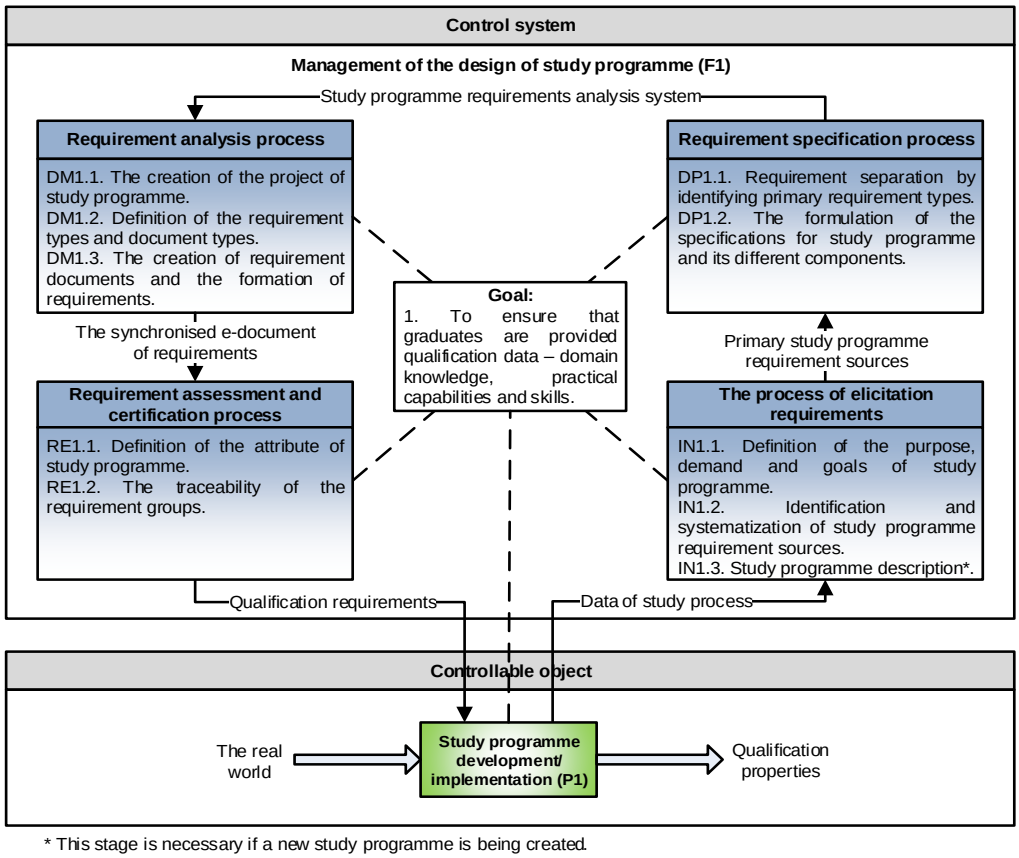


Figure 9. The Formalised Model of the Interaction between the *Function F1 = "Management of the design of study programme"* and the *Process P1 = "Study programme development/ implementation"*

Each interaction between management Function Fi and Process Pj $(Fi \times Pj)$ identified with DVCM is seen as a managed process that implements a specific Management Functional Dependence (MFD).

For example, the formalised model of the interaction between the Function F1 = "Management of the design of study programme" and the Process P1 = "Study programme development/ implementation" (see Fig. 8) is presented in Figure 9.

It is important to mention that any specific EMC does not end, it continues with each new domain process cycle. Thus, the current domain is developed based upon the spiral principle with the renewal of the quality in general. In particular, the steps (rules) and the attributes (data) of the transformation of current domain management information were identified.

4. Educational Domain Management Knowledge Content Identification and Knowledge Updating Computerized Process Model

This section provides a principal scheme of knowledge content identification and educational domain knowledge updating model, on the basis of which a knowledge-based specification is performed. A particular enterprise Management Functional Dependency is the aggregate of the necessary information interactions among specific enterprise management Function components, specified by a particular Elementary Management Cycle.

Construction of a principal scheme of knowledge content identification and domain knowledge updating model. A principal knowledge content identification and domain knowledge updating model scheme is provided in Figure 10.

When analysing the principal knowledge model scheme, it can be stated that management Function model, in this case, becomes a scenario that defines mandatory information interactions, necessary to computerize all tasks that constitute management Function.

The identification of the existing (old) knowledge which are documented before and the new subject area knowledge related with problem domain is presented in two upper DVCM in Figure 11.

Mapping and comparison of the old and the new knowledge allow to construct a modified DVCM producing a set of updated knowledge.

A technique of knowledge analysis is used for comparison between $DVCM^o$ and $DVCM^w$. Comparison of these two (($DVCM^o$ and $DVCM^w$) domain models allow to identify the changes of problem domain content (i.e. identify the actual changes of knowledge about domain) and to construct a derivative new Detailed VCM ($DVCM^N$) encompassing all actual domain knowledge components. Afterwards the $DVCM^N$ is used for identification of new knowledge requirements. Traceability Matrix shows the relation of two knowledge sets (Figure 12).

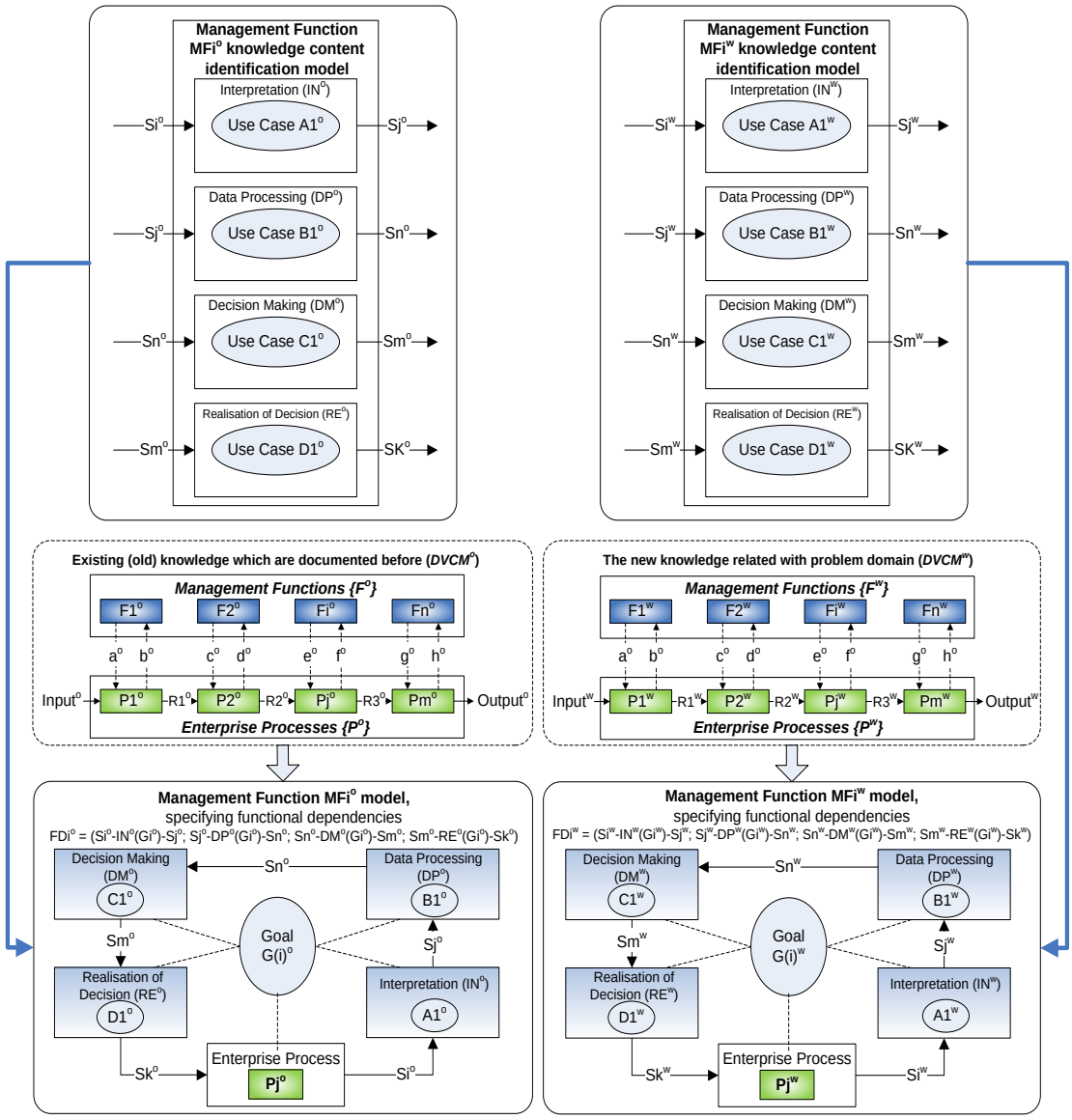


Figure 10. A principal scheme of knowledge content identification and domain knowledge updating model (based on enterprise management function)

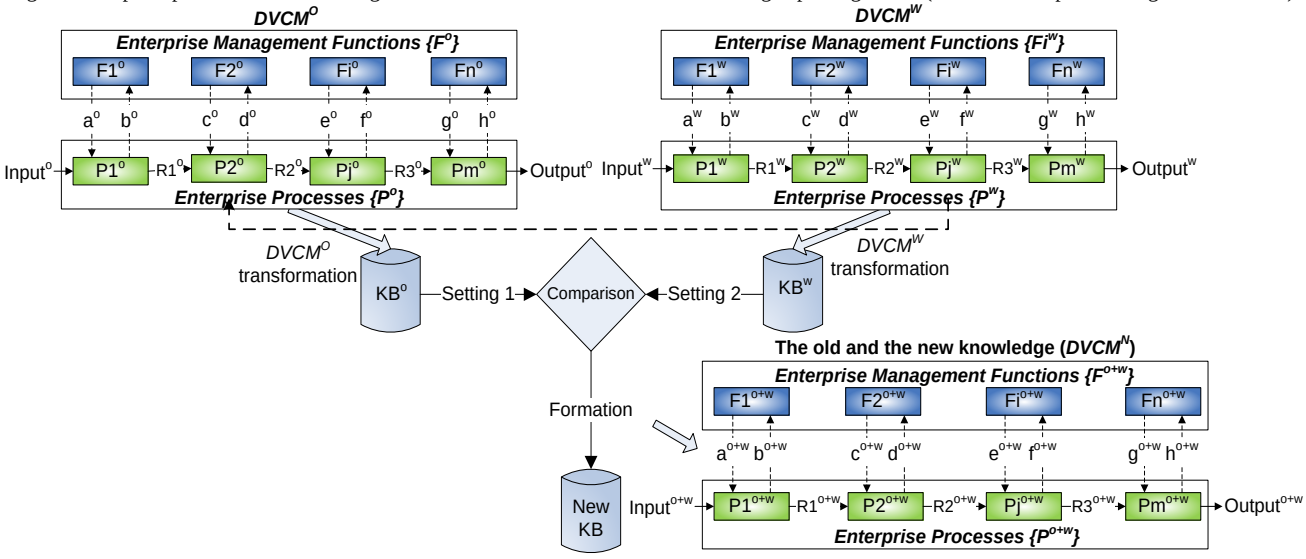


Figure 11. The identification of the existing and the new knowledge (here: $DVCM^o$ – the old knowledge; $DVCM^w$ – the new knowledge; $DVCM^N$ – the renewed domain content model))

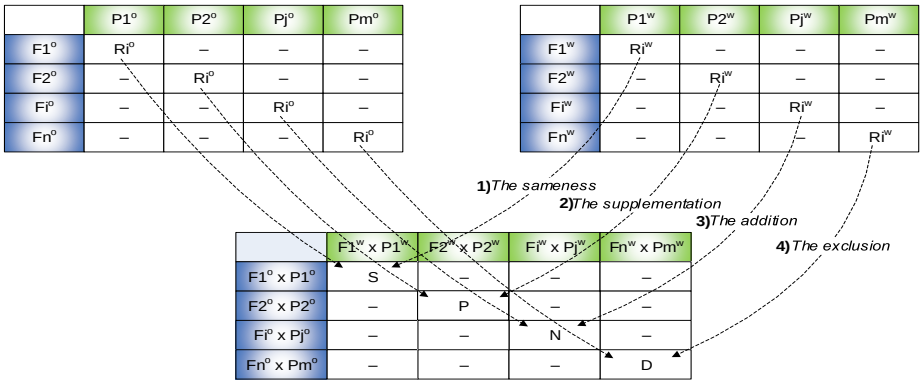


Figure 12. The relations between existing (old) and the new knowledge (requirements) (where $R_1^o, R_2^o, ..., R_i^o$, $R_1^w, R_2^w, ..., R_i^w$ – the existing and new managed process results)

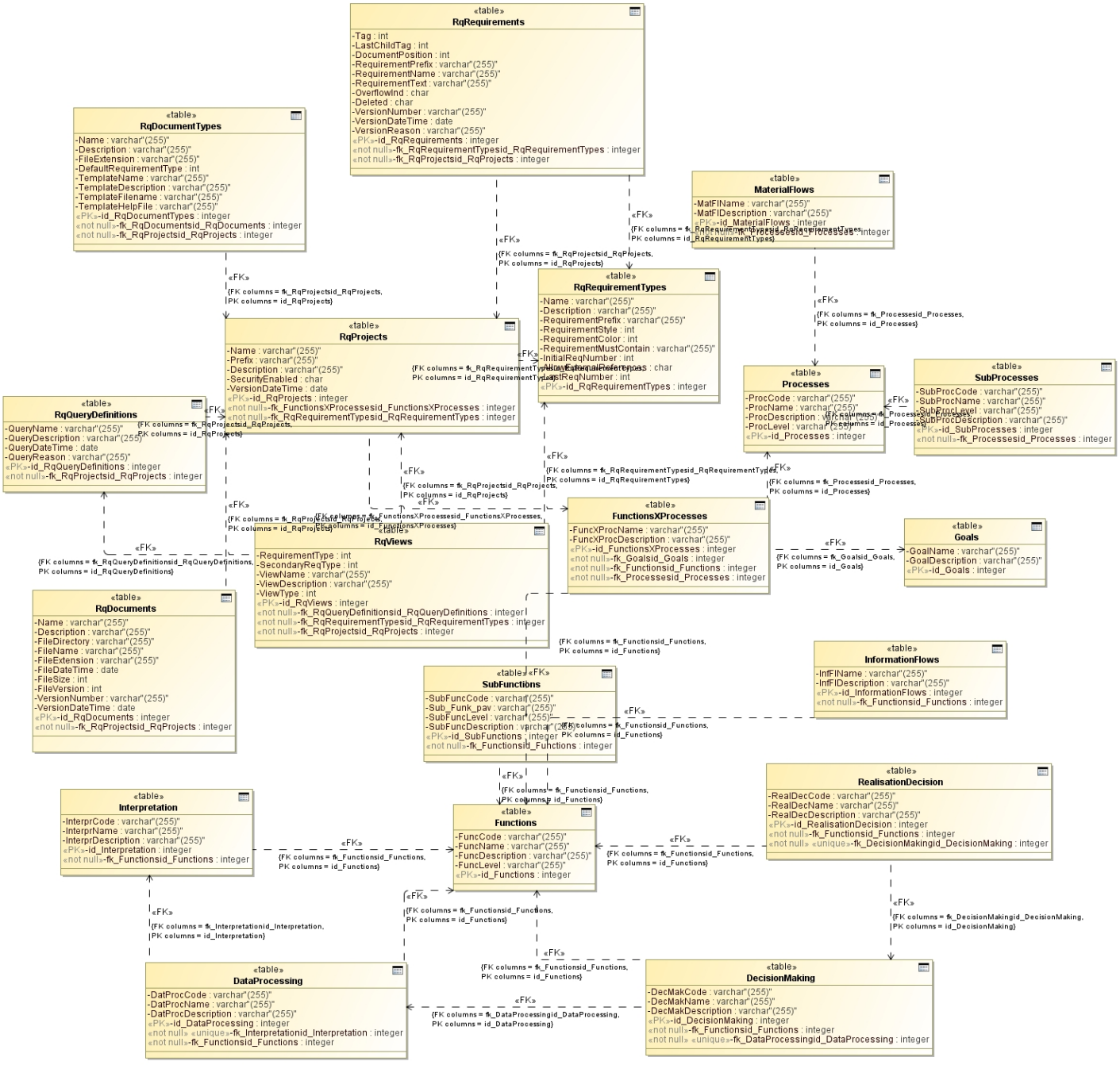


Figure 13. Subject area entity class model (physical data model)

There are several situations that can be identified: 1) the sameness of knowledge (equal requirements, knowledge remains unchanged, denoted by S); 2) the supplementation of existing knowledge with lacking new subject area knowledge (current knowledge only partially covers new requirements and should be supplemented, denoted by P); 3) the addition (inclusion) of new domain knowledge (denoted by N); 4) the exclusion of present knowledge from the current domain (no longer meets subject area requirements, competency becomes less important, denoted by D).

Implementation of the knowledge repository. Implementation of the knowledge repository of proposed model for current domain requirement improvement and knowledge content updating. Subject area entity class model (physical data model) is presented in Figure 13 (using MagicDraw^{TM} CASE tool). The solution is implemented by extending the requirement management system by knowledge-based components.

Requirement management is performed on the basis of IBM Rational RequisitePro^{TM} tool, meanwhile, the structure of a knowledge base is based on Detailed Value Chain and Elementary Management Cycle.

Subject area entity class model of the knowledge repository specifies DVCM and EMC elements which can be a function (consisting of information activities: Interpretation, Data Processing, Decision Making and Decision Realization), a process, a material flow, etc. Thus, the current domain parameters derived from DVCM and EMC together with requirements are stored in the knowledge databases and can be used for solving specific problems of expected application areas.

5. Conclusions

After conducting theoretical and experimental research, the following scientific and practical conclusions were drawn:

After analysing domain modelling methodologies and methods (languages, notations), it was found that a knowledge based enterprise modelling method for identifying organization's operational objectives, as well as the informational interaction between a controlled object and a controlling system forming control feedback, can be based on a Detailed Value Chain Model.

The Detailed Value Chain Model (DVCM) is defined as a problem domain content model. The peculiarity of the DVCM is refinement of the Management Functions {Fi} and Processes {Pj} as an obligatory component types of problem domain knowledge content. Management Functions (F) possess information transformations; while Enterprise Processes (P) are concrete material transformations.

The educational domain management knowledge modelling method was designed, allowing to create a two-level (granular) model for describing knowledge of domain management information interactions: the highest level Detailed Value Chain Model is further elaborated into a set of Elementary Management Cycle models; the chosen hierarchical structure allows to create new structural knowledge models and to update existing ones.

The work goes beyond formal presentation; the method is applied in a specific domain, which requires constant knowledge content updating – higher education study programme construction and updating. The flexibility of the proposed solutions should be noted – the possibility to integrate the created domain knowledge base into a study programme requirement management system, by identifying knowledge management processes and implementing a knowledge subsystem in an existing CASE system environment. Its approval provided results relevant and useful for the members, employees and professionals of academic community in higher education institutions, working on improving existing study programs and seeking to ensure their quality.

Conflict of Interest

The authors declare no conflict of interest.

Important note of applied research

This paper contributes to the applied scientific research at Klaipeda State University of Applied Sciences “Application of Enterprise Management Models in Knowledge-based Information Systems Engineering”. Area: sustainable environmental development; Field: Innovative Technologies; Branch (topic): Smart Information Technology, TMV-027.

References

[1] J. Tekutov, V. Denisovas, S. Gudas, and J. Smirnova, “The Refinement of Curricula Content Based on a Problem Domain Model,” ISMSIT 2018: 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies. Kızılcahamam / Ankara / Turkey, October 19-21, 2018, p. 689–694, 2018.

[2] B. Trask, and A. Roman, “Introduction to Model Driven Development with Examples using Eclipse Frameworks,” ACM Conference on Object-Oriented Programming, Systems, Languages and Applications, Orlando, 2009.

[3] L.C. Christensen, B.W. Johansen, N. Midjo, J. Onarheim, T.G. Syvertsen, and T. Totland, “Enterprise modelling – practices and perspectives,” http://www.idi.ntnu.no/grupper/su/publ/pdf/asme95.pdf. 1995.

[4] S. Gudas, “Knowledge-Based Enterprise Framework: A Management Control View, New Research on Knowledge Management Models and Method,” http://www.intechopen.com/books/new-research-on-knowledge-management-models-and-methods/knowledge-based-enterprise-framework-a-management-control-view. 2012.

[5] S. Gudas, Information systems engineering theory. Vilnius: Vilnius University Press, pp. 25-65, 2012.

[6] A. Caplinskas, “Curricula engineering: application of systems engineering methods to the development of university curricula,” Information technology and control, vol. 22, no. 1, pp. 53–58, 2002.

[7] J. Hagerty, R.L. Sallam, and J. Richardson, “Magic Quadrant for Business Intelligence Platforms,” Gartner Inc., http://www.gartner.com/technology/reprints.do?id=1-1982NPD&ct=120208&st=sb. 2012.

[8] D. Henschen, “BI and Information Management Trends,” InformationWeek, http://www.umsl.edu/~sauterv/DSS/research-2012-bi-and-information-management_9951311.pdf2011. 2011.

[9] A. Andreescu, and M. Mircea, “Managing Knowledge as Business Rules,” Informatica Economică, vol. 13, no. 4, pp. 63–74, 2009.

[10] M.E. Porter, Competitive Advantage: Creating and Sustaining Superior Performance, 1998.

[11] S. Gudas, and A. Lopata, “The enterprise model based on identification of information resources,” Information Sciences, vol. 19, pp. 43–50, 2001.

[12] S. Gudas, “Enterprise knowledge modelling: domains and aspects,” Technological and Economic Development of Economy, vol. 15, no. 2, pp. 281–293, 2009.

[13] A. Caplinskas, and O. Vasilecas, “Modern curricula in information systems: a case study,” Information technology and control, vol. 22, no. 1, pp. 59–63, 2002.

[14] V. Denisovas, S. Gudas, and J. Tekutov, “Study program requirements engineering method and information system,” Information Sciences, vol. 53, pp. 106–126, 2010.

[15] J. Tekutov, S. Gudas, V. Denisovas, and J. Tekutova, “The Study Programme Requirements Enhancement Based on a Problem Domain Model,” INTEL-EDU 2012: 3rd International Workshop on Intelligent Educational Systems and Technology-enhanced Learning, pp. 108–120, 2012.

[16] J. Tekutov, S. Gudas, and V. Denisovas, "The Refinement of Study Program Content Based on a Problem Domain Model," Transformations in Business & Economics, vol. 11, no. 1 (25), pp. 199–212, 2012.
[17] A. Bolt, M. de Leoni, and W.M.P. van der Aalst, "Process variant comparison: Using event logs to detect differences in behavior and business rules," Information Systems, vol. 74, p. 53–66, 2018.
[18] J. Henderson, and N. Venkatraman, "Strategic alignment: A model for organization transformation via information technology," pp. 3223–3290, 1990.
[19] K. Mertins, and R. Jochem, "Architectures, methods and tools for enterprise engineering," International Journal of Production Economics, vol. 98, no. 2, pp. 179–188, 2005.
[20] C.A. Millette, "Status of department of defense architecture framework (DoDAF) implementation within the aeronautical systems center (ASC)," Air force institute of technology, http://www.stormingmedia.us/11/1134/A113434.html. 2005.
[21] OMG consortium, "Business modelling specifications," http://www.omg.org/spec/. 2009.
[22] A.-W. Scheer, and K. Schneider, "ARIS – Architecture of Integrated Information Systems," Handbook on Architectures of Information Systems, Part Three, pp. 605–623, 2006.
[23] I. Wowczko, "Skills and Vacancy Analysis with Data Mining Techniques," Informatics, vol. 2, p. 31–49, 2015.
[24] G.M. Giaglis, "A Taxonomy of Business Process Modelling and Information Systems Modelling Techniques," International Journal of Flexible Manufacturing Systems, vol. 13, no. 2, pp. 209–228, 2001.
[25] A. Morkevičius, and S. Gudas, "Enterprise Knowledge Based Software Requirements Elicitation," Information technology and Control, vol. 40, no. 3, pp. 181–190, 2011.
[26] N.U. Ahmed, and S.K. Sharma, "Porter's value chain model for assessing the impact of the internet for environmental gains," Management and Enterprise Development, vol. 3, no. 3, pp. 278–295, 2006.
[27] R. Bandarian, "Exploiting value chain process concepts in research organisations," International Journal of Value Chain Management, vol. 2, no. 3, pp. 400–416, 2008.
[28] D. Beard, D. Schwieger, and K. Surendran, "A Value Chain Approach for Attracting, Educating, and Transitioning Students to the IT Profession," Information Systems Education Journal, vol. 8, no. 7, pp. 1–12, 2010.
[29] Y-L. Chen, T.-C. Yang, and Z-S. Lin, "A study on the modelling of Knowledge value chain," Knowledge Management, pp. 1–12, 2004.
[30] F. Cummins, "Value Chain Modelling: Linking Customer Value to Business Process Design and 2010 Automation," Cutter IT Journal, vol. 23, no. 2, pp. 23–30, 2010.
[31] S. Gudas, and R. Brundzaitė, "Enterprise knowledge modelling based on modified value chain model," Information Sciences, vol. 35, pp. 179–192, 2005.
[32] S. Gudas, A. Lopata, and T. Skersys, "Approach to enterprise modelling for information systems engineering," Informatics, vol. 16, no. 2, pp. 175–192, 2005.
[33] S. Gui, and Z. Ye, "The Research of Value Chain Operation Model in E-Commerce," International Seminar on Business and Information Management, pp. 238–241, 2008.
[34] M.C. Lee, and M.W. Han, "Knowledge Value Chain Model Implemented for Supply Chain Management Performance," Fifth International Joint Conference on INC, IMS and IDC, pp. 606–611, 2009.
[35] R. Landry, N. Amara, A. Pablos-Mendes, R. Shademani, and I. Gold, "The knowledge-value chain: a conceptual framework for knowledge translation in health," Bulletin of the World Health Organization, vol. 84, no. 8, pp. 597–602, 2006.
[36] A. Najmaei, and Z. Sadeghinejad, "Competitive Strategic Alliances Through Knowledge Value Chain," International Review of Business Research Papers, vol. 5, no. 3, p. 297–310, 2009.
[37] V. Pathak, and K. Pathak, "Reconfiguring the higher education value chain," Management in Education, vol. 24, no. 4, pp. 166–171, 2010.
[38] A. Saha, "Mapping of Porter's value chain activities into business functional units," Management Innovation eXchange, http://www.managementexchange.com/hack/mapping-porter%E2%80%99s-value-chain-activities-business-functional-units. 2012.
[39] G. Stewart, "Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply-chain management," Logistics Information Management, vol. 10, no. 2, p. 62–67, 1997.
[40] E. Sweeney, "Supply Chain Management and the Value Chain. Supply Chain Perspectives," Journal of the National Institute for Transport and Logistics, vol. 10, no. 2, p. 13–15, 2009.
[41] E. A. Williamson, D. K. Harrison, and M. Jordan, "Information Systems Development Within Supply Chain Management," International Journal of Information Management, vol. 24, no. 4, p. 375–385, 2004.
[42] H.K. Wong, "Knowledge Value Chain: Implementation of New Product Development System in a Winery," The Electronic Journal of Knowledge Management, vol. 2, no. 1, pp. 109–122, 2004.
[43] S. Gudas, and A. Lopata, "Towards internal modelling of the information systems application domain," Informatica, vol. 27, no. 1, p. 1–29, 2016.
[44] J. Tekutov, "Model of Domain Knowledge Content Updating Based on Management Information Interactions" PhD dissertation, Dept. of Informatics, Vilnius Univ., 2013.
[45] J. Tekutov, S. Gudas, and V. Denisovas, "The Formation of Requirements Engineering Processes Based on a Problem Domain Model," Proceedings in Conference of Informatics and Management Sciences (ICTIC-2018). Zilina, Slovak Republic, vol. 7, no. 1, p. 93–97, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 141-146 (2018)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Optimal Discrete-time Sliding Mode Control for Nonlinear Systems Subject to Input Constraints

Olfa Jedda[*,1], Ali Douik[2]

[1]*National Engineering School of Monastir, University of Monastir, Monastir, Tunisia*

[2]*National Engineering School of Sousse, University of Sousse, Sousse, Tunisia*

ARTICLEINFO

Article history:
Received: 04 March, 2019
Accepted: 01 July, 2019
Online: 22 July, 2019

Keywords:
Discrete-time sliding mode control
Particle swarm optimization
Control input constraints
Chattering phenomenon
Inverted pendulum system

ABSTRACT

In this paper, an optimal discrete-time sliding mode control is proposed for single-input single-output nonlinear systems with input constraints. The sliding surface is designed on the basis of particle swarm optimization algorithm in order to optimize the system response characteristics while ensuring the follow-up of reference model in presence of constraints. Moreover, the controller is developed such that the elimination of chattering phenomenon, the finite-time convergence and the stability of the closed-loop system are guaranteed. Performed on an inverted pendulum system, simulation results demonstrate the effectiveness of the proposed approach over the discrete-time sliding mode controller using the saturation function and the discrete-time second order sliding mode controller in terms of fast response.

1 Introduction

This paper is an extension of work originally presented in 2018 15th International Multi-Conference on Systems, Signals and Devices (SSD) [1]. Due to the increasing use of computers in control applications, discrete-time sliding mode control (DSMC) has been extensively developed since its appearance with Milosavljevic in 1985 [2]; see, for example, [3–9]. However, the finite sampling time negatively impacts the robustness of continuous-time sliding mode control to external disturbances, parametric uncertainties and modeling errors [10–12]. In fact, the control input remains constant during the sampling period so that it can not be changed when the trajectory of system state crosses the sliding surface. This is at the origin of chattering phenomenon that may badly affect control devices and system performances [13].

Introduced by Gao in 1995 [14], the quasi-sliding mode concept consists to drive the state trajectory to cross the sliding hyperplane in finite-time, to move with a non-increasing zigzag motion and to remain within a specified layer. Thus, the above-mentioned drawback can be reduced but not eliminated. Several approaches was proposed in the literature to overcome it. In [15–17], the saturation function is used as smooth one instead of the sign function which is the origin of discontinuity. In [18–20], discrete-time second order sliding mode control (DSOSMC) was proposed for linear and nonlinear systems. In [21, 22], a piecewise-constant control is used to generate the discrete-time sliding mode. It ensures that the chatter effect is avoided and the system state trajectory convergences to the sliding surface after a finite-time interval.

In the following, an optimal discrete-time sliding mode control (ODSMC) is proposed for single-input single-output (SISO) nonlinear systems with input constraints. Particle swarm optimization (PSO) algorithm is employed to search the sliding vector to which corresponds the optimal response characteristics of the closed-loop system and its behavior follows the reference model. Developed by Eberhart and Kennedy in 1995 [23], PSO algorithm is based on food searching behavior of animals societies having no leaders in their swarm such as bird flocks and fish schools. Its main advantages are the easiness of implementation, the robustness in controlling parameters and the good efficiency in solving nonlinear, nondifferentiable and large search space problems [24, 25]. Regarding the control law, it is designed in such a way that chattering phenomenon is eliminated, the convergence to the sliding manifold is ensured in finite-time and the system dynamics are stable while respecting constraints.

The efficiency of the proposed controller will be demonstrated by applying it to an inverted pendulum system and by comparing it to DSMC controller using the saturation function and to DSOSMC controller with an arbitrary choice of control parameters and without taking into account the

[*]Corresponding Author: Olfa Jedda, olfa_jedda@outlook.com

input constraints.

This paper is organized as follows. The DSMC and DSOSMC controllers are developed for SISO nonlinear systems in sections 2 and 3. Section 4 is devoted for the design of the proposed ODSMC controller. The inverted pendulum system is represented in section 5. Section 6 illustrates numerical simulation results. Concluding remarks are given in section 7.

2 Discrete-time Sliding Mode Control

Consider a class of discrete-time SISO nonlinear system described by

$$\begin{cases} x(k+1) = F(x(k)) + H(x(k))\,u(k) \\ y(k) = Cx(k) \end{cases} \tag{1}$$

where $x(k) \in \mathbb{R}^{n\times 1}$ is the state vector, $u(k) \in \mathbb{R}$ is the control input, and $y(k) \in \mathbb{R}$ is the system output. $F(x(k)) \in \mathbb{R}^{n\times 1}$ and $H(x(k)) \in \mathbb{R}^{n\times 1}$ are vectors of nonlinear functions and C is the output matrix. Let's admit that H and its pseudo-inverse H^{+} are both bounded.

The sliding function is defined as follows

$$s(k) = c^T e(k) \tag{2}$$

where $e(k) = x(k) - x_d(k)$ is the tracking error and $c^T \in \mathbb{R}^{1\times n}$ is the sliding vector chosen such that the sliding dynamic $s(k) \equiv 0$ is asymptotically stable.

The reaching law is given by

$$s(k+1) = (1 - qT_s)\,s(k) - \varepsilon T_s \,\mathrm{sat}(s(k), \varphi) \tag{3}$$

where $0 < 1 - qT_s < 1$, $\varepsilon > 0$, T_s is the sampling period and sat is the saturation function defined as follows

$$\mathrm{sat}(s, \varphi) = \begin{cases} \frac{s}{\varphi} & \text{if } \left|\frac{s}{\varphi}\right| \le 1 \\ \mathrm{sign}(s) & \text{else} \end{cases} \tag{4}$$

with $\varphi > 0$ is the boundary layer width of s and sign is the sign function.

Using the reaching law (3) and the forward expression of the sliding function (2), the control law $u(k)$ is given by

$$\begin{aligned} u(k) = \left(c^T H(x(k))\right)^{-1} (& (1 - qT_s)\,s(k) \\ & - \varepsilon T_s \,\mathrm{sat}(s(k), \varphi) - c^T F(x(k)) + c^T x_d(k+1)) \end{aligned} \tag{5}$$

3 Discrete-time Second Order Sliding Mode Control

The sliding function defined in [19, 20] is given by

$$\sigma(k) = s(k) + \beta s(k-1) \tag{6}$$

with $\beta \in [0, 1[$ in order to ensure the stability of $\sigma(k)$.

The control law is given by

$$u(k) = u_{eq}(k) + u_d(k) \tag{7}$$

where $u_{eq}(k)$ is the equivalent control used to force the system state to evolve on the sliding manifold and $u_d(k)$ is the discontinuous control used to ensure the robustness.

Setting $\sigma(k+1) = \sigma(k) = 0$, the equivalent control $u_{eq}(k)$ is determined from the following relation

$$s(k+1) + \beta s(k) = 0 \tag{8}$$

Hence, It is given by

$$\begin{aligned} u_{eq}(k) = \left(c^T H(x(k))\right)^{-1} (& - c^T F(x(k)) \\ & + c^T x_d(k+1) - \beta s(k)) \end{aligned} \tag{9}$$

The discontinuous control $u_d(k)$ is expressed as follows

$$u_d(k) = u_d(k-1) - \varepsilon T_s \mathrm{sign}(\sigma(k)) \tag{10}$$

4 Optimal Discrete-time Sliding Mode Control

4.1 Design of Sliding Surface

The sliding vector c^T is determined using the PSO algorithm in order to ensure the optimization of the closed-loop system response characteristics and the follow-up of the reference model in the presence of constraints on control input.

PSO algorithm starts with population of np n-dimensional particles. The i^{th} particle of the swarm has a position $x_i = [x_{i,1}, \ldots, x_{i,n}]$ and a velocity $v_i = [v_{i,1}, \ldots, v_{i,n}]$. Its previously best visited position and the global best particle in the swarm are denoted by $pbest_i = [pbest_{i,1}, \ldots, pbest_{i,n}]$ and $gbest$ respectively. The velocity and the position of each particle are updated as follows

$$v_{i,j} = w\,v_{i,j} + \alpha_1 r_1 \left(pbest_{i,j} - x_{i,j}\right) + \alpha_2 r_2 \left(gbest_j - x_{i,j}\right) \tag{11}$$

$$x_{i,j} = x_{i,j} + v_{i,j} \tag{12}$$

with $i = 1, \ldots, np$ and $j = 1, \ldots, n$. w is the inertia weight, $\alpha_{1,2}$ are positive constant so-called the acceleration coefficients and $r_{1,2}$ are random parameters uniformly distributed within $[0, 1]$ at each generation.

PSO is used to generate the sliding vector that minimizes the following cost function

$$cost = \lambda_1 t_r + \lambda_2 t_s + \lambda_3 M_p + \lambda_4 E_{ss} \tag{13}$$

where t_r is the rise-time, t_s is the settling-time, M_p is the overshoot, E_{ss} is the steady-state error and λ_q , $q = 1, \ldots, 4$, are their corresponding weights.

As illustrated in Table 1, the iterative process is repeated until stopping criterion is met.

Table 1: Description of PSO algorithm

Step 1	**Parameter settings**
	Set the population size np, the minimum and maximum bounds $x_{min} = [x_{min,1}, \ldots, x_{min,n}]$ and $x_{max} = [x_{max,1}, \ldots, x_{max,n}]$, and the acceleration coefficients $\alpha_{1,2}$.
Step 2	**Initialization**
	Initialize the particle's position x_i^0, $i = 1 \ldots np$, randomly and uniformly within bounds.
	Initialize the particle's velocity v_i^0 and the particle's best position to its initial position, i.e. $pbest_i^0 = x_i^0$.
	Set $gbest$ equal to the global best particle.
Step 3	**PSO algorithm**
	WHILE stopping criterion is not met DO
	Update the particle's velocity and position
	FOR $i = 1$ to np
	FOR $j = 1$ to n
	$v_{i,j} = w\, v_{i,j} + \alpha_1 r_1 \left(pbest_{i,j} - x_{i,j}\right) + \alpha_2 r_2 \left(gbest_j - x_{i,j}\right)$
	END FOR
	$x_i = x_i + v_i$
	END FOR
	Update the particle's best position
	FOR $i = 1$ to np
	IF $f(x_i) \prec f(pbest_i)$
	$pbest_i = x_i$
	END IF
	END FOR
	Update the global best position
	FOR $i = 1$ to np
	IF $f(pbest_i) \prec f(gbest)$
	$gbest = pbest_i$
	END IF
	END FOR
	END WHILE

4.2 Design of Control Law

The control law is designed such that the chatter effect is avoided, the state trajectory convergences to the sliding surface after a finite-time interval and the closed-loop system is stable while respecting control constraints.

The forward expression of the sliding function (2) can be rewritten as follows

$$\begin{aligned} s(k+1) = s(k) &+ c^T \left(x_d(k) - x_d(k+1)\right) \\ &+ c^T \left(F(x(k)) - x(k)\right) + c^T H(x(k)) u(k) \end{aligned} \tag{14}$$

Setting $s(k+1) = 0$, the equivalent control u_{eq} is expressed by

$$\begin{aligned} u_{eq}(k) = -\left(c^T H(x(k))\right)^{-1} &\left(s(k) + c^T \left(x_d(k) - x_d(k+1)\right)\right. \\ &\left. + c^T \left(F(x(k)) - x(k)\right)\right) \end{aligned} \tag{15}$$

Consider that $\|u(k)\| \leq u_{max}$, the constrained control is given by

$$u(k) = \begin{cases} u_{eq}(k) & \text{if } \left\|u_{eq}(k)\right\| \leq u_{max} \\ \frac{u_{eq}(k)}{\left\|u_{eq}(k)\right\|} u_{max} & \text{else} \end{cases} \tag{16}$$

Suppose that

$$\left\|c^T \left(x_d(k) - x_d(k+1)\right) + c^T \left(F(x(k)) - x(k)\right)\right\| \leq \delta,$$
$$\left\|\left(c^T H(x(k))\right)^{-1}\right\| \leq \eta$$

and

$$u_{max} > \delta\eta \tag{17}$$

with $\delta > 0$ and $\eta > 0$.

It follows that

$$\begin{aligned} &\left\|u_{eq}(k)\right\| \leq \left\|\left(c^T H(x(k))\right)^{-1}\right\| \\ &\left\|s(k) + c^T \left(x_d(k) - x_d(k+1)\right) + c^T \left(F(x(k)) - x(k)\right)\right\| \\ &\qquad \leq \eta \left(\|s(k)\| + \delta\right) \end{aligned} \tag{18}$$

For $\|u(k)\| > u_{max}$, the forward expression of the sliding function (14) is expressed by

$$\begin{aligned} &s(k+1) = s(k) + c^T \left(x_d(k) - x_d(k+1)\right) \\ &+ c^T \left(F(x(k)) - x(k)\right) + c^T H(x(k)) \frac{u_{eq}(k)}{\left\|u_{eq}(k)\right\|} u_{max} \\ &= \left(s(k) + c^T \left(x_d(k) - x_d(k+1)\right)\right. \\ &\left. + c^T \left(F(x(k)) - x(k)\right)\right) \left(1 - \frac{u_{max}}{\left\|u_{eq}(k)\right\|}\right) \end{aligned} \tag{19}$$

For stability analysis, the Lyapunov function is chosen as follows

$$V(k) = \|s(k)\| \tag{20}$$

Thus, the Lyapunov difference is given by

$$\Delta V(k) = \|s(k+1)\| - \|s(k)\| \tag{21}$$

Using (17)

$$\begin{aligned} &\|s(k+1)\| \\ &= \left\|s(k) + c^T \left(x_d(k) - x_d(k+1)\right) + c^T \left(F(x(k)) - x(k)\right)\right\| \\ &\left(1 - \frac{u_{max}}{\left\|u_{eq}(k)\right\|}\right) \\ &\leq \|s(k)\| + \left\|c^T \left(x_d(k) - x_d(k+1)\right) + c^T \left(F(x(k)) - x(k)\right)\right\| \\ &- \frac{u_{max}}{\left\|\left(c^T H(x(k))\right)^{-1}\right\|} \\ &\leq \|s(k)\| + \delta - \frac{u_{max}}{\left\|\left(c^T H(x(k))\right)^{-1}\right\|} \\ &< \|s(k)\| \end{aligned} \tag{22}$$

Hence, $s(k)$ decreases monotonically. From (18), the equivalent control $u_{eq}(k)$ will belong to the constrained domain, i.e. $\left\|u_{eq}(k)\right\| \leq u_{max}$, in finite time and therefore it will bring the system trajectory to the sliding manifold $s(k) = 0$ on which the dynamics of the closed-loop system are stable.

5 Inverted Pendulum System

As shown in Figure 1, the inverted pendulum system consists of a pendulum amounted on a cart at a frictionless pivot point. The pendulum is of mass $m = 0.1$kg and length $2l = 1$m and the cart is of mass $M = 1$kg [26]. u is the force applied to the cart that lies within the range of ± 10N and $g = 9.81\text{m/s}^2$ is the gravity acceleration. The generalized coordinates are x and θ which represents the horizontal movement of cart and the rotation of pendulum respectively [27].

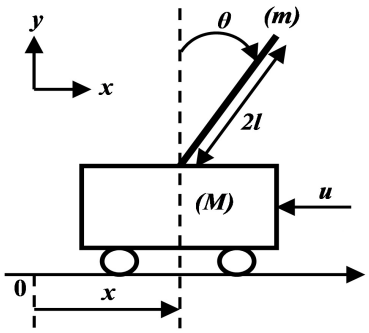


Figure 1: Schematic of the inverted pendulum system

Let $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T = \begin{bmatrix} \theta & \dot{\theta} \end{bmatrix}^T$ be the state vector. The dynamic equations are given by

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x_1, x_2) + h(x_1, x_2)\, u \\ y = x_1 \end{cases} \tag{23}$$

with

$$f(x_1, x_2) = \frac{(M+m)\, g \sin x_1 - mlx_2^2 \sin x_1 \cos x_1}{\frac{4}{3}(M+m)\, l - ml\cos^2 x_1}$$

and

$$h(x_1, x_2) = \frac{\cos x_1}{\frac{4}{3}(M+m)\, l - ml\cos^2 x_1}$$

Using for discretization the Euler forward method defined by

$$\dot{x} \cong \frac{x(k+1) - x(k)}{T_s},$$

the discrete-time model of the inverted pendulum system is expressed as follows

$$\begin{cases} x_1(k+1) = x_1(k) + T_s x_2(k) \\ x_2(k+1) = x_2(k) + T_s f(x_1(k), x_2(k)) \\ \qquad\qquad + T_s h(x_1(k), x_2(k))\, u(k) \\ y(k) = x_1(k) \end{cases} \tag{24}$$

6 Numerical Simulation Results

For $T_s = 0.05$s, the reference model is chosen as follows

$$x_d(k) = \begin{bmatrix} \frac{\pi}{30} \sin(kT_s) & \frac{\pi}{30} \cos(kT_s) \end{bmatrix}^T \tag{25}$$

The initial conditions are given by

$$x(0) = \begin{bmatrix} 0.2 & 0 \end{bmatrix}^T, x_d(0) = \begin{bmatrix} 0 & \frac{\pi}{30} \end{bmatrix}^T \tag{26}$$

The sliding vector and the reaching law parameters for DSMC controller are respectively

$$\begin{gathered} c^T = \begin{bmatrix} 5 & 1 \end{bmatrix}, \\ q = 2,\ \varepsilon = 10,\ \varphi = 0.5 \end{gathered} \tag{27}$$

The parameters of DSOSMC controller are chosen as follows

$$\varepsilon = 0.02,\ \beta = 0.5 \tag{28}$$

Using the PSO algorithm, the cost function to minimize is the following

$$cost = 0.9t_s + 0.1t_r \tag{29}$$

The population size, the acceleration coefficients, and the minimum and maximum bounds of particle's position are given respectively by

$$\begin{gathered} np = 100,\ \alpha_{1,2} = 2, \\ c_{min}^T = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix},\ c_{max}^T = \begin{bmatrix} 10 & 10 \end{bmatrix} \end{gathered} \tag{30}$$

The following linear decreasing inertia weight is used [28, 29]

$$w = w_{max} - (w_{max} - w_{min})\, \frac{iter}{iter_{max}} \tag{31}$$

where $w_{max} = 0.9$ and $w_{min} = 0.4$ are the initial and final values of the inertia weight respectively, $iter$ is the current iteration, and $iter_{max}$ is the maximum number of iterations.

Figure 2 illustrates the evolution of the cost function (29) corresponding to ODSMC controller. It shows that PSO algorithm ensures a rapid convergence of the cost function to its minimum value of 0.236s corresponding to the following sliding vector

$$c^T = \begin{bmatrix} 7.085 & 0.5 \end{bmatrix} \tag{32}$$

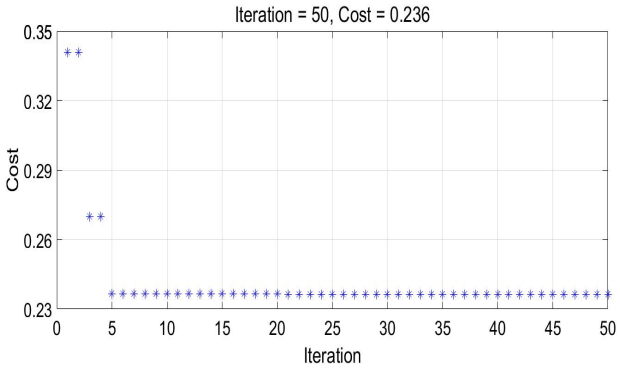
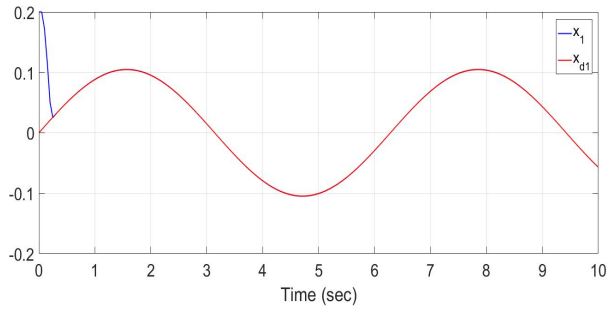
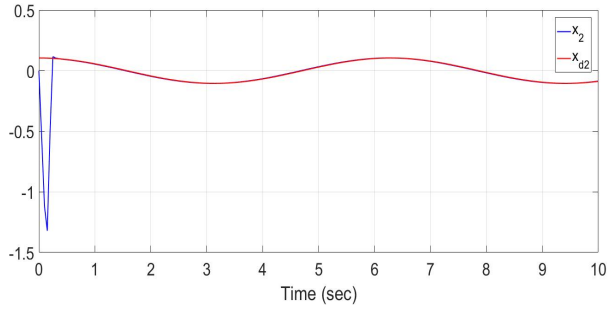


Figure 2: Evolution of the cost function for ODSMC

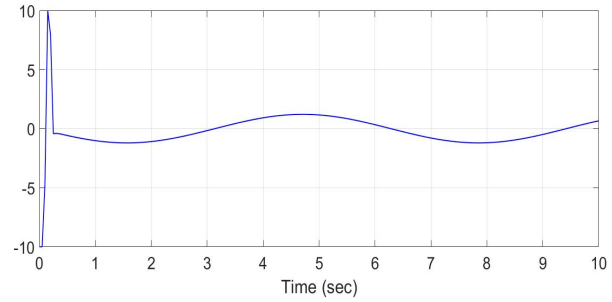
Figure 3 shows numerical simulation results of ODSMC controller using the sliding vector (32). Figures 3a-3b present the state variables $x_{1,2}(k)$ and their corresponding references $x_{d1,2}(k)$. They show that the developed controller ensures the follow-up of the reference model and the stability of the closed-loop system in the presence of control input constraints. Figure 3c depicts the constrained control input $u(k)$. It shows that chatter effect is avoided. Figure 3d illustrates the sliding function $s(k)$. It shows that the state trajectory converges to the sliding surface in finite time.



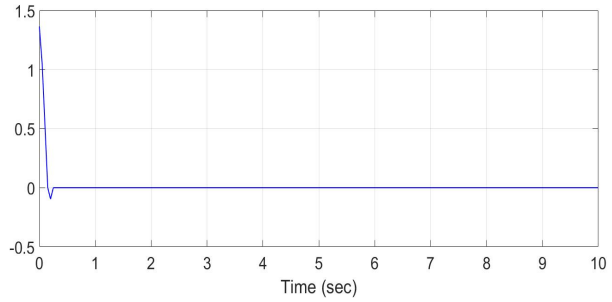
(a) State variable $x_1(k)$ and its reference $x_{d1}(k)$



(b) State variable $x_2(k)$ and its reference $x_{d2}(k)$



(c) Control input $u(k)$

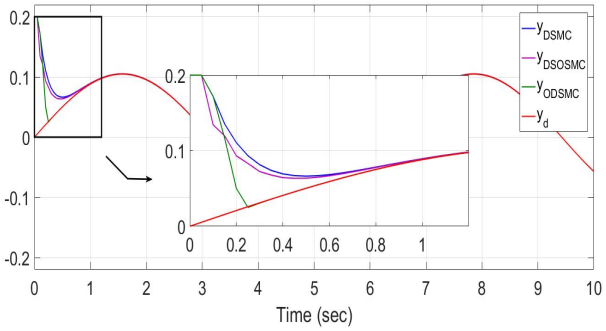


(d) Sliding function $s(k)$

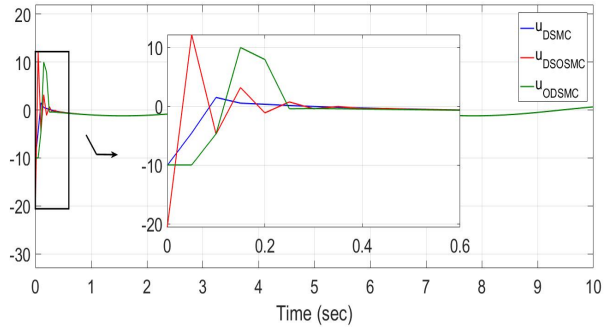
Figure 3: Simulation results for ODSMC controller.

Figure 4 illustrates a comparison between DSMC with saturation function, DSOSMC and ODSMC controllers. Figure 4a presents a comparison of system responses $y(k)$. It shows that all controllers ensure the follow-up of the reference model. Yet, the ODSMC controller ensures the fastest response time while respecting input constraints. Figure 4b depicts a comparison of control inputs $u(k)$. It shows the effectiveness of the applied controllers in avoiding the chattering phenomenon. Moreover, the developed algorithm has lower values of control input in the initial phase than DSOSMC controller.

The results are summarized in Table 2. Actually, It shows that the developed ODSMC ensures the best results with the least rise and settling times of 0.14s and 0.24s respectively, and with lower values of control input than DSOSMC controller whose corresponding minimum and maximum input values are −20.6N and 12.18N respectively. Figure 4 and Table 2 demonstrate that ODSMC controller outperforms DSMC and DSOSMC controllers.



(a) Comparison of system responses



(b) Comparison of control inputs

Figure 4: Comparison between DSMC with saturation function, DSOSMC and ODSMC controllers.

Table 2: Summary table of numerical results

	Settling time (s)	Rise time (s)	u_{min}	u_{max}
DSMC	1.07	0.39	-10.07	1.5
DSOSMC	1.03	0.37	-20.6	12.18
ODSMC	0.24	0.14	-10	10

In this study, Matlab 2016a was used for the implementation of all algorithms.

7 Conclusion

This work presents an optimal discrete-time sliding mode controller for nonlinear SISO systems subject to input constraints. The particle swarm optimization algorithm is employed to determine the sliding vector for which the response characteristics are optimal and the closed-loop system model follows the reference model. Furthermore, the control law which is designed based on the equivalent control concept guarantees the avoidance of chattering phenomenon and the finite-time convergence of system trajectory to the sliding

manifold on which the system dynamics are stable while respecting constraints. Simulation results demonstrate the efficiency of the developed controller in ensuring the fastest response comparing to discrete-time sliding mode controller using the saturation function and to discrete-time second order sliding mode controller. Future work will be to develop the proposed controller to multi-input multi-output nonlinear systems and to verify its effectiveness by experiments.

Conflict of Interest The authors declare no conflict of interest.

References

[1] O. Jedda, A. Douik, "Optimal Discrete-Time Sliding Mode Control for Nonlinear Systems" in 15th International Multi-Conference on Systems, Signals & Devices (SSD), Hammamet Tunisia, 2018. 10.1109/SSD.2018.8570376

[2] C. Milosavljevic, "General Conditions for the Existence of a Quasi-Sliding Mode on the Switching Hyperplane in Discrete Variable Structure Systems" Autom. Remote Control, 14(3), 307-317, 1985.

[3] K. Furuta, "Sliding Mode Control of a Discrete System" Systems & Control Letters, 14(2), 145-152, 1990. https://doi.org/10.1016/0167-6911(90)90030-X

[4] G. Golo, C. Milosavljevic, "Robust Discrete-time Chattering Free Sliding Mode Control" Systems & Control Letters, 41(1), 19-28, 2000. https://doi.org/10.1016/S0167-6911(00)00033-5

[5] S. Z. Sarpturk, Y. Istefanopulos, O. Kaynak, "On the Stability of Discrete-time Sliding Mode Control Systems" IEEE Trans. Autom. Control, 32(10), 930-932, 1987. 10.1109/TAC.1987.1104468

[6] Y. Niu, D. W. Ho, Z. Wang, "Improved sliding mode control for discrete-time systems via reaching law" IET Control Theory Appl., 4(11), 2245-2251, 2010. 10.1049/iet-cta.2009.0296

[7] J. Hu, Z. Wang, H. Gao, L. K. Stergioulas, "Robust H sliding mode control for discrete time-delay systems with stochastic nonlinearities" J. Frankl. Inst., 349(4), 1459-1479, 2012. https://doi.org/10.1016/j.jfranklin.2011.05.018

[8] M. C. Pai, "Discrete-time sliding mode control for robust tracking and model following of systems with state and input delays" Nonlinear Dyn., 76(3), 1769-1779, 2014. https://doi.org/10.1007/s11071-014-1245-0

[9] H. Du, X. Yu, M. Z. Chen, S. Li, "Chattering-free discrete-time sliding mode control" Automatica, 68, 87-91, 2016. https://doi.org/10.1016/j.automatica.2016.01.047

[10] B. Bandyopadhyay, F. Deepak, K. S. Kim, Sliding Mode Control Using Novel Sliding Surfaces, Springer, 2009.

[11] I. U. VADIM, "Survey paper variable structure systems with sliding modes" IEEE Trans. Autom. Control, 22(2), 1977.

[12] K. D. Young, V. I. Utkin, U. Ozguner, "A Control Engineers Guide to Sliding Mode Control" IEEE Trans. Control Syst. Technol., 7(3), 328-342, 1999. 1063-6536(99)03275-3

[13] W. Perruquetti, J. P. Barbot, Sliding Mode Control in Engineering, M. Dekker, 2002.

[14] W. Gao, Y. Wang, A. Homaifa, "Discrete-time Variable Structure Control Systems" IEEE Trans. Ind. Electron., 42(2), 117-122, 1995. 10.1109/41.370376

[15] J.-h. Kim, S.-H. Oh, J. K. Hedrick, "Robust Discrete-time Variable Structure Control Methods" J DYN SYST-T ASME, 122(4), 766-775, 2000. 10.1115/1.1320448

[16] J.-h. Kim, D.-i. D. Cho, "Discrete-time Variable Structure Control using Recursive Switching Function" in Proceedings of the 2000 American Control Conference, Chicago IL USA, 2000. 10.1109/ACC.2000.876673

[17] O. Jedda, A. Douik, "Discrete-time Integral Sliding Mode Control with Anti-windup" Studies in Informatics and Control, 27(4) 413-422, 2018. https://doi.org/10.24846/v27i4y201805

[18] A. Y. Alanis, N. Arana-Daniel, C. Lopez-Franco, E. N. Sanchez, "PSO-gain Selection to Improve a Discrete-time Second Order Sliding Mode Controller" in IEEE Congress on Evolutionary Computation, Cancun Mexico, 2013. 10.1109/CEC.2013.6557672

[19] M. Mihoub, A. S. Nouri, R. B. Abdennour, "The Multimodel Approach for a Numerical Second Order Sliding Mode Control of Highly non Stationary Systems" in American Control Conference, Seattle WA USA, 2008. 10.1109/ACC.2008.4587240

[20] M. Mihoub, A. S. Nouri, R. B. Abdennour, "Real-time Application of Discrete Second Order Sliding Mode Control to a Chemical Reactor" Control Eng. Pract., 17(9), 1089-1095, 2009. https://doi.org/10.1016/j.conengprac.2009.04.005

[21] G. Bartolini, A. Ferrara, V. I. Utkin, "Adaptive Sliding Mode Control in Discrete-time Systems" Automatica, 31(5), 769-773, 1995. https://doi.org/10.1016/0005-1098(94)00154-B

[22] B. Castillo-Toledo, S. Di Gennarob, A.G. Loukianov, J. Rivera, "Discrete time Sliding Mode Control with Application to Induction Motors" Automatica, 44(12), 3036-3045, 2008. doi:10.1016/j.automatica.2008.05.009

[23] J. Kennedy, R. Eberhart, "Particle Swarm Optimization" in Proceedings of IEEE International Conference on Neural Networks, Perth Australie, 1995. 10.1109/ICNN.1995.488968

[24] J. C. Bansal, P. Singh, M. Saraswat, A. Verma, S. S. Jadon, A. Abraham, "Inertia Weight Strategies in Particle Swarm Optimization" in 3rd World Congress on Nature and Biologically Inspired Computing, Salamanca Spain, 2011. 10.1109/NaBIC.2011.6089659

[25] D. P. Rini, S. M. Shamsuddin, S. S. Yuhaniz, "Particle Swarm Optimization: Technique, System and Challenges" Int. J. Comput. Appl., 14(1), 19-26, 2011. 10.5120/ijais-3651

[26] L. X. Wang, "A Supervisory Controller for Fuzzy Control Systems that Guarantees Stability" IEEE Trans. Autom. Control, 39(9), 1845-1847, 1994. 10.1109/9.317109

[27] O. Jedda, J. Ghabi, A. Douik, "Sliding Mode Control of an Inverted Pendulum" In Applications of Sliding Mode Control, Springer Singapore, 105-118, 2017.

[28] J. Xin, G. Chen, Y. Hai, "A Particle Swarm Optimizer with Multistage Linearly-Decreasing Inertia Weight" in International Joint Conference on Computational Sciences and Optimization (CSO), Sanya Hainan China, 2009. 10.1109/CSO.2009.420

[29] Y. Shi, R. C. Eberhart, "Empirical Study of Particle Swarm Optimization" in Proceedings of the 1999 Congress on Evolutionary Computation, Washington DC USA, 1999. 10.1109/CEC.1999.785511

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 147-151 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Multi-Restricted Area Avoidance Scenario Using Hybrid Dynamical Model and Its Predictive Controller

Sutrisno, Widowati*, R. Heru Tjahjana, Sunarsih, Kartono

Department of Mathematics, Diponegoro University, Semarang, 50275, Indonesia

ARTICLE INFO

Article history:
Received: 29 April, 2019
Accepted: 04 July, 2019
Online :21 July, 2019

Keywords:
hybrid system
mixed-logical dynamic
multi-restricted area avoidance,
piecewise-affine
predictive control

ABSTRACT

This article is addressed to show the results of hybrid dynamical modeling in the form of PWA (piecewise-affine) and equivalent MLD (mixed-logical dynamical) model for multi-restricted areas avoidance of an autonomous system. It is a problem of determining the optimal moving trajectory from plant's initial position to some desired position while avoiding some restricted areas (obstacles) between them. In order to calculate the optimal input value capable of generating the optimal trajectory, the model predictive control (MPC) approach was utilized by minimizing an objective function of state/output prediction subject to the formulated hybrid dynamical model. To illustrate the formulated model and its responses, some computational simulations were performed in a three-dimensional state using two/three box-shape restricted areas. From the simulation results, the optimal trajectory was achieved, and the plant avoided the restricted area.

1. Introduction

Dynamical equation models, which comprise of the linear, complex, and hybrid dynamical system, play an important role in engineering the control systems. There are thousands of published research articles developed to analyze the dynamical model of some new engineering systems such as the mobile robot [1, 2], autonomous vehicles, car-like robot, etc. This research deals with an independent system with some known initial state and its corresponding output value with plant's initial position. The plant utilized moves to a decided state known as target position with minimal "effort" where some restricted states are not allowed to be passed through by the plant. The term "effort" in some cases is defined as the shortest path, while an obstacle is a restriction in space movement which should be avoided by the plant. The pioneer mathematical model utilized in this state was developed in [3] by formulating a piecewise affine model which corresponds to the restricted and normal sets. There are some published articles which described the restricted use of some systems, such as vehicles and mobile robots [4–9]. The more complicated problem comprises of several plants which are controlled by applying a multi-agent concept like flocking scheme, which was used in [10,11].

In some cases, the objectives of restricted area avoidance are not only avoiding the obstacle but also determining the optimal trajectory used to determine the final or target point. In this problem, an optimal control method based on mathematical optimization was implemented to solve the technique. For example, a particle swarm algorithm was applied in [12,13]. It is reasonable to utilize an optimization-based method because it will generate the ace result to the problem. Beside of optimal control problem, numerous inconsistencies were solved using the optimization approach which was also used to describe its profitability such as facility location optimization and the colony algorithm for knapsack.

In the system theory, a newly developed strategy is the hybrid dynamical model which comprises of different types of Piecewise-affine (PWA), discrete hybrid automata (DHA) and Mixed Logical Dynamic (MLD) models [14]. To analyze and control a hybrid model, in [15], a toolbox was developed which comprises of some MATLAB functions on model formulation and controlling. For example, the PWA model written in HYSDEL programming language can be converted into MLD using the MATLAB routine "*mld"* in the hybrid system toolbox. Furthermore, the MLD model which consists of trajectory tracking problems tends to be solved by applying a classic control method scheme MPC (model predictive control) and modifying the state prediction along with its corresponding objective function which was carried out in [16,17]. Many research articles applied this control method in agriculture field [18,19], as well as in controlling mechanical vehicles [20], boiler-turbine [21], and spacecrafts [22].

This research therefore aims at solving the problem associated with the restricted area inherent the three-dimensional states. First, the PWA model was formulated to determine whether the dynamical system of the plant is in a normal or restricted area.

*Corresponding Author: Widowati, Jalan Prof. Soedarto, SH. Tembalang, Semarang, Indonesia. Email: widowati@lecturer.undip.ac.id

Next, the PWA system is converted into equivalent MLD using HYSDEL and hybrid toolbox. Furthermore, by using predictive control method for the MLD model, the optimal input was generated to obtain the moving trajectory, which was initialized at the plant's starting point to its target position. Some computational simulations are performed to illustrate and visualize the results.

2. Dynamical System

Let vector $x(k) = \left(x_1(k), x_2(k), \cdots, x_n(k)\right)' \in \mathsf{R}^n$ denotes the state of a plant where k denotes the time instantly. Therefore, the dynamical model of the observed plant is a linear time-invariant system modeled as illustrated in the equation below

$$\left.\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) \end{aligned}\right\} \tag{1}$$

where $u \in \mathsf{R}^p$ and $y \in \mathsf{R}^m$ are input & output vectors respectively, and the notations *A, B, C,* and *D* are real constant matrices. The control method used in this paper is applicable, the controllable and observable assumptions are held by (1).

2.1. Restricted Area Avoidance Scheme

The position to the output vector *y is defined without losing the generality property*. Let the initial position of the plant is obtained by $y(0) = y_0 = Cx(0) = Cx_0$. The value of the output, i.e., plant's position *y* have to maneuver and reach some desired target position denoted by y_f which corresponds to target state x_f, $y_f = Cx_f$, where in the output's domain, some sets such as $R_1, R_2, \ldots, R_r \subset \mathbb{R}^m$ are not allowed to be utilized by *y*. Let $R = \left(R_1 \cup R_2 \cup \cdots \cup R_r\right) \subset \mathsf{R}^m$, *y* is restricted to be in *R,* then $y \notin R$ should be held. To handle this condition, the dynamics of the system is formulated as a hybrid system.

The formula is illustrated as follows, w.l.o.g., let $y(k) \in \mathbb{R}^3$ with two restricted sets, R_1 and R_2 illustrated in Figure 1. The problem is how to determine the optimal trajectory used by the plant to maneuver (or move) from its initial state to the target position. The term "optimal" is interpreted as minimal effort (or energy or work or other similar things) used by the plant. The optimal trajectory shown in Figure 1 illustrates a moving trajectory from the initial to the target point.

The non-restricted sets are known as the normal area where the dynamics of the plant corresponds to (1). Alternatively, the dynamics which corresponds to the restricted set is defined as

$$\left.\begin{aligned} x(k+1) &= x(k) \\ y(k) &= Cx(k) \end{aligned}\right\} \tag{2}$$

which means that the dynamical model is used to prevent the plant from being located in the restricted area. The formulation of the hybrid dynamical model where the plant is prevented from entering the restricted area is illustrated by Figure 2 by assuming $x(k) \in \mathbb{R}^3$ and $y(k) = x(k)$.

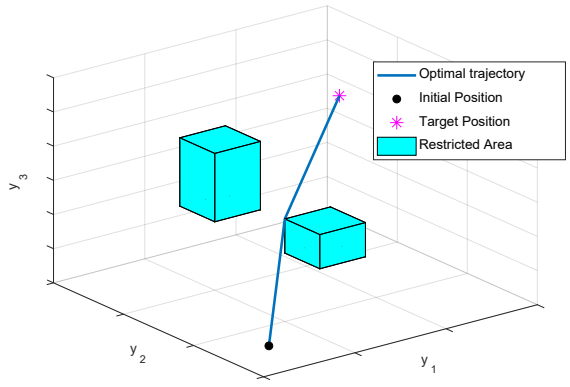


Figure 1: Two box-shape restricted areas illustration

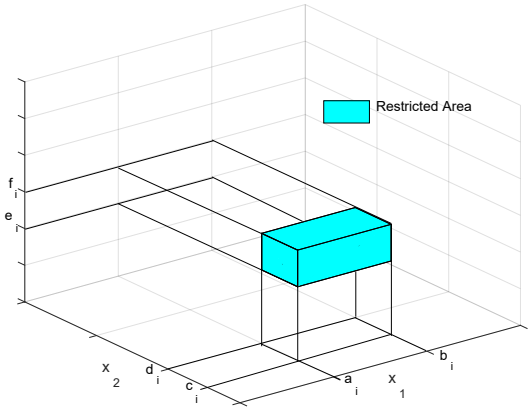


Figure 2: Restricted area labeling

Let the restricted sets (area) R_i, $i = 1, 2, \ldots, k$, and the normal set (area) denoted by *N,* then these sets are written as

$$\begin{cases} R_i : \{(x_1, x_2, x_3) : a_i \le x_1 \le b_i, c_i \le x_2 \le d_i, e_i \le x_3 \le f_i\} \\ N \ : \text{otherwise.} \end{cases}$$

2.2. PWA to MLD Hybrid Model

The hybrid model in the PWA model of the plant for restricted area avoidance purposes is formulated as

$$\left.\begin{aligned} x(k+1) &= \begin{cases} Ix(k), & x \in \text{restricted area} \\ Ax(k) + Bu(k), & x \in \text{normal area} \end{cases} \\ y(k) &= Cx(k) \end{aligned}\right\} \tag{3}$$

with *I* denotes an identity matrix with the appropriate dimension. To apply the predictive control method to this system, the MLD model is, first of all, converted into the form of

$$\left.\begin{aligned} x(k+1) &= Ax(k) + B_1u(k) + B_2\delta(k) + B_3z(k) \\ y(k) &= Cx(k) + D_1u(k) + D_2\delta(k) + D_3z(k) \\ E_2\delta(k) + E_3z(k) &\le E_1u(k) + E_4x(k) + E_5 \end{aligned}\right\} \tag{4}$$

with some initial state $x_0 = x(0)$ where $z(k)$ is an auxiliary state. The notation $\delta(k)$ is a binary valued state describing the mode of the system (mode 0 assuming it is on the normal area and 1 when restricted). The matrices A, B_i, C, D_i and E_i for all i are real constant generated by the conversion process, which is conducted using *mld* MATLAB function embedded in hybrid system toolbox given in [23] by writing the PWA system in HYSDEL, then generating the matrices for the equivalent MLD model.

2.3. Predictive Control Approach

The equation used to determine the optimal input to enable the output vector (position) reach the target point using minimal effort is represented as a terminal state optimal control problem. Furthermore, the predictive control approach is used to obtain the optimal input by letting $x(k|t)$ as the state value predicted at time instant $(k+t)$ which is resulted by applying input value $u(k+t)$ into equation (4) where the corresponding output value is predicted at time instant $(k+t)$, $y(k|t)$. The optimal input will be calculated by solving the minimal value "cost" function of output prediction as follows:

$$\min_{[u,\delta,z]^T} J = \left\| Q_y y(T|t) - y_f \right\|_w + \sum_{k=0}^{T-1} \left[\left\| Q_u u(k) \right\|_w + \left\| Q_y y(k|t) - y_f \right\|_w \right] \tag{5}$$

subject to:

$$\begin{cases} x(0|t) = x(t); y(0|t) = y(t); \\ y(k+1|t) = Cx(k|t) + D_1 u(k) + D_2 \delta(k|t) + D_3 z(k|t); \\ E_2 \delta(k|t) + E_3 z(k|t) \le E_1 u(k) + E_4 x(k|t) + E_5; \\ u_{\min} \le u(k+t) \le u_{\max}, k = 0,1,2,3,...,T-1; \\ x_{\min} \le x(k+t|t) \le x_{\max}, k = 0,1,2,3,...,T; \\ y_{\min} \le y(k+t|t) \le y_{\max}, k = 0,1,2,3,...,T-1; \end{cases}$$

where T is called the horizon control period, Q_u and Q_y are symmetric and positive definite matrices used to weight the input u and output y respectively. These symmetric and positive definite properties are applied to guarantee the objective function J is convex. This is expressed in the notation $\|Qy\|_{w=2} = y^T Qy$ where $\|Qy\|_{w=\infty} = \|Qy\|_\infty$. This predictive control scheme resulting in a mixed integer quadratic optimization problem and in our simulation, *miqp* MATLAB function, which is also embedded in hybrid system toolbox, is utilized to solve. Finally, the optimal values of $u(k)$ for all k are used by the system. For restricted area purposes, the term y_f in (5) is the final/target position where the dynamics of x is (4) which equivalents to (3).

3. Simulation Results

Given a plant with three-dimensional state

$$x(k) = \left[x_1(x), x_2(k), x_3(k) \right]^T \in \mathsf{R}^3$$

and output vector $y(k) = x(k)$ which can be described as the position in a three-dimensional Cartesian coordinate system. Let the initial state be $x_0 = x(0) = \left[1,1,1\right]^T$, which corresponds to the initial position $y(0) = \left[1,1,1\right]^T$.

```
SYSTEM pwa_obs_3d_robot {
INTERFACE { STATE { REAL x1 [-20,20];
                    REAL x2 [-20,20];
                    REAL x3 [-20,20]; }
            INPUT { REAL u [-10,10]; }
            OUTPUT{ REAL y1,y2,y3; }
            PARAMETER { REAL a1;         REAL a2;
                        REAL b1;         REAL b2;
                        REAL c1;         REAL c2;
                        REAL d1;         REAL d2;
                        REAL e1;         REAL e2;
                        REAL f1;         REAL f2; } }
IMPLEMENTATION { AUX { REAL z1,z2,z3;
                       BOOL da1,da2,db1,db2,dc1,dc2,
                            dd1,dd2,de1,de2,df1,df2; }
                 AD  {da1 = x1>=a1;      da2 = x1>=a2;
                      db1 = x1>=b1;      db2 = x1>=b2;
                      dc1 = x2>=c1;      dc2 = x2>=c2;
                      dd1 = x2>=d1;      d2 = x2>=d2;
                      de1 = x3>=e1;      de2 = x3>=e2;
                      df1 = x3>=f1;      df2 = x3>=f2; }
                 DA  {z1 = {IF
                     (da1&~db1)&(dc1&~dd1)&(de1&~df1)
                     THEN x1 ELSE x1+u }; z2 = {IF
                     (da2&~db2)&(dc2&~dd2)&(de2&~df2)
                     THEN x2 ELSE x2+u }; z3 = {IF
                     (da3&~db3)&(dc3&~dd3)&(de3&~df3)
                     THEN x3 ELSE x3+u }; }
                     CONTINUOUS {x1 = z1;
                                 x2 = z2;
                                 x3 = z3; }
                     OUTPUT { y1 = x1;
                              y2 = x2;
                              y3 = x3; } } }
```

Listing Code 1: PWA model (7) with two restricted areas in HYSDEL

Then the dynamic of the plant in the normal area is

$$\left.\begin{aligned} x(k+1) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} u(k) \\ y(k) &= x(k) \\ x(0) &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}\right\}. \tag{6}$$

Suppose there are two restricted sets R_1 and R_2 defined visually as two boxes

$$R_1 : \left\{ (x_1, x_2, x_3) : 3 \le x_1 \le 5, 3 \le x_2 \le 4, 3 \le x_3 \le 5 \right\}, \text{ and}$$

$$R_2 : \left\{ (x_1, x_2, x_3) : 0 \le x_1 \le 3, 7 \le x_2 \le 9, 5 \le x_3 \le 7 \right\}$$

where the rest of state space is normal, then the PWA model of this system is stated as

$$x(k+1) = \begin{cases} x(k) & , x \in R_1 \cup R_2 \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} u(k), & \text{otherwise} \end{cases} \quad (7)$$
$$y(k) = Cx(k).$$

The PWA model in HYSDEL has already been written (See Listing Code 1) and converted it into MLD model (4) resulting in the following matrices $B_2 = D_2 = zeros(3,15)$, $B_3 = C = I_3$, $A = B_1 = D_2 = D_3 = zeros(3,3)$, and E_1, E_2, E_3, E_4, and E_5 because their dimension is sufficiently large.

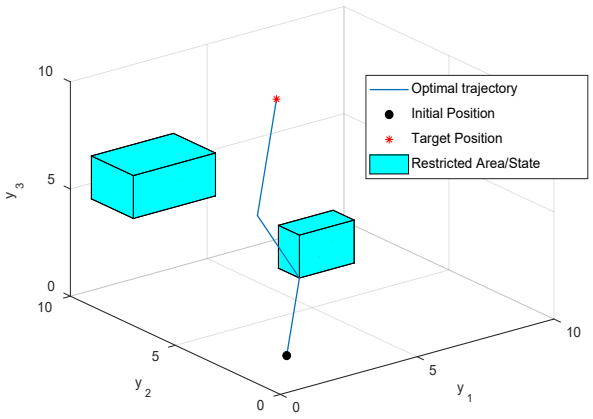


Figure 3: Optimal moving trajectory generated by the controller

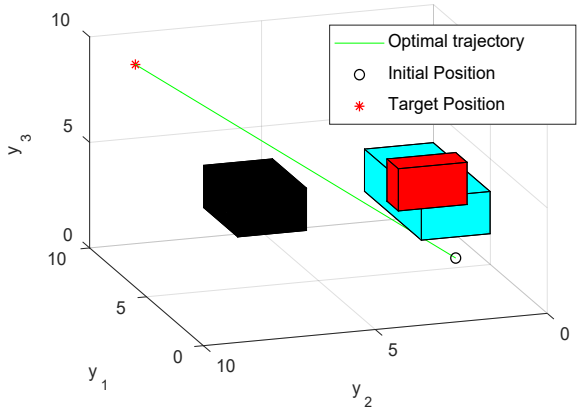


Figure 4: Optimal moving trajectory with three box-shape restricted area

By solving (5) with w=2, the optimal input values are obtained, and by applying them into (7), the optimal moving trajectory from its initial position to its target/final position as show in Figure 3 of the output values are obtained. From Figure 3, it is observed that two restricted areas were restricted and avoided by the system's moving trajectory, as illustrated in the two boxes. For further simulation, another box was added. The result is shown in Figure 4 and is similar to the optimal moving trajectory, which was generated by the controller to prevent the restricted areas.

4. Conclusions

The multi-restricted area which avoids the problem associated with the autonomous linear system was considered dynamic with the region formulated as a hybrid model and the optimal trajectory calculated. From the computational simulation, the obtained system's moving trajectory was generated by the controller, and the given restricted areas were avoided. Further research works will develop the shape of the restricted area into other shapes like polytope to control the inconsistencies associated with irregular shapes. Other control methods will be considered and compared to determine the best in performance.

Conflict of Interest

The authors declare no conflict of interest regarding this work.

Acknowledgment

This work was funded by Universitas Diponegoro under RPI 2019 research grant.

References

[1] E.S. Briskin, Y. V. Kalinin, A. V. Maloletov, N.G. Sharonov, "Mathematical Modelling of Mobile Robot Motion with Propulsion Device of Discrete Interacting with the Support Surface," IFAC-PapersOnLine, **51**(2), 236–241, 2018, doi:10.1016/j.ifacol.2018.03.041.

[2] Y.L. Kuo, "Mathematical modeling and analysis of the Delta robot with flexible links," Computers and Mathematics with Applications, **71**(10), 1973–1989, 2016, doi:10.1016/j.camwa.2016.03.018.

[3] J.J.A.M. Keij, Obstacle Avoidance with Model Predictive Control in a Hybrid Controller, 2002.

[4] Sutrisno, Salmah, E. Joelianto, A. Budiyono, I.E. Wijayanti, N.Y. Megawati, ""Model Predictive Control for Obstacle Avoidance as Hybrid Systems of Small Scale Helicopter," in 3rd International Conference on Instrumentation Control and Automation (ICA), Bali: 127–132, 2013.

[5] Z. Yan, Y. Zhao, S. Hou, H. Zhang, Y. Zheng, "Obstacle Avoidance for Unmanned Undersea Vehicle in Unknown Unstructured Environment," **2013**, 2013.

[6] R. Wang, M. Wang, Y. Guan, X. Li, "Modeling and Analysis of the Obstacle-Avoidance Strategies for a Mobile Robot in a Dynamic Environment," 2015.

[7] D.Q. Bao, I. Zelinka, "Obstacle Avoidance for Swarm Robot Based on Self-Organizing Migrating Algorithm," Procedia Computer Science, **150**, 425–432, 2019, doi:10.1016/j.procs.2019.02.073.

[8] Y. Zhao, X. Chai, F. Gao, C. Qi, "Obstacle avoidance and motion planning scheme for a hexapod robot Octopus-III," Robotics and Autonomous Systems, **103**, 199–212, 2018, doi:10.1016/j.robot.2018.01.007.

[9] Y. Gao, Y. Wu, X. Yang, C. Liao, K. Cheng, H. Luo, "Reactive obstacle avoidance of monocular quadrotors with online adapted depth prediction network," Neurocomputing, **325**, 142–158, 2018, doi:10.1016/j.neucom.2018.10.019.

[10] J. Wang, H. Zhao, Y. Bi, S. Shao, Q. Liu, X. Chen, R. Zeng, Y. Wang, L. Ha, "An Improved Fast Flocking Algorithm with Obstacle Avoidance for Multiagent Dynamic Systems," **2014**(1992), 2014.

[11] Z. Chen, L. Ding, K. Chen, R. Li, "The Study of Cooperative Obstacle Avoidance Method for MWSN Based on Flocking Control," **2014**, 2014.

[12] T. Yang, Z. Liu, H. Chen, R. Pei, "Robust Tracking Control of Mobile Robot Formation with Obstacle Avoidance," **2007**, 2007, doi:10.1155/2007/51841.

[13] G.S. Chyan, S.G. Ponnambalam, "Robotics and Computer-Integrated Manufacturing Obstacle avoidance control of redundant robots using variants of particle swarm optimization," Robotics and Computer Integrated Manufacturing, **28**(2), 147–153, 2012, doi:10.1016/j.rcim.2011.08.001.

[14] A. Bemporad, "Efficient conversion of mixed logical dynamical systems into an equivalent piecewise affine form," IEEE Transactions on Automatic Control, **49**(5), 832–838, 2004, doi:10.1109/TAC.2004.828315.

[15] A. Bemporad, Hybrid Toolbox User's Guide, Lucca, 2012.

[16] F. Borrelli, A. Bemporad, M. Morari, Predictive Control for linear and hybrid systems, Cambridge University Press, Cambridge, UK, 2014.

[17] J.M. Maciejowski, Predictive Control with Constrains, Prentice Hall, USA, 2001.

[18] Y. Ding, L. Wang, Y. Li, D. Li, "Model predictive control and its application in agriculture: A review," Computers and Electronics in Agriculture, **151**, 104–117, 2018, doi:https://doi.org/10.1016/j.compag.2018.06.004.

[19] L. Chen, S. Du, Y. He, M. Liang, D. Xu, "Robust model predictive control for greenhouse temperature based on particle swarm optimization," Information Processing in Agriculture, **5**(3), 329–338, 2018, doi:https://doi.org/10.1016/j.inpa.2018.04.003.

[20] M. Jalali, E. Hashemi, A. Khajepour, S. Chen, B. Litkouhi, "Model predictive control of vehicle roll-over with experimental verification," Control Engineering Practice, **77**, 95–108, 2018, doi:https://doi.org/10.1016/j.conengprac.2018.04.008.

[21] X. Liu, J. Cui, "Economic model predictive control of boiler-turbine system," Journal of Process Control, **66**, 59–67, 2018, doi:https://doi.org/10.1016/j.jprocont.2018.02.010.

[22] P. Li, Z.H. Zhu, "Model predictive control for spacecraft rendezvous in elliptical orbit," Acta Astronautica, **146**, 339–348, 2018, doi:https://doi.org/10.1016/j.actaastro.2018.03.025.

[23] F.D. Torrisi, A. Bemporad, "HYSDEL — A Tool for Generating Computational Hybrid Models for Analysis and Synthesis Problems," IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, **12**(2), 235–249, 2004, doi:10.1109/TCST.2004.824309.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 152-160 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Inference Rule of Collision Risk Index based on Ship Near-Collision via Adaptive Neuro Fuzzy Inference System

Ho Namgung*

Graduate School of Maritime Transportation System, Mokpo National Maritime University, Mokpo, KS005, Republic of Korea

ARTICLE INFO

Article history:
Received: 03 June, 2019
Accepted: 17 July, 2019
Online :21 July, 2019

Keywords:
Near-collision
Ship domain
Collision Risk Index
ANFIS

ABSTRACT

Collision risk index has been studied as the required quantitative values for decision-making of collision avoidance between ships. Recently, inference methods of the collision risk were proposed on the basis of the fuzzy theory because of being possible to collect data in real time. Existing fuzzy inference system was composed of only simulation results using virtual navigation situation. In this study, we obtained the fuzzy inference rule based on ship near-collision data via the adaptive neuro fuzzy inference system. Proposed fuzzy inference rule expressed various collision risk index in order that a ship could avoid collision with an encounter ship at appropriate distance and time. It would support for navigators to make an appropriate decision for collision avoidance with encounter ships.

1. Introduction

Despite the efforts to prevent accidents, marine accidents has been occurring without interruption. According to the statistics of KMST(Korean Maritime Safety Tribunal), 10,991 marine accidents have occurred in the last five years from 2014 to 2018. Out of these accidents, 1,132 marine accidents have occurred in collision resulting from lack of look-out and violation of CORLEGs(International Regulations for Preventing Collisions at Sea) [1]. Accordingly, collision accidents can cause structural hazard, loss of human life and property, and ocean pollution due to oil and cargo spills. Thus, since the demand on system to support the safe navigation of ships has been increased, collision avoidance algorithm has been proposed.

Collision avoidance algorithm starts from assessing the collision risk. In order to assess the collision risk in the early days, the concept of a ship domain proposed by [2] was used to assess the collision risk. This concept is succeeded by [3], [4] and [5] in several shapes, the methods of determining ship domains have evolved with time [6,7,8-10]. Nonetheless, it still doesn't contain time-related information.

Since it has been possible to obtain DCPA(Distance of the Closest Point of Approach) and TCPA (Time to the Closest Point of Approach) in real time, inference method of the CRI(collision risk index) on the basis of the fuzzy theory has been proposed. In [11], the author connected DCPA and TCPA to the CRI using interviews of navigators by showing virtual navigation situation on simulator. In [12], the authors reconstructed the FIS (Fuzzy Inference System) using non-dimensionalized DCPA and TCPA. In [13], the researchers calculated the CRI by including VCD (Variance of Compass Degree) into input parameter proposed by Lee and Rhee.

On the basis of the proposed FIS [11,12,13], various studies have been conducted. In [14], the author designed an estimation algorithm of the collision risk among approaching multiple ships by using the fuzzy theory, and verified performance on the basis of the AIS (Automatic Identification System) maritime traffic data. In [15], the author proposed an evaluation algorithm of the collision risk in order that VTSO (Vessel Traffic Service Operator) was able to analyze the collision risk among ships in advance. In [16], the author proposed a model of predicting ship collision risk based on the FIS considering the general patterns of collision avoidance. Furthermore, the FIS has been applied to collision avoidance system in the field of development for an USV(Unmanned Surface Vehicle). The USV developed for multipurpose of ocean observation and surveillance could flexibly change an action space according to the collision risk, which was inferred by using obstacle information on a basis of fuzzy inference [17]. In [18], the researcher used the FIS as a point of time for generating and sending messages in order that efficient information exchange for collision avoidance was possible between autonomous ship and manned ship.

But the proposed FIS [11,12,13] had the limitation which was relied on the empirical factors of navigators. Hence, in [19], the

*Corresponding Author: Ho Namgung, Email: ngh2009@mmu.ac.kr

author suggested the FIS considering ship's characteristic in virtual navigation situation on simulator without interviews of navigators. Nonetheless, because of membership functions and rules determined by simulation results, the existing FIS still have significant limitation not reflecting on information of actual collision situation between ships.

To overcome the limitation of the existing FIS [11,12,13,19], this study proposed an improved fuzzy inference rule by learning actual near-collision data extracted from the AIS via ANFIS (Adaptive Neuro Fuzzy Inference System). Section II described inference methods of the collision risk using the FIS and decision of near-collision. Section III suggested the fuzzy inference rule obtained by learning ship near-collision data via ANFIS. Subsequently, performance of the proposed FIS comparing with the existing FIS was validated and discussed in section IV. Finally, a summary of the work and the conclusion drew in section V.

2. Materials

2.1. Calculation of the CRI using FIS

The FIS inferred the CRI based on the fuzzy theory using DCPA and TCPA. DCPA referred to a minimum distance through a target ship when own-ship and a target ship were encountered. TCPA was an expected time to arrive at the point where DCPA occurred at the ship's present location. Figure 1 is presented that ships V_1 and V_2 passed the minimum distance between ships at TCPA t_c. The time from current time to t_c was TCPA, and the distance between own-ship and a target ship at estimated time from the t_c was DCPA.

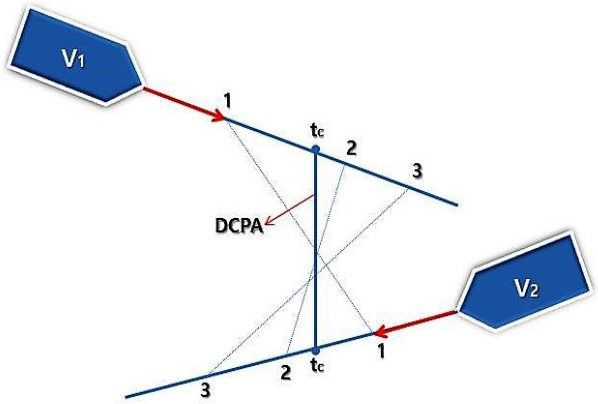


Figure 1: Calculation DCPA and TCPA

Figures 2 and 3 show the fuzzy membership function of TCPA/(L/ V) and DCPA/L, where L was a length of ship and V was a ship speed. In order to increase precise on the FIS, letting DCPA and TCPA to be dimensionless by using a length and a speed of ship was conducted[12]. The CRI can be expressed as values from -1 to 1. Negative values in TCPA mean that a target ship passed through own ship, i.e., safe situation by being out of the collision risk.

Inference rule of state variables used in the FIS are Small (S), Medium (M), B (Big), P (Positive), and N (Negative). Table 1 shows a part where collision risk was determined in order that an input and an output can express the inference rule as a two-dimensional matrix. In other words, it is determined by a condition part of the $i-$th inference rule out of all the inference rules. The CRI at the conclusion as numerals in the fuzzy inference table as show in equation (1).

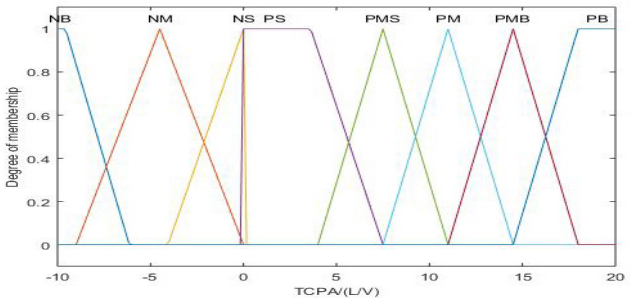


Figure 2: Fuzzy membership function of TCPA/(L/V)

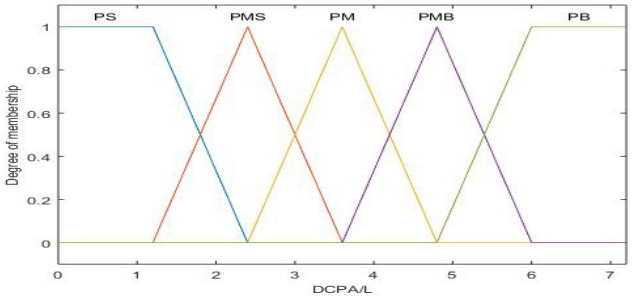


Figure 3: Fuzzy membership function of DCPA/L

$$\text{Collision Risk(CR)} = \frac{\sum_{i=1}^{n} CR_i \cdot a_i}{\sum_{i=1}^{n} a_i} \quad (1)$$

Where,

n = number of reasoning rules,

CR_i = singleton value of conclusion part of $i-$th rule,

a_i = contribution factor of conditional part of $i-$th rule.

Table 1: Fuzzy inference table

Division		TCPA/(L/V)							
		NB	NM	NS	PS	PMS	PM	PMB	PB
DCPA /L	PS	-0.2	-0.6	-1.0	1.0	0.8	0.6	0.4	0.2
	PMS	-0.2	-0.2	-0.6	0.8	0.6	0.4	0.2	0.2
	PM	-0.2	-0.2	-0.2	0.6	0.4	0.2	0.2	0.2
	PMB	-0.2	-0.2	-0.2	0.4	0.2	0.2	0.2	0.2
	PB	-0.2	-0.2	-0.2	0.2	0.2	0.2	0.2	0.2

A point of time for an action of collision avoidance[12,13,19] was determined as follows. In case of a give-away ship, more than 0.6 of the CRI mean an action of collision avoidance. A stand-on ship takes an action of collision avoidance in case of more than 0.8.

2.2. Decision of Ship Near-Collision

Because the number of actual collision accidents was very small, it was significantly difficult to construct a model for evaluating the collision risk based on previous marine accidents [20]. Therefore, near-collision, which was a situation in which there was the danger of collision between ships approaching each other, but with no collision eventually occurring, either due to

deceleration, or evasion by the change of course, was used. In order to detect a number of near-collision, ship domain was utilized for decision of near-collision as criteria not overlapped between ship domains [21,22]. Thus, in this study, near-collision was decided according to proposed methods using ship domain.

Ship domain [2] size with ellipse can be divided into two types as follows. In the sufficient sea area to freely navigate with 10 to 16 knots, the ship domain was composed of 8 L and 3.2 L, where L is ship length. In the constrained sea area such as narrow channel or harbour, the ship domain was composed of 6 L and 1.6 L by reducing speed until 6 to 8 knots. Figure 4 shows ship domain size[2].

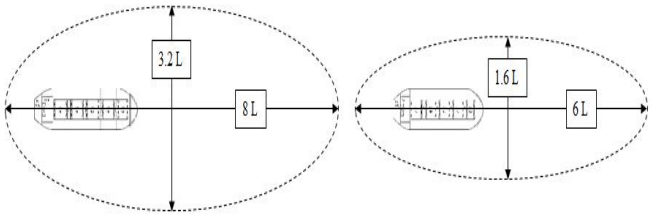


Figure 4: Ship domain size

Ship near-collision based on ship domain with ellipse can be decided in shown as Figure 5.

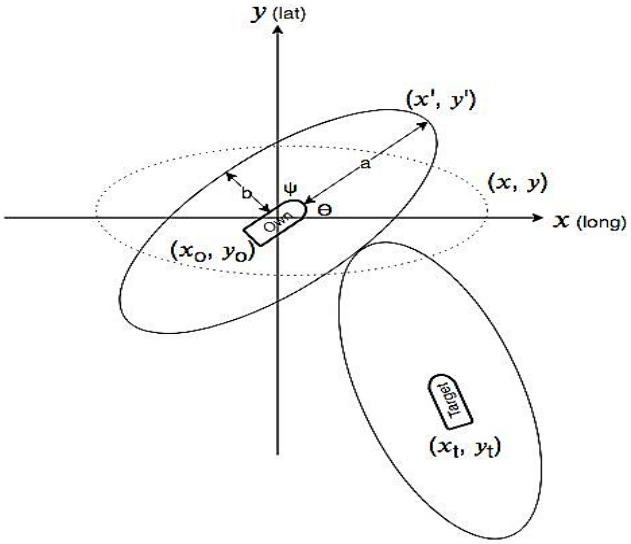


Figure 5: Near-collision between ship's ellipse dimensions

Own ship's ellipse area is parallel to the axis x at the position (x_O, y_O), and can be expressed with a long radius (a) and a short radius (b) in the shape of a dotted line. As an ellipse of position is able to be changed in accordance with own ship's course, rotation of ellipse is necessary corresponding to course. A long radius a and angle of intersection θ of bow direction on the axis x are decided according to ship course angle φ as equation (2). Calculation of rotated own ship's ellipse follows equation (3).

$$\theta = \begin{cases} |90^\circ - \varphi| \ (\varphi \leq 180^\circ) \\ |270^\circ - \varphi| \ (\varphi \geq 180^\circ) \end{cases} \quad (2)$$

$$\frac{(\cos\theta \times (x' - x_O) + \sin\theta \times (y' - y_O))^2}{(a \times L_O)^2} + \frac{(\sin\theta \times (x' - x_O) + \cos\theta \times (y' - y_O))^2}{(b \times L_O)^2} = 1 \quad (3)$$

3. Inference Rule based on Near-Collision via ANFIS

3.1. Procedure for Configuration of the Fuzzy Inference Rule

Procedure for configuration of the fuzzy inference rule presented in Figure 6 can be divided into four steps. A defined action of collision avoidance was defined in first step. Second step designated appropriate distance corresponding to an action of collision avoidance. Third step enlarged an area of the ship domain based on designated distance, and extracted DCPA, TCPA and CRI when ship domain enlarged by level was overlapped. The fuzzy inference rule by learning ship near-collision data via ANFIS was suggested in fourth step.

3.2. Set on the Collision Risk Index

The CRI was set by enlarging the ship domain corresponding to appropriate distance for collision avoidance through definition of an action of collision avoidance by level.

First step defined an action of collision avoidance corresponding to level, and set appropriate distance by level. According to the classification of encounter situations[23,25], an action of collision avoidance was divided into Collision(C), Dangerous(D), Threat(T) and Attention(A). Table 2 defines an action of collision avoidance by level.

Table 2: Definition of an action of collision avoidance by each level

Level	Definition
Collision (C)	· Both ships almost have collision
Dangerous (D)	· Both ships must conduct the best cooperation an action for collision avoidance
Threat (T)	· A give-way ship must conduct an action of collision avoidance · A stand-on ship can take an action for collision avoidance
Attention (A)	· A give-way ship must conduct an action of collision avoidance · A stand-on ship must keep own course and speed

In order to set appropriate distance according to the defined level, it is necessary to take into consideration a give-way ship and a stand-on ship's distance for an action of collision avoidance.COLREGs requested a specified light intensity to show the minimum distance of the ship lighting[23]. The mast light is 6 nm(nautical miles), side light is 3 nm and stern light is 3 nm. Comparing the minimum distance of the ship lighting with an encounter situation for an action of collision avoidance, the mast light and the side light are able to be checked on head-on and crossing situation, and the stern light are able to be checked on overtaking situation.

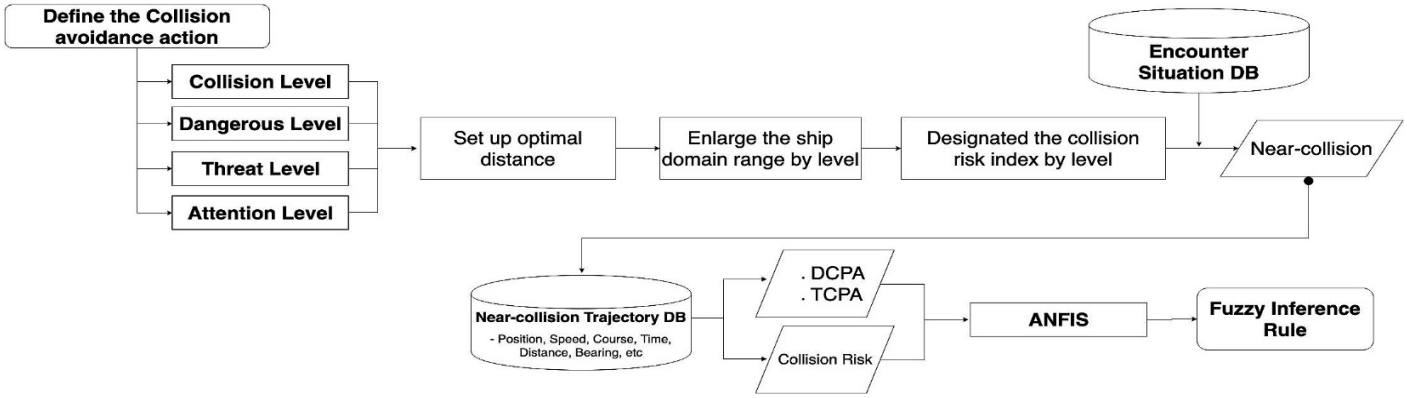


Figure 6: Procedure for configuration of fuzzy inference rule using ship near-collision data

Results of navigator's interviews[24] showed 6 nm, 3 nm and 1nm on the head-on situation, crossing and overtaking situation, and the minimum safe distance. Cockcorft, A. and Lameijer, J.N.F.,(2011) [25] advised the minimum safe distance for collision avoidance as 1 nm in sight of one another and 2 nm in restricted visibility. And, for a collision situation involving two power-driven ships, the minimum distance suggested for an action of collision avoidance were 2 to 3 nm.

Therefore, an action distance of collision avoidance for Collision(C) level was not set due to occurred situation of collision between ships. An action distance for Dangerous(D) level was set as 1 nm in order to pass safely between ships through both ship's cooperation navigation. An action distance for Threat(T) level was set as 2 nm in order that a stand-on ship takes an action of collision avoidance by continuously observing an action of a give-way ship's collision avoidance. At this distance, a give-way ship still has an obligation for collision avoidance. An action distance of Attention(A) was set 3 nm due to initial beginning of the collision risk between ships.

Second step sets the CRI corresponding to enlarged the ship domain based on designated distance by level. In order to enlarge the ship domain, we used the ship domain ($8L \times 3.2\,L$)[2] as standard, considering limited ship speed suggested in each sea area. The standard ship length was set as 70 meters due to having different lengths[2].

In case of overlapping the ship domain ($8\,L \times 3.2\,L$) set in encounter ships, each ship had extra distance 0.3 nm. At this time, the sum of extra distance was 0.6 nm. This was due to actual collision accident was not occurred. However, since distance of the overlapping ship domain was less than the minimum safe distance 1 nm, it could mean that both ships almost had collision. Hence, ship domain ($8L \times 3.2\,L$) was set corresponding to Collision(C) level. Based on this approaching methods, ship domain size was adjusted corresponding to designated distance by level.

The existing FIS had the range of the CRI from -1.0 to 1.0. But in this study, we composed of the range of the CRI from 0.0 to 1.0 except for negative sign. The CRI of Attention(A) level was set 0.0 due to initial beginning of the collision risk between ships. Other levels were designated by dividing 1.0 to the three parts. Accordingly, the CRI of Collision(C) level was 1.0 due to the meaning of having collision between ships, and Dangerous(D) level and Threat(T) level were set as 0.66 and 0.33 due to gradually increasing the CRI by times. Table 3 shows the range of the CRI, ship domain size and the minimum safe distance by proposed level. Overlapping situation between ship domains is presented in Figure 7.

Table 3: The range of the CRI, ship domain and safe distance by proposed level

Level	NM	Ship domain	CRI
C	-	$8\,L \times 3.2\,L$	1.0
D	$0.6 < \text{NM} \leq 1.0$	$13.2\,L \times 5.28\,L$	$0.66 \leq \text{CRI} < 1.0$
T	$1.0 < \text{NM} \leq 2.0$	$26.4\,L \times 10.56\,L$	$0.33 \leq \text{CRI} < 0.66$
A	$2.0 < \text{NM} \leq 3.0$	$39.6\,L \times 15.84\,L$	$0.0 \leq \text{CRI} < 0.33$

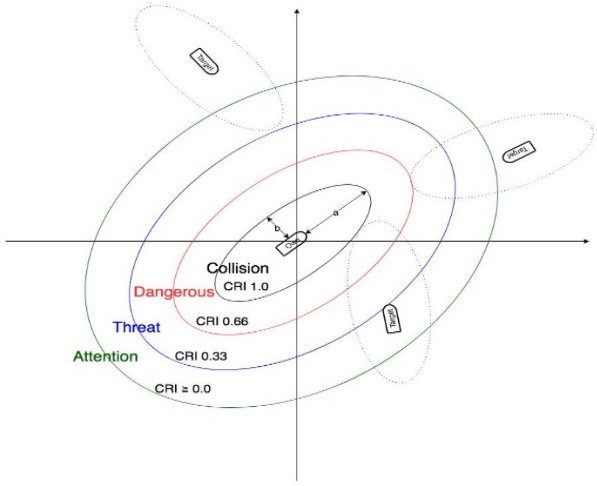


Figure 7: Overlapping situation of ship domain by proposed level

3.3. Ship's Trajectory Data by Level

By setting enlarged ship domain based on Collision(C) level to both ships, ship's trajectory data were extracted when overlapped. This was for comparing and analyzing the information generated from initial beginning of the collision risk to occurring to near-collisions. At this time, ship domain of both ships were set by having the same size in order that enlarged ship domain applied to both ships corresponded to the designated distance when overlapped. To collect ship's near-collision data, Mokpo sea area

was designated. 97 cases out of a total of 154 ships had encounter situations. Results of near-collision by setting ship domain ($8\,L \times 3.2\,L$) as Collision(C) level was a total of 49 cases. Figure 8 is shown as ship's near-collision trajectory.

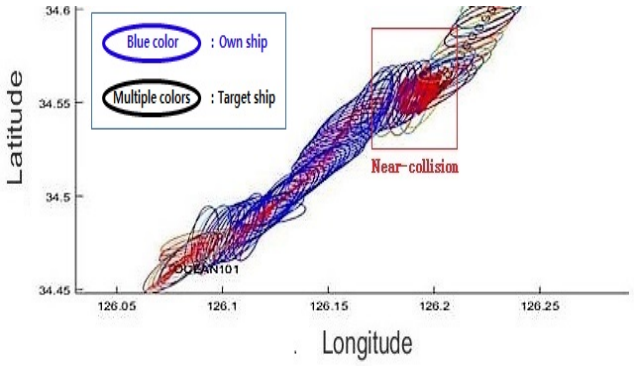


Figure 8: Ship near-collision trajectory

By extracting a total of 196 near-collision ships occurred in each level, we corresponded to the CRI of proposed level. DCPA and TCPA was extracted based on ship's trajectory near-collision data when near-collision by level occurred. At this time, the DCPA was non-dimensionalized by using ship length L, and the TCPA was non-dimensionalized by using ship length L divided by the ship speed V. Table 4 shows the trajectory data of ship near-collision by level.

Table 4: The trajectory data of ship near-collision by level

Level	DCPA/ L	TCPA/(L/V)	CRI
Collision (C)	0.12	0.74	1.0
	0.09	1.93	1.0
	0.17	1.13	1.0
Dangerous (D)	0.48	0.49	0.66
	0.41	2.98	0.66
	0.68	4.41	0.66
Threat (T)	1.07	7.28	0.33
	1.06	9.89	0.33
	0.72	0.81	0.33
Attention (A)	1.29	7.22	0.1
	1.43	3.09	0.1
	2.27	13.3	0.1

3.4. Inference Rule of Collision Risk

3.4.1. Structure of ANFIS

The ANFIS[26] is one of the neuro fuzzy systems for the fuzzy modeling and control proposed by Jang. Configuration of inference rule of the collision risk was as follows. First step designated DCPA/L and TCPA/(L/V) as input parameter, and the CRI as output parameter. At this time, DCPA/L and TCPA/(L/V) were composed of a total 4 of MF(memberships function). Table 5 shows input and output parameter set in MF.

Table 5: Categorized input and output parameter

No.	Input		Output
	DCPA/ L	TCPA/(L/V)	CRI
	Inputmf	Inputmf	Outputmf
1	Collision	Collision	Outputmf 1
2		Dangerous	Outputmf 2
3		Threat	Outputmf 3
4		Attention	Outputmf 4
5	Dangerous	Collision	Outputmf 5
6		Dangerous	Outputmf 6
7		Threat	Outputmf 7
8		Attention	Outputmf 8
9	Threat	Collision	Outputmf 9
10		Dangerous	Outputmf 10
11		Threat	Outputmf 11
12		Attention	Outputmf 12
13	Attention	Collision	Outputmf 13
14		Dangerous	Outputmf 14
15		Threat	Outputmf 15
16		Attention	Outputmf 16

Second step composed of fuzzy inference rule by learning input and output parameters via ANFIS. Figure 9 shows ANFIS structure configured by using MF of input and output parameters.

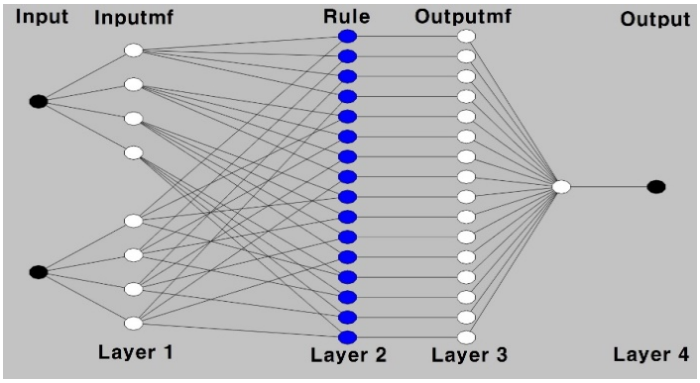


Figure 9: Structure of ANFIS

DCPA/ L and TCPA/(L/V) were designated as premise parameter x, y and CRI was designated as consequent parameter f_i. By using consequent coefficient a, b, c inference rule was composed as shown in equation (4).

$$Rule_i : If\ x\ is\ \mu_{\left(\frac{DCPA}{L}\right)i}\ and\ y\ is\ \mu_{\left(\frac{TCPA}{\left(\frac{L}{V}\right)}\right)i,}$$

$$\text{then } f_i = a_i x + b_i y + c \qquad (4)$$

3.4.2. Inference Rule through Learning Algorithm

The ANFIS can be trained by the backward propagation algorithm and the hybrid learning algorithm. In the backward propagation algorithm, the errors are propagated backward and the premise parameters are updated by gradient descent. At this time, consequent parameters are fixed. In the hybrid learning algorithm, the premise and consequent parameters are updated by least-

squares method and gradient descent. In this study, we used the backward propagation algorithm for maintaining the CRI designated by level. Therefore, surface of CRI through learning by the backward propagation algorithm was represented non-linear as shown in Figure 10.

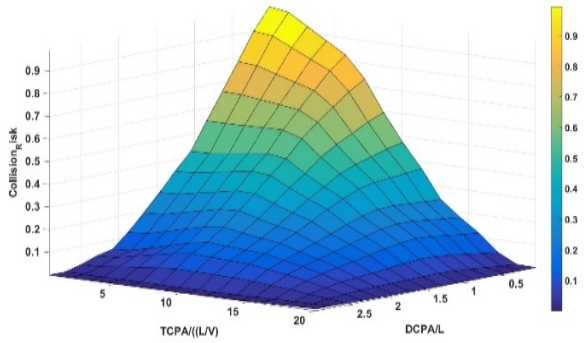


Figure 10: Surface of the CRI

Inference rule of the CRI proposed by learning near-collision data via ANFIS was presented in Table 6.

Table 6: Fuzzy inference rule for membership function

Division		DCPA/*L*			
		C	D	T	A
TCPA/(*L*/*V*)	C	0.994	0.773	0.477	0.021
	D	0.777	0.662	0.401	0.017
	T	0.395	0.423	0.335	0.015
	A	0.062	0.246	0.152	0.011

4. Result and Discussion

4.1. Result

To validate performance of the proposed fuzzy inference rule by comparing with the existing fuzzy rule, two ships navigating at Mokpo sea area were selected from AIS data. One of which was designated as own-ship and the other was designated as a target ship. An encounter situation and near-collision situation are presented in Figures 11 and Figure 12. When the encounter situation occurred within 10 nm, own-ship begun calculating the CRI with DCPA and TCPA. The CRI calculated from four types of inference rule was presented in Table 7 and Figure 13.

At the initial encounter situation between own-ship and a target ship, the FIS[12], FIS-VCD[13], FIS-SC[19] and FIS-NC proposed in section III indicated the CRI 0.29, 0.67, 0.32 and 0.11, respectively.

When it comes to TCPA and distance, the FIS-SC and the FIS-NC were gradually increased until near-collision situation according to reduction of TCPA and distance. Whereas, the FIS and the FIS-VCD were not gradually increased until near-collision situation. In case of the FIS, after it obtained the CRI until 0.94, the same CRI was kept until near-collision situation. In case of the FIS-VCD, unlike the FIS, the FIS-SC and the FIS-NC, it depended on how much the VCD was obtained due to input parameter DCPA, TCPA and VCD. As shown in Table 7, as the VCD got closer at 0, the CRI was increased.

When it comes to a point of time for action of collision avoidance of a give-way ship, distance for action of collision avoidance via the FIS, the FIS-VCD, the FIS-SC and FIS-NC represented about 7.2 nm, about 9.8 nm, about 6.1 nm and about 7.7 nm, respectively. At this time, TCPA represented about 10.7 minutes,15.8 minutes, 7.7 minutes and 11.7 minutes, respectively.

When it comes to a point of time for action of collision avoidance of a stand-on ship, distance for action of collision avoidance via the FIS, the FIS-VCD, the FIS-SC and FIS-NC represented about 5.1 nm, about 7.2 nm, about 1.7 nm and about 6.1 nm, respectively. At this time, TCPA represented about 7.4 minutes,10.7 minutes, 2.5 minutes and 7.7 minutes, respectively.

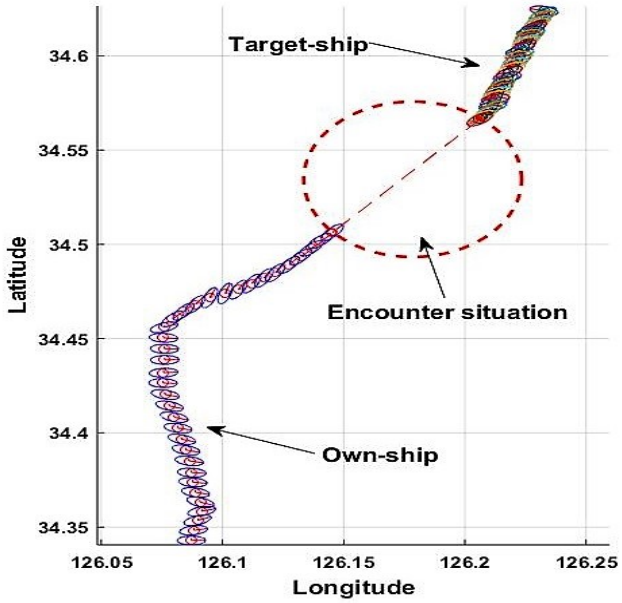


Figure 11: Encounter situation

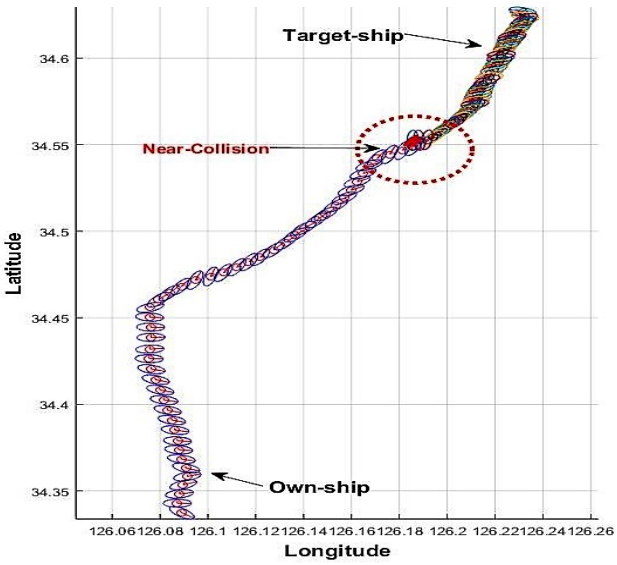


Figure 12: Near-collision situation

Table 7: Comparison with the different CRI according to TCPA and distance

Distance (nm)	TCPA (minute)	VCD (degree)	Collision Risk Index				A point of time for collision avoidance
			FIS	FIS-VCD	FIS-SC	FIS-NC	
9.8	15.8	-	0.29	0.67	0.32	0.11	FIS-VCD(give-way)
9.3	14.5	1.2	0.39	0.71	0.36	0.17	
8.9	13.6	0	0.45	0.73	0.38	0.21	
8.4	13.1	2.2	0.48	0.74	0.41	0.24	
8.1	12.4	6.5	0.51	0.77	0.44	0.31	
7.7	11.7	7.8	0.54	0.78	0.46	0.34	FIS-NC(give-away)
7.2	10.7	0.3	0.62	0.8	0.49	0.42	FIS(give-way), FIS-VCD(stand-on)
6.9	10.1	0.5	0.65	0.81	0.51	0.46	
6.4	8.3	0.3	0.74	0.84	0.58	0.61	
6.1	7.7	0.4	0.77	0.86	0.61	0.67	FIS-SC(give-way), FIS-NC(stand-on)
5.5	7.6	2.8	0.78	0.84	0.62	0.68	
5.1	7.4	0.2	0.8	0.87	0.63	0.71	FIS(stand-on)
4.6	6.8	3.5	0.8	0.83	0.66	0.77	
4.2	6.1	1.7	0.81	0.87	0.68	0.79	
3.8	5.4	2.6	0.83	0.84	0.7	0.82	
3.3	4.9	4.6	0.85	0.81	0.72	0.84	
2.9	4.4	1.3	0.88	0.88	0.75	0.86	
2.5	3.7	1.4	0.93	0.87	0.77	0.89	
2.1	3.1	6	0.94	0.81	0.79	0.92	
1.7	2.5	3.2	0.94	0.83	0.8	0.94	FIS-SC(stand-on)
1.3	2.2	0.1	0.94	0.93	0.84	0.95	
0.9	1.5	0	0.94	0.94	0.83	0.97	
0.4	0.7	2.6	0.94	0.84	0.87	0.98	
0.2	0.1	0.4	0.94	0.91	0.92	0.99	Near-Collision

FIS: Fuzzy Inference System; FIS-VCD: Fuzzy Inference System based on VCD; FIS-SC: Fuzzy Inference System based on Ship's Characteristic; FIS-NC: Fuzzy Inference System based on Near-Collision

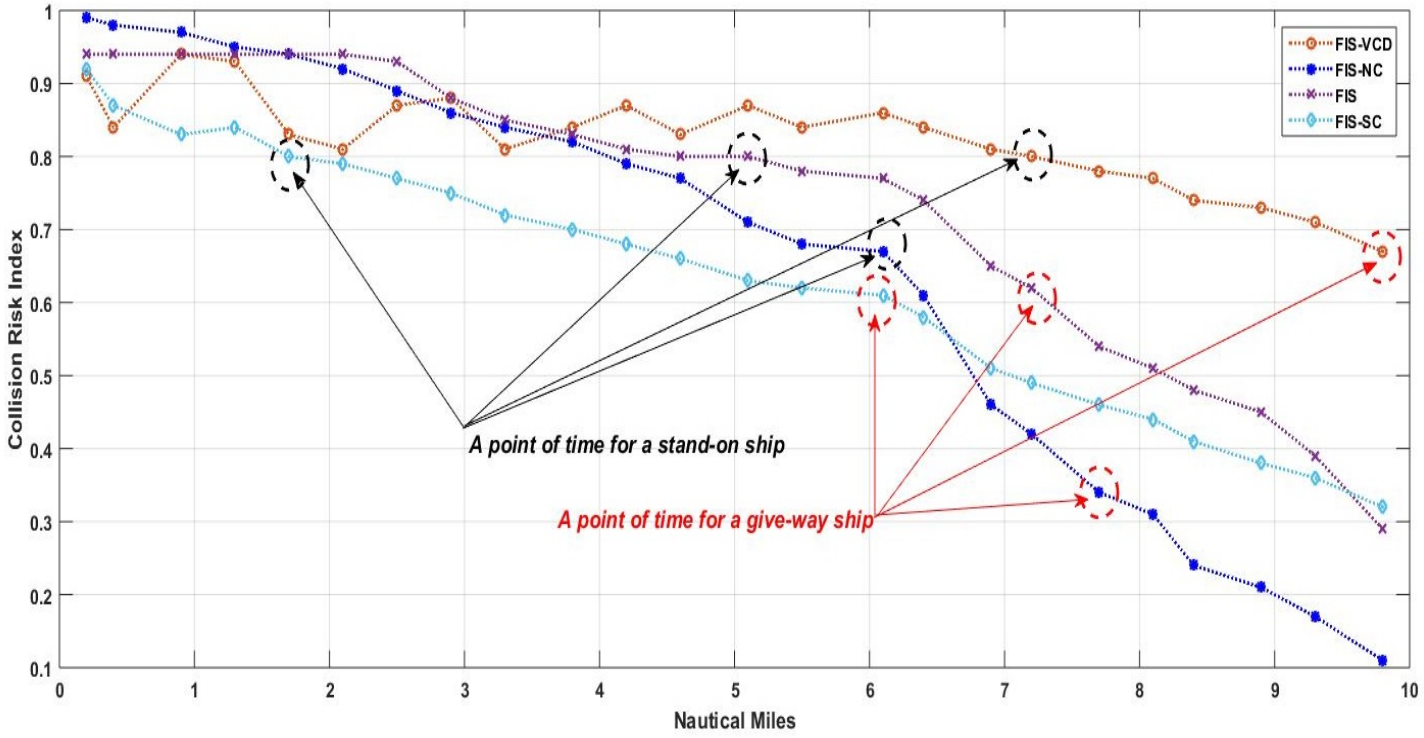


Figure 13: Comparison with the different CRI corresponding to distance

4.2. Discussion

In section 4.1, it could be seen that the exiting rule[12,13,19] and the proposed rule were able to obtain the CRI in the encounter situation between own-ship and a target ship according to values of input parameter. However, a point of time for action of collision avoidance was different according to type of the encounter situation. In case of a give-way ship, the FIS, the FIS-VCD and the FIS-SC defined that own-ship must take an action when the CRI exceeded more than 0.6. In case of a stand-on ship, own-ship must take an action for collision avoidance when the CRI exceeded more than 0.8. On the other hand, when the CRI exceeded more than 0.33, the FIS-NC proposed that a give-way ship must take an action for collision avoidance. In case of a stand-on ship in the FIS-NC, a point of time for collision avoidance was when the CRI was more than 0.66. Therefore, in this section, a point of time for action of collision avoidance via the FIS, the FIS-VCD, the FIS-SC and the FIS-NC were discussed according to the "A guide to the collision Avoidance Rules[25]", as well as a review of the various values obtained from inference rules.

According to [25], it states that ship should take an action for collision avoidance within about 5 to 10 minutes of TCPA or within about 2 to 3 nm of minimum distance. As shown in Table 7 and Figure 13, a give-way ship using the FIS, the FIS-VCD, the FIS-SC and the FIS-NC can obtain outer distance 2 to 3 nm and enough TCPA for taking an action. whereas, in case of a stand-on ship, the only the FIS-SC didn't obtain the requested distance and TCPA. it means that ship using the FIS-SC have no choice but to take into consideration collision risk at all times.

In terms of the CRI expression, even though the FIS had enough distance and TCPA until near-collision situation, it indicated the same CRI from 0.94 to near-collision situation. In case of the FIS-VCD, it relied on the only VCD. For this reason, unlike the FIS, FIS-SC and FIS-NC, the only the FIS-VCD already exceeded more than threshold at the initial encounter situation. On the other hand, the FIS-SC and the FIS-NC expressed the various CRI according to input parameters. It means that the CRI calculated from the FIS-SC and the FIS-NC reflected original information(DCPA, TCPA) well. Accordingly, because priority of multiple ships can be designated via the various CRI, these would be possible to not only take an appropriated action of collision avoidance between ships, but also among multiple ships.

For this reason, a ship using the FIS-NC can take an suitable action in any encounter situations because it reflects all of aspects, i.e., appropriated distance and TCPA according to the various CRI. However, it didn't reflect various factors, e.g., weather condition, ship size, congestion of navigation area and so on.

5. Conclusion

Although the FIS using DCPA and TCPA have been reported previously, they have largely relied on the use of either interviews of navigators or ship's characteristic in virtual navigation situation. These approaches tend to have the limitation not reflecting on information of actual collision situation between ships. Therefore, in this paper, we proposed the fuzzy inference rule based on near-collision via ANFIS by applying an action of collision avoidance corresponding to the CRI by level. For a configuration of fuzzy inference rule, steps was conducted as follows. An action of collision avoidance was defined, and distance to conduct an action of collision avoidance was set; (i). By enlarging ship domain based on designated distance by level, DCPA, TCPA and CRI were extracted when overlapped; (ii). Fuzzy inference rule was proposed by learning near-collision data via ANFIS; (iii). The results of applying proposed fuzzy inference rule to actual navigation area expressed the various CRI corresponding to the required distance and TCPA in order that navigators can make a decision appropriately. Accordingly, the proposed fuzzy inference rule not only overcame these issues, but also offered performance more outstanding than the existing fuzzy inference rule. However, it didn't reflect weather condition, ship size, congestion of navigation area, and so on. In further study, it is required to improve and validate the inference rule of the CRI taking consideration into drawbacks.

Conflict of Interest

The author declares no conflict of interest.

References

[1] Korean Maritime Safety Tribunal, "Annual Report of Marine Accident," 2018.

[2] Fujii, J., Tanaka, K., "Traffic capacity," Journal of Navigation, 24, 543–552, 1971. https://doi.org/10.1017/S0373463300022384

[3] Goodwin, E.M., "A statistical study of ship domains," Journal of Navigation, 28, 328–344, 1975. https://doi.org/10.1017/S0373463300041230

[4] Davis, P.V., Dove, M.J., Stockel, C.T., "A computer simulation of marine traffic using domains and arenas," Journal of Navigation, 33, 215–222, 1980. https://doi.org/10.1017/S0373463300035220

[5] Coldwell, T.G., "Marine traffic behavior in restricted waters," Journal of Navigation, 36, 431–444, 1983. https://doi.org/10.1017/S0373463300039783

[6] Jensen, T.K., Hansen, M.G., Lehn-Schiøler, T., Melchild, K., Rasmussen, F.M., Ennemark, F., "Free flow–efficiency of a one-way traffic lane between two pylons," Journal of Navigation, 66, 941–951, 2013. https://doi.org/10.1017/S0373463313000362

[7] Van Iperen, W.H., "Classifying ship encounters to monitor traffic safety on the North sea from AIS data," TransNav, Int. J. Mar. Navig. Saf. Sea Transp, 9 (1), 51–58, 2015. https://doi.org/10.12716/1001.09.01.06

[8] Pietrzykowski, Z., Uriasz, J., "The ship domain – a criterion of navigational safety assessment in an open sea area," Journal of Navigation, 62, 93–108, 2009. https://doi.org/10.1017/S0373463308005018

[9] Wang, N., Meng, X., Xu, Q., Wang, Z., "An intelligent spatial collision risk based on the quaternion ship domain," Journal of Navigation, 63, 733–749, 2010. https://doi.org/10.1017/S0373463310000202

[10] Wang, N., 2013. "A novel analytical framework for dynamic quaternion ship domains," Journal of Navigation, 66, 265–281, 2013. https://doi.org/10.1017/S0373463312000483

[11] Hasegawa, K., Kouzyki, A., Muramatsu, T., Komine, H., and Watabe, Y., "Ship Auto-navigation Fuzzy Expert System (SAFES)," Journal of the Society of Naval Architecture of Japan, 166, 1989. https://doi.org/10.2534/jjasnaoe1968.1989.166_445

[12] Lee, Han-Jin and Rhee, Key-Pyo, "Development of collision avoidance system by using expert system and search algorithm," Journal of International Shipbuilding Progress 48, 197–212, 2001.

[13] Eun-kyoung Kim, Il-kweon Kang, and Yong-gi Kim, "Collision Risk Decision System for Collision Avoidance," 11(6), 524-527, 2001.

[14] Nam-Sun Son, Sun-Young Kim, and Chan-Soo Lee, "On the Monitoring System of Collision Risks Among Multiple Ships," Annual of Navigation, 19(1), 2012. https://doi.org/10.2478/v10367-012-0012-x

[15] Ahmad C. Bukhari, Inara Tusseyeva, Byung-Gil Lee, and Yong-Gi Kim, "An intelligent real-time multi-vessel collision risk assessment system from VTS view point based on fuzzy inference system," Expert System with Applications, 40, 1220-1230, 2013. https://doi.org/10.1016/j.eswa.2012.08.016

[16] Mira Yi, "Design of the Neuro-Fuzzy based System for Analyzing Collision Avoidance Measures of Ships," Journal of Korean Institute of Intelligent Systems, 27(2), 113-118, 2017. https://doi.org/10.5391/JKIIS.2017.27.2.113

[17] Nam-Sun Son and Sun-Young Kim, "On the sea trial test for the validation of an autonomous collision avoidance system of unmanned surface vehicle, ARAGON," OCEANS 2018 MTS/IEEE Charleston. https://doi.org/10.1109/OCEANS.2018.8604803

[18] Ho Namgung, Jung Sik Jeong, Jo-Cheon Choi, Geun Woong Kim, and Joo-Sung Kim, "Design of Information Exchange Handling System for Safe

Navigation of Autonomous Ship," Journal of Korean Institute of Intelligent Systems, 29(1), 42-49, February, 2019. https://doi.org/ 10.5391/JKIIS.2019.29.1.42

[19] Jin-Hyeong Ahn, Key-Pyo Rhee, and Young-Jun You, "A study on the collision avoidance of a ship using neural networks and fuzzy logic," Applied Ocean Research, 37, 162–173, 2012. https://doi.org/10.1016/j.apor.2012.05.008

[20] Inoue, K. and Kawase, M., "Innovative Probabilistic Prediction of Accident Occurrence," Int. J. Marine Nav. & Safety Sea Trans., 1(1), 19-22, 2007.

[21] Zhang, W., F. Goelandt, P.Kujala, and Y.Wang, "An Advanced Method for Detecting Possible Near Miss Ship Collisions from AIS data," Ocean Eng., 124, 141-156, 2016. https://doi.org/10.1016/j.oceaneng.2016.07.059

[22] Rafal Szlapczynski, "A Framework of a ships domain-based near-miss detection method using mamdani neuro-fuzzy classification," POLISH MARITIME RESEARCH Special Issue, 25, 14-21, 2018. https://doi.org/10.2478/pomr-2018-0017

[23] IMO, International Regulations for Preventing Collisions at Sea, COLREGS, 1972.

[24] Y.S, Lee, J.M. Park, and Y.J. Lee, "A Study on the Initial Action of Navigators to Avoid Risk of Collision at Sea," Journal of Korean Navigation and Port Research, 38, 2014. https://doi.org/10.5394/KINPR.2014.38.4.327

[25] Cockcroft, A. N. and Lameijer, J. N. F., "A Guide to the Collision Avoidance Rules," Butterworth Heinemann, Oxford, UK, 2011.

[26] Jang JR, "Adaptive-network based fuzzy inference system," IEEE Transactions on systems, Man and Cybernetics, 1993.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 156-163 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Linearity Improvement of VCSELs based Radio over Fiber Systems utilizing Digital Predistortion

Muhammad Usman Hadi[*,1], Jacopo Nanni[1], Olivier Venard[2], Geneviève Baudoin[2], Jean-Luc Polleux[2], Pier Andrea Traverso[1], Giovanni Tartarini[1]

[1]*Department of Electronic and Information Engineering, University of Bologna, Bologna, 40136, Italy*

[2]*University of Paris-Est, ESYCOM, ESIEE Paris, France, 2 boulevard Blaise Pascal, Noisy-le-Grand, France*

ARTICLE

Article history:
Received:30 April, 2019
Accepted:18 May, 2019
Online: 10 June, 2019

Keywords:
Digital Predistortion
Generalized Memory Polynomial
VCSELs
Radio-over-Fiber

ABSTRACT

The article proposes a Digital Predistortion (DPD) methodology that substantially meliorates the linearity of limited range Mobile Front Haul links for the extant Long-Term Evolution (LTE) and future (5G) networks. Specifically, the DPD is employed to Radio over Fiber links that contrive of Vertical Cavity Surface Emitting Lasers (VCSELs) working at 850 nm. Both, Memory and Generalized Memory Polynomial models are implied to Single Mode (SM) and Multi-Mode (MM) VCSELs respectively. The effectiveness of the proposed DPD methodology is analyzed in terms of Normalized Mean Square Error, Normalized Magnitude, Normalized phase and Adjacent Channel Power Ratio. The demonstration has been carried out with a complete (Long Term Evolution) LTE frame of 10 ms having 5 MHz bandwidth with 64-QAM modulation configuration. Additionally, the effectuality of the proposed DPD technique is evaluated for varying levels of input power and link lengths. The experimental outcomes signify the novel capability of the implied DPD methodology for different link lengths to achieve higher system linearization.

1. Introduction

This paper is extension of the work originally presented in 2018 IEEE International Topical Meeting on Microwave Photonics (MWP) titled "Experimental evaluation of digital predistortion for VCSEL-SSMF-based Radio-over-Fiber link" [1]. The fifth generation (5G) technology is envisaged to provide faster internet access with low latency, cost effectiveness and pervasive mobile coverage [2-3]. Radio over Fiber (RoF) is a pertinent technology to deal with exorbitant requirement of bandwidth multivariate wireless services both for outdoor and indoor scenarios [4-6], and is regarded as a significant technology for the next generation networks [7].

In particular, RoF technology can provide an essential platform for building centralized/cloud radio access network (C-RAN) which should be able to control the centralized base band units (BBU) coming from different base stations and remote radio heads (RRHs) [8].

The interconnectivity of these BBUs with RRHs is economically viable with the distribution network known as the 'fronthaul' [9-10]. The RoF technology is then a suitable candidate for the fronthauling due to its inherent capillary properties. Figure 1 shows the C-RAN utilizing optical fronthauls (OFHs).

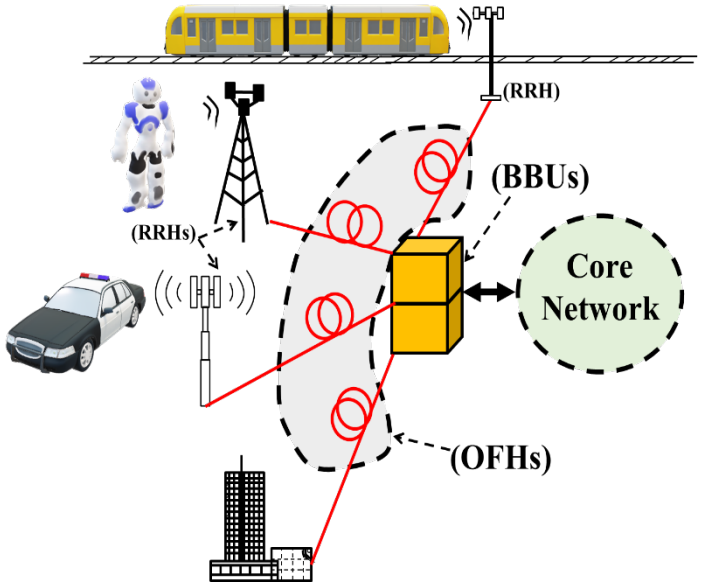


Figure 1: Basic C-RAN architecture showing optical fronthauls (OFHs)

Within these short-medium reach networks, OFHs based on the RoF technology can be realized utilizing Standard Single mode

[*]Muhammad Usman Hadi, Department of Electronic and Information Engineering, University of Bologna, Bologna, Italy, Tel: +390512093951
Email: muhammadusman.hadi2@unibo.it; usmanhadi@ieee.org

Fiber (SSMF) or also Multimode Fiber (MMF) [11]. In both cases, a possible solution which keeps low levels cost and power consumption can be obtained utilizing Vertical Cavity Surface Emitting Lasers also known as VCSELs which emit in the first optical window (wavelength λ=850 nm) as the optical source.

While the use of VCSELs operating at short wavelengths is a relatively consolidated choice finding application e.g. in combinations with Plastic Optical Fibers within in-building networks [12-13] or in combination with MMFs within data centers [14], care must be taken in front of the possible impairments due to the multimodal behavior of SSMFs that operate at the 850 nm [15]. The presence of multimodal behavior would indeed be absent if expensive VCSELs operating at 1.3 μm and 1.55 μm were utilized , in which the major SSMF-related impairments could just be ascribed to optical nonlinear effects and/or to chromatic dispersion [16-17]. However, appropriate countermeasures have been proposed, which can prevent this last impairment cause to be critical [18-19].

Besides the cited OFH, VCSEL-based RoF systems can be found in multivariate scenarios, being utilized for radio astronomic signal transmission [20], multi service indoor wireless signal distribution [21], or machine learning detection [22].

In all these cases, a cardinal issue is represented by the impairments of the opto-electronic devices and particularly the whole VCSEL based Radio over Fiber system. Such nonlinearities, that arise due to relatively stable causes like nonlinear attributes coming from laser diode and perhaps from photodiode [23]. These nonlinearities can also arise due to little dynamics of VCSELs [24], e.g., when high Peak-to-Average Power Ratio (PAPR) signals are transmitted and can have an important role especially for multi-channel transmission. They indeed can cause high in and out of band distortion, which leads to higher interference among near channels.

In order to minimize the nonlinearities in RoF transmission, different methodologies have been proposed that comprises of digital and analog electrical techniques [25-27]. Among them, Digital Predistortion (DPD) has been regarded to be effectual method that leads to reduction in nonlinearities of RoF systems. The primary intension behind this method exploits a digital predistorter (PD) which has inverse and nonlinear profile as that of the Radio over Fiber system. Consequently, when cascaded with the RoF system, it will lead to linearization of the overall cascaded system. DPD linearization technique using memory polynomials based on Volterra series has been applied to Non linear RoF link [26-28]. Similarly, the Digital Predistortion technique based on canonical piecewise-linear (CPWL) function was proposed for intensity modulated/direct detection RoF system [29]. All the aforementioned techniques apply the distortion compensation considering the RoF systems as "black box".

In our previous work reported in [1,30], we demonstrated the novel implementation of DPD implied to considered radio over fiber link which was based on single mode (SM) based VCSEL for LTE applications utilizing Memory Polynomial (MP) architecture. The highlighted results were obtained in back-to-back (B2B) configuration, i.e. utilizing a SSMF fiber patch-cord of a few meters length.

In this paper, the novelty of the current work is summarized as follows: (1) for the first time, the evaluation that was performed in [1] is extended by considering the same SM VCSEL-based RoF system, not in B2B configuration, but equipped with SSMF spans of different lengths; (2) Moreover, also Multimode (MM) VCSEL-based RoF systems are analyzed, utilizing different lengths of MMFs, (3) in addition, to the DPD technique based on Memory Polynomials (MP) utilized in [1], another architecture named as Generalized Memory Polynomial (GMP), is utilized as well, which in general shows an improved effectiveness with respect to MP; (4) the DPD is trained and tested near to threshold to check the efficacy of the proposed technique.

As a performance index evaluation, the link linearity improvements are observed by calculating the Adjacent Channel Power Ratio (ACPR), Normalized Mean Square Error (NMSE), spectral regrowth, Normalized Magnitude (AM/AM) and Normalized phase (AM/PM). The paper is arranged as follows. In Section 2, modeling methodology is highlighted while Section 3 outlines the predistortion learning architecture. Section 4 discusses the estimation algorithm while Section 5 explains briefly the characterization and experimental setup. Section 6 presents the results in detail. Finally, conclusions are drawn in Section 7.

2. Modeling Methodology

The architecture of the predistortion technique is shown in Figure 2 which utilizes Indirect Learning Architecture for the estimation of PD training coefficients [31-34]. Since the statistic of the waveform will remain the same over time, therefore, it can be assumed that system nonlinearity is not time varying and the training of the predistortion can be applied in an off-line practice. This will not only reduce the expenditures and over heads of the predistorter but also avoids the need of an identification algorithm. The identification of DPD is performed in one step. Hence, a linear estimation of PD coefficients result in straightforward identification.

At first, the training phase calculates the predistorter coefficients. At this point, the RoF system output $y(n)$, becomes input $z(n)$ to the Pre-Distorter Training block which is defined as $z(n) = \frac{y(n)}{G}$ where G denotes the gain of the system. The coefficients appraisal can be made using least-squares-based algorithm and is described in Sec. 3. The convergence point of the training is achieved by minimizing the error function. Eventually, the computed training coefficients are employed to the predistorter referred as Digital Pre-Distorter in Figure 2. The predistorter models utilized are derived from the classical Volterra series.

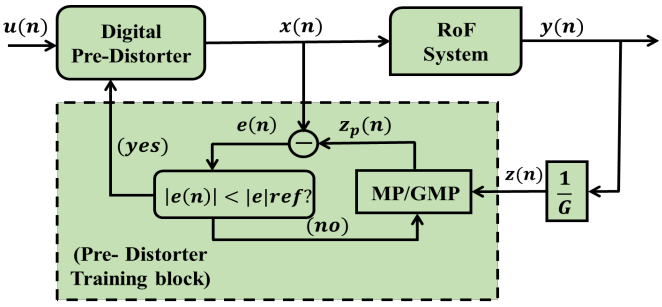


Figure 2. DPD identification for RoF utilizing ILA

3. Digital Predistortion Model

A. Memory Polynomial (MP) Model

Memory Polynomial (MP) model is referred as an inverse non-linear model that has been exploited previously as a powerful model for both, inverse and direct modeling of power amplifier

(PA) nonlinearities. Applying this model for RoF has an additional advantage since memory less model might have problems to characterize the electro-optical (EO) conversion phenomena accurately [34]. The MP model is generally referred as a compromise between memoryless nonlinearity and full Volterra series due to presence of diagonal memory. The output in this case is referred as:

$$z_p(n) = \sum_{k=0}^{K-1} \sum_{q=0}^{Q-1} c_{kq} z(n-q) |z(n-q)|^k$$

Here K represents order of non-linearity, Q is referred as the memory depth, z represents the predistorter input sequence and c_{kq} denotes the model coefficients.

B. Generalized Memory Polynomial (GMP) Model

The use of Generalized Memory Polynomial (GMP) model has been widely utilized for the linearization of PAs [35]. However, GMP has not been yet evaluated for RoF with varying lengths. The GMP model basis functions possess memory for both, for the diagonal terms as well as for the crossing terms i.e., $x(n-q)|x(n-r)|^{k-1}$, where $q \neq r$. The output of the post inverse block $z_p(n)$ modeled with GMP can be expressed as:

$$\begin{aligned} z_p(n) = & \sum_{k=0}^{K_a-1} \sum_{q=0}^{Q_a-1} c_{kq} z(n-q) |z(n-q)|^k \\ & + \sum_{k=1}^{K_b} \sum_{q=0}^{Q_b-1} \sum_{r=1}^{R_b} d_{kqr} z(n-q) |z(n-q-r)|^k \\ & + \sum_{k=1}^{K_c} \sum_{q=0}^{Q_c-1} \sum_{r=1}^{R_c} e_{kqr} z(n-q) |z(n-q+r)|^k \end{aligned} \quad (2)$$

where $z_p(n)$ and $z(n)$ represents the DPD output and input respectively. Similarly, $c_{kq}; d_{kqr}$ and e_{kqr} denotes the complex coefficients for the signal and the envelope; signal and lagging envelope and signal and leading envelope respectively. k represents the index of nonlinearity, and q, r represents the indices of the memory. While K_a, K_b, K_c are the maximum orders of nonlinearity, Q_a, Q_b, Q_c are the memory depths, while R_b and R_c exhibits the lagging and leading delay tap lengths, respectively. GMP has been applied choosing K_a=K_b=K_c= K, Q_a=Q_b=Q_c= Q and $R_b = R_c = R$.

The main objective of the predistorter is to minimize the error ($e(n) = z_p(n) - x(n)$) between the $z_p(n)$ and the input x of the RoF system.

4. Estimation Algorithm

There are number of least squares (LS) algorithm for estimation of model coefficients that take the linear weighting of nonlinear signals [36]. The formulation of the estimation initiates with collecting the coefficients e.g., c_{kq}, d_{kqr} and e_{kqr} in to a $R \times 1$ vector $\boldsymbol{v}$. R represents the total number of coefficients. $\boldsymbol{v}$ means a signal whose time will sample over the same period. Considering (2), coefficients c_{21} denotes the signal $x(n-1)|x(n-1)|^2$. $\boldsymbol{Z}$ characterizes the collection of all such vectors into a Nx R matrix. Once the convergence condition is obtained, the output of the predistorter training block becomes: $z_p(n) = x(n)$ and hence $z(n) = u(n)$. For total samples N, the output can be written in the following way:

$$\boldsymbol{z_p} = \boldsymbol{Zv}$$

Where $\boldsymbol{z} = [z(1),..,z(N)]^T$, $\boldsymbol{z_p} = [z_p(1),..,z_p(N)]^T$, while $\boldsymbol{v}$ displays a $R \times 1$ vector that contains the coefficients c_{kq}, $d_{kqr}$and e_{kqr}.The LS solution will then become a solution for the equation expressed as:

$$[\boldsymbol{Z}^H \boldsymbol{Z}]\hat{\boldsymbol{v}} = \boldsymbol{Z}^H \boldsymbol{x} \quad (4)$$

The LS solution in (4) should minimize the cost function

$$C = \sum_{n=1}^{N} |z_p(n) - x(n)|^2 \quad (5)$$

The advanced variations to Volterra series in form of MP, GMP and others increases the effectiveness of predistortion. However, this advanced variation can come in to existence by increasing the memory depth and nonlinearity order. However, this will extortionate the computational complexity and this has to be weighed against other simpler expediencies. Moreover, the performance would generally be more efficient if the predistortion model has higher number of coefficients. This means that while selecting the model and its complexity, a smart tradeoff between complexity and performance can be made accordingly. The effect of increase in complexity will be shown in Sec. 7.

5. Experimental Setup

The experimental test bed utilized for demonstrating and validating the proposed DPD technique is shown in Figure 3. The baseband LTE signal of 5-MHz with 64 QAM modulation format is emulated through a domestic software on MATLAB compliant with 3GPP release TS 36.104 V15.2.0 [37]. The signal is oversampled at 38.4 MSa/s. After this, the sampled sequence of signals passes through the DPD block, which is then RF-transformed at 800 MHz by a Vector Signal Generator (VSG) (see Figure 3) and is then sent to the optical link.

There are two different optical links that are considered for testing DPD sequentially. In the first, a Single Mode VCSEL (Optowell), that operate at 850-nm, which is followed by a fiber span (SSMF) and a PIN photodiode, having 2.5 GHz bandwidth and responsivity factor of 0.6 A/W. The SSMF for varying lengths has been utilized so that the effect of length on the DPD technique can be observed. In the second a 850-nm MM VCSELs in combination with MMFs was tested, maintaining the same photodetector of the other link.

The reason performing the test with reference to the two RoF configurations is related to the fact that each one of them presents aspects that, according to the particular applicative context, may make it preferable with respect to the other. The SM-VCSEL-SSMF based link can indeed feature convenience in terms of cost and flexibility, because of the lower cost per meter of SSMF with respect to MMF, and because of the huge transmission bandwidth of the SSMF. The MM-VCSEL-MMF based link can take advantage of a typically higher dynamic range and

emitted/coupled power of the MM-VCSEL over MMF with respect to the SM-VCSEL over SSMF case.

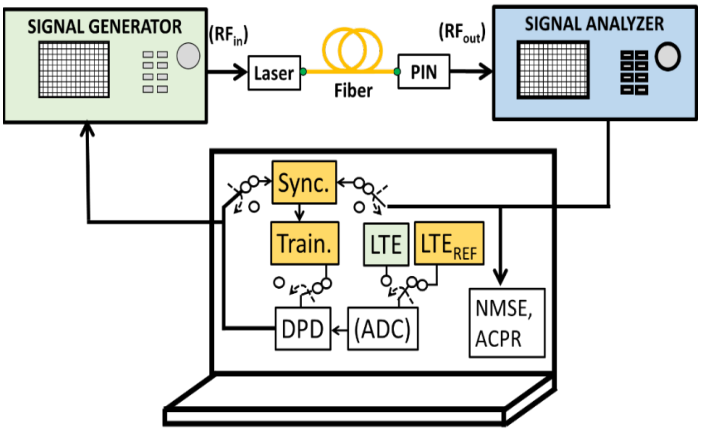


Figure 3. Experimental evaluation comprising of DPD training and application phase.

The reason performing the test with reference to the two RoF configurations is related to the fact that each one of them presents aspects that, according to the particular applicative context, may make it preferable with respect to the other. The SM-VCSEL-SSMF based link can indeed feature convenience in terms of cost and flexibility, because of the lower cost per meter of SSMF with respect to MMF, and because of the huge transmission bandwidth of the SSMF. The MM-VCSEL-MMF based link can take advantage of a typically higher dynamic range and emitted/coupled power of the MM-VCSEL over MMF with respect to the SM-VCSEL over SSMF case.

The process of predistortion is carried out in two steps. The first step is referred as a training phase. In training phase, the parameter identification for predistorter is carried out as referred in Figure 2. During the training phase, reference LTE frames are utilized (see LTE$_{REF}$ block in Figure 3). Primarily, input and output signals are synchronized i.e. time aligned (see Sync. in Figure 3). This block finds the time delay estimation by utilizing the Synchronization Signals (Primary and Secondary) accessible in the LTE framework. Then, the predistorter coefficients are procured through the PD models utilized (see Train. in Figure 3).

In the second step i.e. the testing phase, the training coefficients obtained in previous step are utilized to apply DPD in MATLAB. This means that all switches are turned to opposite position in Figure 3 and the testing is done for LTE frames followed by sampling, pre-distortion, and then uploaded to the Vector Signal Generator (VSG). Finally, they are channelized through the optical link and performance metrics are compared with the one without predistortion. It must be noted that the DPD testing and validation is not only evaluated for the reference LTE signals that were used not only for training, but also for generalized LTE frames.

As mentioned, the quantities utilized to evaluate the effects of the DPD introduced are ACPR and the NMSE. At first, the efficiency of the proposed methods are presented by the ACPR that computes the distortion components outside the useful bandwidth. It is expressed as [1]: (6)

$$ACPR_{dBc} = 10\log_{10}\left[\frac{\int_{ab_l}^{ab_u} S(f)\,df}{\int_{ub_l}^{ub_u} S(f)\;df}\right]$$

where ab_l and ab_u are the lower and upper frequency limits of the adjacent channel; ub_l and ub_u are the frequency bounds of useful bands. $S(f)$ denotes Power Spectral Density (PSD) of the output signal $y(n)$.

NMSE is defined as follows [27]:

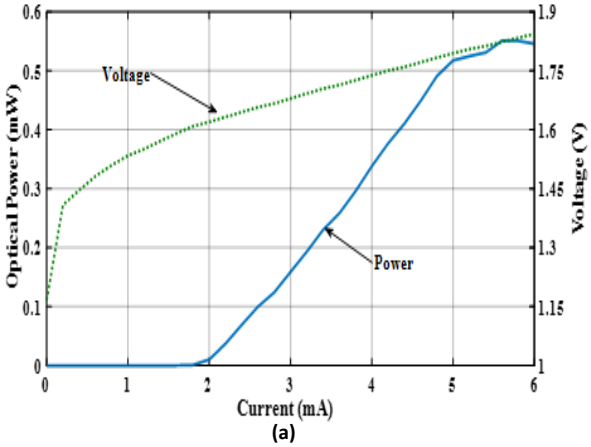
$$NMSE_{dB} = 10log_{10}\left[\frac{\sum_{n=1}^{N}|x(n)-z_p(n)|^2}{\sum_{n=1}^{N}|x(n)|^2}\right] \quad (7)$$

The NMSE value is estimated between $z_p(n)$ and $x(n)$ of the RoF where N embodies the total signal length.

Exploiting these definitions, ACPR and NMSE are calculated for the predistorted and without predistortion output signals. The comparison of the evaluated quantities is presented in the next section. The description of the specifications utilized are given in Table 1.

6. Results and Discussion

The efficacy of the linearization method is appraised and brought into discussion in this section. A primary attention should be given while selecting the parameters of the predistorter such as



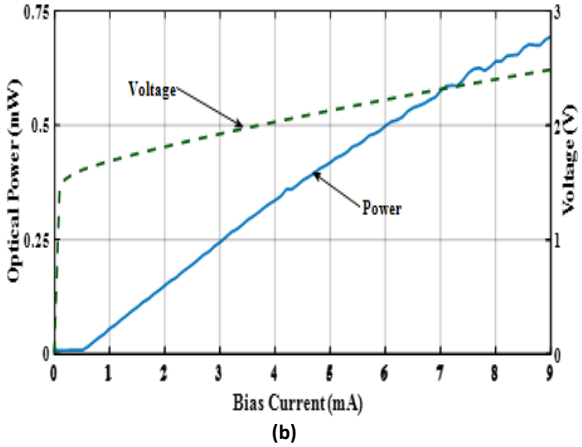


Figure 4. L-I-V characterization utilized in the evaluation of DPD experiments for (a) SM-VCSEL and (b) MM-VCSEL.

memory depth (Q) and nonlinearity order (K) (see Sec. 5 for the details).

Table 1: System Parameters

Optical Link Component	Parameters Values
Laser	**SM-VCSEL** Wavelength= 850 nm $I_{Bias} = 4\ mA$ $I_{Threshold} = 2\ mA$ RIN = -130 dB/Hz **MM-VCSEL** Wavelength= 850 nm $I_{Bias} = 4\ mA$ $I_{Threshold} = 0.8\ mA$ RIN = -125 dB/Hz
Fiber	**SSMF** Maximum Length = 1.5 Km Attenuation = 2.5 dB/km **MMF** Maximum Length = 1 Km Attenuation = 2.5 dB/km
Photo-detector	Responsivity = 0.6 A/W Bandwidth = 2.5 GHz

The L-I-V characteristic curve of SM-VCSEL and MM-VCSEL utilized are illustrated in Figure 4 (a) and (b) respectively. In SM-VCSEL, the maximum saturation current (I_s) is 5 mA while threshold current (I_{th}) is 2 mA. The bias point (I_{bias}) is chosen at 4 mA.

Similarly, In case of MM-VCSEL, the (I_s) is 8 mA and threshold current (I_{th}) is 0.8 mA. The bias point (I_{bias}) is chosen to be 4 mA. The signals having higher PAPR will give rise to higher distortions owing to small dynamics. The PAPR of signal utilized is 9.1 dB.

As explained in Sec. 5, the predistorter complexity is dependent upon the model and order of Q and K chosen. In order to show the trend of this fact, Figure 5 elaborates the experimental NMSE results for different choices of Q and K using both models.

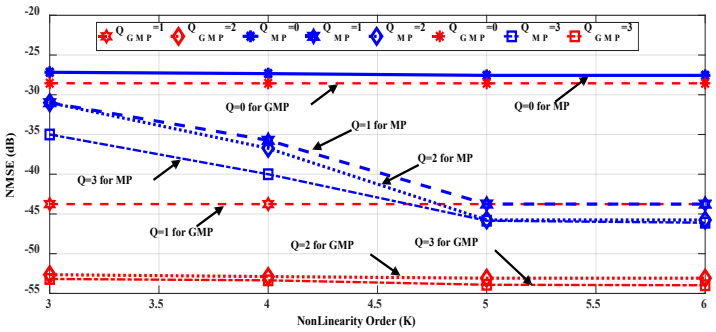


Figure 5. Normalized Mean Square Error results with varying values of memory depth Q and Nonlinearity Order K.

The evaluation in Figure 5 has been done for varying orders of Q (0, 1, 2 & 3) and non-linearity K (3, 4, 5 & 6) by applying a higher RF input power (0 dBm). It can be perceived that GMP results in higher reduction of NMSE in comparison to MP at lower orders of K. From this initial finding, the optimal values $Q = 1,2$ and $K = 3$ are chosen, resulting from a tradeoff between performance achieved and complexity required to the system.

In addition to NMSE, to show the effect of increasing complexity of the proposed distorter, keeping the value of $R = 1$ fixed, Figure 6 represents the ACPR for MP ($Q = 1,2$ and $K = 3$) and GMP ($Q = 1,2$ and $K = 3$) for 1 Km of SSMF. It is perceived that ACPR for MP model doesn't satisfies 3GPP ACPR requirement for MP (both $Q = 1,2$ and $K = 3$) after -5 dBm of input power. While, GMP with $Q = 1$and $K = 3$ satisfies the requirement, however, it is exactly on the borderline at 0 dBm of RF input power.

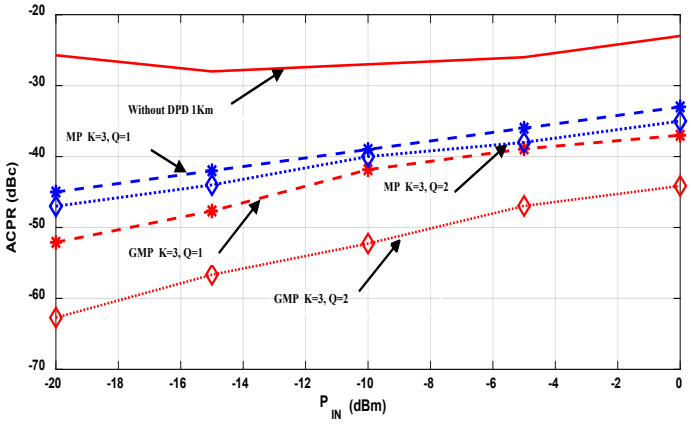


Figure 6. ACPR outcomes vs. input RF power. (Q=1,2, R=1 and K=3 for MP & GMP).

Considering GMP with $Q = 2$ and $K = 3$, the ACPR is well below the requirement set by 3GPP, hence the higher complexity enhances the performance. This demonstration shows that the performance can be enhanced at the cost of higher complexity of the PD model.

Similarly, the Power Spectral Density (PSD) of the received sequence is analyzed when both models are implemented. The length is fixed to 1 Km and RF input power is 0 dBm. The results have been arranged for comparison of MP and GMP with same order of K and Q.

Consider Figure 7(a), the DPD for received output signal is demonstrated for both models by fixing K=3 and Q=1. It is noticeable that GMP has higher reduction in spectral regrowth reduction as compared to MP. Then, in order to highlight the effect of increasing the complexity order, the PSD of GMP and MP are compared for K=3 and Q=2 in Figure 7 (b). This shows that DPD for K=3 and Q=2 is stronger as compared to K=3 and Q=1 due to which spectral regrowth reduction in latter case is higher. Similarly, GMP for K=3 and Q=2 results in ACPR of -44 dBc as compared to -37 dBc when GMP is utilized for (K=3 and Q=1). This substantiates that GMP has better ACPR reduction than MP. This conclusion seems justified due to the fact that GMP carries memory not only in the diagonal but also in the crossing terms.

In light of this discussion and results in the Figure 5, 6 and 7, GMP with optimal values $Q = 2$ and $K = 3$ keeping $R = 1$ fixed have been selected from a tradeoff between performance achieved and complexity required to the system.

In Figure 8 (a) and (b), the ACPR experimental outcomes for several RF input powers (P_{in}) are reported for SM-VCSEL with SSMF and MM-VCSEL with MMF respectively. First consider Figure 8 (a), there are three different cases discussed: Patch cord (Back to back), 1 Km fiber length and 1.5 Km fiber length case. By increasing the length of the fiber, the leakages in adjacent channels rise and therefore the worsening occurs with higher lengths. By employing linearization, it can be seen that fiber length with different lengths have different linearization profiles. For instance, let's consider 1.5 Km length case, the ACPR without linearization at 0 dBm is -18 dBc, and with linearization employed, it is reduced by 22 dBc to -40 dBc.

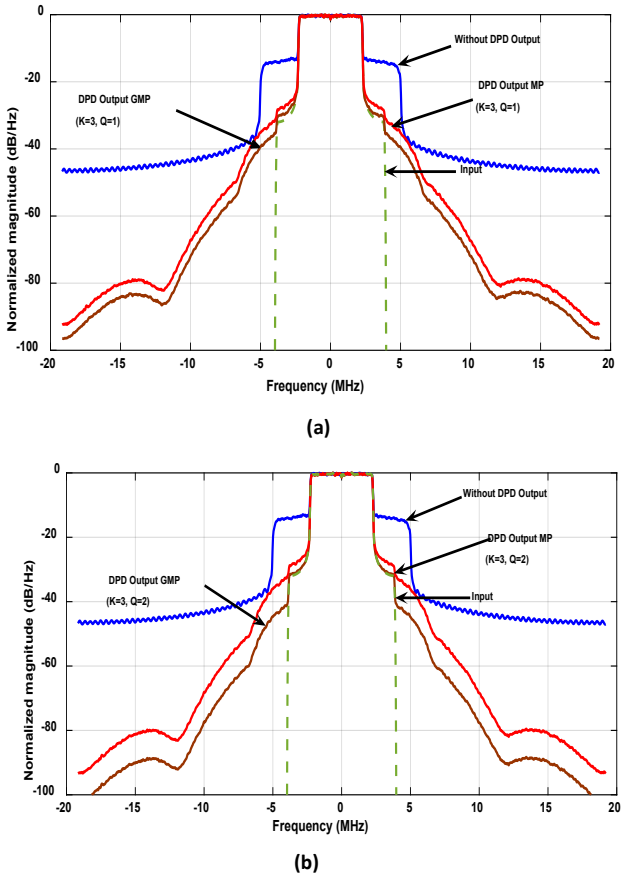


Figure 7. PSD evaluation of output signal without and with DPD (MP/GMP) 0 dBm for 1 Km SSMF with SM-VCSEL having: (a) Q=1, R=1, K=3 and (b) Q=2, R=1, K=3.

In case of Figure 8 (b), four different lengths have been shown up to 1 Km. For 1 Km length, the uncorrected ACPR at 0 dBm is around -22 dBc while the linearization results in a significant reduction to -36 dBc.

It can be deduced from the trend in Figure 8 (a) and (b) that the proposed DPD technique linearizes the length cases as well. Indeed, linearizing the laser nonlinearity is primary aim of the short link lengths, however, in addition to an optical channel consisting in a fiber patch cord (few meters of length), the proposed technique works efficiently for longer fiber lengths.

It must be put into evidence that linearization method has been evaluated for conditions which are very critical such as high PAPR and I_{bias} close to the threshold with optimal values of K and Q. If these critical conditions are relaxed, this will lead to better linearization performance of the predistorter.

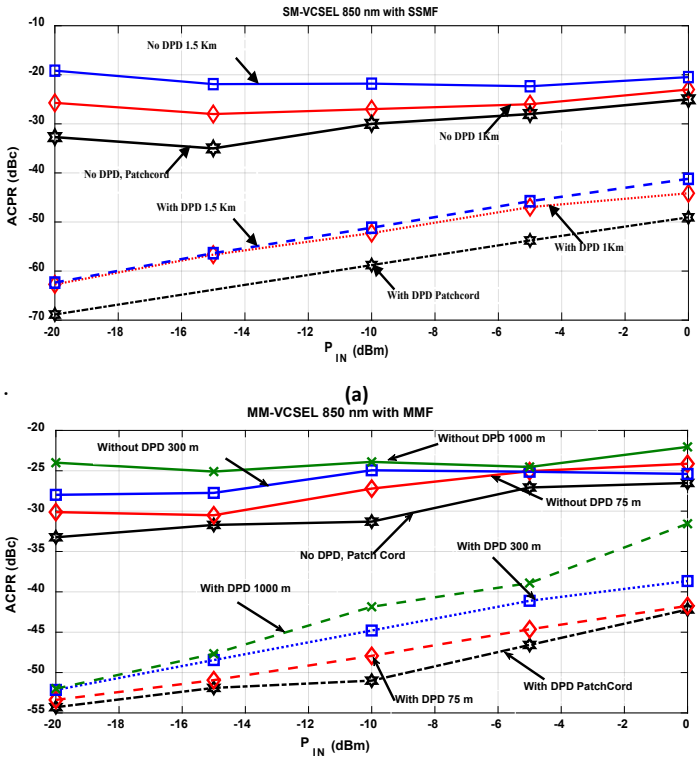
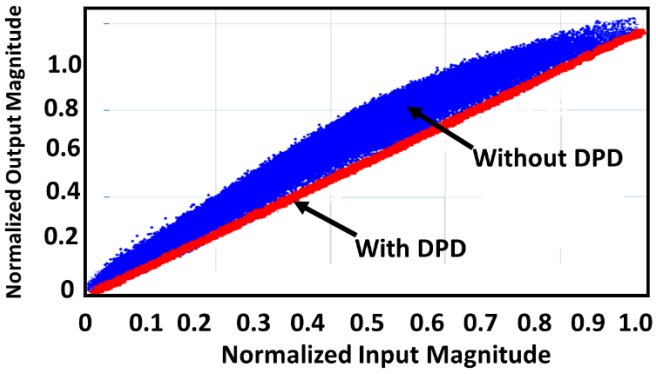


Figure 8. ACPR results vs. varying input signal power using (Q=2 and K=3 for GMP) for (a) SM-VCSEL with SSMF and (b) MM-VSEL with MMF.

Figure 9 represents the AM-AM and AM-PM statistics with and without the linearization employed for the RoF link consisting of 1 Km SSMF and SM-VCSEL. It can be seen that after DPD, the AM/AM is now a straight line and the phase difference is reduced to within 2 degrees.



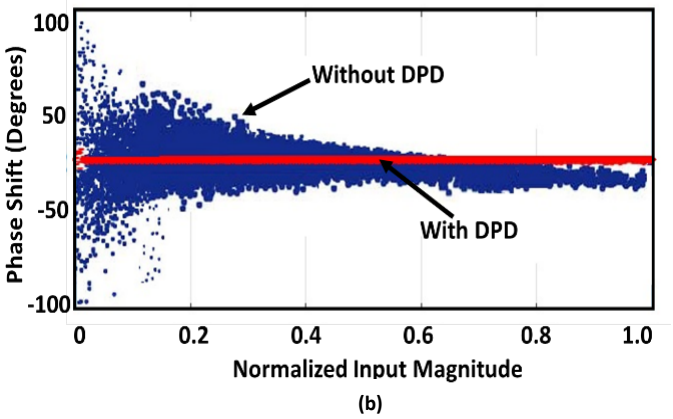


Figure 9. (a) AM-AM and (b) AM-PM plots for SM-VCSEL at 1 Km of SSMF for 0 *dBm* of input RF power with and without DPD.

Table 2 reports the results of the proposed technique, referred to 1 Km of link length with 0 *dBm*. The results are summarized for MP and GMP with K=3 and Q=2 in terms of ACPR and NMSE. The suggested technique linearizes the SM-VCSEL link and MM-VCSEL.

Table 2. Linearization performance for proposed DPD (P_{IN}= 0 *dBm, Link length= 1 Km*)

Model	*SM-VCSEL*		*MM-VCSEL*	
	NMSE (dB)	*ACPR (dBc)*	*NMSE (dB)*	*ACPR (dBc)*
Without DPD	-19.86	-20	-17.63	-22.04
With GMP (K=3, Q=2)	-41.548	-44	-33.18	-36
With MP (K=3, Q=2)	-31.25	-35	-30.14	-29

Indeed, Table 2 suggests that linearization with GMP is better than MP for link length of 1 Km (SM for SM-VCSEL and MM for MM-VCSEL). The reduction in ACPR and NMSE for SM-VCSEL is 22 and 20 dB respectively. While, for MM-VCSEL the improvement in ACPR and NMSE is 14 and 16 dB respectively. This suggests that for SM-VCSEL, the linearization is better than MM-VCSEL for same parameters of data. The improvement for MM-VCSEL is different from SM-VCSEL because the nonlinearities of two lasers are different. It should be noted that MM-VCSEL can achieve higher linearization by selecting appropriate sets of coefficients.

Note that in an applicative scenario, the process of DPD may be positioned at the Central Office (CO), e.g. where BBUs are placed (see Fig. 1) and compensate the RoF Downlink nonlinearities. A periodical re-training of the Digital Predistortion system is in this case necessary, requiring however a negligible time with respect to the time of normal operation of the RoF system. Various methods can be validated for enforcing the periodical training phase. Namely, if a RoF Uplink is applied, the nonlinearities in this case be compensated by a digital post distorter block located at a correspondent CO [38]. By utilizing this method, the RoF downlink nonlinearities can be accumulated at BBU which will decrease the cost and complexity of remote antenna unit.

The time devoted to the periodical re-training of the Digital Predistortion system would in any case remain negligible with respect to the time of normal operation of the RoF system, still allowing its adoption in real applicative scenarios. Similarly, it is possible to utilize an additional photodiode in a base station and feedback the laser nonlinearities and approximating that laser is the main cause of nonlinearity in the RoF link.

Note also that the proposed predistortion method can be applied also to LTE signals of larger bandwidth. The polynomial models nonetheless would require higher values of K & Q with respect to the case presented in the submitted work.

Considering the LTE signal bandwidth and higher modulation format, they would result in a higher PAPR of the transmitted RF signal [39]. At the same time, the increase in bandwidth also determines a correspondent increase in the overall base-band memory of the system to be taken into account by the model.
The polynomial models proposed can still be applied in these operating conditions. They nonetheless would require higher values of the K & Q with respect to the case presented in the submitted work.
These requirements would impact the cost of the DPD implementation, which would be higher due to the higher sampling rate of ADCs, and higher computing capabilities of FPGAs.

7. Conclusions

This article proposes a Digital Pre-Distortion mechanism for linearizing VCSEL based RoF links with different characteristics specifically link lengths. The proposed technique demonstrates a digital predistorter based on MP and GMP. The experiments have been prosecuted for systems based on SM-VCSEL followed by SSMF and on MM-VCSEL followed by MMF. The signal transmitted was a 5-MHz Bandwidth 64 QAM LTE signals and different link lengths have been considered. The performance has been explored in terms of ACPR and NMSE showing that for a link length up to 1 Km, both SM-VCSEL and MM-VCSEL can be linearized in good proportion. Particularly, utilizing SM-VCSEL, GMP results in 22 dBs of reduction in ACPR while MM-VCSEL results in 14 dBs of reduction in spectral regrowth. The results testify an encouraging link performance with low complexity of the predistorter model. It has been demonstrated for the first time that GMPs achieves superior linearization as compared to MP for link lengths up to 1.5 Km.

References

[1] M. U. Hadi et al., "Experimental evaluation of digital predistortion for VCSEL-SSMF-based Radio-over-Fiber link," 2018 International Topical Meeting on Microwave Photonics (MWP), Toulouse, France, 2018, pp. 1-4.
[2] A. Gupta et al., "A Survey of 5G Network: Architecture and Emerging Technologies," IEEE Access, Vol. 3, p. 1206-1232 (2015).
[3] Muhammad Usman Hadi, Hyun Jung, Salman Ghaffar, Pier Andrea Traverso, Giovanni Tartarini, Optimized digital radio over fiber system for medium range communication, Optics Communications, Volume 443, 2019, Pages 177-185, ISSN 0030-4018, https://doi.org/10.1016/j.optcom.2019.03.037.
[4] K. Y. Lau, "RF transport over optical fiber in urban wireless infrastructures," in IEEE/OSA Journal of Optical Communications and Networking, vol. 4, no. 4, pp. 326-335, April 2012.

[5] Y. Shi ; D. Visani ; C.M. Okonkwo ; H. Yang ; H.P.A. van den Boom ; G. Tartarini ; E. Tangdiongga ; A.M.J. Koonen "First demonstration of HD video distribution over large-core POF employing UWB for in-home networks,", 2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference, pp 1-3 (2011)

[6] Y. Shi et al., "Ultrawideband Signal Distribution Over Large-Core POF for In-Home Networks," in Journal of Lightwave Technology, vol. 30, no. 18, pp. 2995-3002, Sept.15, 2012.

[7] Hadi, M., Aslam, N. & Jung, H. (2019). Performance Appraisal of Sigma Delta Modulated Radio over Fiber System. Journal of Optical Communications, 0(0), pp. -Jan. 2019, from doi:10.1515/joc-2018-0227

[8] Hadi, M., Jung, H., Traverso, P., et al. (2019). Digital Radio Frequency Transport over Optical Fiber for 5G Fronthaul Links. Journal of Optical Communications, 0(0), pp. - Apr. 2019, from doi:10.1515/joc-2019-0051

[9] Hadi, M., Hadi, M., Aslam, N., et al. (2019). Experimental Demonstration of MASH Based Sigma Delta Radio over Fiber System for 5G C-RAN Downlink. Journal of Optical Communications, 0(0), pp. -. Retrieved 23 Feb. 2019, from doi:10.1515/joc-2019-0011

[10] J. Nanni et al., "VCSEL-Based Radio-Over-G652 Fiber System for Short-/Medium-Range MFH Solutions," IEEE/OSA J. Lightw. Technol, vol. 36, pp. 4430-4437, Oct.1, 2018.

[11] D. Visani, G. Tartarini, M. N. Petersen, L. Tarlazzi and P. Faccin, "Link Design Rules for Cost-Effective Short-Range Radio Over Multimode Fiber Systems," in IEEE Transactions on Microwave Theory and Techniques, vol. 58, no. 11, pp. 3144-3153, Nov. 2010

[12] D. Visani et al.: Wired and wireless multi-service transmission over 1mm-core GI-POF for in-home networks, Electron. Lett., vol. 47, pp.203-205, Feb. 2011

[13] D. Visani et al., " 3×2N-QAM Constellation Formats for DMT Over 1-mm Core Diameter Plastic Optical Fiber", IEEE Photonics Technology Letters, 23, pp. 768-770 (2011).

[14] J. A. Tatum et al., "VCSEL-based interconnects for current and future data centers," J. Lightw. Technol., vol. 33, no. 4, pp. 727–732, Feb. 15, 2015

[15] J. Nanni et al., "Modal noise in 850nm VCSEL-based radio over fiber systems for manifold applications," Proc. Fotonica AEIT Italian Conference on Photonics Technologies, Turin, 2015, pp. 1-4.

[16] M.Sauer, A. Kobyakov, and A.B. Ruffin, ``Radio-over-fiber transmission with mitigated stimulated brillouin scattering,", IEEE Photonics Technology Letters, vol.19, no.19, pp. 1487--1489, Oct 2007..

[17] G. Tartarini and P. Faccin, ``Efficient characterization of harmonic and intermodulation distortion effects in dispersive radio over fiber systems with direct laser modulation," Microwave and Optical Technology Letters, vol. 46, no. 2, pp. 114--117, 2005.

[18] P. M. Sevila, V. Almenar, and J. L. Corral, "Transmission Over SSMF at 850 nm: Bimodal Propagation and Equalization," *IEEE/OSA J. Lightw.Technol.*, vol. 35, no. 19, pp. 4125–4136, Oct 2017

[19] J. Nanni et al., "Modal Noise Mitigation in 850-nm VCSEL-Based Transmission Systems Over Single-Mode Fiber," in IEEE Transactions on Microwave Theory and Techniques, vol. 64, no. 10, pp. 3342-3350, Oct. 2016

[20] J. Weiss, "Analog Optical RF-Links for Large Radio Telescopes,"IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), San Diego, CA, 2018, pp. 24-27. doi: 10.1109/BCICTS.2018.8551058

[21] D. Visani et al., "Wired and wireless multi-service transmission over 1mm-core GI-POF for in-home networks," in Electronics Letters, vol. 47, no. 3, pp. 203-205, 3 February 2011.

[22] C. Wang et al., "Machine Learning Detection for DMT Modulated 112-Gbps VCSEL-MMF Optical Interconnection," Asia Communications and Photonics Conference (ACP), Hangzhou, 2018, pp. 1-3.

[23] W. P. Ng, T. Kanesan, Z. Ghassemlooy and C. Lu, "Theoretical and Experimental Optimum System Design for LTE-RoF Over Varying Transmission Span and Identification of System Nonlinear Limit," in IEEE Photonics Journal, vol. 4, no. 5, pp. 1560-1571, Oct. 2012.

[24] S. F. Yu, "Nonlinear dynamics of vertical-cavity surface-emitting lasers," in *IEEE Journal of Quantum Electronics*, vol. 35, no. 3, pp. 332-341, March 1999.

[25] Roselli, L., Borgioni, V., Zepparelli, F., Ambrosi, F., Comez, M., Faccin, P., Casini, A.: Analog laser predistortion for multiservice radio-over-fiber systems. IEEE J. Lightw. Technol. 37(5), 1211-1223 (2003)

[26] Zhu, R., Zhang, X.: Broadband predistortion circuit design for electro-absorption modulator in radio over fiber system. In Proceedings of the Optical Fiber Communications Conference and Exhibition(OFC), San Francisco, CA, USA (2014).

[27] F. Fuochi, M. U. Hadi, J. Nanni, P. A. Traverso and G. Tartarini, "Digital predistortion technique for the compensation of nonlinear effects in radio over fiber links," IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), Bologna, 2016, pp. 1-6.

[28] A. Hekkala et al., "Predistortion of Radio Over Fiber Links: Algorithms, Implementation, and Measurements," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 3, pp. 664-672, March 2012.

[29] Mateo, C., Carro, P.L., Garca-Dcar, P., De Mingo, J., Salinas, I.: Radio-overfiber linearization with optimized genetic algorithm CPWL model. Opt. Express. 25(4). pp. 3694-3708, Feb. 2017.

[30] M. U. Hadi, P. A. Traverso, G. Tartarini, O. Venard, G. Baudoin and J. Polleux, "Digital Predistortion for Linearity Improvement of VCSEL-SSMF-Based Radio-Over-Fiber Links," in *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 155-157, Feb. 2019. doi: 10.1109/LMWC.2018.2889004

[31] M. Abi Hussein, V. A. Bohara and O. Venard, "On the system level convergence of ILA and DLA for digital predistortion," 2012 International Symposium on Wireless Communication Systems (ISWCS), Paris, 2012, pp. 870-874.

[32] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," Signal Processing, IEEE Transactions on, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[33] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," Communications, IEEE Transactions on, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[34] Hekkala, A., Lasanen, M.: 'Performance of adaptive algorithms for compensation of radio over fiber links'. Proc. IEEE Wireless Telecommunication Symp., Prague, Czech Republic, April 2009, pp. 1–5

[35] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," Signal Processing, IEEE Transactions on, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[36] L. Guan, and A. Zhu, "Optimized Low-Complexity Implementation of Least Squares Based Model Extraction for Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Microw. Theory Tech*., vol. 60, no. 3, pp. 594-603, Mar. 2012.

[37] 3GPP TS36.104 V15.2.0, Evolved Universal Terrestrial Radio Access (E-UTRA) User Equipment (UE) Radio Transmission and Reception (Release 15), March 2018.

[38] L. C. Vieira and N. J. Gomes, "Experimental demonstration of digital predistortion for orthogonal frequency-division multiplexing-radio over fibre links near laser resonance," in IET Optoelectronics, vol. 9, no. 6, pp. 310-316, 12 2015. doi: 10.1049/iet-opt.2014.0160

[39] M. U. Hadi, "Extending the benefits of LTE to unlicensed spectrum," *2015 International Conference on Information and Communication Technologies (ICICT)*, Karachi, 2015, pp. 1-3. doi: 10.1109/ICICT.2015.7469592

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 161-168 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Smart Meter Data Analysis for Electricity Theft Detection using Neural Networks

Priyamvada Chandel*,[1], Tripta Thakur[2]

[1] *Central Power Research Institute, Bhopal, India*

[2] *Maulana Azad National Institute of Technology, Bhopal, India*

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 11 July, 2019
Online :21 July, 2019

Keywords:
Electricity Theft Detection
Convolutional Neural Network (CNN)
Recurrent Neural Network (RNN)
Bidirectional Long Short Term Memory (BiLSTM)

ABSTRACT

The major problem in electric utility is Electrical Theft, which is harmful to electric power suppliers and causes economic loss. Detecting and controlling electrical theft is a challenging task that involves several aspects like economic, social, regional, managerial, political, infrastructural, literacy rate, etc. Numerous methods were proposed formerly for detecting electricity theft. However, the previous works considered only the one dimensional (1-D) energy consumption data which apprehended the periodicity and were found inefficient in storing and retrieving the memory that resulted in a lower accuracy rate of detection. Hence, this research study intends Convolutional Neural Network combined with Bidirectional Long Short Term Memory based Recurrent Neural Network (CNN-RNN-BiLSTM) for overcoming the aforementioned problems in the detection of electricity theft. The CNN captures the global variables of 1-D data and identifies the non-periodicity and periodicity of 2-D electricity consumption records. RNN-BiLSTM extends the memory storage capacity of the neural network with bidirectional flow of information, thereby allowing learning order dependencies. The proposed method results of the predicted values of the electricity theft show better accuracy rate with reduced time during the training phase and reduced number of epochs. The proposed model helps to discriminate the customers for preventing fraudulent activities in the usage of electric power. The validation of the proposed method is carried out by comparing the method with the existing Support Vector Machine (SVM) and multi-class SVM models. The comparative results prove that the proposed CNN-RNN-BiLSTM model of electricity theft detection works efficiently.

1. Introduction

Electricity has become a basic need in our life. Losses often occur during the transmission, generation, and distribution of electricity. The losses in the electricity supply to consumers, refer to the quantity of electricity introduced into the distribution and transmission grids that are unpaid by the users. The losses of electricity can be classified into technical losses (TLs) and Non-Technical losses (NTLs) [1]. Advanced Metering Infrastructure (AMI) they enable energy companies to obtain active energy, phase current and voltage, apparent energy, and Reactive lead from smart meter deployed at home [2]. Smart meters (SM) collect the active energy from the real-time, and reveal the user habits and behavior at home to forecast the financial loss. For example, if the daily energy is low, then the user is not at home. To disclose the theft detection is addressed in this work. Smart meters allow functions to perceive Non-Technical Loss (NTL), which was very difficult in previous days. Now SM even out functions that often measures the energy consumption (EC), and gives the better perceptive of customer consumption behavior [3].

Non-Technical losses occur due to loads and conditions when the Technical losses calculation become unsuccessful to consider or due to the triggering of the external factors. Non- Technical losses are tedious to evaluate since these losses are not frequently considered by the system operatives and therefore have no verified data existed [4]. One of the fundamental NTLs is the theft of electricity. This theft involves bypassing of the electricity meter,

* Corresponding Author: Priyamvada Chandel, Central Power Research Institute, Bhopal, India. Email Id: priyamvada@cpri.in

hacking the meter, or tampering the meter reading [5]. Electricity theft results in a great revenue fall of power company because of the surging electricity, the dense load of electrical systems, and it brings threats to public safety.

In many developing countries, at many distribution feeders, electricity theft is openly visible. In the past, indiscriminate machine learning methods have been employed for detecting electricity theft. However, convenient variables are created from meter readings, and the performance will be optimal [6]. The existing methods have some concerns that need to be addressed further. For Artificial Intelligence (AI) based methods, due to the difficulty in building a labeled dataset of electricity thefts, the application of classification methods is limited. Since the clustering algorithms are unverified, tampered load profiles with normal figures cannot be identified, ensuing in low accuracy of prediction. For the state-based methods, the measurement data and system information acquisition are much more difficult to obtain [7]. For deep learning approaches, since the scalability of the framework is arbitrary, the network can end up being unstable. Furthermore, the weights fed to the layers of the network have the underfitting of the model that tends to increase the number of epochs for training the model. Epoch is defined as the number of iterations related to the samples of the dataset. For training a network with a large dataset and a reduced number of epochs, memory storage becomes an important concern for the classification model of the network.

2. Literature Review

This section includes a survey of electricity theft detection methods proposed by researchers formerly. The studies can be categorized into hardware-based and non-hardware based solutions to the detection of electricity theft. Non-hardware based solutions include AI and machine learning techniques.

Nizar et al. applied the Extreme Learning Machine (ELM) for electricity theft detection [8]. The ELM-based technique extracts forms of customer behavior using their previous consumption data in kWh and perceives abnormal deeds. The performance of the classifier was compared with OS-ELM and SVM for validation. However, the classification rate of testing phase was found low for this approach.

In [9], the authors recommended a multi-class SVM which was trained to identify a sample of load profiles whether it is malicious or normal. This paper addressed and solved the imbalanced training by generating a synthetic dataset. The algorithm also uses silhouette plots to identify the different distributions in the dataset. One limitation of the machine learning approach is that it is vulnerable to contamination attacks.

In [10], the authors proposed an optimum-path forest (OPF) based unsupervised NTL detection methodology and compared with other popular clustering methods including k-means and GMM. OPF has attained the most accurate results while considering both the applications on two datasets composed of industrial and commercial profiles of irregular and regular consumers. However, clustering techniques are not cost-effective.

In [11], the researcher suggested for distribution lanes from current and real power measurements using linear regression technique; subsequently, NTL of all lane is measured in accordant with the estimated resistance value to find the electricity theft. Limitation of this method is that correlating more than two independent factors with dependent factor becomes complicated in the linear regression model.

In [12], the authors proposed for distribution load transformer is employed in detecting meter tampering using state estimation-based technique. The variation in the measured values and the estimated values is assessed to make suspicious details of consumers with metering issues. Customer meter data in conjunction with intelligent devices measurements in smart distribution grid are used for DSE and NTL detection without additional meter investment at points of delivery.

In [13], the researchers presented a probabilistic methodology NTL assessment in the distribution system. The sensitivity of technical loss when related to the load variation is derived, and the probabilistic distributions of total loss and technical loss are evaluated. This approach is suitable only for countries that have scarce resources and is not suitable for smart grid applications.

In [14], the authors recommended a deep-learning-based real-time technique for the finding of electricity thefts was projected. This paper utilizes real-time data and its state vector estimator (SVE), and computes the occurrence vector and state vector and the topology of the power system, and a detection scheme based on broad belief network assist the SVE identifies the false data injection (FDI). However, possible FDI attacks are not imposed on the network.

In [15], the researchers proposed Principal Component Analysis (PCA) based Theft Detection system to identify energy theft in AMI. This approach was then compared with SVM classifier to validate the supremacy of PCATD performance. But the results obtained from PCA are found to be inconsistent under change of constraints.

The study conducted in [16] detect the anomalous behavior in customer-defined patterns using an SVM classifier. However, the study relied on monthly measurements of the demand, and the detection window is approximately taken for two years. Additionally, the algorithm was found to show low performance while detecting any theft cases that are not viewed as unexpected variations.

Beyond the concept of machine-learning algorithms, techniques based on game-theory have been applied in [17], identified the fraud was executed between the utility company and fraud-committing consumers. The primary problem with these approaches is the struggle in the formulation of practical and realistic utility functions for the players involved.

In [18], the authors proposed a model that exploits the time-series nature of the customers' electricity consumption to implement a Gated Recurrent Unit (GRU)-RNN. As an improvement to RNN, random search analysis in the learning stage has been adopted. In [19], the authors suggested a deep neural network (DNN)-based customer-specific detector, that implements a sequential grid search analysis in its learning stage for tuning hyperparameters. Wide-ranging investigations are performed based on openly accessible originally energy consumption database of 5000 customers. These approaches suffer from inefficient memory for storage and retrieval.

In [20], the author proposed DNN with transfer entropy measure-based anomaly detection for both sensor measurements and innovative sequences, that can be assessed in a data-driven fashion without relying on a model of the underlying dynamic system. The results of the simulation indicated that the method takes a longer time for computation.

In [21], the authors recommended a novel data analysis method for detecting false data injection attack mitigation (FDIA) based on a data-centric model using the margin setting algorithm. The performance of the suggested methodology is presented employing the six-bus power network in a measurement system of the wide-area environment through simulation. However, it handles only a limited data for the period of the identification process.

In [22], the authors proposed to identify cyber-attacks in Cyber-Physical systems using a novel unsupervised approach, combining LSTM-RNN for behavioral-based intrusion detection. This model not only identifies the attacked sensor, but also detects the irregularities in the CPS. The limitation of this method was that the ability to validate the false positives was found less.

In [23], the authors proposed ANN-based intrusion detection. The MLP is trained to employ packet traces of internet, and then it is evaluated based on its capability to resist DoS/DDoS attacks. The study spotlight on the normal and threat patterns classification on an IoT Network. But the mean absolute error obtained for MLP was found to be high.

3. Proposed Methodology

In this section, the Electricity Theft Detection and Implementation using CNN-RNN-BiLSTM model are discussed. The advancement of the methodology proposed such work consists of the following fundamental stages:

- To construct a network for ensuring enhanced storage of memory along with cache memory and for the retrieval of the system.
- To build a balanced system for reducing the training time and epochs.
- To detect electricity theft in the Indian distribution system.

Electricity Theft Detection

Electricity Theft Detection study is categorized into two types: Data-driven solution and Hardware-based solution. The hardware-based solution met with failure due to severe weather conditions and difficulty in maintaining the devices. The data-driven solution has attained more attention

nowadays. The data-driven is planted on data fusion from sensors and AMI. The methodology uses SM data for convolution.

Data attributes

The following are the important attributes of electricity consumption behavior from 41 meters dated 1/1/2017 to 16/2/2018 with the frequency of every 15 minutes that the users consume energy daily. The attributes namely; R, Y, B phase current and voltage, active energy, apparent energy, reactive lag, reactive lead, PF, and label.

CNN

A CNN structure transforms the input work into an output work using a stack of the discrete layer. CNN comprises of an input layer, multiple hidden layers, and output layer; whereas the hidden layer embraced of pooling layer and convolutional layer [24]. The receptive field is the zone of the input matrix that influences a precise unit of the network. The convolutional layer which comprises of the less receptive field and it can extend up to full depth of

input, and it is composed of kernels or filters. Each filter is convolved using height and width of input at onward pass, and it produces the 2D activation map of the filter. The pooling layer reduces the parameters, network computation, and controls the overfitting. The pooling function employed in the proposed method is maximum pooling.

The pooling layer accepts the volume of size $W \times H \times D$, and produces a volume of size $W' \times H' \times D$.'

$$W' = \frac{W-e}{s+1} \quad (1)$$

$$H' = \frac{H-e}{s+1} \quad (2)$$

$$D' = D \quad (3)$$

Where e is the spatial extent and s is the stride. The pooling layer operates independently on every depth of the input slice and resizes it in the spatial dimension. Due to these advantages, the CNN network is used for extracting the convolution features efficiently.

RNN

In RNN the memory formed through recurrent connection. In feedforward networks, inputs are connected, and it is also independent. It exhibits vital behavior for sequential classification in a time sequence. Consider an input classification $X = (X_1, ..., X_T)$, a standard RNN evaluates the hidden vector classification $h = (h_1, ..., h_T)$ and output vector classification $Y= (Y_1, ..., Y_T)$ by repeating the below equations from t = 1 to T [25]:

$$h_t = H(W_{xh}X_t + W_{hh}h_{t-1} + b_h) \quad (4)$$

$$y_t = W_{hy}h_t + b_0) \quad (5)$$

where,

W - weight matrices (e.g., W_{xh} is the input-hidden weight matrix),

b - denote bias vectors (e.g., b_h is the hidden bias vector) and

H - hidden layer function. H is usually a sigmoid function.

Bidirectional LSTM (BiLSTM)

Long Short Term Memory (LSTM) Recurrent Neural Networks (RNN) have been solved the problem of vanishing gradients in RNNs. The LSTM structure composed of memory blocks and block consists of three multiplicative units such as input, forget and output gates and one or more self-associated memory cells. The multiplicative units afford incessant analogs of reading, write, and rearrange functions for cells. An LSTM is similar to RNN, but the main difference is the memory blocks are

used in the hidden layer instead of nonlinear units. To avoid the vanish-gradient issue and to access and store data for a long time of period, the multiplicative gates used in LSTM. Consider that the input gate rests closed (0) the cell will not be overridden by the fresh inputs, and can be set accessible for the next sequence, by opening the output gate.

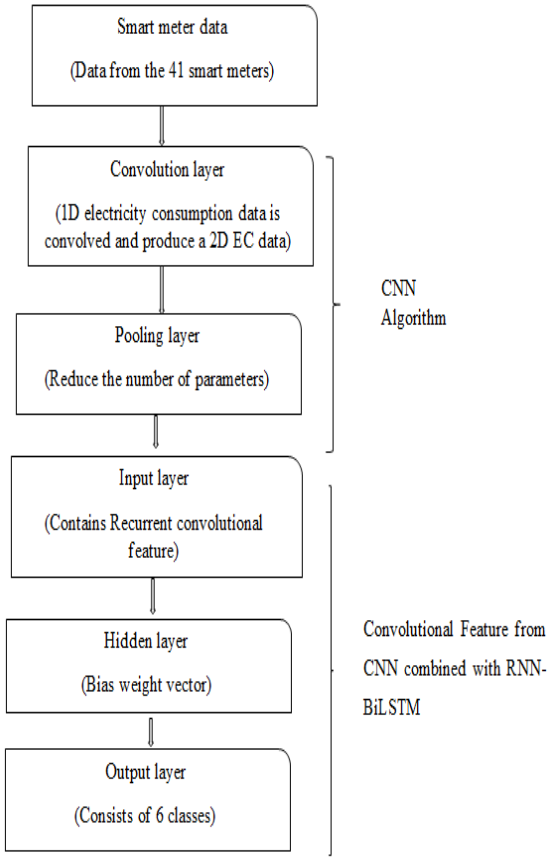


Figure.1: Methodology of CNN-RNN BiLSTM

Traditionally RNN has the problem for vanishing and exploding the gradients. The issue is due to the weight proportional to the gradient of error. During backpropagation, gradients multiply with small values leads to earning smaller values. LSTM does not suffer from any gradients, and it also controls the cell state retains in memory and passes the output. An activation function is a function that defines the output of each node. A sigmoid activation function is used in the design. A sigmoid activation function is an 'S' shaped curve function that ranges between 0 and 1.

$$\mathrm{i} = \sigma(\mathrm{xtUi} + \mathrm{st} - 1\mathrm{Wi}) \quad (6)$$

$$\mathrm{f} = \sigma(\mathrm{xtUf} + \mathrm{st} - 1\mathrm{Wf}) \quad (7)$$

$$\mathrm{o} = \sigma(\mathrm{xtUo} + \mathrm{st} - 1\mathrm{Wo}) \quad (8)$$

where,

xt is the input data sequence, the i-input gate is the gate that allows the input data to enter the network, the f-forget gate is responsible in deciding whether to keep the data or delete the data which is no longer useful, and o-output gate produces the output to be sent to the hidden state. The formula for each gate is formulated using the weight matrix (W), time sequence (U) and a sigmoidal activation function to push the values between a certain interval.

CNN-RNN-BiLSTM

This paper proposes a CNN-RNN-BiLSTM method for the detection of electricity theft. A hybrid wide and deep CNN and RNN is proposed with memory extension rendered by BiLSTM network. The Wide CNN component can capture the convolution features of 1-D electricity consumption data. And the Deep CNN module can recognize the non-periodicity and periodicity of the electricity consumption records.

BiLSTM based RNN offers enhanced memory storage with bidirectional information flow, and make the retrieval and the memory storage and recovery are possible for long term dependencies. The retrieved layer occurs with six classes.

The architecture of the proposed system

Fig 1 shows the flow diagram of the proposed CNN-RNN-BiLSTM prediction model. Data from 41 smart meters are collected and given to the Convolution layer. The convolutional layer convolves the 1-D electricity data and produces 2-D electricity consumption data. And the pooling layer reduces the number of parameters. This working flow of Convolution and Pooling layer is depicted in Fig 2. L is the number of attributes, M is the number of feature values from the convolution layer and pooling layer. Attributes considered in work are phase voltages, phase currents, active energy, apparent energy, reactive lag, reactive lead, and power factor.

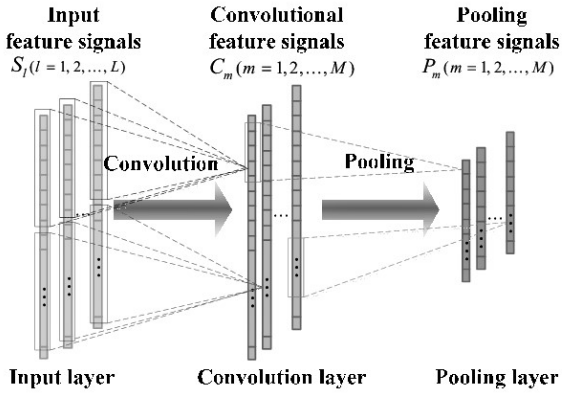


Figure. 2 Convolution and Pooling layer

The length of the input vector f is n, and that of the kernel g is m. The convolution of input vector and kernel function $f * g$ is defined as:

$$(f * g)(i) = \sum_{j=1}^{m} g(j) \cdot f(i - j + \frac{m}{2}) \quad (9)$$

These convolved values passed through max pool layer. The maximum value obtained from pooling layer is chosen as the convolution feature. The convolution features are loaded onto the input layer of RNN. The general structure of RNN is given in Fig 3.

RNN consists of 3 layers, namely, Input layer, Hidden layer, and Output layer. Input cells get convolution features and pass to the hidden or recurrent layer. It consists of neurons x1, x2, x3, etc.,

which gets the input data sequences. The hidden layer contains the hidden units h1, h2, etc., which hold the activation functions that define the output of the neurons in the recurrent layer. The neurons in the output layer y1, y2, etc., will send the predicted values of electricity theft as the output. Each cell of RNN in the proposed prediction model is a BiLSTM memory cell, whose structure is displayed in Fig 4.

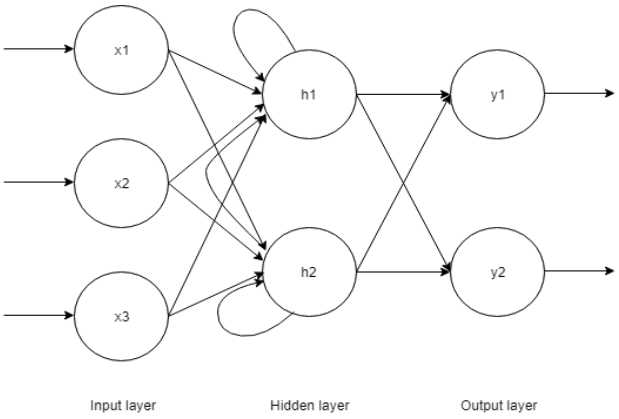


Figure 3. Structure of RNN

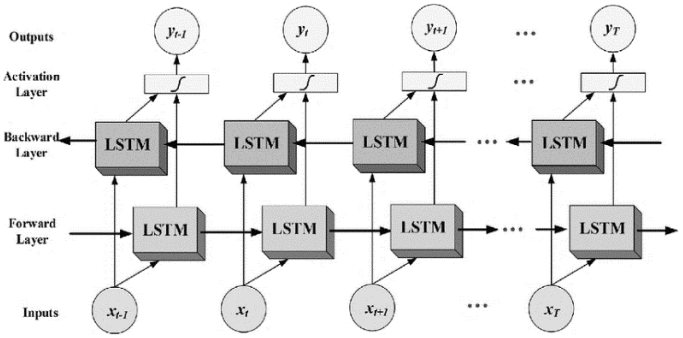


Figure 4. Structure of BiLSTM

The information flow in a BiLSTM is bidirectional, unlike the LSTM. BiLSTM cells acquire bidirectional long term dependencies among time steps or sequential data. These previous long term dependencies are useful for predicting the next state.

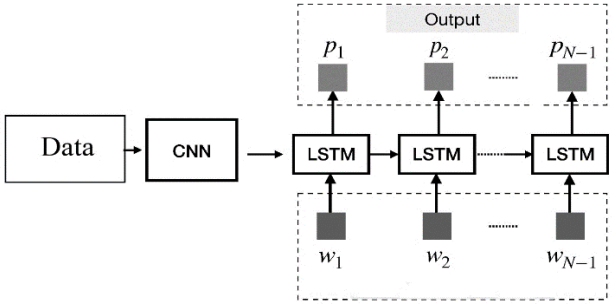


Figure. 5 CNN-RNN-BiLSTM

Fig. 5 shows the architecture of the proposed CNN-RNN-BiLSTM prediction model. Input data is sent to CNN for convolution. Convolution features extracted using CNN are sent to RNN-BiLSTM. The outputs of the model are the detected values of electricity theft.

The voltage and current threshold values for different tamper types are given in Table 1.

Table 1. Occurrence threshold values for different Tamper types

S. NO	TAMPER TYPE	OCCURRENCE THRESHOLDS
1	Link Miss phase-wise	Vx < 60 % Vref and Ix > 5 % Imax
2	CT Bypass	Iavg> 2% IBasic and Ix >2.5% IBasic
3	CT Open phase-wise	Vx > 60% Vref and Ix< 0.5% Imax
4	Current Unbalance	Vavg > 60% Vref and Iavg > 1:
5	Voltage Unbalance	Vavg > 66 % Vref
6	Current Reversal phase-wise	Vavg < 60% Vref and Net PF > -0.5:

Six classes are achieved based on Vref=240V, Ibasic=10A, Imax=40.0A, Iavg=Average of three-phase currents and the tampers are link miss-phase wire, CT bypass, CT open phase-wise, current unbalance, voltage unbalance and current reversal phase wire.

4. Result and Discussion

Evaluation of the proposed energy theft detection using abnormal conditions scheme is done using MATrix LABoratory (MATLAB) The energy usage data comes from Smart Meter Data.

Table 2. Results obtained

Class Labels	Six classes of fault types
No. of Observations	6000
Control classes	[5x1] double
Correct Rate	0.9712
Error Rate	0.0288
Classified Rate	1
Specificity	0.9929
Sensitivity	0.9990
Positive Predictive Value	0.9640
Negative Predictive Value	0.9998
Positive Likelihood	139.7162
Negative Likelihood	0.0010
Prevalence	0.1608

4.1. Experimental Data

Dataset taken for this study is the real-time data communicated by 41 numbers of three phases four-wire smart meters having a unique number, and the data is sent to the Data Collection Unit (DCU) of Chhattisgarh State Electricity Board Raipur. Totally 231637 data for every 15 minutes' interval of time from 1/1/2017 to 16/2/2018 is considered.

4.2. Testing

Then the network is tested using the 6000 data. Each label consists of 1000 data. The tested results are obtained from MatlabR2018a are given in Table 2.

The Fig.6 shows the confusion matrix accuracy of 97.1%.

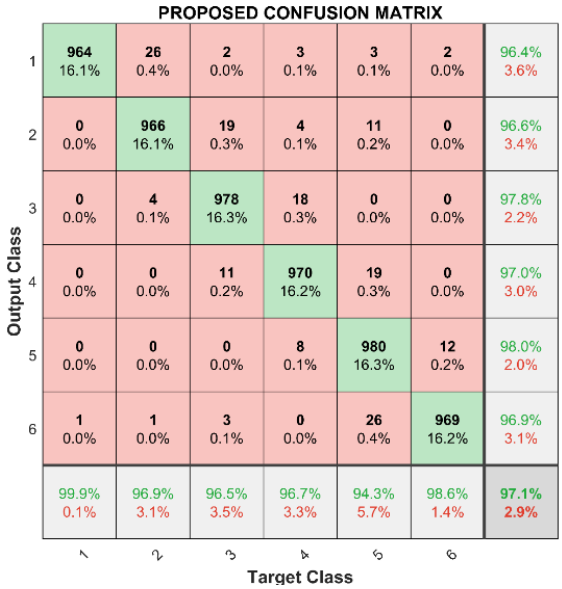


Figure 6. Proposed confusion matrix

4.3. Parameter Study

The Fig.7 shows the accuracy and loss of RNN-BiLSTM.

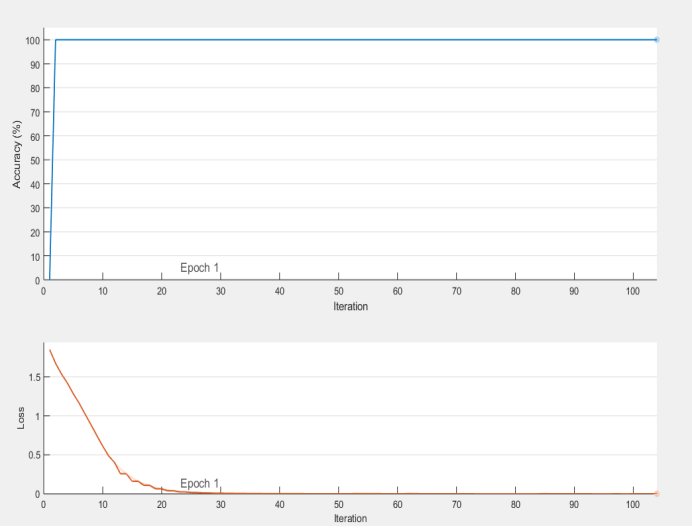


Figure 7. Accuracy and loss of RNN-BiLSTM

First Smart meter(SM) data are taken as label and balanced. Using the Convolutional Neural Network (CNN), the convolution feature of dataset is extracted. In CNN, we used two layers called Convolution and pooling layer to extract the convolution feature. Here, CNN works well for identifying simple patterns within the smart meter data, which will be employed to form higher complex patterns within higher layers. A one dimensional CNN is very efficient to derive the location of feature from shorter (fixed-length) sections of the entire data set and where the location of the feature within the section is not of high significance. The extracted feature from the convolutional feature is used in the RNN-BiLSTM network to predict the class of fault. In RNN-BiLSTM networks, information can be accessed both in forward and reverse direction. These forward and reverse feature in RNN-BiLSTM have to access the past as well as the future information. This has been effectively employed in many practical applications like language translation, future prediction, image captioning, etc., Using the Bidirectional LSTM we feed the learning algorithm with smart meter data once from the beginning to end and once from end to the beginning. There are debates here, but it usually learns faster than one-directional approach, although it depends on the task

Here the loss (error) reduces from 2 to 0, and the accuracy shows increase the value from 0 to 100. Hence it shows our proposed CNN-RNN-BiLSTM shows the efficient accuracy of 100% for the original Training Dataset.

The proposed system is compared with the existing technique and proved the accuracy of 97.1%. Table 3 given below represents the compared result of the proposed with the existing methods, and Fig. 8 represents the accuracy score of the proposed method [27]:

Table.3 Validation of proposed system

Algorithm	Arguments	Accuracy Score
Combined CNN model	100 epochs	0.9267
Single CNN model	100 epochs	0.9218
Simple DNN model	100 epochs	0.9145
Linear SVC	kernel: linear function	0.9178
Random Forest	Max depth: 7	0.9164
Logistic Regression	penalty: L2	0.9141
Proposed Model	Convolution BiLSTM	**0.9712**

The sensitivity, specificity, and accuracy for the proposed CNN-RNN BiLSTM are calculated using the confusion matrix which comprises of False Positive (FP), True Positive (TP), False Negative (FN), and True Negative (TN). Accuracy, sensitivity, and specificity are calculated using the following formulae.

$$Accuracy = \frac{TP+TN}{TN+FN+TP+FP} \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

$$Sensitivity = \frac{TP}{FN+TP} \quad (12)$$

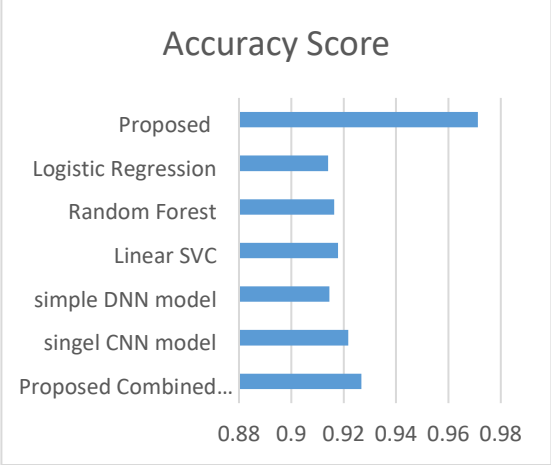


Figure 8. Accuracy score of CNN-RNN BiLSTM

The calculated values are found to be high for the proposed classifier. The specificity, sensitivity, and accuracy obtained for the proposed method are 99.29%, 99.9%, and 97.12% respectively. Fig 9 shows the performance assessment of the proposed framework:

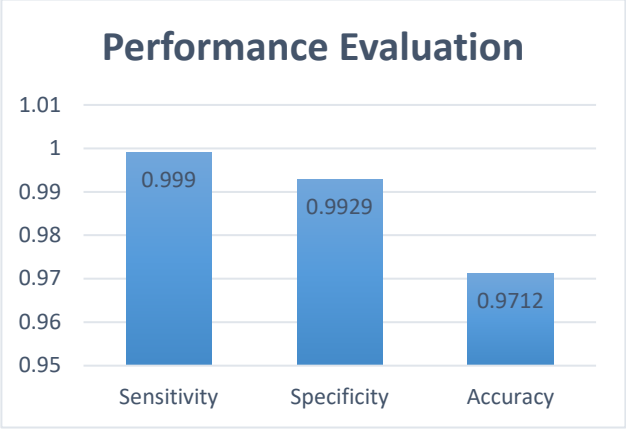


Figure 9. Performance evaluation of CNN-RNN BiLSTM

The performance of the proposed CNN-RNN-BiLSTM is compared with existing methodologies in the literature. Table 4. provides the comparative analysis of the performance parameters of the proposed model with two other models, namely, Support Vector Machine (SVM) and Multi-class SVM.

Table. 4 Performance comparison table

Method	**Accuracy (%)**
SVM [16]	60
Multi-class SVM [9]	94
Proposed	**97.1**

From the table, it is inferred that the proposed CNN-RNN-BiLSTM theft prediction model has better performance compared to the existing models.

5. Conclusion

In this paper, the detection scheme of theft using the abnormal changes in electric values measures in the smart meter is proposed. The proposed electricity theft detection model is based on the combination of CNN and RNN-BiLSTM network, which detects whether the metering data has an abnormal behavior or not. The model predicts the abnormalities in consumer's power consumption behaviors and classifies the behavior based on six classes. The proposed neural network model achieved an accuracy of 100% in training phase and 97.1% in testing phase. The use of technique proposed in this paper will help power utilities to predict theft in lesser time and higher accuracy. Hence in the future, this theoretical approach to the detection of electricity is to be tested for real time practical application. And the use of hybrid spike neural network for electricity theft detection model is presumed to improve the performance of detection effectively.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X. S. Shen, Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology*, *19*(2), 105-120, 2004. DOI: 10.1109/TST.2014.6787363

[2] Y. Sun, L. Lampe, V. W. S. Wong, "Smart meter privacy: Exploiting the potential of household energy storage units," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1–1, 2017. DOI: 10.1109/JIOT.2017.2771370

[3] J. Nezhad, T. K. Wijaya, M. Vasirani, K. Aberer, "SmartD: Smart Meter Data Analytics Dashboard," Proceedings of the 5th international conference on Future energy systems. ACM, pp. 213–214, 2014. DOI: 10.1145/2602044.2602046

[4] J. P. Navani, N. K. Sharma, S. Sapra, Technical and non-technical losses in the power system and its economic consequence in the Indian economy. *International Journal of Electronics and Computer Science Engineering*, *1*(2), 757-761, 2012.

[5] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, S. Zonouz, A multi-sensor energy theft detection framework for advanced metering infrastructures. *IEEE Journal on Selected Areas in Communications*, *31*(7), 1319-1330, 2013. DOI: 10.1109/JSAC.2013.130714

[6] R. Razavi, A. Gharipour, M. Fleury, I.J. Akpan, A practical feature-engineering framework for electricity theft detection in smart grids. *Applied Energy*, *238*, 481-494, 2019. DOI: 10.1016/j.apenergy.2019.01.076

[7] K. Zheng, Q. Chen, Y. Wang, C. Kang, Q. Xia, A Novel Combined Data-Driven Approach for Electricity Theft Detection. *IEEE Transactions on Industrial Informatics*, *15*(3), 1809-1819, 2019. DOI: 10.1109/TII.2018.2873814

[8] A.H. Nizar, Z.Y. Dong, Y. Wang, Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Transactions on Power Systems*, *23*(3), 946-955, 2008. DOI: 10.1109/TPWRS.2008.926431

[9] P. Jokar, N. Arianpoo, V.C. Leung, Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid*, *7*(1), 216-226, 2016. DOI: 10.1109/TSG.2015.2425222

[10] L.A.P. Júnior, C.C.O. Ramos, D. Rodrigues, D.R. Pereira, A.N. de Souza, K.AP. da Costa, J.P. Papa, Unsupervised non-technical losses identification through the optimum-path forest. *Electric Power Systems Research*, *140*, 413-423, 2016. DOI: 10.1016/j.epsr.2016.05.036

[11] D.N. Nikovski, Z. Wang, *U.S. Patent No. 9,945,889*. Washington, DC: U.S. Patent and Trademark Office, 2018.

[12] S.C. Huang, Y. L. Lo, C.N. Lu, Non-technical loss detection using state estimation and analysis of variance. *IEEE Transactions on Power Systems*, *28*(3), 2959-2966, 2013. DOI: 10.1109/TPWRS.2012.2224891

[13] E.A.A. Neto, J. Coelho, Probabilistic methodology for Technical and Non-Technical Losses estimation in the distribution system. *Electric Power Systems Research*, *97*, 93-99, 2013. DOI: 10.1016/j.epsr.2012.12.008

[14] Y. He, G.J. Mendis, J. Wei, Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. *IEEE Transactions on Smart Grid*, *8*(5), 2505-2516, 2017.DOI: 10.1109/TSG.2017.2703842

[15] S.K. Singh, R. Bose, A. Joshi, Energy theft detection in advanced metering infrastructure. In *2018 IEEE 4th World Forum on the Internet of Things (WF-IoT)* (pp. 529-534), 2018. DOI: 10.1109/WF-IoT.2018.8355148

[16] J. Nagi, K.S. Yap, S.K. Tiong, S. K. Ahmed, M. Mohamad, Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE transactions on Power Delivery*, *25*(2), 1162-1171, 2010. DOI: 10.1109/TPWRD.2009.2030890

[17] S. Amin, G.A. Schwartz, A. A., Cardenas, S.S. Sastry, Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure. *IEEE Control Systems Magazine*, *35*(1), 66-81, 2015. DOI: 10.1109/MCS.2014.2364711

[18] M. Nabil, M. Ismail, M. Mahmoud, M. Shahin, K. Qaraqe, E. Serpedin, Deep Recurrent Electricity Theft Detection in AMI Networks with Random Tuning of Hyper-parameters. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 740-745), 2018. DOI: https://arxiv.org/abs/1809.01774

[19] M. Ismail, M. Shahin, M.F. Shaaban, E. Serpedin, K. Qaraqe, Efficient detection of electricity theft cyber-attacks in AMI networks. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-6), 2018. IEEE. DOI: 10.1109/WCNC.2018.8377010

[20] D. Shi, Z. Guo, K.H. Johansson, L. Shi. Causality countermeasures for anomaly detection in cyber-physical systems. *IEEE Transactions on Automatic Control*, *63*(2), 386-401, 2018. DOI: 10.1109/TAC.2017.2714646

[21] Y. Wang, M.M. Amin, J. Fu, H.B. Moussa. A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids. *IEEE Access*, *5*, 26022-26033, 2017. DOI: 10.1109/ACCESS.2017.2769099

[22] J. Goh, S. Adepu, M. Tan, Z.S. Lee. Anomaly detection in cyber-physical systems using recurrent neural networks. In *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)* (pp. 140-145), 2017. IEEE. DOI: 10.1109/HASE.2017.36

[23] E. Hodo, X. Bellekens, A. Hamilton, P.L. Dubouilh, E. Iorkyase, C. Tachtatzis, R. Atkinson, R. (2016, May). Threat analysis of IoT networks using artificial neural network intrusion detection system. In *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1-6). IEEE.DOI: 10.1109/ISNCC.2016.7746067

[24] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *European Conference on Computer Vision*, pp. 791–808, 2016. DOI: https://arxiv.org/abs/1607.08378

[25] A.A. Sharfuddin, M.N. Tihami, M.S. Islam. A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4), 2018. IEEE.

[26] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on, 2016. DOI: 10.1109/SPAWC.2016.7536744

[27] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, B. Yang. Energy Theft Detection with Energy Privacy Preservation in the Smart Grid. *IEEE Internet of Things Journal*, 2019. DOI: 10.1109/JIOT.2019.2903312

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 2, 162-169 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Feature Selection for Musical Genre Classification Using a Genetic Algorithm

Abba Suganda Girsang[1], Andi Setiadi Manalu[1], Ko-Wei Huang[*,2],

[1]*Computer Science Department, BINUS Graduate Program-Master of Computer Science Bina Nusantara University Jakarta, Indonesia 11480*

[2]*Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan*

ARTICLEINFO

Article history:
Received: 11 January, 2019
Accepted: 13 March, 2019
Online: 23 March, 2019

Keywords:
Music genre classification
Feature extraction
Genetic algorithm
Machine learning
Low-level features

ABSTRACT

Music genre classification is an important multimedia research domain, including aspects of music piece representation, distances between genres, and categorization of music databases. The objective of this study was to develop a model for automatic classification of musical genres from audio data by using features from low-level time and frequency domains. These features can highlight the differences between different genres. In the model, feature selection is performed using a genetic algorithm (GA), and the resulting dataset is classified using the k-nearest neighbor (KNN), naive Bayes classifier (NBC), and support vector machine (SVM) learning methods. Tenfold cross-validation is used to obtain the optimal f-measure value. In this study, the data were obtained from the GTZAN genre collection datasets. In the performance evaluation, it was found that the GA-based feature selection strategy can improve the F-measure rate from 5% to 20% for the KNN, NBC, and SVM-based algorithms. In addition, the proposed SVM-GA algorithm can exactly better than other comparison algorithms.

1 Introduction

Music genre classification is used to categorize musical data into suitable categories based on shared characteristics. In the modern era, numerous musical genres are appreciated, such as blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. New technologies and globalization have been beneficial for the evolution of music, leading to the emergence of world music. These developments have increased research interest in music information retrieval (MIR), a field of automatic music management including music recommendation [1], music mood classification [2], and musical instrument classification [3]. As musical genres are most commonly used to manage music databases, a previous study investigated the extraction of audio features and techniques for classifying musical genres [4, 5, 6, 7]. However, genre classification has remained a challenge[8, 9], as it requires systems that are capable of performing automatic grouping [10] and querying retrieving data from large a music data set [8]. The identification of commonalities between genres requires feature extraction [11]. In the case of audio files, combinations of relevant features are used for modeling the musical genre [4]. These features may include timbral texture, rhythmic content, and pitch content [12]. Musical genres have been modeled using different features including short-term Mel-frequency cepstral coefficients (MFCCs) [13, 14], and Daubechies wavelet coefficient histograms (DWCHs), depending on the required application. Features are selected to improve performance and accuracy [15]. Achieving optimal feature selection and extracting appropriate features and dimensions of the features reduction are very challenging tasks [16, 17]. There are many different approaches that have been proposed to classify music genre, such as data mining [18, 19], deep learning strategy [20, 21, 22, 23, 24], and machine learning stretegies such as hidden Markov model, AdaBoost, and support vector machine [25, 26, 27, 28, 29, 30, 31, 32, 33, 34].

In addition to the above, musical genre classification problems can be solved with constant evolutionary algorithms [35]. Among them, a genetic algorithm (GA) based optimization has become a widely used approach for optimizing the selection of relevant features [36, 37, 38, 39, 40].

[*]Corresponding Author: Ko-Wei Huang, Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City, Taiwan Contact No & elone.huang@nkust.edu.tw

https://dx.doi.org/10.25046/aj040221

Among the several methods developed for musical genre classification, the performance of classification algorithms has become the benchmark [41]. In this paper, we propose a model for musical genre classification using low-level time and frequency-domain features for short and mid-term feature extraction. In this model, feature extraction is performed using a GA, and musical genre classification is conducted using a range of machine learning algorithms, including the K-Nearest Neighbor (KNN) [42], Nave Bayes Classifier (NBC) [43], and the Support Vector Machine (SVM) [25] were used for music genre classification. In the performance evaluation, it was found that the GA-based feature selection strategy can improve the F-measure rate from 5% to 20% for the KNN, NBC and SVM-based algorithms. In addition, the proposed support vector machine genetic algorithm (SVM-GA) algorithm can exactly better than other comparison algorithms

The remainder of this paper is organized as follows. Section 2 provides background information and discusses related studies. The proposed algorithm is presented in Section 3, and its performance is evaluated in Section 4. Finally, the conclusions and future research are presented in Section 5.

2 Literature review

The primary challenge in genre identification is selecting features for extraction from audio data [8]. Audio data comprise a series of samples that together represent an audio signal. Classification cannot simply be applied to the audio samples. In previous studies, audio analysis has been prominently performed by extracting numerical values for representative features. In music signals, these features correspond to primary dimensions such as pitch, rhythm, harmony, melody, timbre, and spatial location. Low-level features, such as temporal, energy, and spectral features, are some of the most prominently used features in sound signal analysis. In audio signal analysis, low-level features are generally extracted from the time and frequency domains [44]. Moreover, 11 sets of low-level time- and frequency-domain features are prominently used.

2.1 Time-domain audio features

In general, time-domain features are directly extracted from samples of audio signals. The features include short-term energy, short-term zero-crossing rate, and entropy. These features provide a simple approach to analyzing audio signals, and combining these features with sophisticated frequency-domain features is generally essential.

(1) Energy: The sum of squares of signal values, normalized by the respective frame length.

(2) Zero-Crossing Rate: The rate at which the sign of the signal changes within a particular frame.

(3) Entropy of Energy: The entropy of the normalized energies of subframes, which can be interpreted as a measure of abrupt changes in the signal.

2.2 Frequency-domain audio features

The discrete Fourier transform (DFT) is extensively used in audio signal analysis, because it provides a convenient representation of the frequency content distribution of a signal as a sound spectrum. Numerous audio features are based on the DFT of a signal. These are termed frequency-domain (or spectral) audio features.

(1) Spectral Centroid: The center of gravity of a spectrum.

(2) Spectral Spread: The second central moment of a spectrum.

(3) Spectral Entropy: The entropy of the normalized spectral energies for a set of sub-frames.

(4) Spectral Flux: The squared difference between normalized magnitudes of the spectra of two successive frames.

(5) Spectral Rolloff: The frequency below which 90% of the magnitude distribution of a spectrum is concentrated.

(6) MFCCs: MFCCs are used to capture short-term, spectral-based features. The logarithm of the amplitude spectrum based on a short-time Fourier transform is derived for each frame. Frequency bins are then grouped and smoothed using Mel-frequency scaling to ensure that they agree with predetermined concepts. MFCCs are generated by decorrelating Mel-spectral vectors by using a discrete cosine transform.

(7) Chroma Vector: A chroma vector is a 12-element representation of spectral energy, wherein bins represent 12 equal-tempered pitch classes of Western music (semitone spacing).

(8) Harmonic: Harmonic features represent beats, and the harmonic ratio and fundamental frequency are two harmonic features.

Although studies have been conducted on music genre classification, the most accurate results have been achieved using the GTZAN dataset. A comprehensive set of features was proposed for the direct modeling of music signals [4]. When these features were employed for genre classification by using KNNs and Gaussian Mixture models (GMMs), an accuracy rate of 61% was achieved. A novel feature extraction method was proposed for musical genre classification, wherein DWCHs are combined with machine learning classification algorithms, including SVMs and linear discriminant analysis [14]. An accuracy of 78.5% was obtained with this method. Feature integration methods and late information fusion were examined in a previous study [12], by using majority voting for classification of all short-time features. A novel feature integration technique using an autoregressive (AR) model was proposed. This approach was reported to exhibit superior performance to that obtained by the use of mean variance features. Another study [45] addressed musical genre classification from a multilinear perspective. Multiscale spectro-temporal modulation features were extracted on the basis of an auditory cortical processing model, which provided an accuracy of 78.2%. Nonnegative matrix factorization was used to obtain a novel description of the

timbre of musical sounds[13]. A spectrogram was factorized, which provided a characteristic spectral analysis. Gaussian Mixture Models were then applied, to achieve a reported accuracy of 74%. An ensemble approach [37], involving the use of multiple feature vectors and time and space decomposition strategies was also reported in a previous study. In the study, time decomposition was performed using feature vectors, which were extracted from music segments obtained from the beginning, middle, and end of the original signal. The study also employed four machine learning algorithms, namely NBC, decision trees, SVMs, and multilayer perceptron (MLP) neural nets, which provided a maximum accuracy of 65.06%. In an alternative approach [46], a method based on the classification accuracy of an SVM classifier was proposed for selecting training instances. These instances comprised feature vectors representing short-term, low-level characteristics in an audio signal. This method was reported to provide an accuracy of 59.6%. An invariance of MFCCs to musical key and tempo was explored in a previous study[47]. The study indicated that MFCCs encode both timbral and key information. An accuracy of 69.3% was achieved by applying GMMs. The acoustic signal feature can be derived using a mathematical model that represents the acoustic signal [39]. In this approach, features were selected using a GA. Classification was then performed by generating an adjusted KNN classifier. The reported accuracy was 67.6%. In a previous study, a subspace cluster analysis process was used to automate the construction of a classification tree[48]. Experimental results validated the tree-building algorithm, and the study provided a new research direction for automatic genre classification. Several machine learning algorithms were employed, such as KNNs, J48, logistic regression (LOG), LibSVM with a radial basis function kernel, MLP neural networks, and sequential minimal optimization (SMO) SVM with a polynomial kernel. This method was reported to achieve an accuracy of 72.9%. Another study[49] proposed automatic musical genre classification by using spectral, time-domain, tonal, rhythmic, sound effect, and high-level descriptors. An analysis was conducted using KNNs, an SVM, and a GMM, with a resulting accuracy of 79.7%.

3 Methods

3.1 Feature extraction

Feature extraction is a crucial step in audio analysis, because it is also required in other pattern recognition and machine learning tasks. Extraction is performed in the following two steps: short- and mid-term processes. In short-term feature extraction, the input signal is divided into brief windows (or frames), and the number of features is computed for each frame, generating a sequence of short-term feature vectors from the entire signal. In mid-term extraction, statistical analysis is generally applied to these short-term feature sequences to obtain the signal. Several statistics, including mean, median, standard deviation, standard deviation by mean, max, and min, are derived for each short-term feature sequence. The data are then normalized to avoid the introduction of anomalies and minimize redundancy

3.2 Machine learning methods (naive Bayes classifier (NBC), K-nearest neighbor (KNN), support vector machine (SVM)) combined with genetic algorithm (GA)

Fig. 1 shows the system design. GA optimization is combined with the KNN, NBC, and SVM learning methods. The fitness function uses the largest f-measure value derived using machine learning, and is evaluated using 10-fold cross-validation. The system then derives the maximum fitness.

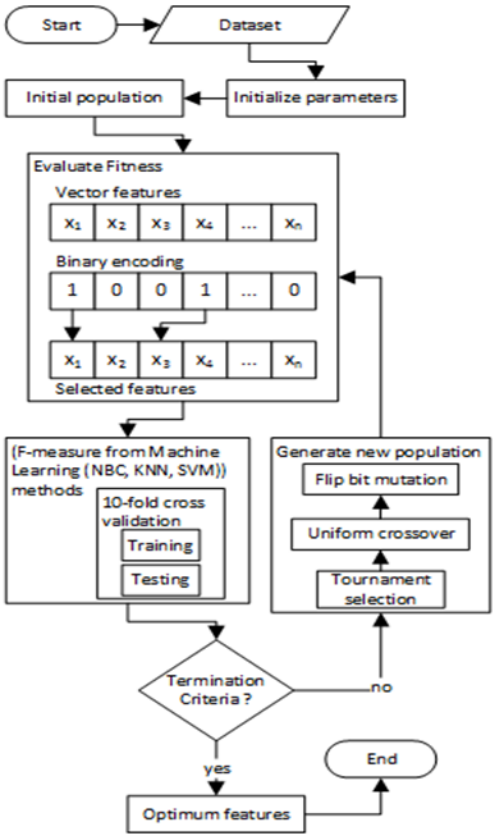


Figure 1: Machine learning methods (NBC, KNN, SVM) combined with GA

(1) The dataset is a feature vector containing normalized data. All necessary parameters are initialized. The initial population is generated and encoded in binary $\{0,1\}$. Individual solutions are randomly generated to form the initial population $P = \{S_1, S_2, S_3, \cdots, S_n\}$; $S_1 = S_2 = S_3 = \cdots = S_n = \{G_1, G_2, G_3, \cdots, G_n\}$. Here S is a solution and G_n represents features, where $G_n \in \{0,1\}$; feature labeled 0 are not used, whereas those labeled 1 are used.

(2) The fitness of each solution (S) is evaluated using the highest f-measure value for musical genre classification conducted through the KNN, NBC, and SVM learning methods, following the 10-fold cross-validation method. The highest fitness value indicates the optimal classification result.

(3) After all fitness values are obtained, tournament selection is used to select parents. This prominently used method probabilistically selects genomes for recombination; the selection is based on fitness. The algorithm

randomly selects genomes from the population to compete in tournaments of user-selected size T. The genome with the highest fitness value in each tournament is selected for recombination.

(4) A uniform crossover is primarily used for problems, wherein elements in the genome are independent of each other. In this case, the locations of genes relative to each other in the genome do not influence fitness. The probability of crossover per gene is set by the user. The uniform crossover process generates a solution in the (t+1)th generation by randomly selecting genes from each of the winners, corresponding to the relevant locations.

(5) A matrix of 1s and 0s is generated, and mutation is applied individually to each element. The probability of mutation per gene is set by the user. The selected values are interchanged such that 1s become 0s (and vice versa).

(6) Crossover and mutation then generate a new population (solution), whose fitness value is reevaluated.

(7) The process is stopped when the required criteria are satisfied; otherwise, the fitness is reevaluated.

(8) The GA then identifies the optimal solution, which in this case is a set of optimal features.

3.3 Evaluation

For the proposed method, selected audio files were used as validation data in this study. Each file was tested through cross-validation for ensuring its allocation to the accurate class. The result was presented in a confusion matrix summarizing accurately and inaccurately classified cases. The f-measure value was calculated from the matrix to determine the optimal classification. The harmonic mean of precision and recall was used as the f-measure value, with the highest f-measure value indicating the most accurate classification.

4 Experimental results

4.1 Environment setting

The data used in this study comprised instrumental music files from the GTZAN genre collection datasets[4]. The data were obtained from the following source:[1]. The dataset comprised 1000 30-s music files in an audio format at a sampling rate of 22050 Hz in 16-bit mono. Samples were classified into the following 10 musical genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. The raw data were obtained from audio data files. The data were preprocessed by converting the data files into a numerical form before the execution of feature extraction. A total of 11 sets of low-level time- and frequency-domain features were extracted.

All experiments were performed over 15 independent runs and conducted using MATLAB 2016b with the parallel toolbox. The experiments were conducted on a standard laptop with an Intel i5-3320M 2.60-GHz CPU and 4-GB RAM.

4.2 Without feature selection

4.2.1 Experimental results obtained using KNN

As presented in Table 1, the KNN method performed data execution in less than 1 second in all experiments, with the F-measure mean (Fmean) time being 0.87s and a standard deviation (Fsd) time of 0.03s, yielding an Fmean of 66.2%. The difference between the best F-measure (Fbest) and the worst F-measure (Fworst) was not significant, at 1.4%. Moreover, the Fsd value was negligible, at 0.5%. For classification, the f-measure range was 65.5% to 66.9%. In the results shown, the performance of KNN is less than 70%, but very quickly classifies the music data, owing to the fact that KNN does not need the training process..

[1] http://marsyasweb.appspot.com/download/data_sets

Table 1: Results obtained by KNN without feature selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
66.9%	0.67	65.5%	0.83	66.2%	0.83	0.5%	0.03

Table 2: Results Obtained by NBC without Feature Selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
59.1%	9.07	56.8%	8.90	58.0%	8.93	0.6%	0.08

Table 3: Results Obtained by SVM without Feature Selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
77.0%	14.37	75.3%	12.36	76.2%	13.16	0.005%	1.69

4.2.2 Experimental results obtained using NBC

Table 2 shows the NBC results. The Fmean time for data execution was 8.93s, with an Fsd time of 0.08s for each experiment and an Fmean of 58%. The difference between Fbest and Fworst was not significant, at 2.3%. Fsd was also very small, at 0.6%. For classification, the f-measure range was approximately 56.8% to 59.1%. In the results show, the performance of NBA is less than 60% and even slower than the KNN.

4.2.3 Experimental results obtained using SVM

Table 3 shows the SVM results. For all experiments, the Fmean and Fsd times for data execution were 13.16 and 1.69 s, respectively, with an Fmean of 0.762 (76.2%). The difference between Fbest and Fworst was 0.017 (1.7%), which was nonsignificant. Moreover, the Fsd value was 0.005 (0.5%), which was negligible. The f-measure range for classification was 0.7530.770 (75.3%77.0%).

4.3 With GA feature selection

4.3.1 Parameter Setting

The previous result shows that KNN is the best algorithm for F-measure and computation times. Thus, in this section we will try to import the GA-based feature selection strategy into these three algorithms. The basic parameter settings of all experiments are listed in Table 4. All experimental results are collected from 15 independent runs, each of 100 iterations.

Table 4: GA parameter settings

Parameter	Setting
iteration	100
population	15
crossover rate	90%
mutation rate	1%
selection operator	tournament
crossover operator	uniform
tournament size	2
Number of features	204

4.3.2 Experimental results obtained using KNNGA

From Table 5, it can be seen that KNN-GA required less than 5s for data execution in each experiment. The Fmean time of 4.74 s and Fsd time of 0.43 s gave an Fmean of 71.3%. The difference between Fbest and Fworst was not significant, at 1.2%. Fsd was also very small, at 0.3%. For classification, the approximate f-measure range was 70.8% to 72%. . In summary, we found that the GA-based feature selection will improve the 5% F-measure rates with the KNN approach.

4.3.3 Experimental results obtained using NBCGA

From Table 6 it can be seen that, for data execution, the NBC-GA required a fairly long Fmean time of 61.46 s. With the Fsd time of 8.50 s for each experiment, this gave an Fmean of 66.8%. The difference between Fbest and Fworst was not significant, at 0.9%. The Fsd was also insignificant, at 0.3%. The f-measure range for classification was approximately 66.4% to 67.3%. In summary, we found that the GA-based feature selection will improve the 10% F-measure rates with the NBC approach.

4.3.4 Experimental results obtained using SVMGA

As can be seen from Table 7, SVM-GA required a fairly long Fmean time for data execution, of 125.614s. With an Fsd time of 44.29s from each experiment, the Fmean was 79.5%. The difference between Fbest and Fworst was not significant, at 1.1%. The Fsd was also very small, at 0.3%. The f-measure range for classification was approximately 79.0% to 80.1%. In summary, we found that the GA-based feature selection will improve about the 3-5% F-measure rates with the SVM approaches.

4.3.5 Convergence Rates of KNN-GA, NBC-GA, and SVM-GA

The graphs in Figs 2, 3, and 4 reach steady state because they include a stop criterion at the 100th iteration when the process ended.

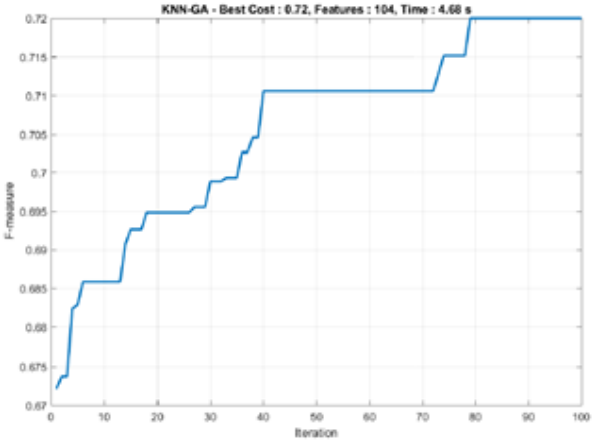


Figure 2: Convergence rates of KNN with Genetic Algorithm

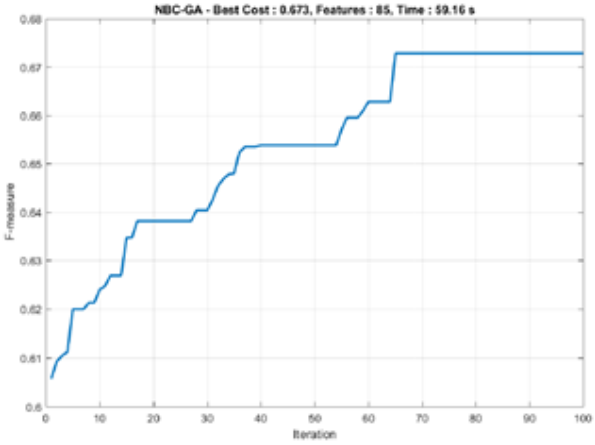


Figure 3: Convergence rates of NBC with Genetic Algorithm

Table 5: Results Obtained by KNN with Feature Selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
72.0%	4.68	70.8%	4.66	71.3%	4.74	0.003%	0.43

Table 6: Results Obtained by NBC with Feature Selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
67.3%	59.16	66.4%	57.52	66.8%	61.46	0.3%	8.50

Table 7: Results Obtained by SVM with Feature Selection

Fbest		Fworst		Fmean		Fsd	
Value	time(s)	Value	time(s)	Value	time(s)	Value	time(s)
80.1%	121.19	79.0%	123.46	79.5%	141.05	0.3%	44.29

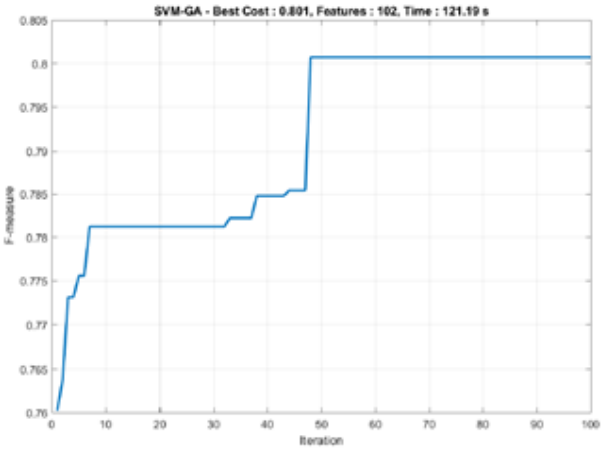


Figure 4: Convergence rates of SVM with Genetic Algorithm

4.4 Comparison of F-measure values

The performance is assessed based on Eq. 1.

$$\Delta F_M = \left(\frac{F_{new} - F_{orig}}{F_{orig}} \right) \times 100\% \quad (1)$$

Here, ΔF_M denotes the improvement of the F-measure values, Δt_{avg} denotes the improvement of computation time, F_{new} denotes the F-measure values of the KNN, NBA, and SVM algorithm performing the genetic algorithm feature selection, and F_{orig} denotes the KNN, NBA, and SVM algorithms without performing any feature selection strategy, and t_{avg} is the computation time in seconds.

Table 8 shows that the average increase in the Fmean value was approximately 7.24% for KNN-GA, approximately 13.27% for NBC-GA, and 4.23% for SVM-GA. However, in order to perform the feature selection strategy, the computation time was increased to between 500% and 871%. In summary, the feature selection increases the quality of the solution, but with a massive increase in computation time, especially with SVM-GA, which requires spending more than 10 times the computation time..

4.5 Comparison of results in this study with those in previous studies

Previous results showed the ability of a GA to improve the quality of result classification by reducing features. The existing algorithm performs poor quality classification (approximate Fmean 59%66%). Moreover, by feature selection using the GA, the quality classification can be improved until Fmean is 67%80%. In this section, we compare the proposed algorithm with other algorithms to verify the performance or proposed algorithms. Table 9 compares the accuracy of different methods with our proposed algorithms, KNN-GA, NBA-GA, and SVM-GA, using the highest measure achieved by each. All studies used the same GTZAN Genre Collection Data Sets [4]. The KNN-GA method from the current study produced an Fbest value of 72%, which is better than the comparison algorithms of [4, 39, 47, 48]. The NBC-GA method produced an Fbest value of 67.3%, which exceeded that reported in [4]. The SVM-GA method produced an Fbest value of 80.1%, which exceeded previous studies. In summary, the GA combined with the SVM method can obtain a better solution than other algorithms.

5 Conclusions and Future work

In this study, automatic musical genre classification was conducted by combining GA optimization with three machine learning methods. This approach entails using low-level time- and frequency-domain features. The features are extracted using short- and mid-term processes. The optimal classification accuracy levels achieved in the study were determined to be comparable to those achieved by other state-of-the-art musical genre classification algorithms. However, the feature selection strategy will consume a massive amount of computation time. As part of our future research, we aim to investigate three aspects: 1) applying other meta-heuristic methods such as particle swarm optimization (PSO) to select low-level features; 2) combining deep-learning techniques such as CNN with a meta-heuristic algorithm to improve the accuracy of classification results; and 3) decreasing the computation time by using feature reduction mechanisms to make the algorithm perform more effectively.

Table 8: Comparison of f-measure values

Measure	KNN-GA		NBA-GA		SVM-GA	
	ΔF_M	$\Delta\, t_{avg}$	ΔF_M	$\Delta\, t_{avg}$	ΔF_M	$\Delta\, t_{avg}$
Fbest	7.08	598	12.18	552	3.87	743
Fworst	7.49	461	14.46	546	4.68	898
Fmean	7.15	444	13.17	588	4.15	971
Average	7.24	501	13.27	562	4.23	871

Table 9: Comparison with previous studies

Algorithm	Feature Selection	Accuracy
Tzanetakis et al. [4]	none	61.0%
Li et al. [14]	none	78.5%
Holzapfel et al. [13]	none	74.0%
Panagakis et al. [45]	none	78.2%
Li et al. [47]	none	69.3%
Karkavitsas et al. [39]	Hybrid GA	67.6%
Ariyaratne et al. [48]	none	71.3%
Martins de Sousa et al. [49]	none	79.7%
KNN-GA	GA	**72.0%**
NBA-GA	GA	**67.3%**
SVM-GA	GA	**80.1%**

6 Acknowledgement

This work was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C., under grants MOST 107-2218-E-992-303-.

References

[1] Markus Schedl. The lfm-1b dataset for music retrieval and recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pages 103–110. ACM, 2016.

[2] Xiao Hu, Kahyun Choi, and J Stephen Downie. A framework for evaluating multimodal music mood classification. *Journal of the Association for Information Science and Technology*, 68(2):273–285, 2017.

[3] DG Bhalke, CB Rama Rao, and Dattatraya S Bormane. Automatic musical instrument classification using fractional fourier transform based-mfcc features and counter propagation neural network. *Journal of Intelligent Information Systems*, 46(3):425–446, 2016.

[4] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[5] R Hillewaere. Computational models for folk music classification. *Vrije Universiteit Brussel*, 2013.

[6] Débora C Corrêa and Francisco Ap Rodrigues. A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, 60:190–210, 2016.

[7] Izaro Goienetxea, José María Martínez-Otzeta, Basilio Sierra, and Iñigo Mendialdua. Towards the use of similarity distances to music genre classification: A comparative study. *PloS one*, 13(2):e0191417, 2018.

[8] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, 2011.

[9] Marius Kaminskas and Francesco Ricci. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(2):89–119, 2012.

[10] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141, 2006.

[11] Dalibor Mitrovi, Matthias Zeppelzauer, and Christian Breiteneder. Features for content-based audio retrieval. volume 78, pages 71–150. Elsevier, 2010.

[12] A. Meng, P. Ahrendt, and J. Larsen. Improving music genre classification by short time feature integration. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 5, pages v/497–v/500 Vol. 5, 2005.

[13] A. Holzapfel and Y. Stylianou. Musical genre classification using nonnegative matrix factorization-based features. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):424–434, 2008.

[14] Tao Li, Mitsunori Ogihara, and Qi Li. A comparative study on content-based music genre classification. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 282–289, 2003.

[15] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16 – 28, 2014.

[16] I. Vatolkin, W. Theimer, and G. Rudolph. Design and comparison of different evolution strategies for feature selection and consolidation in music classification. In *2009 IEEE Congress on Evolutionary Computation*, pages 174–181, 2009.

[17] Babu Kaji Baniya, Joonwhoan Lee, and Ze-Nian Li. Audio feature reduction and analysis for automatic music genre classification. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 457–462. IEEE, 2014.

[18] Mangesh M Panchwagh and Vijay D Katkar. Music genre classification using data mining algorithm. In *2016 Conference on Advances in Signal Processing (CASP)*, pages 49–53. IEEE, 2016.

[19] Nimesh Ramesh Prabhu, James Andro-Vasko, Doina Bein, and Wolfgang Bein. Music genre classification using data mining and machine learning. In *Information Technology-New Generations*, pages 397–403. Springer, 2018.

[20] Arjun Raj Rajanna, Kamelia Aryafar, Ali Shokoufandeh, and Raymond Ptucha. Deep neural networks: A case study for music genre classification. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 655–660. IEEE, 2015.

[21] Il-Young Jeong and Kyogu Lee. Learning temporal features using a deep neural network and its application to music genre classification. In *Ismir*, pages 434–440, 2016.

[22] Sergio Oramas, Oriol Nieto, Francesco Barbieri, and Xavier Serra. Multi-label music genre classification from audio, text, and images using deep features. *arXiv preprint arXiv:1707.04916*, 2017.

[23] Yandre MG Costa, Luiz S Oliveira, and Carlos N Silla Jr. An evaluation of convolutional neural networks for music classification using spectrograms. *Applied soft computing*, 52:28–38, 2017.

[24] L Rafael Aguiar, MG Yandre Costa, and N Carlos Silla. Exploring data augmentation to improve music genre classification with convnets. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[25] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

[26] Wei Chai and Barry Vercoe. Folk music classification using hidden markov models. In *Proceedings of international conference on artificial intelligence*, volume 6. sn, 2001.

[27] Guodong Guo and Stan Z Li. Content-based audio classification and retrieval by support vector machines. *IEEE transactions on Neural Networks*, 14(1):209–215, 2003.

[28] Changsheng Xu, Namunu Chinthaka Maddage, and Xi Shao. Automatic music classification and summarization. *IEEE transactions on speech and audio processing*, 13(3):441–450, 2005.

[29] Alexios Kotsifakos, Evangelos E Kotsifakos, Panagiotis Papapetrou, and Vassilis Athitsos. Genre classification of symbolic music with smbgt. In *Proceedings of the 6th international conference on PErvasive technologies related to assistive environments*, page 44. ACM, 2013.

[30] Kamelia Aryafar, Sina Jafarpour, and Ali Shokoufandeh. Automatic musical genre classification using sparsity-eager support vector machines. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 1526–1529. IEEE, 2012.

[31] Loris Nanni, Yandre MG Costa, Alessandra Lumini, Moo Young Kim, and Seung Ryul Baek. Combining visual and acoustic features for music genre classification. *Expert Systems with Applications*, 45:108–117, 2016.

[32] AB Mutiara, R Refianti, and NRA Mukarromah. Musical genre classification using support vector machines and audio features. *TELKOMNIKA TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, 14:1024–1034, 2016.

[33] Hareesh Bahuleyan. Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*, 2018.

[34] Aldona Rosner and Bozena Kostek. Automatic music genre classification based on musical instrument track separation. *Journal of Intelligent Information Systems*, 50(2):363–384, 2018.

[35] B. Xue, M. Zhang, W. N. Browne, and X. Yao. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626, 2016.

[36] Il-Seok Oh, Jin-Seon Lee, and Byung-Ro Moon. Hybrid genetic algorithms for feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1424–1437, 2004.

[37] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner. Feature selection in automatic music genre classification. In *2008 Tenth IEEE International Symposium on Multimedia*, pages 39–44, 2008.

[38] Dragan Matic. A genetic algorithm for composing music. volume 20, pages 157–177, 2010.

[39] George V. Karkavitsas and George A. Tsihrintzis. Automatic music genre classification using hybrid genetic algorithms. In *Intelligent Interactive Multimedia Systems and Services*, pages 323–335, 2011.

[40] Chih-Fong Tsai, William Eberle, and Chi-Yuan Chu. Genetic algorithms in feature and instance selection. *Knowledge-Based Systems*, 39:240 – 247, 2013.

[41] Bob L. Sturm. Classification accuracy is not enough. *Journal of Intelligent Information Systems*, 41(3):371–406, 2013.

[42] Muhammad Atif Tahir, Ahmed Bouridane, and Fatih Kurugollu. Simultaneous feature selection and feature weighting using hybrid tabu search/k-nearest neighbor classifier. *Pattern Recognition Letters*, 28(4):438–446, 2007.

[43] Jingnian Chen, Houkuan Huang, Shengfeng Tian, and Youli Qu. Feature selection for text classification with naïve bayes. *Expert Systems with Applications*, 36(3):5432–5435, 2009.

[44] Theodoros Giannakopoulos and Aggelos Pikrakis. *Introduction to Audio Analysis: A MATLAB Approach*. Academic Press, 1st edition, 2014.

[45] I. Panagakis, E. Benetos, and C. Kotropoulos. Music genre classification: a multilinear approach. In JP Bello, E Chew, and D Turnbull, editors, *International Symposium Music Information Retrieval*, pages 583 – 588, 2008.

[46] M. Lopes, F. Gouyon, A. L. Koerich, and L. E. S. Oliveira. Selection of training instances for music genre classification. In *2010 20th International Conference on Pattern Recognition*, pages 4569–4572, 2010.

[47] Tom L. H. Li and Antoni B. Chan. Genre classification and the invariance of mfcc features to key and tempo. In *Advances in Multimedia Modeling*, pages 317–327, 2011.

[48] H. B. Ariyaratne and D. Zhang. A novel automatic hierachical approach to music genre classification. In *2012 IEEE International Conference on Multimedia and Expo Workshops*, pages 564–569, 2012.

[49] J. Martins de Sousa, E. Torres Pereira, and L. Ribeiro Veloso. A robust music genre classification approach for global and regional music datasets evaluation. In *2016 IEEE International Conference on Digital Signal Processing (DSP)*, pages 109–113, 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 164-170 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

An Adaptive Fuzzy-Sliding Mode Controller for The Bridge Crane

Trinh Luong Mien*

Department of cybernetics, Faculty of electrical and electronic engineering, University of transport and communications,

No.03, Caugiay, Langthuong, Dongda, Hanoi, Vietnam.

ARTICLE INFO

Article history:
Received:27 April, 2019
Accepted:27 May, 2019
Online: 10 June, 2019

Keywords:
Crane
Anti-swing
Fuzzy logic
Adaptive control
Sliding mode control

ABSTRACT

This article presents a bridge crane nonlinear dynamic model in 2-dimensional space, and then given a novel adaptive fuzzy-sliding mode controller based on combining sliding mode control with fuzzy logic and Lyapunov function. Firstly, the article proposes an intermediate variable to link signal between two slide surfaces, related to trolley movement and payload swing. Then the fuzzy controller, compensative controller and parameter adaptive update law for the bridge crane are defined. The asymptotic stability of the proposed bridge crane control system is proven based on Lyapunov stability theory. Simulation results show that the adaptive fuzzy-sliding mode controller ensures the trolley follows the input reference with the short settling time, eliminating steady error, and anti-payload swing, anti-disturbance.

1. Introduction

The bridge crane is a lifting machinery which commonly used in industry and transportation to move goods from one location to another. The bridge crane consists of three main components: hoist, trolley and beam, corresponding to three movements: lifting and lowering the cargo of the hoist (mounted on the trolley), horizontal movement of the trolley (on the beam), and vertical movement of the beam (on the factory frame). Goods (refer to as payload) are linked to cargo hoist by the cable and hanger. This connection causes the vibration of the payload when the crane moves. This is undesirable - causing unsafe operation of the crane, reducing the accuracy in controlling the cargo transport of crane, reducing the crane operation.

In some cases of actual crane operation, it is often focused on controlling the exact position of the trolley, after lifting the payload. In this case, the bridge crane is studied as a trolley movement in the 2D space, created by the horizontal movement of the trolley and the verticality of the cable. Common works on position control of the trolley are often based on state feedback control, PID control [1, 2], LQR control [3], PID combined with fuzzy [4-7], PID adjusting parameters [8] and obtain certain results. In [1], the state feedback controller is combined with the integral part to eliminate the steady error. The work [4] presents a controller PID combined with the fuzzy-sliding mode control. These controllers are designed based on the linear model of the crane and the existing of payload oscillation.

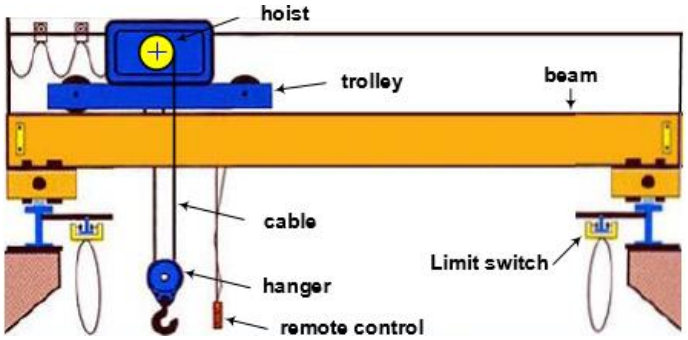


Figure 1. Bridge crane movement in 3D: trolley-0x, beam-0y, hoist-0z.

In recent times, the researches have taken into account the nonlinearity of the crane based on back stepping [9], nonlinear control [10, 11], sliding mode control [12], sliding mode control combined with fuzzy logic [13], partial linear feedback control [14], sliding mode control combined with partial linear feedback [15], sliding mode control combined with adaptive fuzzy control [16], or neural network [17]. These controllers have enriched the crane control strategy, ensuring that trolley follows the reference trajectory with small static error and small payload oscillation. However, these control algorithms are quite complicated, require a large amount of computation, the deployment of them on hardware devices is difficult and not considering the impact of

* Trinh Luong Mien. Tel: +84 904 684 595. Email: mientl@utc.edu.vn

noise in the working environment of the crane, such as the impact of the wind.

This article presents a nonlinear mathematical model of a crane in two-dimensional space (2D bridge crane model). On that basis, an adaptive fuzzy-sliding mode controller is designed for 2D crane model based on the sliding mode control principle combined with fuzzy logic and Lyapunov function, ensuring that the trolley moves quickly, following set reference and eliminating payload oscillation, anti-disturbance. The rest of this article is organized as follows. Section 2 introduces the bridge crane dynamic model. Section 3 deals with the design of controllers for the bridge crane 2D. Section 4 presents the simulation results. Finally, section 5 presents conclusions.

2. Bride crane dynamic model

The motion of the bridge crane in the two-dimensional space *Oxz* as shown in Figure 2, where: *x(t)* is the position of the trolley moving in the *Ox* direction, *α(t)* is the oscillation angle of the payload. Assuming that the cable is a rigid rod and connected with a cargo hanger, the cable length is constant and the cable mass is negligible; ignore the trolley friction when moving; consider payload as the point $P(x_p,z_p)$ with mass m_p; trolley with mass m_t; F_x is the force driving a trolley in the *Ox*-direction, g is the gravitational acceleration.

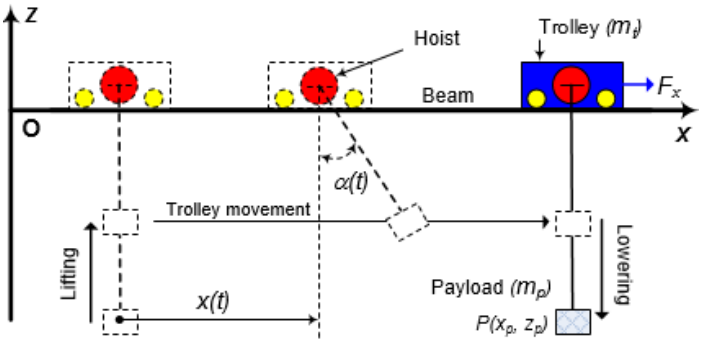


Figure 2. Bridge crane trolley movement in 2D

The dynamic model of bridge crane is defined based on energy conservation law. The total kinetic energy K and the potential energy T of the system are determined as follows:

$$K = K_{trolley} + K_{payload} = \frac{1}{2} m_t \dot{x}^2 + \frac{1}{2} m_p (\dot{x}_p^2 + \dot{z}_p^2)$$
$$; \; x_p = x + l\sin\alpha \; , \; z_p = -l\cos\alpha \qquad (1)$$
$$T = T_{trolley} + T_{payload} = -m_p g l \cos\alpha$$

Lagrange function is determined by the following formula:

$$L = K - T \qquad (2)$$

The mathematical equation describing crane dynamics is determined from the Euler-Lagrange equation, with $q = [x, \alpha]^T$, $\tau = [F_x, 0]^T$, as shown below

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}(q, \dot{q})\right) - \frac{\partial L}{\partial q}(q, \dot{q}) = \tau \qquad (3)$$

After transforming, we obtain the mathematical equation describing the dynamics of crane:

$$\begin{cases} (m_t + m_p)\ddot{x} + m_p l(\ddot{\alpha}\cos\alpha - \dot{\alpha}^2 \sin\alpha) = F_x \\ \ddot{x}\cos\alpha + l\ddot{\alpha} + g\sin\alpha = 0 \end{cases} \qquad (4)$$

The system of equations (4) is separated into 2 subsystems, describing 2D bridge crane dynamics in the state space, including: system A describes the trolley movement; system B describes the payload oscillation.

$$A: \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f_1(x) + g_1(x)u \end{cases} \qquad (5)$$

$$B: \begin{cases} \dot{x}_3 = x_4 \\ \dot{x}_4 = f_2(x) + g_2(x)u \end{cases} \qquad (6)$$

where:

$$x_1 = x(t); x_2 = \dot{x}(t); x_3 = \alpha(t); x_4 = \dot{\alpha}(t); u = F_x$$

$$f_1(x) = \frac{m_p l x_4^2 \sin x_3 + m_p g \sin x_3 \cos x_3}{m_t + m_p \sin^2 x_3}$$

$$g_1(x) = \frac{1}{m_t + m_p \sin^2 x_3}$$

$$f_2(x) = -\frac{(m_t + m_p) g \sin x_3 + m_p l x_4^2 \sin x_3 \cos x_3}{(m_t + m_p \sin^2 x_3) l}$$

$$g_2(x) = -\frac{1}{(m_t + m_p \sin^2 x_3) l}$$

The 2D bridge crane model is a nonlinear system. When the trolley moves, the payload oscillation will appear, so that as the proposed controller not only ensures the trolley follows the reference trajectory but also quenching payload oscillation, in order to ensure safe operation and accurate cargo transportation.

3. Adaptive fuzzy-sliding mode controller design for the bridge crane

3.1. Sliding mode controller

In this research, mathematical model of 2D bridge crane is described by equations (5) and (6), corresponding to system A, system B.

The goal here is to synthesize a controller $u = u(\underline{x})$ so that the outputs $\underline{x}(t) = [x, \alpha]^T$ of the crane (5) and (6) follows on the input reference $\underline{x}_d(t) = [x_d, \alpha_d]^T$ i.e. the steady error:

$$\underline{e}(t) = [e_1(t), e_2(t)]^T \to \underline{0}$$

with $e_1 = x_1 - x_{1d}$;

$e_2 = x_3 - x_{3d};\ x_{1d} = x_d; x_{3d} = \alpha_d, \alpha_d = 0.$

According to the sliding control principle, we determine the general sliding surface for systems A and B:

$s_{1b} = \dot{e}_1 + \lambda_1 e_1 = \dot{x}_1 - \dot{x}_{1d} + \lambda_1 (x_1 - x_{1d}) = x_2 + \lambda_1 x_1 - (\dot{x}_{1d} - \lambda_1 x_{1d}$ (7)

$s_{2b} = \dot{e}_2 + \lambda_2 e_2 = \dot{x}_3 - \dot{x}_{3d} + \lambda_2 (x_3 - x_{3d}) = x_4 + \lambda_2 x_3 - (\dot{x}_{3d} - \lambda_2 x_{3d}$ (8)

where: λ_1, λ_2 are positive constants.

To ensure the control target for bridge crane (5) and (6), we define the sliding surface s_1 for system A to ensure precise control of trolley position and sliding surface s_2 for system B to ensure the general control target for bridge crane, that is: controlling the trajectory of the trolley, quenching the payload oscillation of the payload, the anti-disturbance. Therefore, we need to use the intermediate variable z to link the signal between the sliding surface s_1 and the sliding surface s_2. On that basis, the author defines the sliding surface s_1, s_2 as follows:

$s_1 = s_{1b} = \dot{e}_1 + \lambda_1 e_1 = \dot{x}_1 - \dot{x}_{1d} + \lambda_1 (x_1 - x_{1d}) = x_2 + \lambda_1 x_1 - (\dot{x}_{1d} - \lambda_1 x_{1d}$ (9)

$s_2 = \dot{e}_2 + \lambda_2 (e_2 - z) = \dot{x}_3 - \dot{x}_{3d} + \lambda_2 (x_3 - x_{3d} - z) = x_4 + \lambda_2 (x_3 - z) - (\dot{x}_{3d} - \lambda_2 x_{3d}$ (10)

Intermediate variable z are defined as follows:

$$z = \bar{z}.sat(s_1/\Phi_z) \quad (11)$$

where: $|z| \le \bar{z};\ 0 < \bar{z} < 1,\ \bar{z}$ is the upper limit of |z|, Φ_z is the boundary limit s_1.

$$sat(s_1/\Phi_z) = \begin{cases} sign(s_1/\Phi_z) & if \quad |s_1/\Phi_z| \ge 1 \\ s_1/\Phi_z & if \quad |s_1/\Phi_z| < 1 \end{cases}$$

Because of $\bar{z} < 1$ it can be effective in reducing. When s_1 reduced z is also decreases. When $s_1 \to 0,\ z \to 0,\ e_2 \to 0$ then $s_2 \to 0$ and achieves the control target for crane (5) and (6). Therefore, by adding intermediate variables z now s_1 and s_2 at the same time tends to 0.

According to the sliding control principle, when $\dot{s}_2 = 0$ we get the sliding control law for bridge crane, it looks like this:

$$u_{eq} = \frac{-\lambda_2 x_4 + \lambda_2 \dot{z} - f_2 + \ddot{x}_{3d} + \lambda_2 \dot{x}_{3d}}{g_2} \quad (12)$$

Control law (12) for bridge crane is a complex nonlinear function, for which $\dot{z}$ is not defined. Therefore, in this article, the author proposes adaptive fuzzy-sliding mode control law to imitate the sliding control law (12) based on fuzzy logic control and Lyapunov function.

3.2. Fuzzy controller based on sliding mode control principle

Define fuzzy controller with input s_2 and output u_f with fuzzy control rules *i-th as below*:

Rule i: IF s_2 is F_i THEN u_f is θ_i, i=1,2..n (13)

where: F_i is the input fuzzy set with the membership function $\mu_{Fi}(.)$ and θ_i are the changing singleton output values.

Defuzzification with centroid method, we obtain:

$$u_f(s_2, \theta) = \frac{\sum_{i=0}^{n} \xi_i.\theta_i}{\sum_{i=0}^{n} \xi_i} = \theta^T \xi \quad (14)$$

where: $\xi_i = \mu_{Fi}(s_2)$ is the reliability of the *i-th* fuzzy control rule.

$$\theta = [\theta_1, \theta_2, ..., \theta_n]^T;$$

$$\xi = [\xi_1 / \sum_{i=1}^{n} \xi_i, \xi_2 / \sum_{i=1}^{n} \xi_i, ..., \xi_n / \sum_{i=1}^{n} \xi_i]^T \quad (15)$$

According to the fuzzy approximation theory, there is an optimal fuzzy controller u_f^* of the form:

$$u_f^* = (\theta^*)^T \xi \quad (16)$$

where: θ^* is the optimal parameter set as follows:

$$\theta^* = \arg\min_{|\theta| \le M_\theta} \left\{ \sup_{|s_2| \le M_{s2}} |u_f - u_{eq}| \right\},\ \forall t \quad (17)$$

where: M_θ, M_{s2} are the appropriate values when designing.

The control signal (16) is closest to (12), in other words, the deviation *d(t)* is the smallest:

$$d(t) = u_f^* - u_{eq} \text{ với } 0 \le d(t) \le D \quad (18)$$

Here the limit *D* is a positive number, but it is uncertain in practice. Called $\hat{D}$ the estimated value of the limit D. Then the estimate of the limit *D* is $\tilde{D}$ is defined as follows: $\tilde{D} = D - \hat{D}$

Therefore, based on (12), the author proposed a new control law for bridge cranes (5) and (6) with the following form:

$$u = u(s_2, \theta, \hat{D}) = u_f(s_2, \theta) + u_c(s_2, \hat{D}) \quad (19)$$

Here $u_f(s_2, \theta)$ the fuzzy controller (14) has the closest value to (12); and $u_c(s_2, \hat{D})$ is the compensated controller to compensate for the deviation (if any) between control laws (14) and (12). The compensation controller $u_c(s_2, \hat{D})$ is determined based on the Lyapunov stability theory, ensuring the stable asymptotic bridge crane control system.

3.3. Adaptive laws based on Lyapunov function

Next, the author summarizes the compensation controller $u_c(s_2,\hat{D})$ and determines the adaptive law to update the parameters θ, $\hat{D}$ so that (19) optimal operation is closest to (12), which means $\theta(t)\to\theta^*$ and $\hat{D}(t)\to D$. Set $\tilde{\theta}=\theta^*-\theta$ and select Lyapynov function with the following form:

$$V(s_2,\tilde{\theta},\tilde{D})=\frac{1}{2}s_2^2+\frac{1}{2\eta_1}\tilde{\theta}\tilde{\theta}+\frac{1}{2\eta_2}\tilde{D}^2 \quad (20)$$

According to (10) sliding surface derivative s_2 and from (19) combined with (12), (15), (18), we obtain:

$$\begin{aligned}\dot{s}_2&=\dot{x}_4+\lambda_2(\dot{x}_3-\dot{x}_{3d})-\ddot{x}_{3d}-\lambda_2\dot{x}_{3d}\\&=\lambda_2x_4-\lambda_2\dot{x}_{3d}+f_2+g_2u-\ddot{x}_{3d}-\lambda_2\dot{x}_{3d}\\&=\lambda_2x_4-\lambda_2\dot{x}_{3d}+f_2+g_2(u_c+u_f-u_f^*+u_f^*)-\ddot{x}_{3d}-\lambda_2\dot{x}_{3d}\\&=g_2(u_f-u_f^*+u_c+d)\end{aligned} \quad (21)$$

The derivative (20) over time, combined with (14), (18), (21) we have:

$$\begin{aligned}\dot{V}(s_2,\tilde{\theta},\tilde{D})&=s_2\dot{s}_2+\frac{1}{\eta_1}\tilde{\theta}\dot{\tilde{\theta}}+\frac{1}{\eta_2}\tilde{D}\dot{\tilde{D}}\\&=s_2g_2(u_f-u_f^*+u_c+d)+\frac{1}{\eta_1}\tilde{\theta}\dot{\tilde{\theta}}+\frac{1}{\eta_2}(D-\hat{D})(-\dot{\hat{D}})\\&=s_2g_2(-\tilde{\theta}\xi+u_c+d)+\frac{1}{\eta_1}\tilde{\theta}\dot{\tilde{\theta}}-\frac{1}{\eta_2}\dot{\hat{D}}(D-\hat{D})\\&=\frac{1}{\eta_1}\tilde{\theta}(\dot{\tilde{\theta}}-\eta_1g_2s_2\xi)+s_2g_2u_c+\frac{1}{\eta_2}\hat{D}\dot{\hat{D}}+s_2g_2d-\frac{1}{\eta_2}D\dot{\hat{D}}\end{aligned} \quad (22)$$

According to Lyapunov stability theory, the bridge crane control system is stable, the necessary condition is $\dot{V}\le 0$. From there according to (22), we determine the compensation controller $u_c(s_2,\hat{D})$ according to (23) and the parameter $\hat{D}$ update law according to (24), the parameter θ update law as follows (25).

$$u_c=-\hat{D}sign(s_2g_2) \quad (23)$$

$$\dot{\hat{D}}=-\dot{\tilde{D}}=\eta_2|s_2g_2| \quad (24)$$

$$\dot{\tilde{\theta}}=-\dot{\theta}=\eta_1g_2s_2\xi \quad (25)$$

3.4. Prove asymptotic stability of bridge crane control system

Replacing (23,24,25) into (22) we obtain:

$$\dot{V}(s_2,\tilde{\theta},\tilde{D})=s_2g_2d-|s_2g_2|D\le 0 \quad (24)$$

According to (24), Lyapunov (20) function, $\dot{V}\le 0$, which ensures that $s_2,\hat{\theta},\hat{D}$ is limited. Thus, when time reaches infinity function *V* goes to 0:

$$\lim_{t\to\infty}V=0 \quad (25)$$

This also ensures that when $t\to\infty$ then $s_1\to 0, s_2\to 0$, $\theta\to\theta^*,\hat{D}\to D$, i.e. the crane control systems (5) and (6) are asymptotically stable, the output is tracked to the input reference.

Thus, the crane control system described by (5) and (6) uses the new control law proposed in the article, determined by (19), (14), (23), (24), (25) based on combination of sliding mode control, fuzzy logic and adaptive law based on Lyapunov function, ensuring precise control of the trolley position according to the set reference with eliminated payload oscillation, anti-disturbance.

The block diagram of adaptive fuzzy-sliding control system for 2D crane is shown as Figure 3, including: sliding surface block, fuzzy control block and compensation control block, parameters adaptive update block.

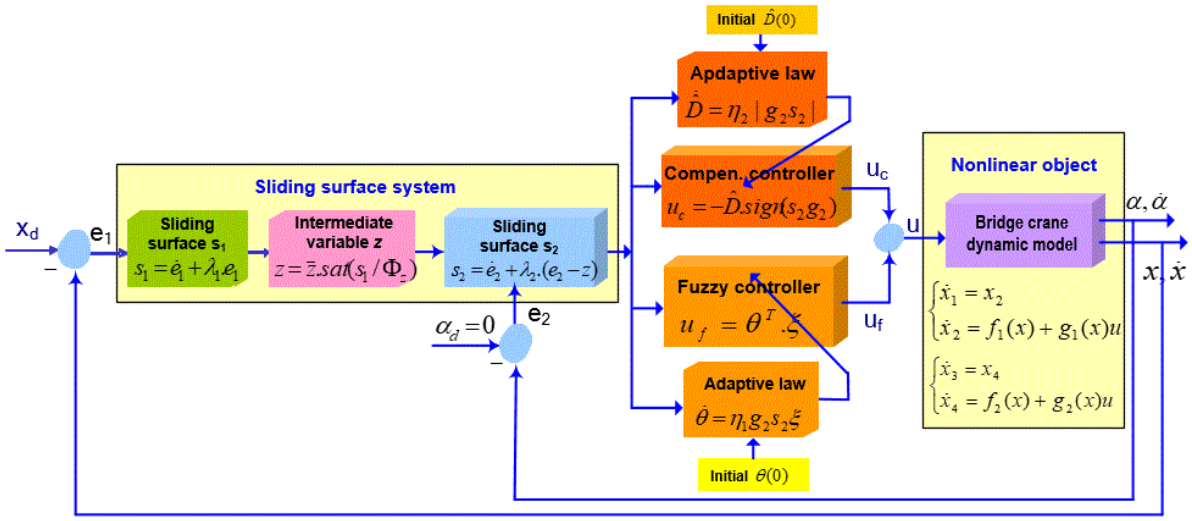


Figure 3. The block diagram of adaptive fuzzy-sliding control system for 2D crane

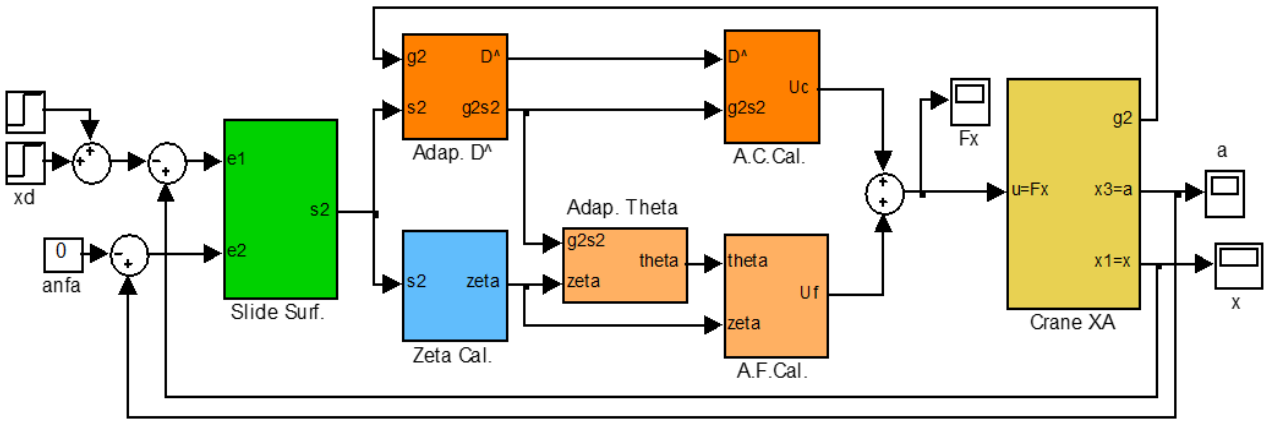


Figure 4. Simulation diagram of the bridge crane adaptive fuzzy-sliding mode control system

4. Simulation results

Simulation parameters of the crane are selected based on physical crane model at UTC laboratory [1].

According to the principle of sliding mode control [12] and crane parameters [1], we choose the sliding surface parameters as follows:

$$\lambda_1 = 0.6, \lambda_2 = 100, \eta_1 = 50, \eta_2 = 0.001, \bar{z} = 0.99, \Phi_z = 5.$$

Building a 2D crane control system on Matlab, using the adaptive fuzzy-sliding control law (AFS) proposed in the paper, we obtained the simulated system diagram as shown in Figure 4, including function blocks: Slide Surf block: calculate the sliding surface of the system; ZetaCal block: calculate ξ; Adap D^ block: calculate the update value $\hat{D}$ for the offset compensation controller; Adap. Theta block: calculate updated value θ for fuzzy controller; AF Cal. block: output of fuzzy controller; AC Cal. block: the output of the controller compensates for deviation; Crane XA block for bridge crane dynamic model.

Based on experience in crane operation and fuzzy control principle, with the above crane parameter [5], the author performs fuzzy input s_2 of the controller (14) by {NB, NM, NS, ZE, PS, PM, PB} having triangular integrated functions with domain $s_2 \in$[-20; +20] and fuzzy output u_f with {UN, UM, US, UZ, FS, FM, FB} having singleton functions with the physical value domain $u_f \in$[0;100].

Simulation scenario of the bridge crane adaptive fuzzy-sliding control system includes the following cases:

+ Case 1: Let the trolley run to position 0.5m and then the trolley runs to position 1.0m and changes coefficients λ_1, λ_2;

+ Case 2: Let the trolley run to position 0.5m and then the trolley runs to position 1.0m and changes coefficients η_1, η_2;

+ Case 3: Let the trolley run to position 0.5m and then the trolley runs to position 1.0m and changes coefficients $\bar{z}, \Phi_z$;

+ Case 4: The input reference is a slope;

+ Case 5: There is impact disturbance;

+ Case 6: Use AFS controller and PID controller [1].

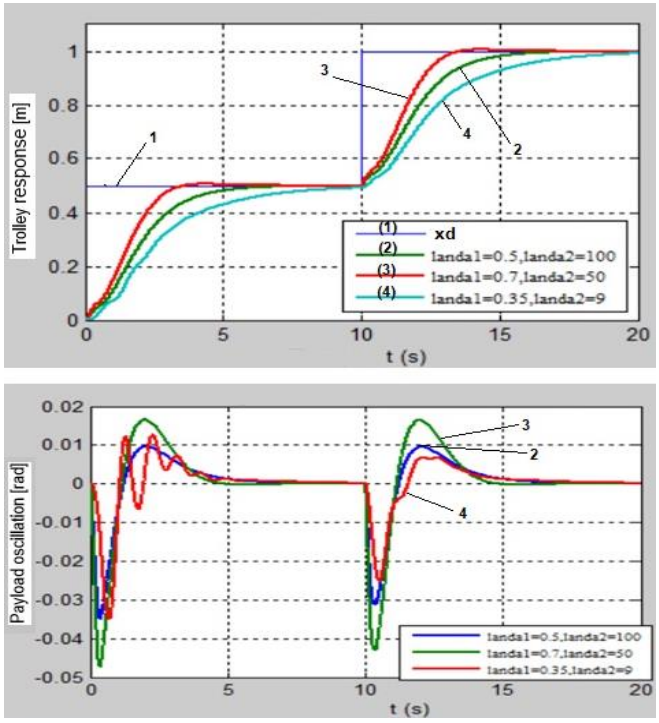


Figure 5. Crane response with different λ_1, λ_2

Table 1. The proposed controllers' performance

Index \ Controller	PID	AFS
Rise time	Small, ~2.5s	Small, ~2.5s
Over shoot	Large, 20.5%	Not
Steady time	Large, ~7s	Small, ~5s
Steady error	<5%, but large with disturbance	~0, or very small
Anti-swing payload	<10°, but large with disturbance	~0, or very small
Anti-disturbance	<10°, but large with loud disturbance	~0, or very small

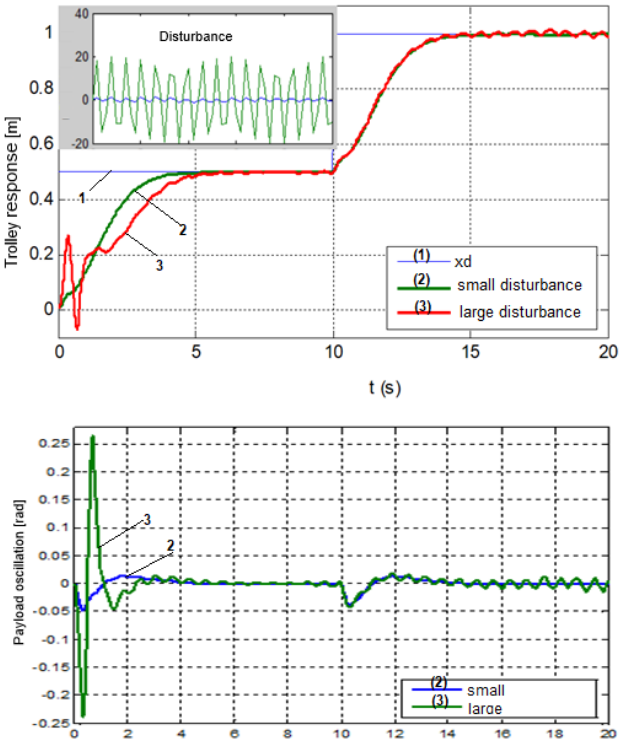


Figure 6. Crane response as affected by disturbance

Simulation results show that the quality of the bridge crane control system as using AFS controller is better than as using PID controller. The application of the AFS controller to the crane obtained the following results: quick trolley's position response, short settling time, eliminating steady error, anti-swing (or very small) payload, anti-disturbances (or very small).

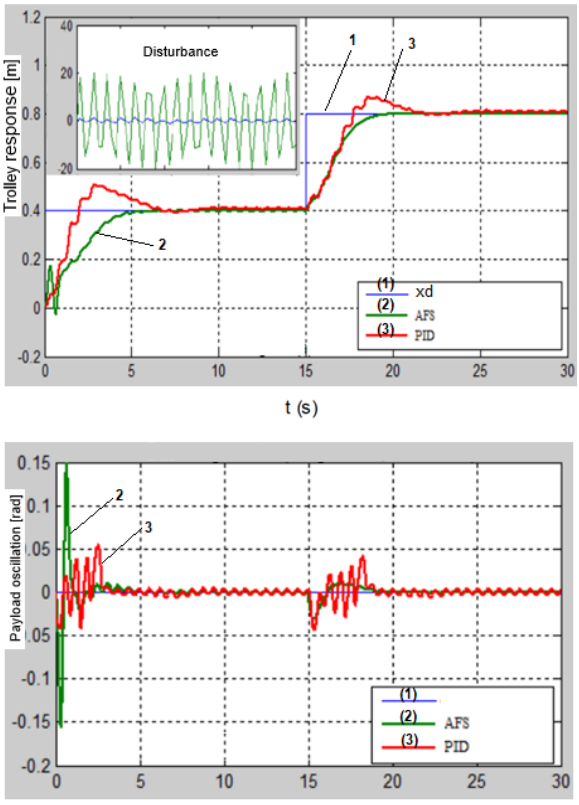


Figure 7. Crane response with AFS and PID controller

The simulation results also show that the selection of the coefficient of the sliding surface affects the quality of the adaptive control system of fuzzy-sliding crane, specifically as follows: when increasing λ_1, the trolley position response is better but greater is the payload oscillation; when decreasing λ_2 or increasing it too large, the payload swings strongly; when η_1 is too small or when η_2 is too large, the trolley position response has vibration and greater payload oscillation; when Φ_z decreases or $\overline{z}$ decreases too small, the trolley position response has vibration and the payload Will swing larger.

5. Conclusion

The article has proposed a new adaptive fuzzy-sliding mode controller (AFS) for 2D bridge crane based on combining sliding mode control with fuzzy logic and Lyapunov function. The AFS controller ensures the objective of controlling the position of the trolley tracking according to the reference trajectory, eliminating the payload oscillation and anti-disturbance (small amplitude). The quality of the bridge crane AFS control system is better than PID controller when crane is affected by large disturbance. The adaptive fuzzy-sliding mode control algorithm allows simple installation on hardware control devices with a normal amount of calculation. The AFS controller allows the crane to operate safely and reliably in harsh environments, such as the harbor.

The success of the AFS controller for 2D bridge crane will continuously study to be applied for the object in real time and for 3D crane with the random disturbances.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Trinh Luong Mien, Nguyen Van Tiem, *Design the state feedback controller for bridge crane*, The 5th national metrology conference, 20-21/05/2010, 720-725, 2010.

[2] Khalid L. Sorensen, William Singhose, *A controller enabling precise positioning and sway reduction in bridge and gantry cranes*, Control Engineering Practice 15, 2007.

[3] Jafari J., Ghazal M., Nazemizadeh M., *A LQR Optimal Method to Control the Position of an Overhead Crane*, International Journal of Robotics and Automation, 2014.

[4] Le Xuan Hai, Thai Huu Nguyen, et al., *Anti-sway tracking control of overhead crane system based on pid and fuzzy sliding mode control*, Journal of Science and Technology 55(1), 116-127, 2017.

[5] Amanpreet Kaur, Priyahansha, Shashiprabha Kumari, *Tanvi Singh, Position Control of Overhead Cranes Using fuzzy Controller*, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2014.

[6] Pal A.K., Mudi R.K., *An Adaptive PD-Type FLC and Its Real Time Implementation to Overhead Crane Control*, International Association of Scientific Innovation and Research, 2013.

[7] Shebel Asad, Maazouz Salahat, *Design of Fuzzy PD-Controlled Overhead Crane System with Anti-Swing Compensation*, S. ASAD ET AL Engineering, 2011.

[8] M. Nazemizadeh, *A PID Tuning Method for Tracking Control of an Underactuated Gantry Crane*, Universal Journal of Engineering Mechanics, 2013.

[9] Yonggang Li, Shuqing Zhou, Hongqiu Zhu, *A backstepping controller design for underactuated crane system*, The 30th Chinese Control and Decision Conference (2018 CCDC), 2018.

[10] Mahan Mahrueyan and Hamid Khaloozadeh, *Designing a Nonlinear Optimal Anti-Sway Controller for Container Crane Systems*, International Conference on Circuits, System and Simulation, 2011.
[11] Fang Y., Zergeroglu E., *Nonlinear coupling control laws for overhead crane system*, IEEE/ASME Transaction on Mechatronics, 2003.
[12] Shibly Ahmed AL-Samarraie, Bashar Fateh Midhat, *Sliding Mode Controller Design for a Crane Container System*, IJCCCE, 2014.
[13] Diantong Liu, Jianqiang Yi, Dongbin Zhao, Wei Wang, *Adaptive sliding mode fuzzy control for a two-dimensional overhead crane*, Mechatronics 15, 2005.
[14] Hahn Park, Dongkyoung Chwa, and Keum-Shik Hong, *A Feedback Linearization Control of Container Cranes: varying rope length*, International Journal of Control Automation and Systems, 2007.
[15] Tuan Anh Le, Soon-Geul Lee, Sang-Chan Moon, *Partial feedback linearization and sliding mode techniques for 2D crane control*, Transactions of the Institute of Measurement and Control 2014, Vol. 36(1) 78–87, 2014.
[16] Trinh Luong Mien, *Adaptive fuzzy sliding mode control for gantry crane as varying rope length*, International Journal of Engineering and Technology, Vol. 8 No. 4 Aug-Sep2016, 1784-1791, 2016.
[17] Chen Zhi-mei, Meng Wen-jun, Zhang Jing-gang, *Intelligent NN anti-swing control for bridge crane*, J. Cent. South Univ., 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 174-187 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Model-Driven Engineering Infrastructure and Tool Support for Petrochemical Industry Automation

Thaise Poerschke Damo[*,1], Leandro Buss Becker[1], Fabio Paulo Basso[2]

[1]*Federal University of Santa Catarina (UFSC), Automation and Control Systems Department, 88066-040, Brazil*

[2]*Federal University of Pampa (UNIPAMPA), Campus Alegrete, 97546-550, Brazil*

ARTICLEINFO

Article history:
Received: 03 June, 2019
Accepted: 16 July, 2019
Online: 30 July, 2019

Keywords:
Model-Driven Engineering
Object-Orientation
Industrial Automation

ABSTRACT

The definition of equipment and components of physical plants is a necessary step towards the development of simulation, control, and supervisory applications for the petrochemical industry. Often it happens that the same plant/equipment is (re)modeled on each application, causing a waste-of-time on repetitive (re)design, besides introducing potential inconsistencies between the models. Moreover, even though each software platform normally offers a different view of the same plant, it is desirable to have some kind of interoperability between them. This paper presents a solution for such issues named M4PIA, which consists in a Model-Driven Engineering (MDE) tool support developed mainly for petrochemical industry automation. M4PIA allows representing industrial plants by means of different and interchangeable object-oriented models, providing means to perform automatic code generation from a plant specification for different software platforms. Currently, our work involves using M4PIA in high-level automation manouvers and plant simulations. Evaluations studies performed with M4PIA shows that it covers most features expected from a MDE tool suite. Besides, the use of M4PIA is expected to result in less development time and costs, while it increases efficiency, maintainability, and reliability of the developed applications.

1 Introduction

There is no doubt that todays industrial environments are highly automated, at least in the first two layers of the ISA95 automation pyramid. The most basic layer, named *Field Level*, includes devices, actuators, and sensors. The *Control level* is immediate above, and accounts with PLCs and PIDs to control such devices. Programming PLCs and PIDs is typically done by means of integrated development environments (IDEs) that comply with international standards IEC 61131-3 [1] and IEC 61499 [2]. Rarely this kind of system does not make use of SCADA (supervisory control and data acquisition) in a third *Supervisory* layer.

The IEC 61131-3 standard [1] addresses PLC programming, including four programming languages and sequential function charts (SFC). It helps migrating software developed for one PLC type to PLCs of other vendors. Targeting distributed systems, the IEC 61499 standard [2] replaces the cyclic execution model of IEC 61131 by an event-driven execution model, which allows an explicit specification of the execution order of function blocks (FB). Both standards do not address equipment specification semantics.

To cope with procedures for automating continuous process operations there is the ISA 106 standard [3], which is typically deployed at the *Supervisory level*. Examples of commercial tools therefore are *Exapilot* from Yokogawa and *GenSym G2* from Ignite Technologies. The present work stands at this layer, and makes use of a proprietary tool with similar goals named MPA [4].

Another common need in automation systems, specially when advanced control techniques come to play, is performing simulations. Examples of tools for such purposes are Labview, Matlab/Simulink, and Modelica-based (like Dymola and OpenModelica). These tools provide powerful simulation cores, and also provide a vast library of components that allow designers to build virtually every system or system element they want to. It happens, however, that there is no specific semantics for equipment specification.

Considering an automation plant, the lack of semantics for equipment specification brings difficulties for representing this same automation plant in the different software de-

[*]Thaise Poerschke Damo, DAS/CTC/UFSC, +554837217606 & thaisedamo@gmail.com

velopment tools used in all automation levels. Such repeated representation is something quite common, for instance, to implement different perspectives in the automation domain: processes operation, monitoring, simulation, and control.

The work presented in this paper targets the integration of different perspectives in petrochemical industry automation through a tool support built on the Model-Driven Engineering (MDE) paradigm [5]. MDE is built on widespread techniques in Systems and Software Engineering, presenting the following main characteristics: (1) uses models in all the phases of software development to improve understanding; (2) raises the abstraction level of software system specifications, hiding platform-specific details; (3) develops Domain-Specific Languages (DSLs) [6] and frameworks to suit a domain; and (4) applies transformations to automate repetitive activities and improve product quality derived from Software Engineering (e.g., source code, libraries, processes, etc.) [7]. In other words, MDE proposes representation of diverse artifacts through modeling and tool support.

This paper presents M4PIA (Model-Driven Engineering for Petrochemical Industry Automation), which consists of an infrastructure and tool support to help the design of advanced control applications for the Petrochemical Industry. An evaluation of the proposal is also presented. Therefore, the remainder parts of this paper are organized as follows: Section 2 presents background information that motivates this research, highlighting the application domain and related works. Section 3 details the proposed M4PIA infrastructure and tool support. Section 4 shows the evaluation method employed for assessment. Section 5 details M4PIA assessment and presented obtained results. Section 6 highlights our conclusions and future works perspective.

2 Background

2.1 Advanced Control in the Petrochemical Industry

This work targets the design of advanced control strategies for the Petrochemical Industry. The advanced control does not intend to control the automation plant devices (valves, motors, etc) located in the *Field Level* of the automation plant. This low-level control should be done by the regulatory control in the upper, *Control level*, that executes on the PLCs. The advanced control stands at the *Supervisory level*.

For example, let us analyze the design cycle of an advanced control system for oil extraction platforms. Initially, designers should create a software-in-the-loop (SIL) environment, where the phenomena to be controlled (e.g. gas compression) is "executed" within a simulation tool (Labview, Simulink, Modelica) that communicates with the control software executing in the same hardware platform that will be used in the final system. Here, the automation plant elements are represented twice: (1) in the simulation model and (2) in the control software.

Before deploying the control system under design in the operational environment (the oil extraction platform), tests are typically performed in a hardware-in-the-loop (HIL) environment. At this point, instead of communicating the control system directly with the simulation tool, it must interact with the same hardware (PLCs) and software (drivers) that are used in the operational environment - here the PLCs communicate with the simulator. The communication between control software and PLCs is typically done by means of some kind of bridge, for instance OPC drivers. Using this kind of structure brings the need of creating a third plant representation, which stands for the supervisory software. A fourth plant representation could be used if an specific simulator is adopted in the HIL environment, which is something common in the oil industry. In theory, if the HIL implementation passes all tests, the control system is ready to be implemented in the operational environment.

Given such scenario, the problem under consideration in our work concerns providing a solution to optimize the long time needed to model the equipment in different software development tools (and languages). It also should reduce possible inconsistencies generated when moving from one model to the other. It is important to provide interoperability among such plant representations.

The development starting point of the proposed infrastructure was the analysis of a system called MPA (acronym for Automated Procedures Module). MPA software platform is widely used by the development team of Petrobras in the operation and control of several oil platforms and refineries. It can be used for simulation purposes and also to control the operation of real systems. In order to maximize the gains from using a MDE approach, it was also analyzed the system called EMSO (Environment for Modeling, Simulation and Optimization), a platform used for modeling and simulating chemical processes.

2.1.1 MPA Operation and Control Software

MPA [4] software was developed for oil platforms automation in order to support the development and execution of industrial control and automation applications. This software was developed by Tecgraf/PUC-Rio Institute under request of Petrobras S.A.

Automation is performed through operation maneuvers of plant equipment. Basically, MPA consists of an execution server and a configuration and management application. In the latter, industrial plants are modeled using object-orientation and diagrams are used to define the maneuvers in the respective plant. The server is responsible for executing the configured operation maneuvers in the diagrams and handles the equipment interacting with the supervisory system through OPC (OLE for Process Control) communication bridges.

Currently, processes equipment are modeled directly in LUA programming language as classes in the application's pre-configuration phase. In this phase, attributes and methods of each class are defined, i.e., of each type of equipment. One equipment can be used to compose another, being modeled as its attribute. Equipment classes are described in a pre-configuration file and loaded at the MPA configuration stage. The developer uses such info to define the plant equipment instantiation and to model the execution diagrams, in flow language, which describes the operating maneuvers of plant equipment, i.e., the sequence of equipment functions execution. Plant and flow files are saved separately in other file formats.

The infrastructure proposed in this paper is able to simplify the pre-configuration phase by generating code automatically, reducing the time spent in conceptual modeling for new plants or even guiding the development of similar supervision and control applications, reusing existing models by including, extending, or modifying them.

2.1.2 EMSO Simulation Software

EMSO [8] is a tool for modeling and simulating dynamic processes based on equations. It is part of a project maintained by a consortium of Brazilian universities and national petrochemical companies.

The EMSO platform is composed by a graphical interface and its own object-oriented modeling/programming language. Such language was created from the combination of the best modeling characteristics found in existing languages, resulting in a simpler language with a better code reusability.

EMSO modeling language is composed of three major entities: models, devices, and flowsheets. Model is the mathematical description of a device. Device is an instance of the model and represents a real process equipment. Flowsheet represents the process to be analyzed, which is composed of a set of devices. A model development in EMSO consists of defining a class of a real equipment or a part of a process. Model can be composed of parameters, variables, equations, initial conditions, and sub-models and can be based on a pre-existing one, only adding new functionalities.

The graphical user interface allows developers to create models, build flowcharts, check project consistency, execute simulations and visualize the results. The software has several consistency analysis of the model and the whole process to be simulated, including: initial conditions consistency, measurement units, and system of equations solvability. The platform is multithread, which allows real-time simulations to run more than one simulation at the same time.

The proposed solution aims to support equipment models by generating automatic codes and to maintain a correct correspondence between plants to be controlled by other applications and their simulated processes in EMSO for tuning tests and validation of control proposals.

2.2 Challenges Related with MDE Usage

Using MDE in the present work comes from the fact that it is a software development methodology with emphasis on the domain specification models, allowing to improve productivity, system understanding, and its maintenance and evolution. MDE paradigm proposes applications described through models at different levels of abstraction using standards such as the UML (Unified Modeling Language). More than just conceptual design, the produced models can be interpreted by automation tools that can generate schemas, code skeletons, and tests for multiple platforms. Definitions of more abstract layers provide formal basis for structuring lower-layer models. Thus, it is possible to facilitate design decisions and to build artifacts automatically from models, reducing development times and costs.

Additionally, MDE has been used to support software systems development in both academia [9]–[11] and industry [12]–[19]. Among the benefits credited to MDE, Mohagheghi et al. [20] highlight the following: shortened development time and increased productivity; improved software quality; automation through generation of code and other artifacts; provision of a common framework for software development across the company and lifecycle phases; maintenance and evolution; improved communication and information sharing among stakeholders. MDE is especially interesting for scenarios involving systems that should be made available on multiple platforms [21].

However, developing software systems with MDE is not trivial. Besides the development of domain specific languages, it requires an automated process, including transformation scripts that connects MDE resources, such as refinement tools and model-based operations [22]. In special, these works need specific assessment for their pieces, such as for DSLs, as well as for their integration promoted by model transformations. In this sense, Mohagheghi et al. [14] go beyond and also expose the criteria that led companies to adopt MDE. Such criteria involve the abstraction level that hides details, communicating with non-technical staff, as well as model-based simulation, execution, and test. In other words, a feature analysis is required.

2.3 Related Works

The idea to use object-orientation towards developing industrial automation systems is not new. For instance, in [23], the authors presented an environment that allowed to make object-oriented modeling of plant equipment ir order to further make automatic code generation. Even though the proposal was conceptually very interesting, it did not make use of standard modeling languages (at that time UML was emerging), so that it could not be adapted to current standards. More up-to-date standards were used by Thramboulidis in [24], but the foccus of that work was not on the static-structure of the plants but on their their dinamic behavior, specially targeting implementation on PLCs. In fact that was an anticipation of the recent IEC-61131-3 standard [1], which covers the use of object-oriented programming on PLCs.

Other related works only covered partially the issues presented in this paper. In chemical engineering domain, for instance, Becker [25] presented an UML-based framework for building integration tools that help developers in change propagation for maintaining inter-document consistency, focusing on relationships between flow sheets, that describe chemical processes and its simulation models. MindCSP solution [26] focuses on specifying the autonomic behavior of cyber-physical systems by its sensor-actuator networks and the autonomic control loops (monitor/analyze/plan/execute). Another example of approach that supports development of industrial control applications is the AUKOTON process [27], based on the UML Automation Profile, uses profile's concepts to represent the requirements, functionality and structure of the control applications, including requirements of stakeholders, alarm events, and control algorithms such PID, fuzzy and MPC. Other relevant applications of MDE are referenced in a state-of-the-art review of software engineering in industrial automation [28]. However, so far there is no other related work on MDE application with focus

on the generic definition of industrial plants equipment for sharing between different software: control, operation, and simulation platforms. This gap is fulfilled by the proposal presented in this paper.

3 Proposed M4PIA Infrastructure

The proposed M4PIA infrastructure is composed by: 1) domain specific languages, which intents to abstract three levels that hides details from implementation; 2) two sets of model transformations supporting automated model integration between these three DSLs; and 3) two sets of code generators. Everything is developed under the same Eclipse ecosystem, including the Eclipse Modelling Framework (EMF) [29], the Acceleo model transformation engine for code generator [30], and QVTo as a mean to make transformations between different metamodels [31].

A model in MDE represents part of a function, structure, or behavior of a system. The domain knowledge necessary to design M4PIA was acquired by the analysis of a significant set of applications designed on different software platforms from the petrochemical industry, seeking to identify typical requirements and behaviors of these systems. Understanding different applications provide the ability to differentiate general structures, the invariant domain aspects, from the specific structures, the variable aspects. Successive analysis and synthesis steps converged to the generic domain design model that was created.

In a model-driven engineering method, a high level of abstraction is typically used in the application development. The system development process in MDE starts with the highest abstraction model and decreases through properly defined transformations (model refinement). There are two MDE abstraction levels used in the present work: a platform independent viewpoint, the Platform Independent Model (PIM) and specific platform viewpoints, the Platform Specific Model (PSM). The latter contains details and characteristics of specific implementations platforms.

In a new application development its PIM model is created according to the provided metamodel (constructed based on domain conceptual analysis), then the PSM model is automatically generated by a Model-to-Model transformation (M2M), based on the PIM model and according to the PSM metamodel. Lastly, the platform-specific source code is generated from the PSM through an automatic Model-to-Text transformation (M2T). This process is illustrated in Figure 1.

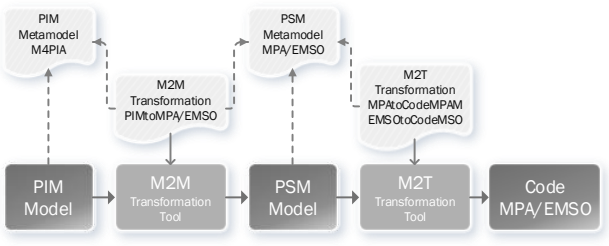


Figure 1: Development cycle using M4PIA

M4PIA infrastructure is built as support for the MDE of equipment class definition, so that applications for operation, control, and simulation platforms for the petrochemical industry can be created. The proposed solution was implemented in the EMF (*Eclipse Modeling Framework*) tool [29], in *Eclipse Oxygen* environment. The first and fundamental element built is the M4PIA metamodel, a PIM metamodel that represents the entire domain of the desired applications to be created, independently of its implementation platform. Then the PSM metamodels were constructed, defining the specificities of each supported platform. Based on metamodels, it was defined the M2M transformations from PIM model to PSM models and, at last, the M2T transformations for the automatic source code generation from each PSM model.

3.1 M4PIA Metamodel

The identification and analysis of petrochemical industry applications allowed the definition of a set of generic elements, capable of being shared by a wide range of industrial automation systems. This set of elements are materialized by means of a metamodel named *M4PIA*. Our aim is that M4PIA metamodel can contribute substantially for the implementation of new systems from such application domain, targeting reusability and facilitating automatic code generation.

The developed M4PIA metamodel is presented here through a class diagram designed using the *Eclipse Modeling Framework* and its *Ecore* metamodel, similar to UML. The proposed class diagram is shown in Figure 2. The representation shows a set of classes, interfaces, and collaborations, with their respective relationships, expressing results of the structure and requirements analysis of the problem domain and its components. Model constraints were specified using OCL (Object Constraint Language). The main metamodel elements are detailed next.

An entity is defined as the most elementary unit of the proposed model. The `Entity` class is considered abstract and only provides a basic structure for the more specialized classes in the hierarchy. A `Project` is instantiated and can be composed of several `Files`. These can be imported as libraries, `ImportedFile`, or be an entity group to have their source code generated by the infrastructure, `GeneratedFile`.

An `Equipment` is a type of entity that can contain attributes and methods, defined as specific classes to reflect physical characteristics and functions performed by the equipment in a plant.

The abstract class `Variable` represents a logical variable and can be `NonTyped` or `Typed`. Typed variables can be `EquipmentType` or `BasicType`, more specifically `Real`, `Integer`, `Boolean`, or `String`.

The `Method` class is a `Function` specialization and has exclusive connection to an `Equipment`. The `Function` class represents a logical entity that can either represent a high-level operation (in the physical domain of the plant) or a low-level (in the application domain). A function can have multiple `Variable` instances associated, acting as parameters or results, as well as may have a language and an associated code, which textually describes the instruction desired for the interpreter.

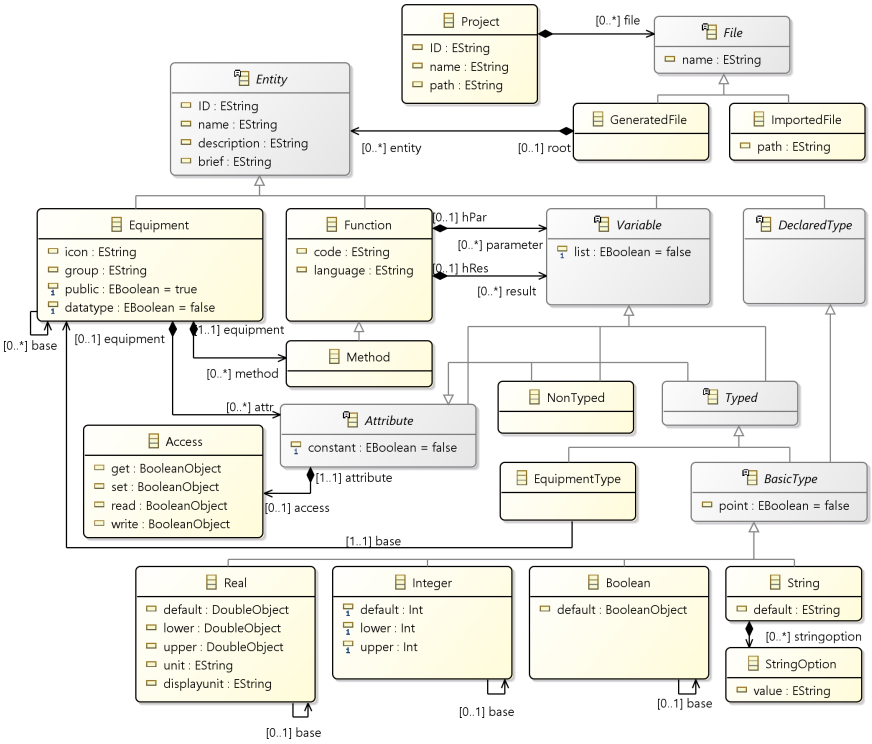


Figure 2: Elements of the proposed M4PIA metamodel

An equipment `Attribute` is a specific case of Variable which is allowed, through the relationships hierarchy, to assume a type that can be another `Equipment`. This aspect guarantees to the metamodel the ability to reproduce scenarios where there are natural recursions in relationships between equipment, i.e., an equipment or a machine that has other equipment as an attribute. For example, a specific type *Compressor* can be used as one of the attributes of a *Compression System*, both devices being instantiated by the `Equipment` class.

Furthermore in the `Attribute` context, there is a permission management for read or write operations, as highlighted in the `Access` class. New data types can be defined through the `DeclaredType` entity, based on basic types, and can be used as the basis for variables and attributes as well. For example, it is possible to declare a new type *Positive Real* and set a real type attribute *Position* of an equipment *Valve* based on the *Positive Real* type.

3.2 MPA and EMSO Metamodels

Given that the generic M4PIA metamodel was obtained through the analysis of the specific software platforms, the specific metamodels for each software are similar to the generic one and the main differences between them are further highlighted.

The PSM for MPA, or simply MPA metamodel, is depicted in Figure 3. The `Component` element corresponds to `Entities` in M4PIA. In addition to `Equipment`, the `Class` component can also be specified as `PointClass`, considered on the platform as a data type. `Variable` and `Attribute` classes are not abstract and must have an associated `Type`. Possible types are `BasicType`, point classes, and equipment. MPA basic types instances are imported automatically at the beginning of each project, they are: *Real*, *Integer*, *Logical*, and *Textual*. Four instances of native MPA point classes are also imported: *Real Point*, *Integer Point*, *Logical Point*, and *Textual Point*, each with its attributes.

In the MPA platform, codes are not exclusive related with functions, classes and files can also be composed by `Codes`. File codes are used to import components of the `DLL`, declaring the available `DLLFunctions` and their `DLLParameters`.

The second PSM developed is the EMSO metamodel, which is presented in Figure 4. Since EMSO is a simulation platform, mathematical models of the equipment, or `Devices`, are instantiated through the `Model` class. Functions are always related to a model as `Equations`. In this specific domain, `Variable` is an `Attribute` specialization and represents the variable attributes of the models and `Parameter` the constant ones. Variables cannot assume a textual type, but textual parameters can be instantiated through a `Switcher`. `Flowsheet` entities with their `Devices` and `Connections`, parameter `Estimation` and process `Optimization` were not considered primordial for the equipment description domain and are not specified at this stage of the project.

3.3 M2M and M2T Transformations

For each PSM supported by M4PIA infrastructure it is necessary to define the respective transformations. Transformations can be either model-to-model (M2M) or model-to-text (M2T). In M2M transformations the elements specified in the PIM model are mapped to elements from a specific PSM. M2M are implemented in our work using the *Eclipse QVTo* tool, an implementation of the standard

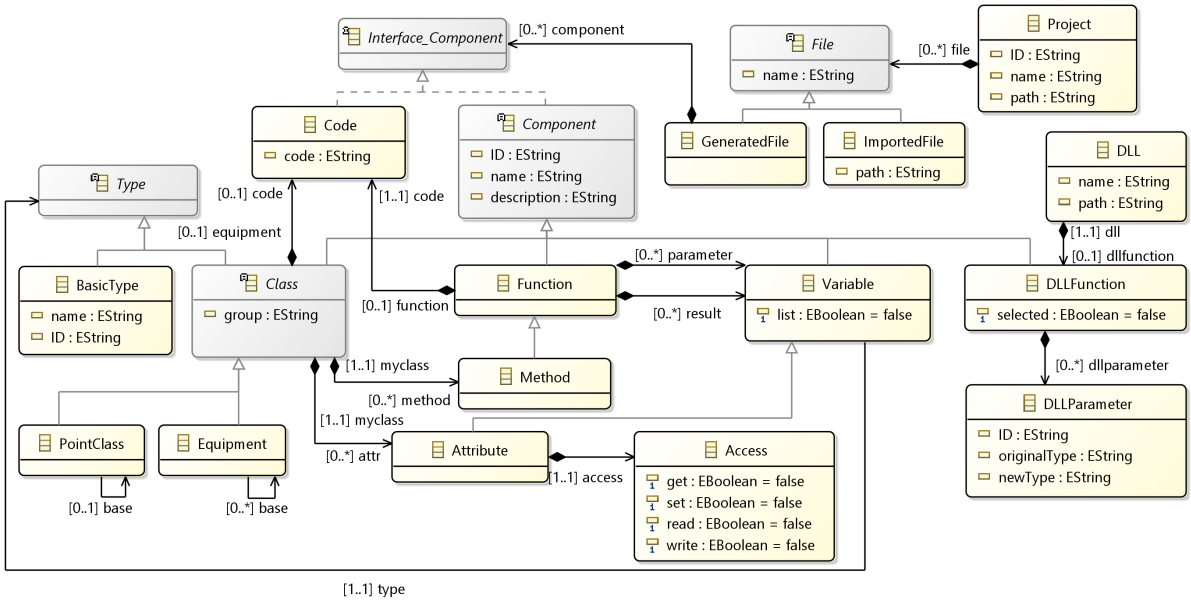


Figure 3: Elements of MPA metamodel

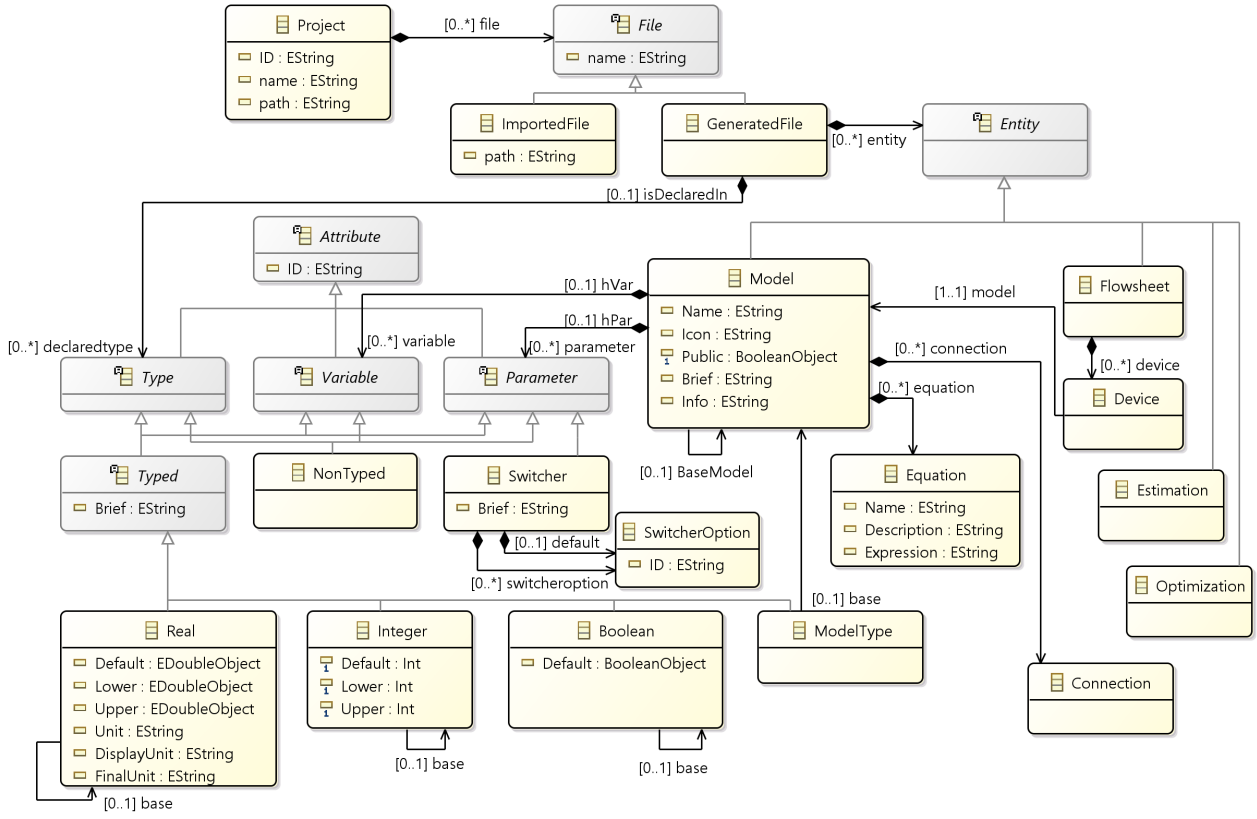


Figure 4: Elements of EMSO metamodel

OMG model transformation language QVTo (Operational Query/View/Transformation). The M2T transformations are developed specifically for each target platform. They are target in this work by using the *Eclipse Acceleo* tool for templates definitions and have as output the application source code ready to be used.

The explanation regarding each PSM metamodel provided in the previous section provide readers a good overview about the adopted M2M transformation rules, i.e., how the elements of the M4PIA metamodel (Fig. 2) map to the MPA metamodel (Fig. 3) and to the EMSO metamodel (Fig. 4). The M2T transformations are target in this work by using the *Eclipse Acceleo* tool for templates definitions and have as output the application source code ready to be used.

4 Evaluation Method

This section describes the evaluation method used to allow analyzing the completeness and applicability of the proposed infrastructure. Our goal is to capture the perception of M4PIA users in performing a characterization of the features associated with this MDE-based approach. By achieving this goal, it would allow us to understand whether the concepts introduced in M4PIA infrastructure really have potential in terms of technology, thus providing evidence of quality attributes associated with all the developed tools.

This evaluation, therefore, attempts an empirical study of type "feature analysis-experiment" [32], in a quantitative and qualitative experiment format. It allows providing evidences of suitability, very common for assessing tool support in computer science.

4.1 Context Selection

The context of this study is the academy, a first trial of in-vitro approach, with half of participants working in industry. The experiment captures the perceptions of individuals to measure the value of a technology in a real problem. It is applied with nine master students and two undergrad students from the Federal University of Pampa (Unipampa), all them with good knowledge on tool selection and software acquisition. This context configures a good sample, since students were introduced in the MDE theme, with desirable features from these tools. Therefore, it characterizes a good scenario to evaluate two main quality attributes for suitability that could be associated with the proposed tool support: completeness and applicability.

4.2 Experiment Design

No randomization and balancing: The subjects are not randomly placed into groups (not a randomized block design), so that only one design approach is performed by the same number of subjects. Thus, the goal is not to compare DSLs or code generators.

Blocking: The selected subjects for this quasi-experiment had different backgrounds in modeling systems, but no background in modeling petrochemical systems. Thus, to minimize the effect of eventual differences in experiences, the subjects received the same training. We also defined whether a subject is inexperienced or experienced, applying a survey, prior to the experiment, to qualify the subject background on the problem domain.

Instrumentation: It was provided a configured version of the Eclipse IDE with M4PIA plugins. The subjects installed this version in their notebooks in Windows or Linux. Then, we provided video tutorials in mp4 format, composed by: 1) A long duration video (75 minutes) depicting the proposed infrastructure, the needs and the proposed solution; 2) A second video (15 minutes) demonstrating how to use the Eclipse IDE, and 3) A third Video (25 minutes) demonstrating how to design a model using the proposed approach. Six other video-tutorials complement the other quantitative part of this experiment, focusing in evaluating specific design and codification tasks. However, they are irrelevant for the present evaluation, which focuses in presenting the feature analysis experiment.

4.3 Variables

Two sets of **independent variables** are used: (1) EMSO, MPA, and M4PIA; (2) Academic Experience and Industrial Experience. Besides, two **dependent variables** are used: Completeness and Applicability.

4.4 Data Analysis

Applicability: a survey to measure the quality attributes from the experience along design and transformation tasks. **Completeness**: a coverage analysis of the main features considered important in MDE including the following.

4.4.1 Desirable Features from MDE Tools

Feature 1: Platform independence: Also known as PIM support [33], it evaluates the capacity to model the PIM application at a high abstract level.

Feature 2: Multiview representation: Also known as PSM support [33], it evaluates the capacity to automatically generate a PSM model for a chosen target platform.

Feature 3: Platform adherence: Also defined in [33], it states whether multiple target platforms are supported.

Feature 4: Evolution: Also known as Application evolution (Changeability) [33] [34], observes if model changes should be done in PIM level and implies new execution of transformations.

Feature 5: Interoperability: Relates to model import/export (tool interoperability) [33], observes the support for exporting models in XMI format.

Feature 6: Flexibility: From [33], accounts for the flexibility to change the transformation process and model refinements.

Feature 7: Correctness: From [33], accounts for model validation in order to check consistency with its metamodel.

Feature 8: Expressiveness: Also called completeness [33] [35], observe if all domain concepts are needed.

Feature 9: Traceability: From [33], accounts for models traceability.

Feature 10: Reverse engineering: From [33], accounts for Code-to-model and PSM-to-PIM transformations.

Feature 11: Built on standards: From [33], accounts for the use of standarized technologies.

Feature 12: Well defined concepts: Adopted from [34], observes if there are no contradictions between concepts of the models.

Feature 13: Code generation: Feature from [35], accounts for how complete is the generated application code.

4.4.2 Metrics

Metrics for estimating the completeness of the features:

$$CFx = 1 - ((44 - Cov)/44)$$

Where CFx is the completeness for a feature x, 44 is the maximum value rank for each feature, and Cov is the computed by:

$$Cov = (TD + PD + NA + PA + TA)$$

Where Cov is the coverage for a feature, as the sum of the following values posted by each respondent: **TD** is 0 (zero) and replaces the survey value for "Totally disagree"; **PD** is 1 (one) and replaces the survey value for "Partially disagree"; **NA** is 2 (two) and replaces the survey value for "Not agree or disagree"; **PA** is 3 (three) and replaces the survey value for "Partially agree", and; **TA** is 4 (four) and replaces the survey value for "Totally agree";

Thus, at the end of the quasi-experiment, a questionnaire was filled out in a Google Forms, were the participants should rate different criteria, where those for feature analysis are built on a five point Likert scale (0=worst, 4=best) for the analysis of the proposed infrastructure.

5 M4PIA Assessment

In order to evaluate the applicability of the proposed solution, this section presents the impressions noticed by the 11 users previously mentioned. Such impressions were collected after the users modeled a simplified *Gas Compression System* (GCS). The model was further used to generate the control software in the MPA tool and the simulation software in the EMSO tool. All related development, including the PIM model design and the transformation executions, was done within the Eclipse environment.

The simplified GCS consists of one *surge tank* (a *knock-out drum* with a *flaring valve*), one *output header*, and two *one-stage compressors*. A *stage of a compressor* has one *heat exchanger*, one *suction drum*, one *compression element*, and one *recycle valve*. The *output header* has one *exportation valve* and two *gas-lift output valves*. Such system is illustrated in Figure 5. The present case study has 11 equipment, with a total of 36 attributes and 2 methods.

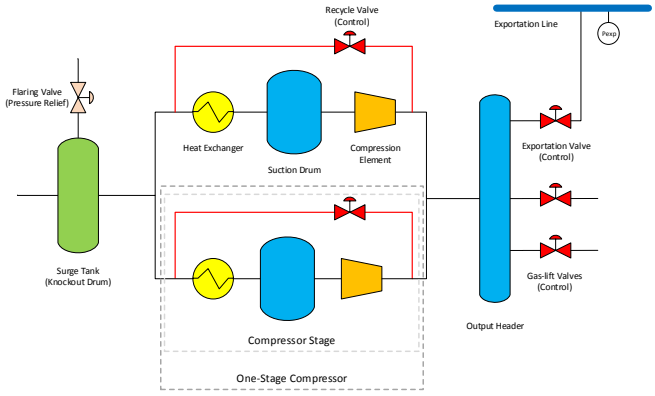


Figure 5: Simplified Gas Compression System

The work in the Eclipse tool can be summarized as follows. After creating a new project, the designer must create a PIM using the Eclipse tree structure. To illustrate how the designer interacts with Eclipse, Figure 6 shows the instantiation of three equipments from the GCS: `Compression Element`, `Compressor`, and `Compression System`. The former contains the following attributes: `Flow`, `Suction/Discharge Pressures`, and `Suction/Discharge Temperatures`. The latter has as attributes: `Power`, `Surge Tank`, `Output Header`, and a `List of Compressors`. It also has the method `Calculate Power of the Compression System`.

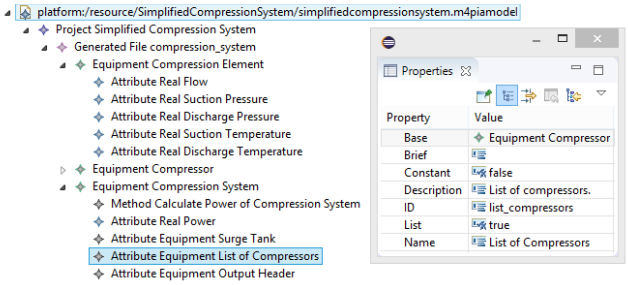


Figure 6: Illustration of the EMF tree structure

To model an equipment on both MPA and EMSO software platforms, the developers should seek out for generalization. For example, a *compression system* may have one or more *compressors* and a *compressor* may have one or more *stages*. Hierarchy is also highly desired, for example, a *valve* can be specialized in *pressure relief* or *control valve*, just as a *tank* can be specialized in *surge tank* or just a *knockout drum*. Given that in EMF the PIM is developed using a tree structure, as shown in Figure 6, the UML class diagram from Figure 7 was created to better illustrate the designed PIM.

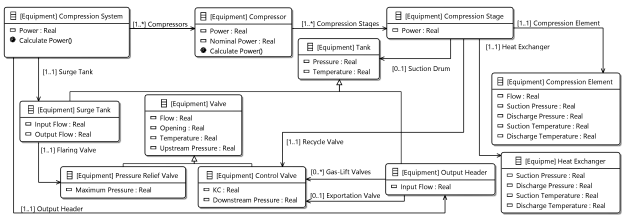


Figure 7: Class Diagram of the Simplified Gas Compression System PIM Model

After the PIM model development, the M2M transformations for MPA and EMSO platforms were performed. The PSM models were refined according to the specificity of each platform and it was possible to automatically generate the necessary source code to implement the compression system on the target platforms by applying the M2T transformations. In total, 245 effective lines of code were automatically generated for MPA and 117 for EMSO. One should recall that effective LOC means that comments and blank lines are not counted.

5.1 Preliminary Analysis

The system developed on this study can be considered small if compared to other existing applications, which can be five times bigger. In order to better illustrate the impact that the use of M4PIA can have on the development of the applications of this domain it is shown some data from a native MPA application example, which aims the detection and treatment of hydrate plug formation in gas pipelines of a floating production storage and offloading (FPSO). The numbers presented in Table 1 allow us to observe that an application with 50% more equipment classes (right on the table) can have twice the total amount of attributes and more than 12 times the number of methods. Observing the numbers regarding LOCs, the FPSO application is above 1K LOC. So it is not difficult to conclude that generating such code automatically – like M4PIA infrastructure does – can potentially save many working hours.

Table 1: Comparison of Domain Applications

	GCS	FPSO
Class of Equipment	11	17
Total Amount of Attributes	36	74
Total Amount of Methods	2	25
Effective Lines of Code (*LOC*) - MPA	245	1297
Instantiated Equipment	19	621
Instantiated Control Points	73	2428

5.2 Feature Analysis

In the following it is present data collected through a feature analysis, which allowed us to understand participants perceptions about the overall M4PIA infrastructure. The goal is to understand initial expectations of the overall approach introduced in the first video, so that their perception could help on the execution of the controlled experiment.

Figure 8 provides the answers to the analysis of Feature 1 - The approach allows to: model the application at platform independent level (PIM), at a high level of abstraction of the target problem used in the implementation (simulation, supervision, operation and control of industrial plants of the petrochemical industry. In this sense, nine from eleven students agree that this is an important feature provided in M4PIA infrastructure.

Figure 9 provides the answers to the analysis of Feature 2 - The approach allows: automatic generation of the platform-specific model (PSM) for the chosen target problem (simulation, supervision, operation and control of industrial plants in the petrochemical industry). Thus, nine from eleven students agree that this is an important feature provided by the infrastructure.

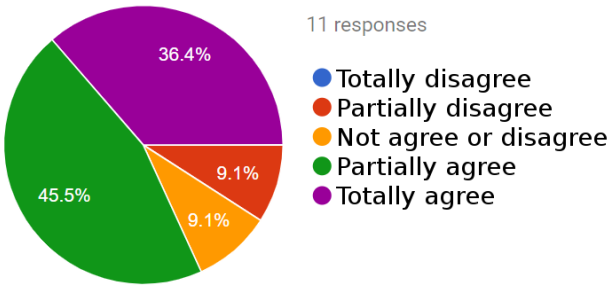


Figure 8: Analysis of Feature 1: Platform independence

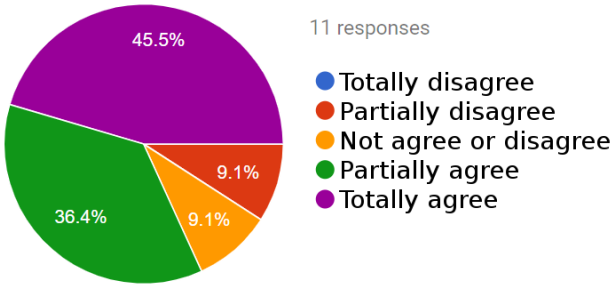


Figure 9: Analysis of Feature 2: Multiview representation

Figure 10 provides the answers to the analysis of Feature 3 - The approach allows: automatic generation of code for multiple target platforms. Seven participants agree that this is an important feature provided by the infrastructure, where five of them only partially agree. Thus, feature 3 should be planned for improvement.

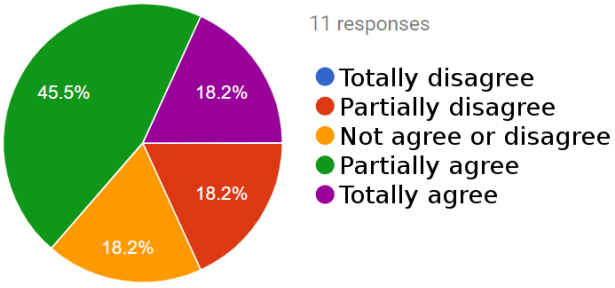


Figure 10: Analysis of Feature 3: Platform adherence

Figure 11 provides the answers to the analysis of Feature 4 - The approach supports the evolution of applications through model to model and model to code transformations. In this sense, nine from eleven students agree that this is an important feature provided by the M4PIA infrastructure. In fact, evolution is one of the greatest benefits promoted by MDE approaches for automated integration [14], and M4PIA is on a solid ground to allow this feature for their future users.

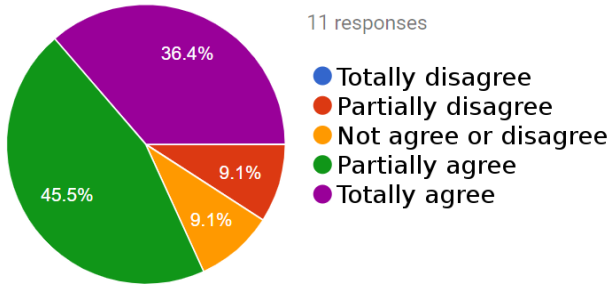


Figure 11: Analysis of Feature 4: Evolution

Figure 12 provides the answers to the analysis of Feature 5 - The approach supports model interoperability. That is, it is evaluated whether these are exported and imported in a standard way, such as those provided by OMG. Seven agree that this is an important feature provided by the infrastructure, where five of them only partially agree. Three answers were uncertain, and this is because OMG supports lots of other standards not handled in M4PIA.

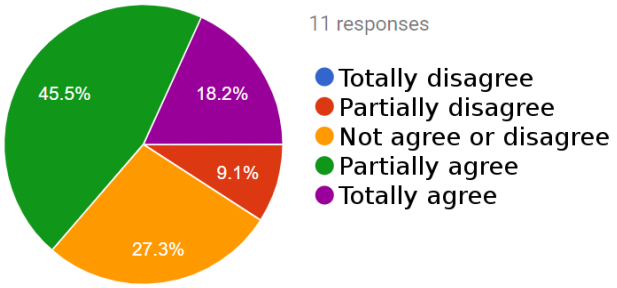


Figure 12: Analysis of Feature 5: Interoperability

Figure 13 provides the answers to the analysis of Feature 6 - The approach supports flexibility to change the transformation process and model refinements. Eight respondents agree that this is an important feature that should be provided by the infrastructure. Three answers were also uncertain, and this is because they did not scaled the problem through the demonstration.

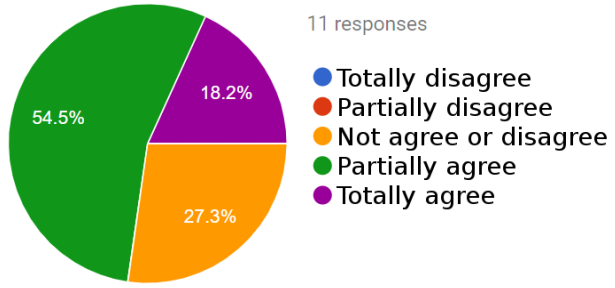


Figure 13: Analysis of Feature 6: Flexibility

Figure 14 provides the answers to the analysis of Feature 7 - The approach supports the validation of the models to verify consistency with their respective metamodels. Eight respondents agree with that statement, concluding that M4PIA provides this as an important feature for designers. Consistency in design is an essential feature in a MDE-based process that needs the correct code generation. It can only be ensured with a rich set of OCL rules [36].

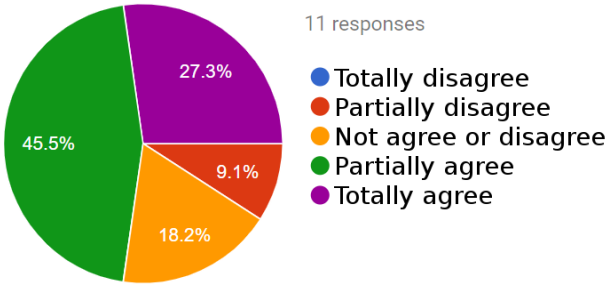


Figure 14: Analysis of Feature 7: Correctness

Figure 15 provides the answers to the analysis of Feature 8 - The approach is expressive and allows to express all the concepts of the represented problem domain. Nine from eleven respondents agree that this is a design feature associated with M4PIA. Thus, since the ability of express domain concepts is a requirement for DSL development, we conclude that the provided metamodels satisfy the modeling needs of the presented context.

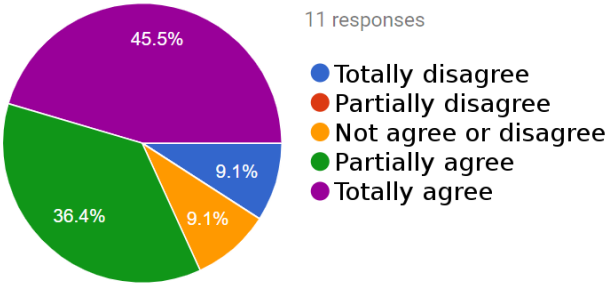


Figure 15: Analysis of Feature 8: Expressiveness

Figure 16 provides the answers to the analysis of Feature 9 - The approach supports the traceability of models. Nine from eleven respondents agree that this is a design feature associated with M4PIA. Trace is promoted through model-to-model transformations and are used for configuration management issues [37].

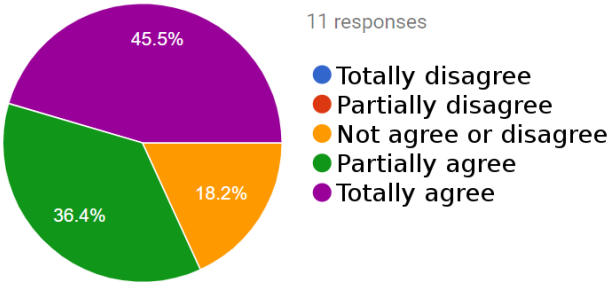


Figure 16: Analysis of Feature 9: Traceability

Figure 17 provides the answers to the analysis of Feature 10 - The approach supports reverse engineering from code to model. This statement was detected as an issue from the current version of M4PIA. first experimental trial. This feature is only implemented as a preliminary version of the infrastructure developed in ANTLr parser, not include in the provided Eclipse experimental package. In order to configure a reverse engineering into the Eclipse package, we are starting the development of reverse engineering using

MODISCO. So the answers for this feature analysis were also confused.

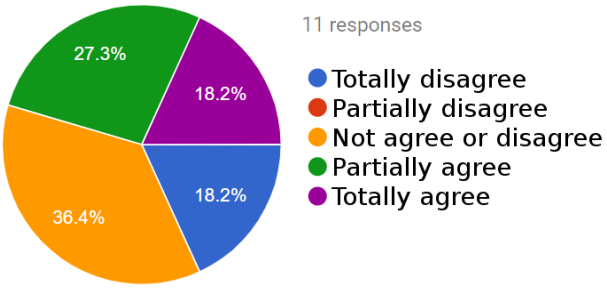


Figure 17: Analysis of Feature 10: Reverse engineering support.

Figure 18 provides the answers to the analysis of Feature 11 - The approach is heavily based on standards for building DSLs and model transformations. Nine from eleven respondents agree that this is an important feature associated with M4PIA. One of the answers reasoned that the adopted technologies is not a benefit.

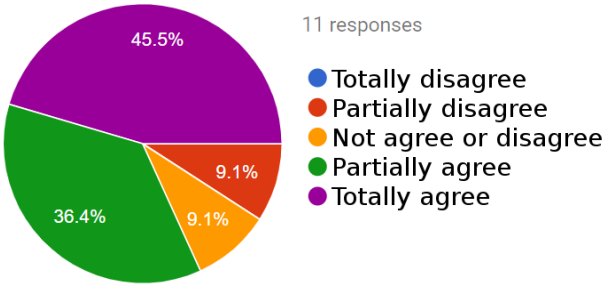


Figure 18: Analysis of Feature 11: Built on standards

Figure 19 provides the answers to the analysis of Feature 12 - There are no contradictions between the concepts of models in the M4PIA infrastructure. Here only five agree that the metamodels are well characterized, while five are still confused and suggested that they would have a stronger position with a better characterization of the context from the Petrochemical industry.

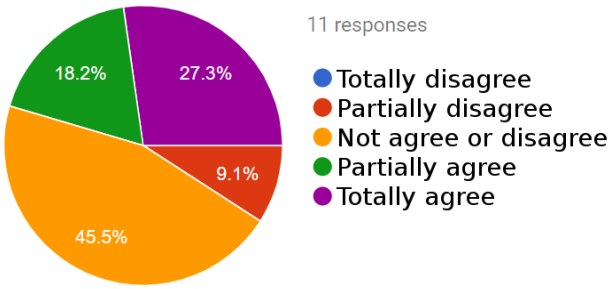


Figure 19: Analysis of Feature 12: Focused domain specific entities

Figure 20 provides the answers to the analysis of Feature 13 - The complete application code can be generated. six from eleven respondents agree that this is an important feature associated with M4PIA. Almost half the respondents or disagree or did not have sufficient information to conclude about that.

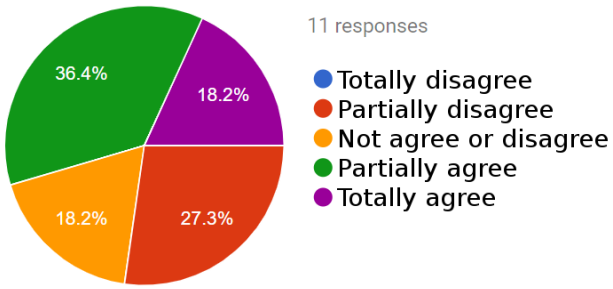


Figure 20: Analysis of Feature 13: Code generation

5.3 Final Remarks

Figure 21 presents the coverage analysis of the features after applying the metric. The highest possible score for each feature is 44. Figure 23 shows the dispersion bloxplot graph. Therefore, these graph demonstrates that M4PIA infrastructure is considered relevant for the stated problem.

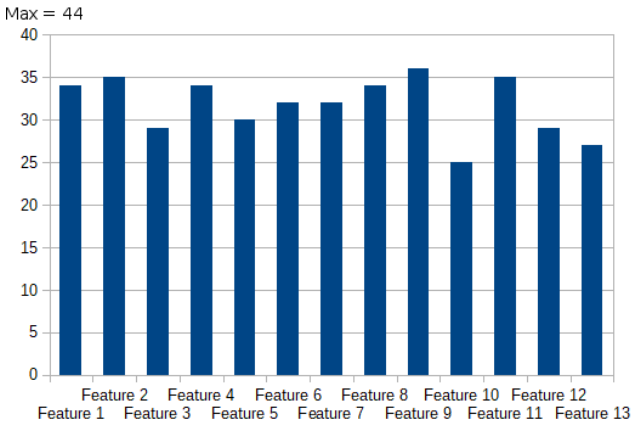


Figure 21: Coverage analysis of all the features

Finally, Table 2 presents the analysis of completeness of each analyzed feature. This data is also shown in Figure 22. The average of results from coverage is 72.73%, a good index suggesting the relevance to the MDE context. Thus, this suggests that we are on the way to achieve benefits reported by other MDE studies as well [14].

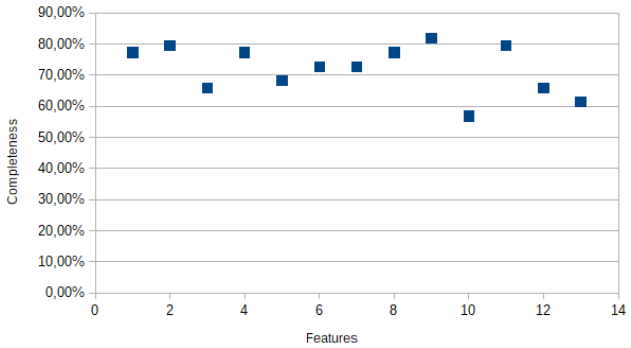


Figure 22: Completeness analysis of all the features.

5.4 Threats to Validity

As discussed by [32], studies of applicability using feature analysis experiment are usually target of the following

Table 2: Analysis of completeness after the feature analysis

	Feat 1	Feat 2	Feat 3	Feat 4	Feat 5	Feat 6	Feat 7	Feat 8	Feat 9	Feat 10	Feat 11	Feat 12	Feat 13
Uncovered	22.73%	20.45%	34.09%	22.73%	31.82%	27.27%	27.27%	22.73%	18.18%	43.18%	20.45%	34.09%	38.64%
Covered	77.27%	79.55%	65.91%	77.27%	68.18%	72.73%	72.73%	77.27%	81.82%	56.82%	79.55%	65.91%	61.36%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

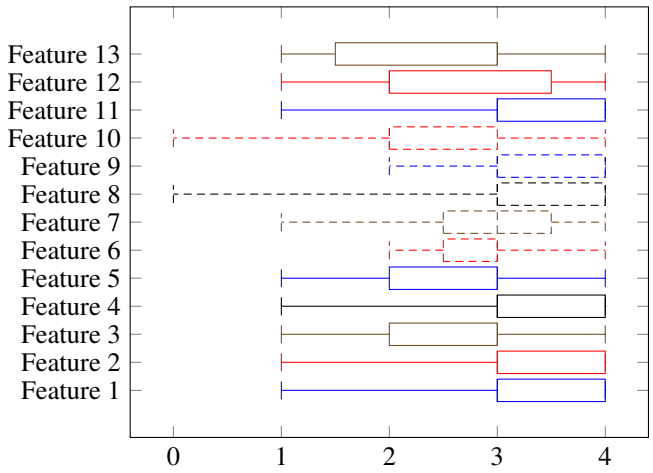


Figure 23: Boxplot of the analyzed features

threats to validity: 1) Benefits are difficult to quantify, 2) Benefits directly observable from the task output, 3) Relatively small learning time, and 4) Tool/method user population very varied. In the following, we present other threats conforming to [38].

5.4.1 Construct Validity

Feature analysis presents social threats to construct validity. In order to mitigate the Hypothesis guessing, before executing the experiment we: 1) presented the context of the problem; 2) presented the context of MDE tools; and 3) presented our goals, where we would evaluate their intuitiveness in using the M4PIA approach for each planned task along three different phases. These experimental tools were also important to mitigate the other two threats: Evaluation apprehension and Experimenter expectations. In order to mitigate the apprehension threat, an external executor for this quasi-experiment was used (the third author). This makes the study less biased, avoiding the interest in positive results, which could occur when the study is executed by the tool creators. The experiment executor also ensured the selection of subjects without any expectation about the evaluated tool, making clear their responsibility in feature analysis to rank M4PIA accordingly to the thirteen best MDE practices.

5.4.2 Internal Validity

This study was planned as a controlled experiment to be executed in three distinct phases: 1) the first one is to test the applicability of M4PIA; 2) the second one is the manual execution of the overall process, scoping design and codification tasks, performed without M4PIA; and 3) the last is the execution of the process with M4PIA. Along the execution of the first phase in the first trial, we observed that the internal validity threat called “mortality” would happen for the next phases due to the scarce available time from subjects. In order to mitigate this threat, and thus ensuring we would not loose subjects, we decided that participants could execute the experiment at home along three weeks. In this sense, they followed the video tutorials for execution of each requested task in a quasi-experiment rather than a fully controlled experiment, as initially planned. Therefore, in order to ensure the internal validity, the independent variable “time to accomplish each task” was not used.

5.4.3 External Validity

The results observed for feature analysis of M4PIA infrastructure are not generalized yet. For a such, we would need to repeat this study with, at least, three other subject profiles: 1) experts in physical plants presented on the development of simulation, control, and supervisory applications for the petrochemical industry; 2) experts both in physical plants and MDE; and 3) experts both in physical plants, MDE and software acquisition. Likewise, the threat “interaction of selection and treatment” was not mitigated, and will not gonna be while this experiment is not repeated with different subject profiles. Since our subjects are not experts in the problem domain, i.e. they are trainees in physical plants adopted by petrochemical industry, but they are experts in Software Engineering tool selection and software acquisition discipline, this study is also target of the threat “interaction of setting and treatment”. We mitigated this threat through nine video-tutorials, shared in the experimental package, which by the way are considered of high instructive quality.

5.4.4 Conclusion Validity

This study presents an associated threat, which is its low statistical power, as the evaluation presented was applied to only eleven participants. It happens that, for a feature analysis, the random heterogeneity of the subjects compensates the low statistical power, as it provides different perspectives

when evaluating the tool. Likewise, the subjects present two distinct evaluators profiles: half working on companies as professional software engineers and half characterized as students, where two of them are finishing the undergraduate school in Software Engineering. Thus, our conclusions are moderated and drawn accordingly to the statistical power provided.

6 Conclusions and Future Works

The current paper presented M4PIA, an MDE solution that facilitates the design and development of equipment classes for simulation, control, and operation of applications devoted for the petrochemical industry. It contains three metamodels, which is a rich material in respect to formalizing the representation of the elements that constitute petrochemical plants and the related control structure. These metamodels, together with the developed model transformation engines, serve to provide automatic code generation.

The conducted evaluation showed that the majority of the 13 features considered important in the scope of MDE are properly covered by the proposed M4PIA. Solutions to improve those features that were not highly ranked, like reverse engineering, are already under development. Moreover, as future work it is possible to state that M4PIA should evolve towards allowing even non-programmers (like plant engineers/designers) to describe plant structures and thereby providing additional help to develop the required applications. Therefore, a graphical modeling language might be adopted.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment Firstly we thank Cenpes/ Petrobras for the finance support that made this work possible. Thanks are also given to: (i) Tecgraf/PUC-Rio Institute team for the outstanding support in respect to MPA-related issues; (ii) Prof. Rafael de Pelegrini Soares (UFRGS) for his support in respect to the EMSO platform; (iii) the students from UNIPAMPA/Alegrete that did not measured efforts to perform the proposed evaluation protocol.

References

[1] International Electrotechnical Commission, "Iec 61131: Programmable controllers – part 3: Programming languages," standard, 2003.

[2] International Electrotechnical Commission, "Iec 61499: Function blocks standard," standard, 2005.

[3] The International Society of Automation, "Isa 106: Procedure automation for continuous process operations," standard, 2013.

[4] E. Satuf, S. F. Pinto, and B. Q. Dias, "Sistema automático de alinhamento para a plataforma de rebombeio autônomo PRA-1," in *V Congresso Rio Automação*, (Rio de Janeiro, RJ, Brasil), Inst. Brasil. de Petroleo, Gás e Biocombustíveis - IBP, 2009.

[5] R. B. France and J. M. Bieman, "Multi-view software evolution: A UML-based framework for evolving object-oriented software," in *ICSM*, pp. 386–395, 2001. https://10.1109/ICSM.2001.972751.

[6] M. Vöelter and I. Groher, "Handling variability in model transformations and generators," in *Companion to OOPSLA 2007*, pp. 1–8, ACM, 2007.

[7] P. Mohagheghi, *Evaluating Software Development Methodologies Based on their Practices and Promises*, pp. 14–35. Frontiers in Artificial Intelligence and Applications, 2008. https://doi.org/10.3233/978-1-58603-916-5-14.

[8] R. P. Soares and A. R. Secchi, "EMSO: A new environment for modelling, simulation and optimization," in *13th European Symposium on Computer Aided Process Engineering*, pp. 947–952, Elsevier Science Publishers, 2003. https://doi.org/10.1016/S1570-7946(03)80239-0.

[9] D. Batory, E. Latimer, and M. Azanza, "Teaching model driven engineering from a relational database perspective," pp. 121–137, 2013. https://doi.org/10.1007/978-3-642-41533-3_8.

[10] V. Aranega, A. Etien, and S. Mosser, "Using feature model to build model transformation chains," pp. 562–578, 2012. https://doi.org/10.1007/978-3-642-33666-9_36.

[11] R. S. P. Maciel, R. A. Gomes, A. P. F. Magalhães, B. C. da Silva, and J. P. B. Queiroz, "Supporting model-driven development using a process-centered software engineering environment," *Autom. Softw. Eng.*, vol. 20, no. 3, pp. 427–461, 2013. https://doi.org/10.1007/s10515-013-0124-0.

[12] P. Mohagheghi, M. A. Fernandez, J. A. Martell, M. Fritzsche, and W. Gilani, "Mde adoption in industry: Challenges and success criteria," vol. 5421, pp. 54–59, 2009. https://doi.org/10.1007/978-3-642-01648-6_6.

[13] J. Hutchinson, J. Whittle, M. Rouncefield, and S. Kristoffersen, "Empirical assessment of mde in industry," in *33rd International Conference on Software Engineering*, ICSE '11, pp. 471–480, 2011. https://doi.org/10.1145/1985793.1985858.

[14] P. Mohagheghi, W. Gilani, A. Stefanescu, M. A. Fernandez, B. Nordmoen, and M. Fritzsche, "Where does model-driven engineering help? experiences from three industrial cases.," *Software & Systems Modeling*, vol. 12, pp. 619–639, july 2013. https://doi.org/10.1007/s10270-011-0219-7.

[15] R. Capilla, J. Bosch, P. Trinidad, A. Ruiz-Cortés, and M. Hinchey, "An overview of dynamic software product line architectures and techniques: Observations from research and industry," *Journal of Systems and Software*, vol. 91, no. 0, pp. 3–23, 2014. https://doi.org/10.1016/j.jss.2013.12.038.

[16] R. Monteiro, G. Zimbrao, and J. Moreira de Souza, "Collaborative evolution process in mdarte: Exchanging solutions for information systems development among projects," in *Computer Supported Cooperative Work in Design (CSCWD), 2014 IEEE 18th International Conference on*, pp. 569–574, May 2014. https://doi.org/10.1109/CSCWD.2014.6846907.

[17] M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, and G. Reggio, "Relevance, benefits, and problems of software modelling and model driven techniques-a survey in the italian industry," *J. Syst. Softw.*, vol. 86, pp. 2110–2126, aug 2013. http://dx.doi.org/10.1016/j.jss.2013.03.084.

[18] G. Liebel, N. Marko, M. Tichy, A. Leitner, and J. Hansson, "Assessing the state-of-practice of model-based engineering in the embedded systems domain," pp. 166–182, 2014. https://doi.org/10.1007/978-3-319-11653-2_11.

[19] J. Whittle, J. Hutchinson, M. Rouncefield, B. Håkan, and H. Rogardt, "A taxonomy of tool-related issues affecting the adoption of model-driven engineering," *Software & Systems Modeling*, pp. 1–19, 2015. http://dx.doi.org/10.1016/10.1007/s10270-015-0487-8.

[20] P. Mohagheghi and V. Dehlen, *Where Is the Proof? - A Review of Experiences from Applying MDE in Industry*, pp. 432–443. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. https://doi.org/10.1007/978-3-540-69100-6_31.

[21] B. Selic, "On software platforms, their modelling with uml 2, and platform-independent design," in *8th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing*, ISORC'05, pp. 15–21, 2005. https://doi.org/10.1109/ISORC.2005.40.

[22] L. Rose, E. Guerra, J. Lara, A. Etien, D. Kolovos, and R. Kolovos, "Genericity for model management operations," *Software & Systems Modeling*, vol. 12, no. 1, pp. 201–219, 2013.

[23] L. B. Becker and C. E. Pereira, "SIMOO-RT-an object-oriented framework for the development of real-time industrial automation systems," *IEEE Transactions on Robotics and Automation*, vol. 18, pp. 421–430, Aug 2002. https://doi.org/10.1109/TRA.2002.802933.

[24] K. C. Thramboulidis, "Using UML in control and automation: a model driven approach," in *Industrial Informatics, 2004. INDIN '04. 2004 2nd IEEE International Conference on*, pp. 587–593, June 2004. https://10.1109/INDIN.2004.1417414.

[25] S. M. Becker, T. Haase, B. Westfechtel, and J. Wilhelms, "Integration tools supporting cooperative development processes in chemical engineering," in *Proc. of Integrated Design and Process Technology (IDPT-2002), Pasadena, California*, Citeseer, 2002.

[26] C. Vidal, C. Fernández-Sánchez, J. Díaz, and J. Pérez, "A model-driven engineering process for autonomic sensor-actuator networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 3, p. 684892, 2015. https://doi.org/10.1155/2015/684892.

[27] D. Hästbacka, T. Vepsäläinen, and S. Kuikka, "Model-driven development of industrial process control applications," *Journal of Systems and Software*, vol. 84, no. 7, pp. 1100–1113, 2011. https://doi.org/10.1016/j.jss.2011.01.063.

[28] V. Vyatkin, "Software engineering in industrial automation: State-of-the-art review," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1234–1249, 2013. https://doi.org/10.1109/TII.2013.2258165.

[29] D. Steinberg, F. Budinsky, E. Merks, and M. Paternostro, *EMF: Eclipse Modeling Framework*. Pearson Education, Inc., 2 ed., 2008.

[30] "Acceleo mda generator," 2013. Av. at `http://www.acceleo.org`.

[31] OMG, "Meta object facility version 2.0 query/view/transformation specification," tech. rep., OMG, 2005.

[32] B. Kitchenham, S. Linkman, and D. Law, "Desmet: a methodology for evaluating software engineering methods and tools," *Computing Control Engineering Journal*, vol. 8, pp. 120–126, June 1997.

[33] "An Evaluation of Compuware OptimalJ Professional Edition as an MDA Tool," tech. rep., Department of Computer Science - The University of York and King's College London, 2003.

[34] P. Mohagheghi, V. Dehlen, and T. Neple, "Definitions and approaches to model quality in model-based software development – A review of literature," *Information and Software Technology*, vol. 51, no. 12, pp. 1646–1669, 2009. https://doi.org/10.1016/j.infsof.2009.04.004.

[35] C. F. Lange and M. R. Chaudron, "Managing model quality in UML-based software development," in *Software Technology and Engineering Practice, 2005. 13th IEEE International Workshop on*, pp. 7–16, IEEE, 2005. https://doi.org/10.1109/STEP.2005.16.

[36] P. Andersson and M. Höst, "Uml and systemc - a comparison and mapping rules for automatic code generation," vol. 10, pp. 199–209, 2008. https://doi.org/10.1007/978-1-4020-8297-9_14.

[37] J. Vara, V. Bollati, A. Jiménez, and E. Marcos, "Dealing with traceability in the mdd of model transformations," *Transactions on Software Engineering*, vol. 40, no. 6, pp. 555–583, 2014. http://dx.doi.org/10.1109/TSE.2014.2316132.

[38] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer, 2012.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 183-189 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Quantitative Traffic Congestion Analysis Approach in Ahmedabad

Tsutomu Tsuboi*

General Manager, Global Business Development Office, Nagoya Electric Works Co., Ltd., 490-1294, Japan

ARTICLE INFO

Article history:
Received:13 April, 2019
Accepted:24 May, 2019
Online: 10 June, 2019

Keywords:
Traffic Flow
Traffic Congestion
Social Loss

ABSTRACT

This study is the extension of the previous study about "Traffic Service Quantitative Analysis Method under Developing Country" in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). In the previous study, it is introduced how to make quantitative calculation for traffic congestion by traffic parameters and its characteristics curve such as traffic volume (q) to inverse of vehicle average speed (=1/v). In order to identify the traffic congestion condition, it is focused on vehicle speed ratio which is average speed (v_{ave}) to its free speed (v_f). And the threshold level is 2/3 (=v_{ave}/v_f). This 2/3 value comes from Viscous fluids model between parallel flat plates by using similarity of the viscous fluids flow and the traffic flow in India which is introduced at the CODATU XVII and UMI Conference 2017. This threshold value definition needs more traffic theory back up because its similarity between viscous fluids flow and traffic flow comes from the basis of traffic flow measurement results. In this extension study, it focuses on the occupancy of one of typical traffic parameter for traffic congestion. When it is compared between the traffic occupancy measurement data and Speed Ratio, the Speed Ratio 2/3 is the level of the occupancy 30%, which is used as traffic congestion condition. According to daily based traffic volume and average vehicle speed, the congestion condition is occurred by not only those traffic parameter, it is also considered about time zone traffic condition and its actual traffic condition. One of measurement point is always congested in the early evening even if its traffic volume is smaller than the morning traffic congestion. Therefore it is important to analyze traffic condition by time zone. As the result, it clarifies the relationship between Speed Ratio and occupancy. As the result, there are two types of traffic congestion in Ahmedabad city traffic.

1. Introduction

This study focuses on traffic flow and congestion analysis especially in the under developing country (India). In the previous study, it is proposed the traffic service quantitative traffic congestion analysis method in the International Conference on Advances in Computing, Communications and Informatics (ICACCI) [1]. From the actual measurement traffic data study during more than one month in Ahmedabad city of Gujarat State in India, the major traffic parameters are extracted from the traffic fundamental diagram such as traffic density (k) to vehicle average speed (v) and traffic density (k) to traffic volume (q), which introduced at the International conference CODATU 2017 [2].

As for obtaining the traffic parameters and analysis, the detail is explained in the next section. The previous study of CODATU introduces the unique traffic characteristics which is useful for using Envelopment Observation method because there is clear boundary in those characteristic and it is appropriate way to define the traffic parameters from their unique measurement data. The another phenomenon from the measurement data is vehicle speed ratio which is the ratio of average vehicle speed (v_{ave}) to free speed (v_f). The value of this speed ratio (SR) is valid for understanding congestion level of each road. When S.R. becomes lower than 0.65, the condition of road becomes congested. Author found similarity of this phenomena between the Indian traffic flow model and viscous fluids model between parallel flat plates.

In this extension of the research, more detail analysis is added between SR and occupancy of traffic parameter which is used as traffic congestion level judgement from traffic engineering. In addition, the relationship between parameter occupancy and traffic density is described in order to understand traffic congestion condition. As for the traffic congestion analysis, it also describes

* Tsutomu Tsuboi, 29-1 Mentoku Shinoda Ama, Aichi, +81-902738-3496, t_tsuboi@nagoya-denki.co.jp.

the time zone base analysis in this paper. From this time zone base analysis, we found there are two patterns for Ahmedabad traffic congestion condition. One is heavy traffic density congestion which follows the traffic flow theory and the other is unheavy traffic density congestion which is unique in India. Therefore it is important for considering what is the traffic congestion definition. when the traffic information is provided from the traffic information sign board so called VMS or Various Message Sign on the road.

2. Related Studies

2.1. Traffic Flow Theory and Measurement Analysis of Ahmedabad city

In the previous study for Ahmedabad traffic flow analysis at CODATU conference, the traffic fundamental diagram is introduced such as *k-v* curve (traffic density (*k*) to average vehicle speed (*v*)), *k-q* curve (traffic density (*k*) to traffic volume (*q*)), and *q-v* curve (traffic volume (*q*) to vehicle speed (*v*)). This analysis has been continued since Ahmedabad ITS or Intelligent Transport Systems business project from October 2014. The project is providing Ahmedabad city to provide its traffic condition for the drivers through VMS. The ITS system consist of four VMSs and 14 traffic monitoring cameras. The traffic flow condition is captured by the traffic monitoring cameras and its result of analysis is shown at the VMS for considering change route or using other transportation choice. The Ahmedabad ITS system configuration is shown Figure 1.

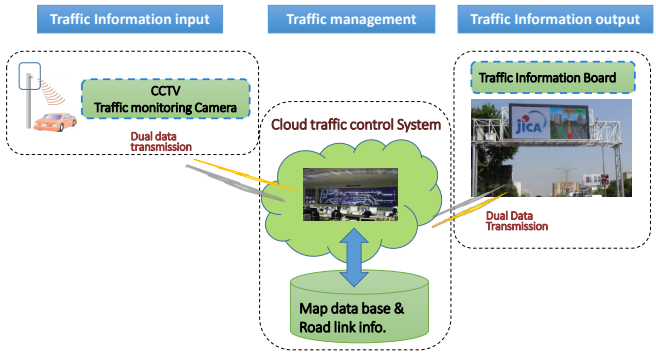


Figure 1: Ahmedabad ITS system configuration

In the CODAU study, authors found the unique characteristics of Ahmedabad traffic flow measurement data. There is clear boundary in each traffic fundamental diagram. The *k-v* curve and *k-q* curve at Camera#1 are shown in Figure 2.

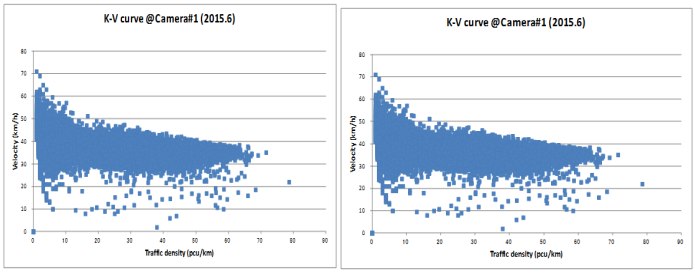


Figure 2: Traffic Flow Characteristics at Camera#1

From Figure 2, there is clear boundary in each characteristics curve by envelop observation. The other characteristics at other location are similar results which are omitted in this paper. When these boundary are taken as traffic flow characterizes, it is able to adjust traffic flow characteristics curve from traffic flow equation to the envelop line. The Figure 3 shows the example of adjustment result between traffic flow theoretical curve from its equation and the measurement data at Camera#1.

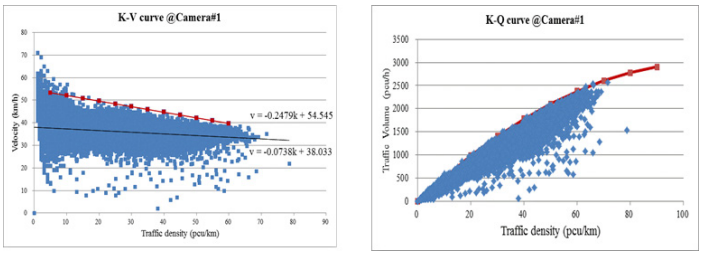


Figure 3: Traffic Flow Characteristics at Camera#1

In terms of traffic equation form the theory, traffic density (*k*) to traffic speed (*v*) equation is given by (1) and traffic density (*k*) to traffic volume (*q*) equation is given by (2). The equation (1) is so called Greenshields [3] curve.

$$v = v_f\left(1 - \frac{k}{k_j}\right) \tag{1}$$

v_f: free speed

k_j: jam traffic density

$$q = -\frac{v_f}{k_j}\left(k - \frac{k_j}{2}\right)^2 + \frac{v_f k_j}{4} \tag{2}$$

The theoretical curve illustration is shown in Figure 4.

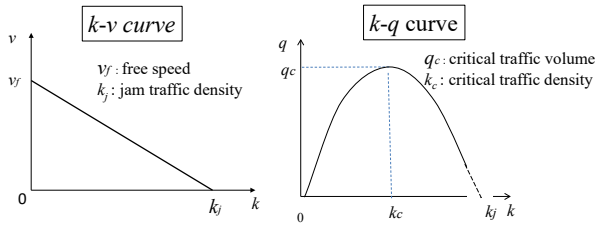


Figure 4: Theoretical Traffic Flow Curve

In the advanced country traffic flow, the actual measurement plots including Japan [4] are aligned with theoretical curve and or near boundary in Figure 4. Therefore the traffic characteristics curve from measurement in Ahmedabad is quite different from those of the advanced countries (refer to Figure 2).

2.2. Quantitative Traffic Congestion Study in Ahmedabad

The traffic congestion problem in developing countries become more serious situation these days, especially in India. Therefore it is crucial to establish for analysis method about the traffic congestion quantitatively. In the International Conference on Advances in Computing, Communications and Informatics or ICACCI, T.Tsuboi proposes quantitative traffic congestion analysis based on the actual traffic measurement in Ahmedabad. In this paper, it is introduced the traffic congestion by vehicle speed ratio (SR) which is the ratio of the average vehicle speed (v_{ave}) to the free speed (v_f). In ICACCI paper, it is defined that the

threshold level of traffic congestion is 0.65. The comparison SR at each traffic camera location is shown in Figure 5. According to Figure 5, the Camera#2 is most small value. And the actual daily traffic flow (traffic volume and average vehicle speed) at Camera#2 in June 2015 is shown in Figure 6 (a) and (b). From Figure 6(b), the average vehicle speed is lower than 20km/h, which shows the traffic congestion occurred [5].

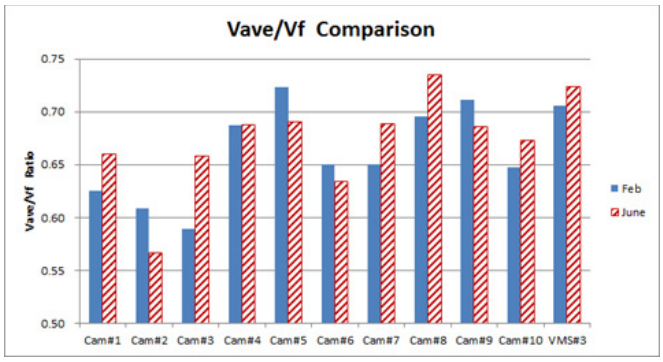
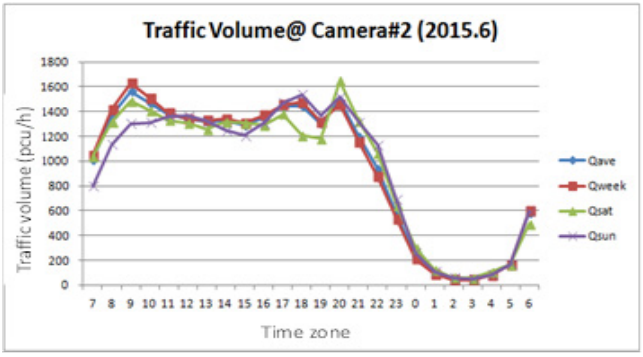
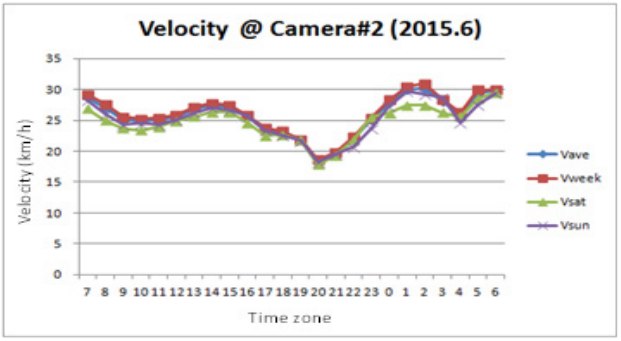


Figure 5: Traffic Speed Ratio Comparison in Ahmedabad



(a) Traffic Volume Time Zone condition at Camera#2



(b) Traffic Speed Time Zone condition at Camera#2

Figure 6: Daily Traffic Flow condition at Camera#2

From Figure 6 (b), it is a clear difference that Camera #2 vehicle velocity ratio is lower than other locations. According to traffic flow theory, most of measurement data of q–k curve is located at shadow area in Figure 7. And free traffic flow area is bottom part of q – k curve; jam traffic flow area is top of q – k curve. But our measurement data of q – k curve in Camera # 2 is located inside q – k curve (refer to Figure 7). The speed ratio of inside curve area is lower than 0.65 which means that its traffic condition becomes congested.

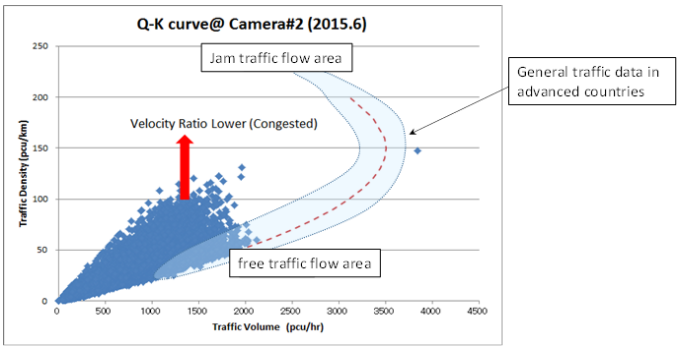


Figure 7: q – k Curve at Camera#2

In order to identify that inside curve area is under congested condition, we divide between two areas. One area has speed ratio equal and less than 065 and the other area has speed ratio more than 0.65. Figure 8 shows those two areas. According to Figure 8, the speed ratio equal and lower than 0.65 (≈ 2/3) is inside envelope line area. According to the previous traffic flow analysis, the threshold between free flow and congested flow occurs at the average vehicle velocity (v_{ave}) ≤ 2(v_f) /3 [2]. Therefore the traffic flow capacity is able to obtain by our envelope data analysis and it is able to be described by the traffic flow equation. This is the first time to show emerging countries traffic analysis and brings basic traffic flow parameters by boundary traffic analysis. And vehicle velocity ratio of average speed by free speed is good reference of explaining traffic jam condition; the critical value of speed ratio is 0.65 in this paper.

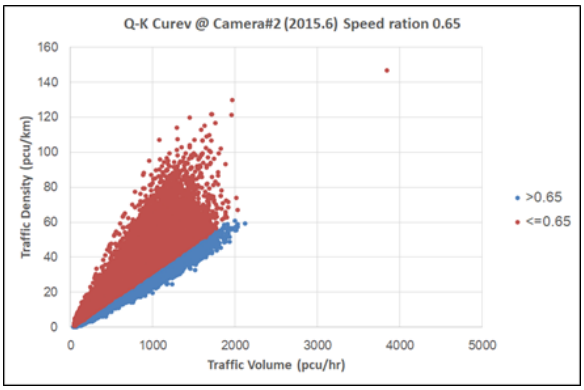


Figure 8: q – k measurement data by vehicle Speed Ratio at Camera#2

In terms of traffic congestion analysis, the congestion social loss analysis theory [6] is used. From the traffic flow theory, traffic volume (q) is mathematical product by vehicle velocity (v) multiplying to traffic density (k) shown in (3).

$$q = v \times k \tag{3}$$

Since the vehicle velocity (v) is inverse of travel time (t), then the vehicle velocity is described as $v = 1/t$. Therefore traffic volume (q) is a function of travel time (t). Assuming that the traffic user's opportunity n per unit time is constant, travel time (t) is proportional to the cost of traffic congestion. In another word, the relationship between traffic volume (q) and travel time (t) means the traffic supply curve of the relationship between traffic

service supply and cost. In order to analyze traffic volume (q), differentiating e (3) by (t) provides (4).

$$\frac{dq}{dt} = -\left(v\frac{dk}{dv} + k\right)\frac{1}{t^2} \tag{4}$$

When vehicle velocity (v) becomes large volume (travel time becomes small value), traffic density (k) becomes small value and then $dk/dv < 0$. Therefore he value in parenthesis of (4) is negative and then dq/dt becomes positive value. This relationship is illustrated in Fig.8. The supply curve line rises to the right from travel time ($1/v_f$) to ($1/v_c$), which means private cost curve for transportation. And when (v) becomes small value (travel time (t) becomes large value), the value in parenthesis of (4) becomes positive and the supply curve drops to the right, which corresponds to private cost curve drops to the right.

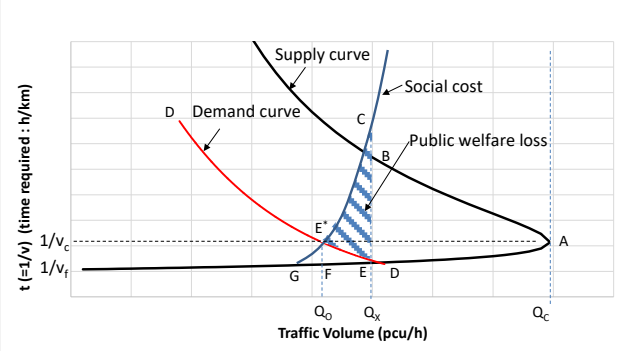


Figure 8: Road Traffic Service Market

In Figure 8, the point A is called hyper congestion condition at critical traffic volume (q_c) when traffic volume becomes larger beyond point A, travel time takes longer because of traffic congestion. There are two travel time (t) except point A. For instance, there are point E and point B at a certain traffic volume (Q_x). In case of point B, the travel time takes longer than that of at point E. If traffic demand curve D-D is given, the point E is cross point between demand curve D-D and supply curve which provides balance condition between traffic demand and supply. If social cost curve is given, the point E* becomes balance point between traffic demand and social cost. Then area CE*E provides traffic service cost loss caused by traffic congestion because infrastructure should cover at the level of traffic volume (Q_x) at the point E and social cost rises at the point C where its traffic volume (Q_x) is same as that of at the point E. Therefore area CE*E is defined as the public welfare loss which means traffic congestion social loss.

In order to calculate the public welfare loss from measurement data, it is necessary to define condition for each curve of Figure 8 from the actual measurement analyzed data. The condition of traffic congestion social loss is as follows:

- Supply curve is $q-v$ curve form envelopment observation.
- The critical speed (v_c) is used at SR=0.65 because this point is congestion threshold.
- Demand curve D-D is used from actual measurement traffic data.
- The social cost curve is unknown under this condition. Therefore BE* liner curve is taken.
- The congestion social loss is defined as area BE*E instead of area CE*E.

The Figure 9 shows the congestion social loss calculation example at Camera#2 with this quantitative social loss analysis. The result of Social loss analysis at all locations shows in Table 1. According Table 1, it is clear relationship between Speed Ratio (SR) and Social loss value.

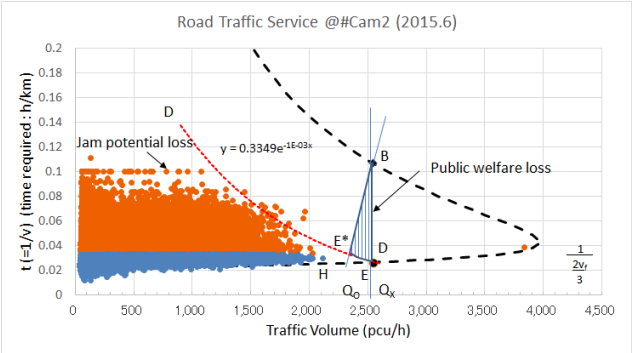


Figure 9: Road Traffic Service Market at Camera#2

Table 1: Social Loss and Speed Ratio

Location	SR(v_{ave}/v_f)	Social Loss
Camera#1	0.66	3.3
Camera#2	0.57	8.1
Camera#3	0.66	3.1
Camera#4	0.69	3.2
Camera#5	0.69	2.0
Camera#6	0.63	2.2
Camera#7	0.69	1.3
Camera#8	0.73	0.2
Camera#9	0.69	1.6
Camera#10	0.67	2.4
VMS#3	0.72	0.7

3. Traffic Congestion Analysis

In the previous section, the envelopment observation method is used for getting traffic flow equation from its measurement. Based on the each traffic flow equation, new quantitative social loss of traffic congestion analysis is introduced. In terms of traffic congestion judgement, the average vehicle speed is used. This average vehicle speed is different at each road and its condition. Therefore it is necessary to have some relative vehicle speed compared with normal driving speed which is taken as free speed in this paper. As for traffic congestion judgement, the average speed is not only parameter but also the traffic occupancy for example.

3.1. Traffic Occupancy

The traffic occupancy is used for one of parameters which shows the level of traffic congestion condition. There are two traffic occupancies in the theory, time occupancy and space occupancy. In real traffic management, time occupancy is often used for the freeway [7]. In this section, time occupancy is used.

The time occupancy is defined the percentage ratio between total measurement time (t) of the vehicles to a certain block of road section under certain time frame. The formula of time occupancy is provided by (5).

$$\mathrm{OC} = \frac{1}{T}\sum_{i} t_i \times 100\ (\%) \tag{5}$$

where T is time of measurement, (t_i) is detected time of vehicle (i)[8].

When number of existing vehicle a certain section is (N), average length of vehicle is ($\bar{l}$) in (6) is given.

$$OC = 100\frac{q}{v}\bar{l} = 100k\bar{l} \tag{6}$$

Therefore occupancy (OC) is proportional to traffic density (k) and traffic volume (q).

From the one month measurement data of Camera#1 and #2 in June 2015, traffic density (k) to occupancy (OC) relationship are shown in Figure 10. According to Figure 10, the relationship between (k) and (OC) is proportional.

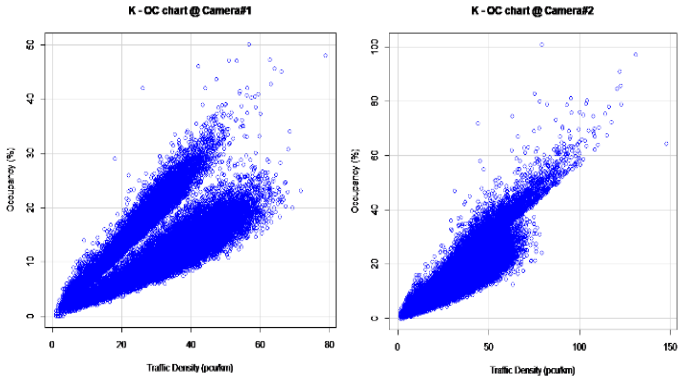
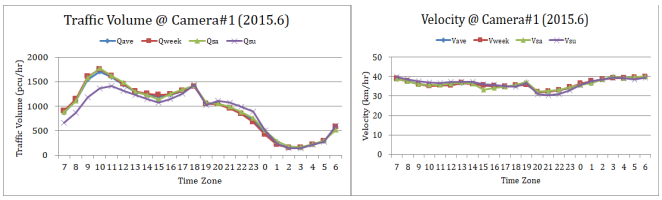


Figure 10: Example of k – q curve at Camera#1 and Camera#2

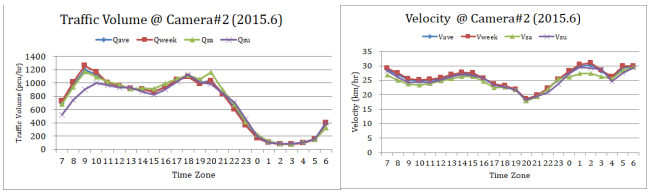
From Figure 10, it seems that the measurement data is plotted into two groups. In the section 3.3, time zone analysis is introduced because traffic congestion is dependent with time of a day (refer to Figure 6) [9].

3.2. Daily Traffic Condition

In this section, a daily traffic changes of traffic volume and velocity is described. In Figure 11, there are two cases of daily based traffic condition at Camera#1 and #2. The graph shows one month traffic volume changes from 7:00 am to next day 6:00 am and average vehicle velocity changes. There are two peaks of traffic volume in each case. One peak occurs between 7:00 am to 10:00 am and the second peak occurs between 18:00 to 20:00. When vehicle average speed is concerned, there is not so much changes of velocity in Camera#1 but different speed at Camera#2. In case of Camera#2, there is a drop during the second peak volume of traffic. And each graph shows classification by four cases—total one month average, weekday, Saturday, and Sunday. According to Figure 11, there is traffic jam in Camera#2 between 18:00 to 20:00 and the average vehicle speed goes down to less than 20 km/h which is confirmed the speed under heavy traffic congestion through the recorded traffic monitoring video data.



(a) Daily Traffic Condition at Camera #1



(b) Daily Traffic Condition at Camera#2

Figure 11: Daily Traffic Condition at Camera#1 and Camera #2

3.3. Occupancy and Traffic Density

As described in section 3.1, time zoning is adapted into Figure 10. In this section, six time zone is defined as shown in Table 2

Table 2: Social Loss and Speed Ratio

Zone Name	Time Zone
T1	7:00 — 10:59
T2	11:00 — 14:59
T3	15:00 — 18:59
T4	19:00 — 22:59
T5	23:00 — 2:59
T6	3:00 — 6:59

There are 6 time zones such as T1 is defined time frame from 7:00 am to 10:59 am which is most congested time frame as we see in Figure 11. And T4 is defined time frame from 19:00 to 22:59 which is the second congested time frame. In case of using this Time Zone classification for k - OC curve, Time Zone based k - OC curve is shown in Figure 12. In this paper, we focus on the characteristics at Camera#1 and #2.

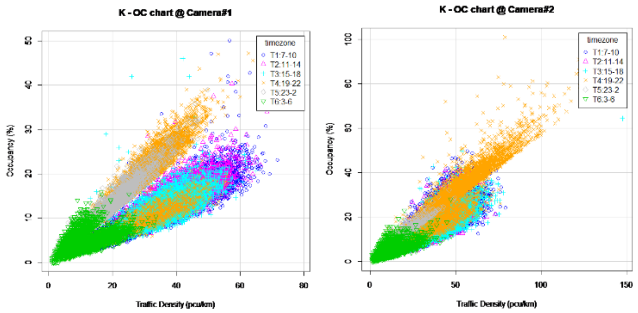


Figure 12: Time Zone based k – OC curve at Camera#1 and Camera#2

When the Time Zone based analysis, it is clear two groups data plot in Camera#1, where there is no traffic congestion condition compared with that of Camera#2. From the result of Camera#1 analysis, the upper group in k – OC curve comes from majority of

T4 (19:00-22:59). The lower group in k –OC curve comes from T1, T2, and T3. In another words, upper part of T4 is relative lower density compared with lower part of T4. This means there are two types of traffic congestion condition. Therefore one type of traffic congestion is defined as low density traffic congestion (herein after it is called as Low density Congestion) and the other one is higher traffic density congestion (herein after it is called as High density Congestion).

The next section describes traffic volume to occupancy and average vehicle speed to occupancy in order to understand the traffic congestion condition with other traffic parameters.

3.4. Occupancy and Traffic Volume and Speed

In Figure 11, it is shown about the daily traffic condition by traffic volume and average vehicle speed. In this section, it adapts Time Zone basis analysis for q-OC curve and v-OC curve in order to understand relationship between traffic volume to occupancy and average vehicle speed to occupancy.

The Traffic volume (q) to Occupancy (OC) Time Zone basis relationship at Camera#1 and #2is shown in Figure 13.

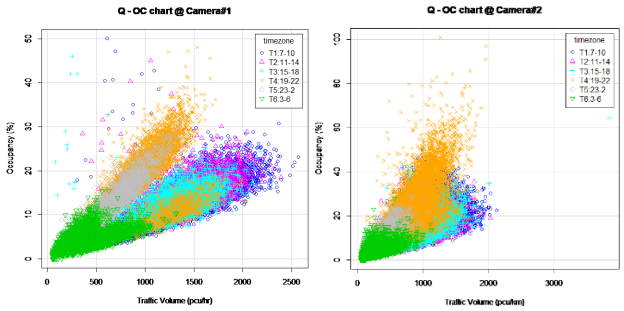


Figure 13: Time Zone based q – OC curve at Camera#1 and Camera#2

From the q – OC curve at Camera#1 in Figure 13, it shows more clear about difference between Low density congestion area and High density congestion area. The Low density congestion occurs in T4. It looks same trend of characteristics at Camera#2 but there is no clear two groups separation between Low density congestion and High density congestion.

The next Figure 14 shows average vehicle speed (v) to Occupancy (OC) relationship.

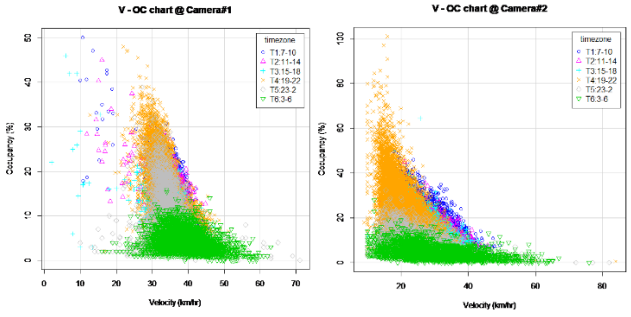


Figure 14: Time Zone based v – OC curve at Camera#1 and Camera#2

From Figure 14, it is clear that traffic congestion at lower than 30km/h occurs above 30% Occupancy condition.

Here is a question why Low density congestion occurs. It is considered at the next discussion section.

4. Discussion of Low density congestion

According to all traffic measurement location in Ahmedabad, the location Camera#2 is the most congested condition location (refer to Figure 5). And traffic condition during T4 time zone becomes most congested as described in previous section. The traffic volume at T4 of Camera#2 is the second peak of traffic volume in a daily condition as shown in Figure 6 (b) and Figure 11 (b). The highest peak of traffic volume occurs at T1 and T2. Therefore it is possible to define that High density congestion occurs at T1 and T2 and Low density congestion occurs at T4.

There is one more traffic parameter which is called Headway. The Headway is defined the distance between the front of driving vehicle group and the front of the second driving vehicle group (refer to Figure 15).

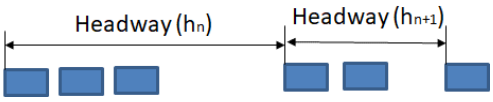


Figure 15: Headway definition

It is relative true the relationship between the small value of Headway and traffic congestion because its traffic density is large. The Headway ($\bar{h}$) and the traffic density relationship is described (7) from the theory.

$$\bar{h} = \frac{1}{k} = \frac{v}{q} \tag{7}$$

where $\bar{h}$ is average of headway. From (7) and (6), (8) is given.

$$\bar{h} = \frac{1}{k} = \frac{100\bar{l}}{OC} \tag{8}$$

From (8), the headway ($\bar{h}$) is inverse proportional to occupancy (OC) with the average of vehicle length ($\bar{l}$). The average of vehicle length is difficult to get from real measurement. Therefore the relationship between the average headway ($\bar{h}$) to occupancy (OC) is not exactly inverse relation. The Figure 16 shows $\bar{h}$ - OC curve at Camera#1 and #2. The vertical axis is legalism for clear display.

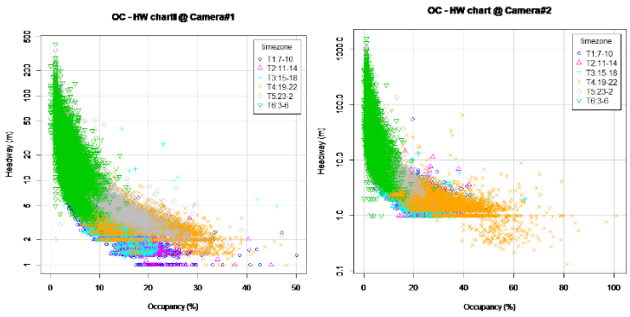


Figure 16: Time Zone based

From Figure 16, the traffic congestion occurs at T4 Time Zone and its Occupancy is above 30%. In this analysis, it is difficult to clarify between Low density congestion and High density congestion. The difference of Low density congestion and High density congestion becomes clear from traffic volume to occupancy or traffic density to occupancy relationship.

When it check the Time Zone of Low density, T4 and time period is from 18:00 to 22:59. After discussion with local traffic

police, they observe that the Low density congestion may be caused by limitation of parking space in the city. In the evening after work, the residents of Ahmedabad stop by restaurants and or shopping place and they make parking their vehicle at not permitted place such as along the street. Therefore this parking issues have potential for creating Low density congestion. If so, there is a solution for Low density congestion by preparing appropriate parking space for restaurant and shopping. The reason why the Low density congestion is detected by the traffic monitoring cameras is assumed that the collecting data is based on number of driving vehicles, not parking vehicles. For the traffic monitoring cameras, they take the vehicles as a part of side roads, which means parking vehicles make roads just narrow path and is unable to count the vehicles number as their traffic density. This situation tells that it is important to judge the traffic monitoring data by not only numerical data from the monitoring camera but also actual visual data.

In terms of quantitative social loss analysis method, the demand curve in Figure 9 is used from the measurement data, which is described as the condition of the quantitative social loss analysis method. When Figure 9 is mapped by time zone, it becomes Figure 17.

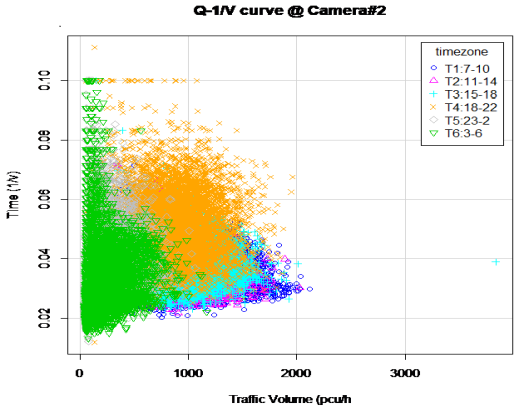


Figure 17: Time Zone based q-1/v curve at Camera#2

From Figure 17, Low density congestion area at T4 is located inside the demand curve. Therefore it is able to confirm proposed quantitative social loss analysis is valid.

5. Summary

In this study, there are three features for Indian traffic flow analysis.

- Feature 1: Introduce envelop observation method based on traffic flow data to provide traffic flow fundamental characteristics by traffic flow equation. It is able to get traffic flow parameters from this equation i.e. jam density (k_j), free speed (v_f).
- Feature 2: Propose quantitative social loss of traffic congestion method by only using traffic measurement data. As for critical speed (v_c), it is defined by Speed Ratio (SR)—average speed (v_{ave}) to free speed (v_f)—0.65 or 2/3 of (v_f), which is based on traffic flow analysis in India.
- Feature 3: There are two types traffic congestion condition in India, Low density congestion and High density congestion. It is presumed that Low density congestion mainly is caused by parking vehicle problem. Therefor Indian Traffic congestion occurs not only by large number of vehicles but also other infrastructure issues.

All this study analysis is based on one month of real traffic flow data in Ahmedabad city. It is necessary to have more different timing, location i.e. place, city, country. It is also important to have traffic video data in order to evaluate traffic congestion condition and understand congestion reason. The above all research items are future work.

Conflict of Interest

There is no conflict of interest in this study.

Acknowledgment

This study also underwent the program ID:JPMJSA1606 of the International Science and Technology Cooperation Program (SATREPS) challenges for global challenges in 2016.

Special appreciation to Mr. Kikuchi.C and Mr. Mallesh.B of Zero-Sum ITS India for providing traffic data in Ahmedabad.

References

[1] T. Tsuboi, “Traffic Service Quantitative Analysis Method under Developing Country” IEEE Xplore, 2018. DOI;10.1109/ICACCI.2018.8554495.

[2] T. Tsuboi, N. Yoshikawa, “Traffic Flow Analysis in Emerging Country (India)” in the CODATU XVII and UMI Conference, Hyderabad, India, 2017.

[3] B.D. Greenshields, A Study of Traffic Capaci, Proc.H.R.B., 14, pp.448-477, 1935.

[4] H. Kubota, T. Oguchi, K. Takahashi.” Traffic Engineering•Traffic Planning,” Rikoh Tosho Book, pp.24-25, April.2010.

[5] F. He, X. Yan, Y.L.L.Ma, “A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index”, Procedia Engineering 137 pp. 425 – 433, 2016.

[6] T. Kuroda, et al: Urban and Regional Economics, Yuhikaku Book, pp. 265-268, 2008M. H. Nehrir, C. Wang, Modeling and Control of Fuel Cells: Distributed Generation Applications, Wiley-IEEE Press, 2009.

[7] F.B. Lin, C.W. Sy, H.H. Huang, “Uniform Criteria for Level of Service analysis of Freeways.” Journal of Transportation Engineering, ASCE, 122(2), pp.123-130, 1996

[8] McMaster University, Department of Civil Engineering and Department of Geography, “TRAFFIC STREAM CHARACTERISTICS BY FRED L. HALL”, p.10

< https://www.fhwa.dot.gov/publications/research/operations/tft/chap2.pdf>

[9] T.Tsuboi, “Traffic Congestion Visualization by Traffic Parameters in India”, International Conference on Innovative Computing and Communication (ICICC), 2019

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 186-192 (2019)

www.astesj.com

Special Issue on Clean and Green Energy

ASTESJ

ISSN: 2415-6698

The Impact of Using Upgraded Biogas on Generator Performance

Wilson Babu Musinguzi*, Ibrahim Luqman Mpungu

Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Kyambogo University, P.O. Box 1 Kampala, Uganda

ARTICLE INFO

Article history:
Received: 27 April, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Upgraded Biogas
Thermal Characterization
Efficiency

ABSTRACT

This research focussed on assessing the impact of using upgraded biogas from chicken waste on the performance of a gasoline generator. In the study, the characteristics of raw biogas were determined and a biogas upgrading device was developed before assessing the generator performance on upgraded biogas. A Geotech GA 2000 plus gas analyser was used for biogas characterization and a slightly modified Elemax SH 2900 gasoline generator (5 HP) was used. The findings revealed that biogas from chicken waste comprised of 57% CH_4 and 41% CO_2 with small amounts of H_2S, O_2 and NH_4. During biogas upgrading, the highest solvent to biogas flow rate used was 0.80. NaOH excelled with 93.0% CO_2 reduction followed by KOH at 82.5% and lastly water at 63.0%. Using upgraded biogas improved the generator's brake thermal efficiency from 14.2% for raw biogas to 17.6% for upgraded biogas. A locally fabricated scrubbing device significantly upgraded the biogas quality.

1. Introduction

Biogas is a renewable energy gaseous fuel produced from anaerobic digestion of biological wastes such as cattle dung, vegetable waste, sheep, poultry droppings, municipal solid waste, industrial waste water, land fill, etc. in the absence of oxygen and the presence of anaerobic microorganisms [1]. The process is achieved at elevated temperature range of 30 to 60°C. The produced biogas comprises of mainly methane (CH_4) and carbon dioxide (CO_2) with smaller amounts of hydrogen sulphide (H_2S) and ammonia (NH_4). Biogas is an excellent fuel for a large number of applications including production of heat and steam, electricity production, vehicle fuel and production of chemicals. In Uganda, the most common applications for biogas are cooking and lighting. Conventional gas burners and gas lamps can be modified to suit biogas by changing the air to gas ratio.

In this study, Okweru Poultry Farm was used as a case study. Apparently, biogas from chicken droppings was being used for lighting the chicken houses, heating the brooder and cooking by the farm workers. The researchers explored alternative uses for the biogas produced with focus on electricity generation using a gasoline generator.

Although biogas can be used as a fuel, it has an extremely low energy density on volume basis. This is because of the presence of CO_2 at higher quantities that reduce its calorific value, flame velocity and flammability range compared with other fuels [2]. The high CO_2 levels in biogas can be lowered through biogas upgrading. Global upgrading technologies include: absorption, membrane, absorption and cryogenic technology. Applying these techniques to small scale biogas plants presents complexity and high costs of investment. This leads to utilization of biogas without upgrading for basic applications in developing countries.

The utilization of biogas in stationary engines is a typical case where contaminants have to be removed for better performance. In dual fuel operation, increasing the ratio of biogas increases the brake specific fuel consumption while decreasing brake thermal efficiency and the brake engine power of a Compression Ignition (CI) engine [3]. It was reported in [4] that an improved thermal efficiency from 26.2% to 30.4%, when there is 21% reduction of CO_2 in biogas used in a Spark Ignition (SI) operation while a 15% CO_2 decrease in the biogas increased the thermal efficiency to 22% for dual fuel operations. However, there are other factors that influence the performance of a gas engine.

Various factors responsible for the low thermal efficiency of engines run on biogas were reported in [2]. These include; the long pilot ignition delay and high self-ignition temperature of biogas. These factors delay the fuel combustion process while the low burning velocity due to presence CO_2 leads to incomplete combustion. This increases the brake specific energy consumption and exhaust gas temperature of dual fuel modes.

*Corresponding Author: Wilson Babu Musinguzi, Faculty of Engineering Busitema University, P.O. Box 236 Tororo, Uganda, Tel: +256774076181, Email: Wilson.Musinguzi@gmail.com

This research aimed at ascertaining the performance improvement of a generator when upgraded biogas was used. This was carried out by analysing the characteristics of raw biogas and later developed a biogas upgrading unit. The performance of a generator was done using Model Elemax SH 2900 gasoline generator. Biogas upgrading reduces the levels of CO_2 which improves the burning velocity and calorific value of the gas among others. Through several experiments, the generator brake power output, brake specific fuel consumption and brake thermal efficiency were tested at different load profiles.

2. Materials and Methods

The experimental research included the determination of the biogas characteristics, biogas upgrading device design, fabrication and testing. Also experiments to assess generator performance were carried out for both raw and upgraded biogas. This yielded the primary data utilized in the analysis.

2.1. Characteristics of Raw Biogas

In line with biogas characterisation, a Geotech GA 2000 plus Gas Analyser was used to measure the percentage composition of the biogas constituents. The biogas pressure was measured in kPa using a pressure gauge while a thermocouple was employed for temperature measurement in degrees Celsius. The gas flow rate was measured in litres/minute using an Omega HHF11A Air flow meter.

2.2 Designing the Biogas Upgrading Device

The factors that influence the upgrading process include the dimension of the scrubbing tower, gas pressure, raw biogas composition, water flow rate and purity of the water in use. According to [5], it is possible to produce high quality CH_4 enriched gas from biogas by chemical absorption, where a packed bed column and bubble column are used to provide liquid to gas contact.

Therefore, a low-cost upgrading system was designed to receive raw biogas from the bottom and force it through packed marbles to reduce its CO_2 content. According to [6], for purification, the height to diameter ratio should be in the range 10:1 to 130:1. For experimental purposes, a device of height to diameter ratio of 10:1 was made and utilized in this study. Based on recommendations by [7], the column was divided into three parts. The top part is half the height of the bottom part while the middle part contains the packing. Marbles were used as packaging material and they occupied 70% of the column height. Based on the data above, the upgrading device was designed with dimensions as shown in Figure 1 while Figure 2 shows the assembly of the device.

2.3 Generator Performance

In internal combustion engines, the combustion of a fuel occurs with an oxidizer (usually air) in a combustion chamber that is an integral part of the working fluid flow circuit [2]. Petrol engines can run fully on biogas whereas diesel engines require combination of biogas and diesel. In the diesel engine, the primary gaseous fuel is inducted with air into the engine cylinder and a small amount of diesel, called pilot fuel is injected for promoting combustion [8].

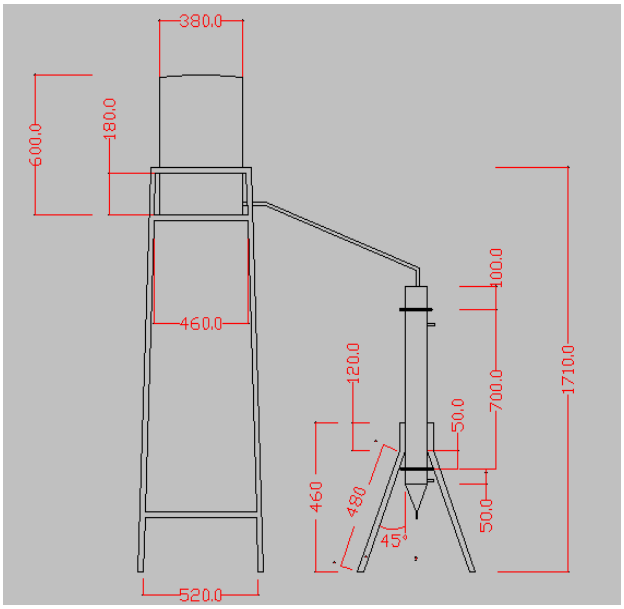


Figure 1: Design of the upgrading device (units in mm)



Figure 2: Assembly of the biogas purification device

To assess the generator performance, an Elemax SH 2900 gasoline generator was used. Elemax SH 2900 generator was used because it has a spark ignition engine which can easily ignite the biogas using the spark plug unlike compression ignition engines. Therefore, biogas could be used as a fuel without blending it as it is with compression ignition engines. Table 1 gives the generator specifications.

The gasoline generator was slightly modified to enable it run on biogas. This was done by inserting a wooden piece holding a 10mm metal duct in between the generator engine and carburettor. This duct was to direct biogas to the generator engine without necessarily changing its entire make-up. Therefore, as the generator run, the fuel could be swiftly changed from gasoline to biogas.

Table 1: Elemax SH 2900 Generator Specifications

S/N	Parameter	Specification
1	Model	Honda GX160 SH 2900 Generator
2	Engine type	4-stroke OHV
3	Rated power	3.6 kW @ 3600rpm
4	Cooling system	Air cooled
5	Ignition system	Transistorized magnets
6	Generator AC output	2400VA @ 220V, 50Hz

During experiments, biogas was collected in a polythene bag of capacity 1m^3. The generator was manually started on petrol and later biogas introduced. The supply of petrol was deliberately and steadily reduced as biogas was steadily introduced while regulating the air intake. Eventually, petrol was completely cut off and the generator ran on biogas. The experimental setup is shown in Figure 3.

The experiments were conducted under different loading conditions. Compact fluorescent lamps were used as the test loads. These test loads were grouped in percentage of the total load which was made of 16 lamps connected in series and parallel. Test load one comprised of 25% of the total load, load two was made of 50% while loads three and four were made of 75% and 100% of the total load respectively.

For each test load, various measurements were taken after stabilization. These include; current, voltage and fuel consumption. An ammeter connected in series was used for current measurement while a voltmeter connected in parallel with the test load was used for voltage measurement. Current was measured in amps while voltage was measured in volts. An air flow meter was used to measure the biogas flow for a given period of time. The results were used to determine the biogas fuel consumed by the generator for each test load given the period it was run. A ball valve connected to the pipe on the biogas bag was used to control the flow rate. The results obtained for each test load were used in calculation of Brake Power, Brake Specific Fuel Consumption (BSFC) and Brake Thermal Efficiency (BTE), using the formulae obtained from [2].

Engine brake power output (Pb):

$$\text{Voltage developed (V) x Current produced (I)} \quad (1)$$

Brake specific fuel consumption (BSFC):

$$\frac{\text{Fuel Consumption (FC)}}{\text{Engine brake power output (pb)}} \quad (2)$$

Brake thermal efficiency (BTE):

$$\frac{3600 \text{ x pb}}{\text{FC x LHV}} x\,100\% \quad (3)$$

Where LHV: Lower Heating Value of the gas.

Table 2 below shows the thermal content of biogas that was considered when calculating the Brake Thermal Efficiency. Since CH_4 is the only combustible gas in biogas, it was used when considering the lower heating value of the gas. From the biogas characterization done and the results obtained from the upgrading process, the heating values of biogas were obtained from Table 2 below based on their CH_4 content.

Table 2: Methane Yield from Animal Waste [9]

Animal	Typical experimental yield/kg manure	CH_4%	CO_2%	Thermal content MJ/m^3
Cattle	200 – 350 L	57.5%	46.5%	23
Poultry	550 – 650 L	70.0%	30.0%	28
Pig	400 – 500 L	65.0%	35.0%	26



Figure 3: Experimental Setup for the Generator Running *on Biogas*

3. Results and Discussions

3.1 Characterization of Biogas from the Digester

Biogas from chicken droppings was got from a 30m^3 Fixed Dome Digester. It comprised of 57.3% CH_4, 41.6% CO_2, 0.2% O_2, 332 ppm of H_2S and a balance of 0.4%. These were read at a digester temperature of 28.1^{o}C, pressure of 16kPa and maximum flow rate of 15litres/minute. Biogas from poultry wastes can comprise 70% CH_4 (Abdul Kareem, 2005). However, the variations exhibited can be attributed to numerous factors including the operating temperature of the digester, level of substrate mixing, the pH, carbon to nitrogen ratio (C:N) and toxins in the feedstock which may be as a result of the vaccines administered to the chicken. Biogas production can be increased by operating the digester at thermophilic temperature ranges. The said digester was run at mesophilic temperature ranges expected to be between 25-35^{o}C. Factors such as, optimum substrate mixing, neutral pH and optimum C: N ratios can affect the biogas yield quantities and component distribution.

3.2 Experimental Results obtained from the designed Biogas Upgrading Device

To establish the performance of the fabricated apparatus, different experiments were carried out to determine the device efficiency. This was done using three solvents including; H_2O, NaOH and KOH. At solvent to biogas flow rates of 0.20, 0.40, 0.60 and 0.80, the data captured was relating to biogas composition, CH_4 enrichment and CO_2 reduction. The comparative data obtained for analysis of the three solvents using average values is given in Table 3. The findings in table 3 indicate that at the same ratio of solvent to biogas flow rate, the CO_2 reduction was lowest for water, followed by KOH and highest for NaOH. On the other hand, the same ratios of the flow rate for the solvent to biogas, highest CH_4 enrichment was generated in NaOH, followed by KOH and lowest for H_2O. This is because solutions of NaOH and KOH have enhanced scrubbing capabilities for CO_2 removal. Figure 4 illustrates the percentage CH_4 enrichment using the various solvents at varying flow rate ratios.

Graphically, for the three solvents, best performance was exhibited at a ratio of NaOH to biogas flow rate of 0.80 where CO_2 reduction clocked 93% and the CH_4 enrichment level was 60.6%. The device thus performed better at higher flow rate ratios of 0.80 than at low flow rate ratios of 0.20 for all solvents. This is because there is more CO_2 absorption when the volume of solvent interacting with a given biogas volume is relatively high. Although better results are obtained at high flow rates, further increasing flow rate led to flooding of the device. Higher columns can properly cater for higher flow rates hence delaying flooding. The height of the column can be technically increased by packing. Packaging materials greatly increase the surface area of the column and hence interaction time of the gas and solvent thereby delaying flooding. Packaging further allows for efficient contact between the water (solvent) and gas phases in a counter current absorption process [7]. It is recommended that almost 70% of the column height should be filled with packaging material for good results [10]. However, the kind of packaging material defines the impact on the performance of the column. This is because packaging material affects the uniform distribution of the solvent and biogas as they flow across the column cross section.

3.3 Characteristic Generator Performance

Having developed and tested the performance of the biogas upgrading unit, control experiments were conducted to establish the performance of biogas in electricity generation. This was done by replacing the generator fuel from gasoline with biogas. Both raw and upgraded biogas was used and varying results were obtained.

3.3.1. Experimental Results of the Generator Performance on Raw Biogas

In this experiment, raw biogas was used in the generator. Ammeter, voltmeter and flow meter readings were recorded and used in Formulae 1 to 3, used to come up with data in Table 4. From the biogas characterization done, the CH_4 content was recorded at 57.3%. Therefore, the heating value of the raw gas was assumed to be 23MJ/m^3 or 20MJ/kg as earlier noted in Table 2.

3.3.2 Experimental Results of the Generator Performance on Upgraded Biogas

In this set of experiments, biogas was upgraded before it was used in the generator. After the biogas upgrading process, the CH_4 content was enriched to 78% using plain water as the solvent at water to gas flow rate ratio of 0.80. Therefore, the heating value of the upgraded biogas was assumed to be 28MJ/m^3 or 24.3MJ/kg since it had a higher CH_4 content than that quoted for poultry wastes with 70% CH_4 in Table 2. Ammeter, voltmeter and flow meter readings were recorded and used in Formulae1 to 3, and used to come up with data in Table 5. From Figure 5, the highest brake power is achieved at 100% loading of the generator with upgraded biogas recorded at 0.239 kW while the lowest brake power was attained at 25% electric load with raw biogas recoded at 0.060 kW. The current consumed increases with increase in the connected load. Therefore, the more loads connected, the higher the current consumed and hence an increase in brake power. This is in line with [2] where it is stated that increase in loading increases the combustion quality of the fuel and hence the power output of the generator. This is because the generator output depends on its fuel burning efficiency. From Figure 6, Brake Specific Fuel Consumption (BSFC) decreased with percentage electric load. This was because BSFC decreases with increase in electric load as it depends on fuel consumption directly and brake power inversely which increases with increase in brake load [2]. BSFC is higher for raw biogas at 2.07 kg/kWh as compared to that of upgraded biogas at 1.45 kg/kWh. This is because a larger volume of biogas was required for the same load for raw biogas than for upgraded biogas.

Upgraded biogas had a higher calorific value of 24.3 MJ/kg due to the high CH_4 content of 78% and hence the less volume of biogas required for the same load. It is well known that the calorific value of the fuel has an impact on the amount of fuel required for running the engine. On the other hand, Brake Thermal Efficiency (BTE) of upgraded biogas is higher than that from raw biogas. This is because a drop of CO_2 in biogas for dual fuelling increases the thermal efficiency. The highest value of upgraded biogas BTE was 17.6% compared to 14.2% of raw biogas and the lowest value for upgraded biogas was 10.2% as compared to 8.7% of raw biogas. BTE increases with increase in brake power. It can be noticed from Figure 6 that BTE increased with increased load. This is the same pattern demonstrated by brake power in Figure 5 since it's directly proportional with the percentage load. BTE further depends on brake power, fuel consumption rate and calorific value of biogas. Since upgraded biogas had a higher calorific value of 24.3MJ/kg compared to raw biogas' 20 MJ/kg and lower fuel consumption, it greatly contributed to the higher efficiency of upgraded biogas.

Table 3: Comparison of the Three Solvents on CH_4 Enrichment and CO_2 Reduction

Flow Rate Ratio	H_2O		KOH		NaOH	
	CH_4 enrichment	CO_2 reduction	CH_4 enrichment	CO_2 reduction	CH_4 enrichment	CO_2 reduction
0.2	18.3	42.8	26.4	54.8	43.5	76.4
0.4	23.2	48.6	34.0	65.4	51.1	84.9
0.6	30.7	58.7	40.1	72.8	57.6	91.8
0.8	36.8	63.0	48.9	82.5	60.6	93.0

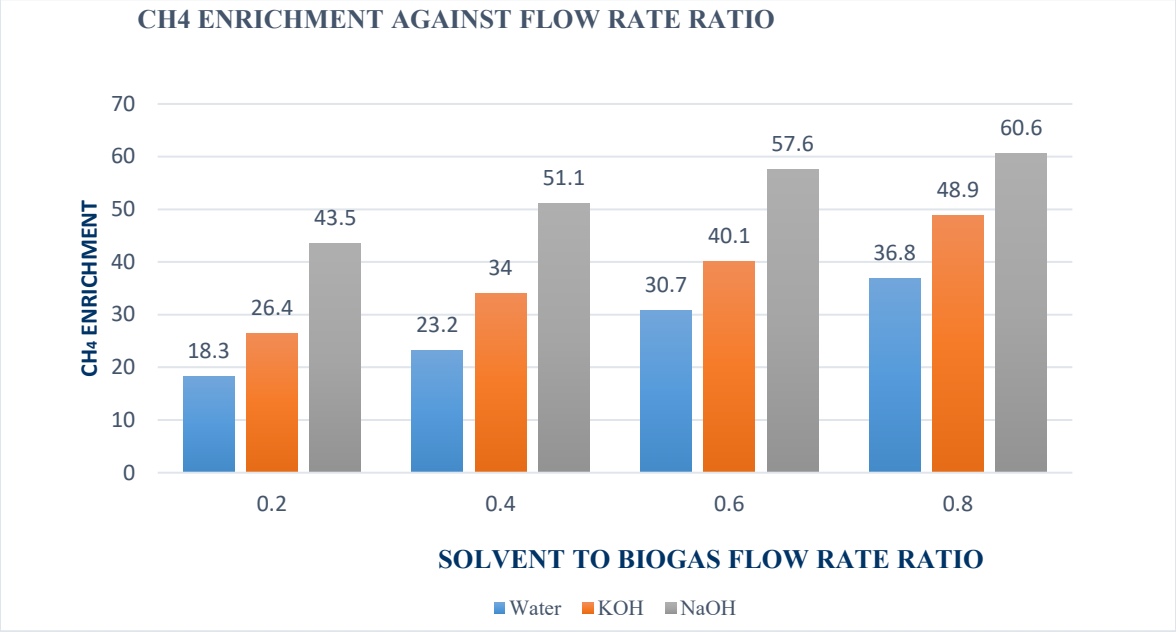


Figure 4: Comparison of the Three Solvents on CH_4 Enrichment

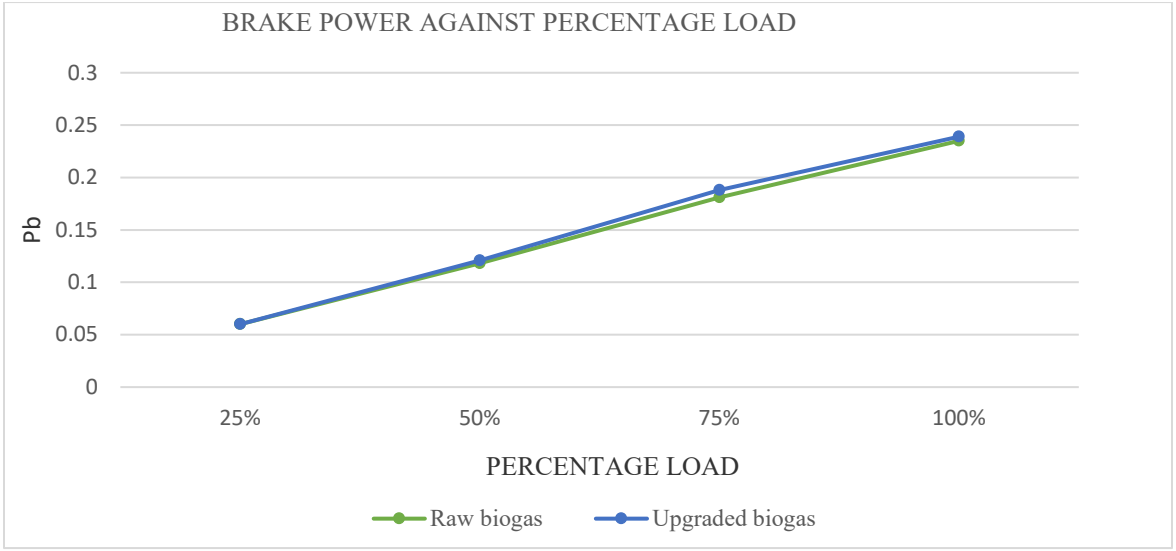


Figure 5: Variation of Brake Power against Percentage Load for both Raw and Upgraded Biogas

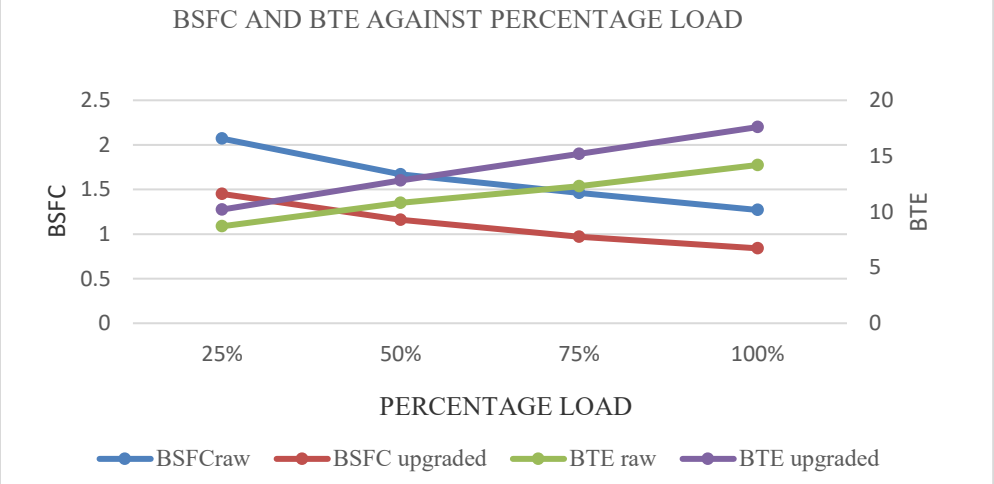


Figure 6: Variation of Brake Specific Fuel Consumption and Thermal Efficiency with Percentage Load

Table 4: Generator Performance on Raw Biogas

S/N	ELECTRIC LOAD	25%	50%	75%	100%
1.	Voltage (V)	222.2	227	226.3	223.8
2.	Current (A)	0.27	0.52	0.80	1.05
3.	Pb (kW)	0.060	0.118	0.181	0.235
4.	Fuel consumption (kg/h)	124.1	197.3	264.9	297.9
5.	BSFC (kg/kWh)	2.07	1.67	1.46	1.27
6.	BTE (%)	8.7	10.8	12.3	14.2

Table 5: Generator Performance on Upgraded Biogas

S/N	ELECTRIC LOAD	25%	50%	75%	100%
1.	Voltage (V)	225.9	232.7	235	227.6
2.	Current (A)	0.27	0.52	0.80	1.05
3.	Pb (kW)	0.061	0.121	0.188	0.239
4.	Fuel consumption (kg/h)	88.6	140.0	183.2	201.2
5.	BSFC (kg/kWh)	1.45	1.16	0.97	0.84
6.	BTE (%)	10.2	12.8	15.2	17.6

During the start of experiments, the generator was run on petrol to heat up and later biogas introduced. This was because it was hard burning biogas at the generator start due to the high heating temperatures required. However, this wasn't the case when the generator had been running because the generator would have heated up already. This further highlighted the impact of biogas' low heating value. BTE increases with increased load due to increase in combustion zone temperature. Increase in load increases the combustion zone pressure and heat release rate. This improves the fuel combustion quality hence increasing the power output of the generator. Therefore, the higher the generator power output, the better the BTE.

4. Conclusion

The performance of the biogas upgrading device not only depends on the biogas flow rate but also depends on the dimensions of the scrubbing tower, biogas pressure, packaging material and purity of the water (solvent) used in the process. Upgraded biogas had a higher calorific value of 24.3 MJ/kg due to the high CH_4 content of 78% and hence the less volume of biogas required for the same load. Increase in load increases the combustion zone pressure and heat release rate. This improves the fuel combustion quality hence increasing the power output of the generator. Therefore, the higher the generator power output, the better the BTE. Using upgraded biogas improved the generator's brake thermal efficiency from 14.2% for raw biogas to 17.6% for upgraded biogas.

Conflict of Interest

The authors declare that although they may have conferred with others in carrying out the research and preparation of this paper, and drawn upon the range of sources cited in this work, this paper is their original piece of work and has not been published in any other journal.

Acknowledgment

The authors gratefully acknowledge the funding and support received from RELOAD, a Germany organization dedicated to finding solutions to post-harvest losses, and Dr. Peter Okweru of Okweru Poultry Farmers for availing his farm to act as the study area where all experiments were carried out.

References

[1] S.S. Kapdi, V.K. Vijay, S.K. Rajesh, R. Prasad, "Biogas scrubbing, compression and storage: perspective and prospectus in Indian Context" Ren. Energ., 30(8), 1195–1202, 2005. http://dx.doi.org/10.1016/j.renene.2004.09.012

[2] A.V. Reddy, T. Sharath-Kumar, D,K, Tharun-Kumar, B. Dinesh, Y.V.S. Sai-Santosh, "Energy and exergy analysis of I.C. engines" Eng. and Sci., 3(5), 7– 26, 2014. ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

[3] N.H.S. Ray, P.R. Swain, M.K. Mohanty, "An investigation on performance characteristics of C.I. engine using biogas and diesel in dual fuel mode" Sci., Eng. and Technol. Research, 3(6), 1716 – 1723, 2014. ISSN: 2278 – 7798

[4] D. Barik, S. Sah, S. Murugan, "Biogas production and storage for fueling internal combustion engines" Emerging Technol. and Advanced Eng., 3(3), 193–202, 2013. ISSN 2250 – 2459

[5] L.B. Allegue, J. Hinge, "Biogas and bio-syngas Upgrading," Report, Danish Technological Institute, Kongsvang Allé 29 Dk–8000 Aarhus C., 2012.

[6] A.K. Dubey, "Wet scrubbing of carbon dioxide," Annual Report of Ciae, Bhopal, India, 2000.

[7] R. Chandra, V.K. Vijay, P.M.V. Subbarao, "Vehicular quality bio-methane production from biogas by using an automated water scrubbing system," ISRN Ren. Energ., 1-6, 2012.

[8] N.H.S. Ray, M.K. Mohanty, R.C. Mohanty, "A study on application of biogas as fuel in compression ignition engines," Innovations in Eng. And Technol., 3(1), 239 – 245, 2013. ISSN: 2319 – 1058

[9] A.S. Abdul-Kareem, "Refining biogas produced from biomass: an alternative to cooking gas," Leonardo Sci., 7, 1–8, 2005. ISSN 1583-0233

[10] D. Barik, S. Sah, S. Murugan, "Biogas production and storage for fueling internal combustion engines" Emerging Technol. and Advanced Eng., 3(3), 193–202, 2013. ISSN 2250 – 2459

[11] N. Tippayawong, P. Thanompongchart, "Biogas quality upgrade by simultaneous removal of CO_2 and H_2S in a packed column reactor," Energ., 35(12), 4531–4535, 2010. DOI: 10.1016/j.energy.2010.04.014.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 188-196 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Well Balanced Multi-value Sequence and its Properties Over Odd Characteristic Field

Md. Arshad Ali[*,1], Yuta Kodera[1], Md. Fazle Rabbi[2], Takuya Kusaka[1], Yasuyuki Nogami[1], Satoshi Uehara[3], Robert H. Morelos-Zaragoza[4]

[1]*Graduate School of Natural Science and Technology, Okayama University, Okayama, 7008530, Japan*

[2]*Faculty of CSE, Hajee Mohammad Danesh Sci. and Tech. University, Dinajpur, 5200, Bangladesh*

[3]*Faculty of Environmental Engineering, The University of Kitakyushu, Fukuoka, 8080135, Japan*

[4]*Department of Electrical Engineering, San Jose State University, San Jose, CA 95192, United States*

ARTICLEINFO

Article history:
Received: 28 May, 2019
Accepted: 12 July, 2019
Online: 30 July, 2019

Keywords:
Pseudo random sequence
Trace function
Power residue symbol
Correlation
Linear complexity
Distribution of bit patterns

ABSTRACT

The authors propose a well balanced multi-value sequence (including a binary sequence). All the sequence coefficients (except the zero) appear almost the same in number, thus, the proposed sequence is so called the well balanced sequence. This paper experimentally describes some prominent features regarding a sequence, for instance, its period, autocorrelation, and cross-correlation. The value of the autocorrelation and cross-correlation can be explicitly given by the authors formulated theorems. In addition, to ensure the usability of the proposed multi-value sequence, the authors introduce its flexibility by making it a binary sequence. Furthermore, this paper also introduces a comparison in terms of the linear complexity and distribution of bit patterns properties with their previous works. According to the comparison results, the proposed sequence holds better properties compared to our previous sequence.

1 Introduction

Pseudo random sequences of having random numbers are crucial components of many cryptographic applications, for instance, key generation, session keys, masking protocol, navigation, radar ranging, and so on [1, 2, 3]. The security of these cryptographic systems deliberately depends on the randomness and unpredictability regarding the sequence. By using the non-linearity features of some mathematical functions, a pseudo random sequence of having excellent randomness characteristics can be generated. The major substances for randomness are independency of values (or lack of correlation), unpredictability (or lack of predictability), and uniform distribution (or lack of bias) [4]. Therefore, a prominent pseudo random number generator is essential to generate pseudo random sequence having good randomness property.

Most renowned pseudo random number generators are the Mersenne Twister (MT) [5], Blum-Blum-Shub (BBS) [6], Legendre sequence [7], and M-sequence [8]. Among those, the former two pseudo random number generators (MT and BBS) are well known considering their applications in cryptography rather than the theoretical aspect. On the other hand, the M-sequence and Legendre sequence are prominent geometric sequences regarding the theoretical aspect. As a result, the authors attracted in the pseudo random sequence generation research area by observing the theoretical prospect on the M-sequence and Legendre sequences.

A well balanced pseudo random signed binary sequence proposed in our previous work [9]. It is generated by utilizing a primitive polynomial, trace function, and Legendre symbol. The period and autocorrelation properties of the well balanced signed binary sequence were explained based on some experimental results. This work is actually an extension of previous works on the signed binary sequence by introducing additional two parameters k and non-zero scalar A (where k and A are responsible for generating multi-value sequence and extending the sequence period to its maximum value, respectively). It should be noted that the k-th power residue symbol is actually an extension of the Legendre symbol, therefore, this power residue symbol includes the case of the well balanced signed binary sequence. Furthermore,

[*]Corresponding Author: Md. Arshad Ali, Graduate School of Natural Science and Technology, Okayama University, 3-1-1, Tsushima-naka, Kita, Okayama, 7008530, Japan, +81-8042661986, arshad@s.okayama-u.ac.jp

this work is also an extension of our previous work on multi-value sequence [10] by considering additional two properties (linear complexity and distribution of bit patterns) and introducing its flexibility by making it binary sequence, whereas, previous multi-value sequence introduced along with its autocorrelation and cross-correlation properties (based on experimental observations only).

In this paper, the authors propose a well balanced multi-value sequence (including a binary sequence). Let $f(x)$ be a primitive polynomial of degree m and $\omega \in \mathbb{F}_q$ be its zero. Then, the sequence

$$\mathcal{S} = \{s_i\} \mid s_i = \mathrm{Tr}\left(\omega^i\right), i = 0, 1, 2, \ldots, q-2, \ldots$$

becomes a maximum length sequence whose period is $q-2$. Here, $\mathrm{Tr}(\cdot)$ is a trace function which maps an element of the extension field $\mathbb{F}_q$ to an element of the prime field $\mathbb{F}_p$. In brief, the proposed well balanced multi-value sequence generation procedure is as follows: in the beginning, a primitive polynomial generates maximum length sequence of vectors, then the $\mathrm{Tr}(\cdot)$ maps vectors to scalars, next a non-zero prime field scalar $A \in \{1, 2, \ldots, p-1\}$ added to the scalars, and finally k-th power residue symbol maps the scalars to a well balanced multi-value ($k+1$ values) sequence.

From the viewpoint of auto and cross-correlation, there are a lot of considerations to use multi-value sequence in communications [11, 12]. However, there are few papers regarding the usage of pseudo random binary sequence with a long period, high linear complexity, and good distribution of bit patterns in security applications. To make attention to the usability of the proposed sequence, the authors introduce the flexibility of their proposed well balanced multi-value sequence to make it more worthy. To do so, the authors explain how to transform their proposed sequence into a binary sequence (along with its linear complexity and distribution of bit patterns properties) due to the extensive usage of binary sequence in numerous applications (especially in cryptography).

All our previous works on sequence generation (both binary and multi-value) utilizes a mapping function during the sequence generation procedure. As a result, there exists a big difference between the appearance of sequence coefficients, which leads the distribution of bit patterns ununiform. On the other hand, the proposed sequence is a $k+1$ values well balanced multi-value sequence without applying any kind of mapping function. Therefore, all the sequence coefficients (except the 0) appear almost the same in number, thus, it is called a well balanced multi-value sequence. This balanced characteristic in the sequence coefficients contributes to low correlation (both autocorrelation and cross-correlation), high linear complexity, and almost uniform distribution of bit patterns, whereas, a suitable pseudo random sequence for cryptographic applications asks for such kinds of features.

This paper experimentally explains some prominent features regarding a sequence, for instance, its period, autocorrelation, and cross-correlation. The authors formulate theorems by which the value of the autocorrelation and cross-correlation can be explicitly given. This is one of the major contributions of this paper. Furthermore, to emphasize the usability of the proposed sequence, the authors introduce its flexibility by making it a binary sequence. In addition, a comparison result regarding the linear complexity and distribution of bit patterns properties are also included in this paper. According to the comparison results, the proposed sequence in this paper holds better properties compared to our previous sequence.

Notations

In this paper, the notation p denotes an odd characteristic prime, m be a extension degree, and q denotes the power of p, for instance, $q = p^m$. In addition, k is a prime number as well as a factors of $p-1$, such as $k \mid (p-1)$. $\mathbb{F}_q^* = \mathbb{F}_q - \{0\}$ stands for multiplicative group of $\mathbb{F}_q$ excluding the zero.

2 Preliminaries

This section briefly introduces a few mathematical fundamentals which are related to this research work such as primitive polynomial, trace function, and k-th power residue symbol. In addition, the multi-value sequence also introduced along with its properties.

2.1 Primitive Polynomial

A polynomial $f(x)$ of degree m over the prime field $\mathbb{F}_p$ is said to be irreducible if it cannot be factorized into smaller degree polynomials (including the scalar factor), then $f(x)$ is said to be an irreducible polynomial. Let e be an smallest positive integer and $f(x) \mid (x^e - 1)$. If $x = p^m - 1$, then the polynomial $f(x)$ is said to be a primitive polynomial.

Let ω be an arbitrary element in the extension field $\mathbb{F}_q$. If $f(\omega) = 0$, then ω is said to be the root of the primitive polynomial. In addition, ω becomes a primitive element in $\mathbb{F}_q$ and all the non-zero elements can be generated by the power of ω^i such as

$$\omega^0, \omega^1, \omega^2, \ldots, \omega^{q-2}.$$

The primitive element ω has a multiplicative order of $q-1$. An extension field $\mathbb{F}_q$ and its base field $\mathbb{F}_p$ holds the following property [13].

Property 1 Let ω be a generator of $\mathbb{F}_q^*$, $\omega^{(q-1)/(p-1)}$ becomes a non-zero element in prime field $\mathbb{F}_p$ and is also a generator of $\mathbb{F}_p^*$. □

2.2 Trace Function

A trace function is defined to find the sum of conjugates. Let $\mathbb{F}_q$ be an extension field and X be one of the elements (vector) of $\mathbb{F}_q$. On the other hand, let x be a prime field $\mathbb{F}_p$ element (scalar). The trace of X over $\mathbb{F}_q$ is the sum of conjugates of X with respect to $\mathbb{F}_q$. It is defined as follows:

$$x = \mathrm{Tr}(X) = \sum_{i=0}^{m-1} X^{p^i}. \tag{1}$$

Aforementioned the trace function $\mathrm{Tr}(\cdot)$ sums the conjugates in the extension field $\mathbb{F}_q$ and maps them as the prime

field $\mathbb{F}_p$ elements. As a result, it has a linearity property, which shown in the following equation.

$$\mathrm{Tr}\left(\alpha X + \beta Y\right) = \alpha \mathrm{Tr}\left(X\right) + \beta \mathrm{Tr}\left(Y\right), \tag{2}$$

where α, β are prime field $\mathbb{F}_p$ elements and X, Y are extension field $\mathbb{F}_q$ elements.

Property 2 Let $i = 0, 1, 2, \ldots, p-1 \in \mathbb{F}_p$. Then for each i the number of elements in the extension field $\mathbb{F}_q$ whose trace with regard to the prime field $\mathbb{F}_p$ becomes i be given by $q/p = p^{m-1}$. □

2.3 *k*-th Power Residue Symbol

The k-th power residue symbol with $(k > 2)$ is a generalization of the Legendre symbol to k-th powers [14]. Let a be an arbitrary element in the prime field $\mathbb{F}_p$, then the k-th power residue symbol $\left(a/p\right)_k$ can be defined as follows [15]:

$$\begin{aligned}\left(\frac{a}{p}\right)_k &= a^{(p-1)/k} \bmod p \\ &= \begin{cases} 0 & \text{if } a = 0, \\ 1 = \epsilon_k^0 & \text{else if } a \text{ is a } k\text{-th PR in } \mathbb{F}_p^*, \\ \epsilon_k^i & \text{otherwise } a \text{ is a } k\text{-th PNR in } \mathbb{F}_p^*. \end{cases}\end{aligned} \tag{3}$$

Throughout this paper, k is a prime number as well as a factor of $p-1$, such as $k \mid (p-1)$. According to the definition of the k-th power residue symbol $\left(a/p\right)_k$, a is called as the k-th Power Residue, when it has a k-th root in the base field $\mathbb{F}_p$. On the other hand, a is called as k-th Power Non-Residue. In addition, here ϵ_k is a primitive k-th root of unity belongs to $\mathbb{F}_p$ and it holds the relation $0 \le i < k$.

In Eq. (3), the value of the exponent i will be within the range of $0 \sim k-1$, since $\epsilon_k^k = \epsilon_k^0 = 1$. The k-th power residue symbol translates the scalars generated by the trace function $\mathrm{Tr}\left(\cdot\right)$ to a multi-value sequence. Thus, the sequence coefficients will be $\{0, \epsilon_k^i\}$, where $i \in \{0, \ldots, k-1\}$. In this paper, an alternate representation of the exponent i in Eq. (3) is as follows:

$$i = \log_{\epsilon_k}\left(\left(a/p\right)_k\right) = \log_{\epsilon_k}\left(a^{(p-1)/k} \bmod p\right). \tag{4}$$

Furthermore, the k-th power residue symbol holds the following property.

Property 3 For each i from 0 to $k-1$, the number of non-zero elements in $\mathbb{F}_p$ such that

$$\left(\frac{a}{p}\right)_k = \epsilon_k^i \tag{5}$$

is given by $(p-1)/k$. □

2.4 Multi-value Sequence and Its Properties

In this section, the proposed multi-value sequence introduced along with its period, autocorrelation, cross-correlation, linear complexity and distribution of bit patterns properties.

2.4.1 Notation

Throughout this paper, the proposed multi-value (more specifically, $k+1$ values) sequence $\mathcal{S}$ will be denoted as follows:

$$\mathcal{S} = \{s_i\}, i = 0, 1, 2, \ldots, n-1, \ldots, \tag{6}$$

where n denotes the period of the proposed sequence $\mathcal{S}$. In addition, here $s_i = s_{n+i}$.

2.4.2 Autocorrelation and Cross-correlation

The autocorrelation of a sequence is a measure for how much a sequence differs from its each shift value. In addition, the period and other patterns regarding a sequence can be obtained by evaluating the autocorrelation property [16]. Let $\mathcal{S} = \{s_i\}$ be a sequence and x be the shift value, then the autocorrelation $\mathrm{R}_{\mathcal{S}}(x)$ of $\mathcal{S}$ can be calculated by using the following equation as,

$$\mathrm{R}_{\mathcal{S}}(x) = \sum_{i=0}^{n-1} \tilde{\epsilon}_k^{s_{i+x} \times s_i}, \tag{7}$$

where $\tilde{\epsilon}_k$ is a primitive k-th root of unity over the complex number $\mathbb{C}$ [17] and it follows that

$$\mathrm{R}_{\mathcal{S}}(0) = \sum_{i=0}^{n-1} \tilde{\epsilon}_k^0 = n. \tag{8}$$

Furthermore, the cross-correlation property is as important as the autocorrelation property. It defines the similarities between two completely different sequences. If multiple sequences are used in an application (more specifically in any security application), then it is important to analyze their cross-correlation property to evaluate how much similar these sequences to each other. Considering this point, the cross-correlation value is preferred to be low [18, 19]. Let $\hat{\mathcal{S}} = \{\hat{s}_i\}$ and $\mathcal{S} = \{s_i\}$ be two sequences and x be the shift value, then the cross-correlation $\mathrm{R}_{\mathcal{S}}(x)$ between $\hat{\mathcal{S}}$ and $\mathcal{S}$ can be calculated by using the following equation as,

$$\mathrm{R}_{\hat{\mathcal{S}},\mathcal{S}}(x) = \sum_{i=0}^{n-1} \tilde{\epsilon}_k^{\hat{s}_{i+x} \times s_i}, \tag{9}$$

where $\tilde{\epsilon}_k$ is a primitive k-th root of unity over the complex number $\mathbb{C}$ [20, 21].

2.4.3 Linear Complexity

The linear complexity regarding a sequence is a measure of unpredictability by the length of the shortest Linear Feedback Shift Register (LFSR). In the literature, this length of the LFSR is referred to as the linear complexity [22]. The Berlekamp-Massey algorithm is an efficient method of determining the linear complexity of a sequence [23]. The forward unpredictability can be confirmed by the linear complexity property.

To calculate the linear complexity of a sequence $\mathcal{S} = \{s_0, s_1, \ldots, s_{n-1}\}$, at first, the sequence $\mathcal{S}$ needed to be represented by the polynomial expression $\mathcal{S}(x)$ as follows:

$$\mathcal{S}(x) = \sum_{i=0}^{n-1} s_i \cdot x^i, \tag{10}$$

where n denotes the period of the sequence $\mathcal{S}$. If we consider a binary sequence, then the sequence coefficients $s_i \in \{0, 1\}$, in other words, s_i belongs to $\mathbb{F}_2$. On the other hand, in case of multi-value sequence (k-values sequence), $s_i \in \{0, 1, 2, \ldots, k-1\}$, furthermore, $s_i \in \mathbb{F}_k$. After translating the sequence into polynomial, the linear complexity is evaluated by utilizing the equation in below (over $\mathbb{F}_2$ or $\mathbb{F}_k$).

$$\mathrm{LC}(\mathcal{S}) = n - \deg(\gcd(x^n - 1, \mathcal{S}(x))). \tag{11}$$

In the above equation, $\deg(f(x))$ denotes the degree of the primitive polynomial $f(x)$.

2.4.4 Distribution of Bit Patterns

The distribution of bit patterns is an important measure to judge the randomness of a sequence. As a reference, an M-sequence is well known for its uniform distribution of bit patterns. A uniform distribution of bit patterns means all the bit patterns (1-bit pattern, 2-bit patterns, 3-bit patterns, and so on) appear the same in number. Assume an M-sequence of having a period of 15 as follows and its bit distribution is shown in Table 1.

$$\mathcal{S}_{15} = \{1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0\}.$$

Table 1: Uniform distribution of bit distribution of an M-sequence [1].

n	$\mathrm{H_w}\left(b^{(n)}\right)$	$Z\left(b^{(n)}\right)$	$D_{S_{15}}\left(b^{(n)}\right)$
1	0	1	7
	1	0	8
2	0	2	3
	1	1	4
	2	0	4
3	0	3	1
	1	2	2
	2	1	2
	3	0	2

In an M-sequence, except 0 all the bit patterns appear same in number, therefore, the distribution of bit patterns of M-sequence is known as uniform.

It should be noted that the randomness and bit patterns hold a strong relationship with each other. In other words, the more uniform distribution of bits in a sequence, the sequence is more random.

3 Proposed Multi-value Sequence

The authors propose a well balanced multi-value sequence $\mathcal{S}$ by combining the features of the trace function and k-th power residue symbol. Assume that in the extension field $\mathbb{F}_q$, ω be a primitive element. Furthermore, A is a non-zero scalar which belongs to the prime field $\mathbb{F}_p$. Then, the proposed sequence $\mathcal{S}$ is defined as follows:

$$\mathcal{S} = \{s_i\},\ s_i = \left(\mathrm{Tr}\left(\omega^i\right) + A \middle/ p\right)_k, \tag{12}$$

here k is a factor of $p-1$, such as $k \mid (p-1)$. The sequence coefficients s_i in Eq. (12) can be described as the exponent of ϵ_k, such as ϵ_k^e. For instance, let $p = 7$ and $k = 3$, then the sequence coefficients in this example becomes $s_i \in \{0, 1, 2, 4\}$ and the 3-rd primitive root in $\mathbb{F}_7$ is equal to 2 or 4. In addition, let us fix 2 as a 3-rd primitive root. Then all of the non-zero sequence coefficients can be represented as an exponent of primitive root 2, this relation is developed as,

$$\epsilon_3\{1, 2, 4\} = \{2^0, 2^1, 2^2\} = \{\epsilon_3^0, \epsilon_3^1, \epsilon_3^2\}.$$

At first, the authors will focus on the autocorrelation and cross-correlation properties regarding the proposed sequence. It should be noted that the autocorrelation and cross-correlation are very close to each other. The main difference between them is the cross-correlation is calculated between two different sequences and the autocorrelation is focused in a single sequence. Thus, in the beginning, let us focus on the cross-correlation property. As mentioned earlier, using two different sequences of having the same period the cross-correlation is calculated. Let, $\mathcal{S}$ and $\hat{\mathcal{S}}$ be two different sequences which are defined as follows:

$$\mathcal{S} = \left\{ s_i \mid s_i = \left(\mathrm{Tr}\left(\omega^i\right) + A \middle/ p\right)_k \right\}, \tag{13a}$$

$$\hat{\mathcal{S}} = \left\{ \hat{s}_i \mid \hat{s}_i = \left(\mathrm{Tr}\left(\omega^i\right) + \hat{A} \middle/ p\right)_k \right\}. \tag{13b}$$

Here, A and $\hat{A}$ are non-zero elements in $\mathbb{F}_p$ can be represented by a generator $g \in \mathbb{F}_p$ such as

$$\hat{A} = g^h A, \tag{14}$$

here h satisfies the relation $0 \le h \le p-2$ and g needs to be given by $\omega^{(p^m-1)/(p-1)}$. When the value of $h = 0$, that is $\hat{A} = A$ which means $\hat{\mathcal{S}}$ and $\mathcal{S}$ becomes the same sequence. Thus, the cross-correlation becomes the autocorrelation of $\mathcal{S}$. After inspecting several experimental results, it was found that the value of the cross-correlation explicitly given by the following theorem.

Theorem 1 The cross-correlation between the two sequences $\hat{\mathcal{S}}$ and $\mathcal{S}$ is defined by the Eq. (13) is as follows:

$$\mathrm{R}_{\hat{\mathcal{S}},\mathcal{S}}(x) = \begin{cases} \left(p^m - 1 - p^{m-1}\right)\tilde{\epsilon}_k^{f_k(g^h)}, & \text{if } x = h\bar{n}, \\ -p^{m-1}\left(\tilde{\epsilon}_k^{f_k(g^j)}\right) - \tilde{\epsilon}_k^{f_k(g^h)}, & \text{else if } x = j\bar{n}, \\ -\tilde{\epsilon}_k^{f_k(g^h)}, & \text{otherwise,} \end{cases} \tag{15}$$

where $\bar{n} = n/(p-1) = (p^m - 1)/(p-1)$, h satisfies the condition in Eq. (14), and $0 \le j \ne h \le p-2$. □

If the value of $h = 0$, then $\hat{\mathcal{S}} = \mathcal{S}$ which actually means they becomes the same sequence. In this case, the cross-correlation in Eq. (15) becomes the autocorrelation after replacing the value $h = 0$.

[1] The notations $n, \mathrm{H_w}\left(b^{(n)}\right), Z\left(b^{(n)}\right)$, and, $D_S\left(b^{(n)}\right)$ means length of a bit pattern $b^{(n)}$, Hamming weight of the bit pattern $b^{(n)}$, number of zeros in $b^{(n)}$, and appearance of $b^{(n)}$ in numbers in a sequence period, respectively.

Theorem 2 The autocorrelation of a sequence $\mathcal{S}$ is given as follows:

$$\mathrm{R}_{\mathcal{S}}(x) = \begin{cases} p^m - 1 - p^{m-1}, & \text{if } x = h\bar{n}, \\ -p^{m-1}\left(\tilde{\epsilon}_k^{f_k(g^j)}\right) - 1, & \text{else if } x = j\bar{n}, \\ -1, & \text{otherwise}, \end{cases} \tag{16}$$

Corresponding to the above autocorrelation equation, the period of the proposed well balanced multi-value sequence undeniably given by $p^m - 1$. □

In the next section, the authors will introduce experimental observation regarding the period, autocorrelation and cross-correlation properties.

4 Example and Discussion

This section experimentally observes the proposed multi-value sequence properties such as its period, autocorrelation, and cross-correlation along with some examples. In this section, the notation $\mathcal{S}_2$ denotes the proposed sequence with the parameter $A = 2$. The proposed sequence in this paper is a multi-value sequence, thus its correlation is calculated over the complex number $\mathbb{C}$. To represent the absolute value of a complex number x, this section uses the notation $|x|$.

4.1 $p = 7, m = 2, k = 3$, and $A = 2, 3$

Assume, $x^2 + 4x + 5$ be a primitive polynomial over $\mathbb{F}_7$. Then, the generated sequence $\mathcal{S}_2$ having a period of 48 ($p^m - 1 = 7^2 - 1 = 48$) is shown as follows:

$$\begin{aligned} \mathcal{S}_2 = \{&2,4,1,4,4,1,1,4,4,2,2,2,4,2,1,2,2,0,4,0,4,4,2,0, \\ &0,1,2,1,4,2,4,1,1,1,0,1,4,0,2,1,1,2,1,2,4,1,0,2\}. \end{aligned} \tag{17}$$

The autocorrelation of this generated sequence $\mathcal{S}_2$ is calculated by the Eq. (7) and autocorrelation graph of $\mathcal{S}_2$ is shown in Figure 1.

$$|\mathrm{R}_{\mathcal{S}_2}(x)| = \begin{cases} 41 & \text{if } x = 0 \\ 6 & \text{else if } x = 8, 16, 32, 40 \\ 8 & \text{else if } x = 24 \\ 1 & \text{otherwise} \end{cases}.$$

To confirm the well balanced property in the proposed multi-value sequence, the authors introduce sequence coefficients appearance in the following Table 2. According to the table, it was found that in every case all of the sequence coefficients (except the 0) appears almost the same in number. This is one of the positive properties of the proposed sequence, thus, it is called as a well balanced sequence.

Table 2: Appearance of the sequence coefficients.

sequence coefficient	number of appearance
0	7
1	14
2	14
4	13

On the other hand, $\mathcal{S}_3$ is given as follows. It should be noted that the sequence $\mathcal{S}_3$ is different from the sequence $\mathcal{S}_2$, but both of them having the same period.

$$\begin{aligned} \mathcal{S}_3 = \{&4,1,4,1,2,4,0,1,1,2,4,2,2,4,4,2,2,1,1,1,2,1,4,1, \\ &1,0,2,0,2,2,1,0,0,4,1,4,2,1,2,4,4,4,0,4,2,0,1,4\}. \end{aligned} \tag{18}$$

The autocorrelation of this generated sequence $\mathcal{S}_3$ is calculated by the Eq. (7) and autocorrelation graph of $\mathcal{S}_2$ is shown in Figure 2.

$$|\mathrm{R}_{\mathcal{S}_3}(x)| = \begin{cases} 41 & \text{if } x = 0 \\ 6 & \text{else if } x = 8, 16, 32, 40 \\ 8 & \text{else if } x = 24 \\ 1 & \text{otherwise} \end{cases}.$$

The cross-correlation of $\mathcal{S}_2$ and $\mathcal{S}_3$ becomes as follows and the cross-correlation graph shows in Figure 3.

$$|\mathrm{R}_{\mathcal{S}_2,\mathcal{S}_3}(x)| = \begin{cases} 6 & \text{if } x = 0, 8, 24, 32 \\ 8 & \text{else if } x = 16 \\ 41 & \text{else if } x = 40 \\ 1 & \text{otherwise} \end{cases}.$$

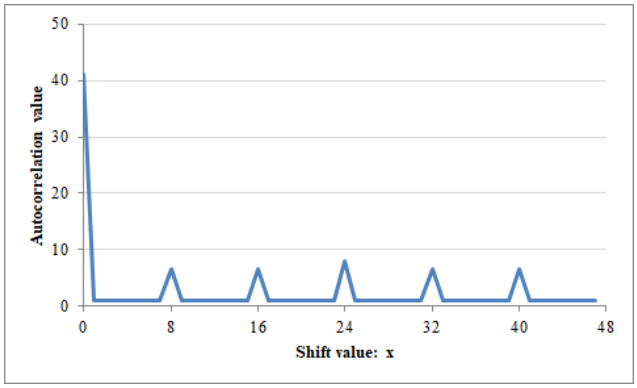


Figure 1: $|\mathrm{R}_{\mathcal{S}_2}(x)|$ with $p = 7, m = 2, k = 3$, and $A = 2$.

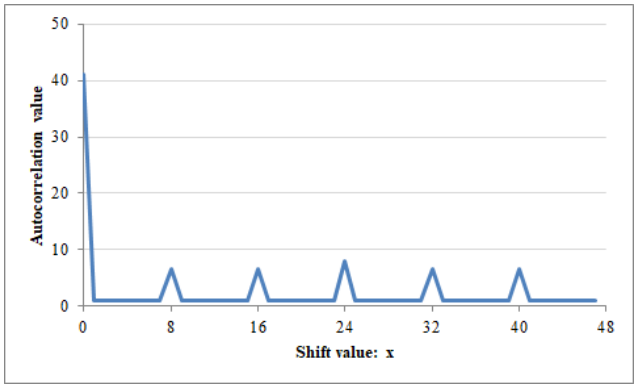


Figure 2: $|\mathrm{R}_{\mathcal{S}_3}(x)|$ with $p = 7, m = 2, k = 3$, and $A = 3$.

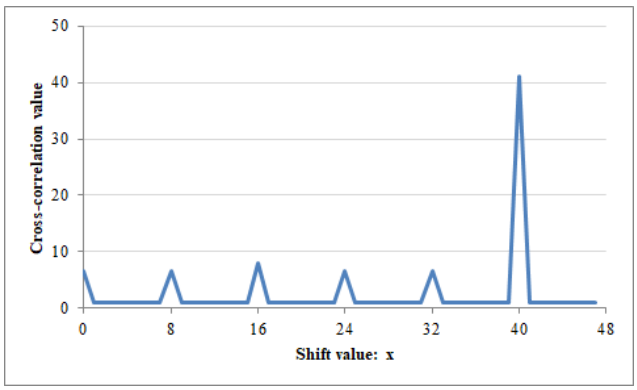


Figure 3: $|R_{S_2,S_3}(x)|$ with $p = 7, m = 2$, and $A = 2, 3$.

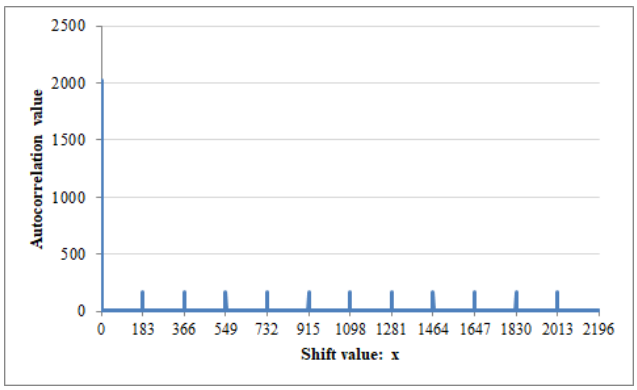


Figure 5: $|R_{S_7}(x)|$ with $p = 13, m = 3, k = 3$, and $A = 7$.

4.2 $p = 13, m = 3, k = 3$, and $A = 6, 7$

Assume, $x^3 + 6x^2 + 3x + 7$ be a primitive polynomial over $\mathbb{F}_{13}$. Then, the period of this generated sequence becomes 2196. Here, Figure 4 and Figure 5 respectively represents the autocorrelation graph of $\mathcal{S}_6$ and $\mathcal{S}_7$. Their cross-correlation graph is shown in Figure 6.

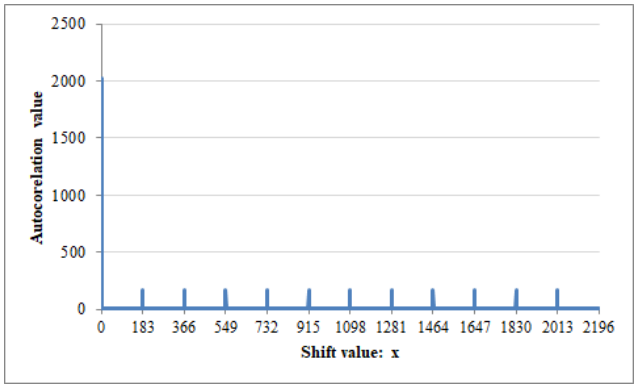


Figure 4: $|R_{S_6}(x)|$ with $p = 13, m = 3, k = 3$, and $A = 6$.

After observing the cross-correlation graphs, it was found that in each graph the number of peaks exactly is given by $p - 1$. Only a single peak has the maximum value, as an example, in Figure 3 the maximum peak value is 41 that corresponds to the first case ($x = h\bar{n}$) of Eq. (15). Remaining $p - 2$ smaller peaks conforms to the second case ($x = j\bar{n}$). Except all of these peak values, the remaining parts of the cross-correlation graph consistently having a constant value of 1, which confirms the final case of Eq. (15). It should be noted that only changing the value of non-zero scalar parameter $A \in \mathbb{F}_p$, different sequences can be generated. It is also observed that by changing all the other parameters such as primitive polynomial $f(x)$, extension degree m, non-zero scalar A, and prime factor k does not have any impact in the correlation (both autocorrelation and cross-correlation) evaluation.

Alike the cross-correlation, autocorrelation also have a $p-1$ number of peaks. Among them, only one peak hold the maximum value, while other peaks have small values and remaining parts holds a constant value of 1 and all of these values explicitly given by the Eq. (16).

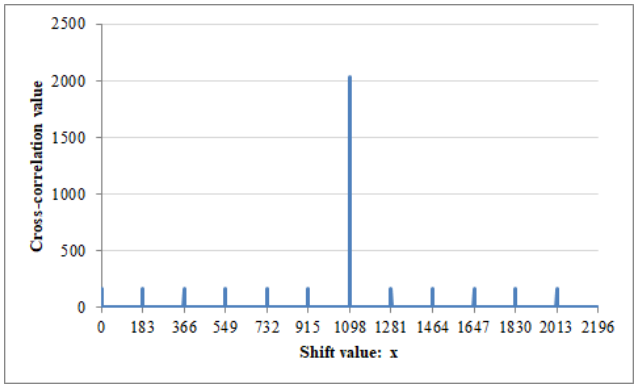


Figure 6: $|R_{S_6,S_7}(x)|$ with $p = 13, m = 3$, and $A = 6, 7$.

5 Flexibility of the Proposed Sequence and Its Application

Although nowadays multi-value sequence does not have enough application except the binary sequence especially in security applications. Therefore, the authors emphasize the flexibility of their proposed sequence to make it more worthy. To do so, the authors in this section, explains how to transform their proposed sequence into a binary sequence. In addition, this section also describes a comparison with our previous work [24] in terms of the linear complexity and distribution of bit patterns like crucial properties from the experimental viewpoint.

5.1 Proposed Binary Sequence

Binary sequences are extensively used in numerous applications (especially in cryptography). Although the authors proposed sequence is a multi-value sequence, it can be easily mapped into a binary sequence by setting the parameter value $k = 2$ and using the mapping function $M_2(\cdot)$. As mentioned previously, the proposed multi-value sequence is a well balanced sequence, in other words, all of the sequence coefficients (except the 0) appears same in number. To maintain the same property, the authors utilized the following algorithm (Algorithm 1) to make a uniform binary sequence from the well balanced multi-value sequence.

To make it binary sequence, a mapping function (intro-

duced in the following algorithm) is defined as,

$$M_2(s) = \begin{cases} 0 & \text{if } s = 0 \text{ or QR}, \\ 1 & \text{otherwise}, \end{cases} \tag{19}$$

By utilizing the parameters $p = 7$, $k = 2$, $f(x) = x^2 + 5x + 5$, and $A = 2$ the multi-value sequence becomes as,

$$\begin{aligned} \mathcal{S}_2 = \{&2,2,4,4,0,1,1,2,2,2,2,1,1,2,1,4,4,2,4,4,4,4,2,1,0,1,4, \\ &1,1,0,4,1,2,4,1,4,4,4,2,1,1,2,0,4,4,0,0,2,4,1,4,2,0\}. \end{aligned} \tag{20}$$

Algorithm 1 Generating procedure of the proposed uniform binary sequence

1: generate primitive element $\omega \in \mathbb{F}_{p^m}$
2: initialize flag $f \leftarrow 0$
3: **for** $i = 0$ to $p^m - 2$ **do**
4: compute ω^i
5: $a \leftarrow \left(\mathrm{Tr}\left(\omega^i\right) + A/p\right)$
6: **if** $a = 0$ and $f = 0$ **then**
7: $a \leftarrow 1$, $f \leftarrow 1$
8: **else**
9: $a \leftarrow p - 1$, $f \leftarrow 0$
10: **end if**
11: $s_i = M_2(a)$
12: **end for**

After using the above algorithm, the well balanced multi-value sequence in Eq. (20) can be transformed into a binary sequence as follows.

$$\begin{aligned} \mathcal{S}_2 = \{&0010000101110011111110001 \\ &1001001111000101001101 00\}. \end{aligned} \tag{21}$$

5.2 Comparison with Our Previous Work

Before applying any sequence in some security application, a lot of sequence properties needs to be well studied such as its period, autocorrelation, cross-correlation, linear complexity, distribution of bit patterns, and so on. The authors already discussed the former three properties. Additionally, this paper also includes a comparison with our previous work [24] regarding the linear complexity and distribution of bit patterns like crucial properties from the experimental viewpoint. It should be noted that from here on the authors previous sequence proposed in [24] will be called as NTU (Nogami-Tada-Uehara) sequence.

Linear Complexity

The linear complexity is an important measure to judge the unpredictability of a sequence. Thus, before recommending a sequence for any security application, its linear complexity needs to be well-studied. The linear complexity of the proposed sequence (binary case) and NTU sequence (previous sequence) having a period of 2400 are shown in Figure 7 and Figure 8, respectively. According to the comparison result, it is found that in both cases the linear complexity reaches to their maximum value. Since the M-sequence has the minimum linear complexity [25], on the other hand, the Legendre sequence has the maximum linear complexity [8]. It should be noted that the proposed sequence for being a well balanced sequence, its linear complexity reaches to its maximum value.

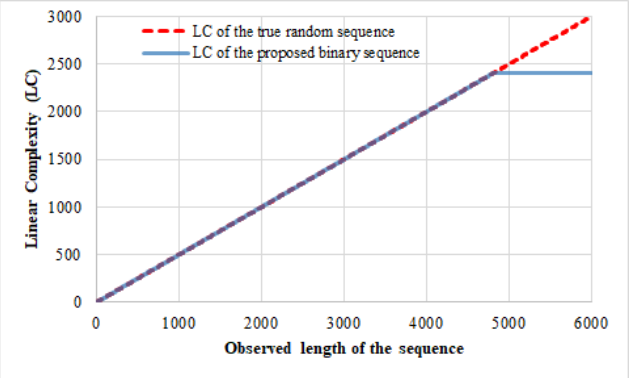


Figure 7: Linear complexity of the proposed sequence (binary case).

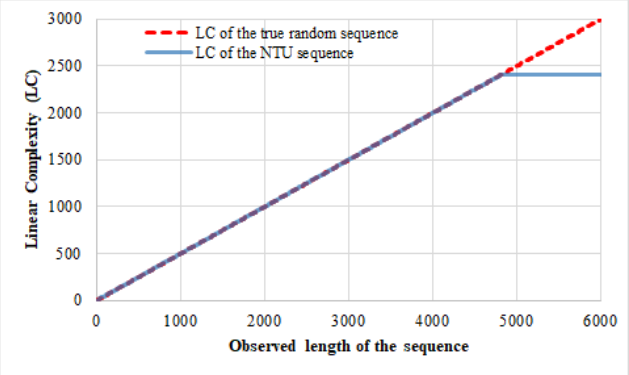


Figure 8: Linear complexity of the NTU sequence.

Distribution of Bit Patterns

The randomness of a sequence can be evaluated by observing the distribution of bit patterns of it. The bit pattern of the proposed sequence (binary case) and NTU sequence (previous sequence) having a period of 117648 are shown in Table 3. In the following table, the notations $n, b^{(n)}, Z\left(b^{(n)}\right)$, and $D_S\left(b^{(n)}\right)$ means length of a bit pattern, specific bit pattern, number of zeros in $b^{(n)}$, and appearance of $b^{(n)}$ in numbers in a sequence period, respectively. According to the comparison result, it is found that the distribution of bit pattern of the proposed binary sequence is almost uniform compared to NTU sequence.

The authors applied $k = 2$ and $M_2(\cdot)$ mapping function to make a uniform binary sequence from the well balanced multi-value sequence. It should be noted that after applying such a mapping function and uniformization algorithm, the sequence properties remains almost the same. For instance, only a small amount of change in the peak values regarding the autocorrelation, linear complexity remains the maximum, and distribution of bit patterns becomes almost uniform. In other words, they exhibit almost the same properties. It means the authors proposed sequence possesses a great flexibility.

Table 3: A comparison of distribution of 4-bit patterns between the proposed sequence (binary case) and NTU sequence.

n	$b^{(n)}$	$Z(b^{(n)})$	(proposed sequence) $D_S(b^{(n)})$	(NTU sequence) $D_S(b^{(n)})$
1	0	1	58824	67227
	1	0	58824	50421
2	00	2	28852	38415
	01	1	29972	28812
	10	1	29972	28812
	11	0	28852	21609
3	000	3	13927	21951
	001	2	14925	16464
	010	2	15066	16464
	011	1	14906	12348
	100	2	14925	16464
	101	1	15047	12348
	110	1	14906	12348
	111	0	13946	9261
4	0000	4	6680	12543
	0001	3	7247	9408
	0010	3	7391	9408
	0011	2	7534	7056
	0101	2	7402	7056
	0010	3	7664	9408
	0111	1	7631	5292
	0100	3	7275	9408
	1000	3	7247	9408
	1001	2	7678	7056
	1010	2	7675	7056
	1011	1	7372	5292
	1100	1	7523	5292
	1101	1	7383	5292
	1110	1	7275	5292
	1111	0	6671	3939

As far the authors know, there are a lot of considerations to use multi-value sequence in communications from the viewpoint of correlation [11, 12]; however there are few papers regarding the usage of pseudo random sequence with a long period, high linear complexity, and good distribution of bit patterns in security applications. The most typical security application of the pseudo random binary sequence will be the XOR-based stream cipher. First of all, in such an application, the same key is used for both encryption and decryption. Thus, each user should have a different key. In this case, these keys should have a minimum cross-correlation property compared to each other. Under this circumstance, it is important to discuss the cross-correlation property between several sequences along with linear complexity and distribution of bit patterns properties. The authors briefly introduced a use case in the following section of their proposed well balanced sequence in this paper, to emphasis on its usability.

5.3 Application

One of the most common applications of the pseudo random sequence (binary case) is in a stream cipher. Basically, a stream cipher is divided into two classes: block cipher and stream cipher. Among these a block cipher uses the same key for both encryption and decryption of each block (≤ 64 bits) of data. On the other hand, in case of a stream cipher, encryption and decryption are performed by the bit wise $\oplus$ (XOR) operation with a key stream. Here, the authors restrict the discussion of their proposed pseudo random binary sequence in a stream cipher. An image of the stream cipher is shown in Figure 9. Few important considerations during the design of a stream cipher are the key (which used for both encryption and decryption) should have a long period, good randomness, and unpredictability properties due to the usage of the same key in both encryption and decryption. Here, the encryption is carried out by applying a bit-wise $\oplus$ (XOR) operation between the plain-text of byte stream M and encryption key K. Then, the cipher-text C is transmitted through a network. On the other hand, during the decryption, after the bit-wise $\oplus$ operation between the cipher-text C and the same key K, we will get the original plain-text M. In a stream cipher, a lot of sequences are assigned to several users, respectively. If these sequences have some correlation, then it will make some security vulnerabilities. Under this circumstance, it is important to observe the cross-correlation property between several sequences. Additionally, its linear complexity and distribution of bit patterns needs to be high and uniform, respectively to confirm its randomness. Although the authors proposed sequence is a well balanced sequence, it can be easily mapped into a binary sequence with a long period, typical auto and cross-correlation, high linear complexity, and almost uniformly distributed bit patterns features. After observing the experimental and comparison results, it can be concluded that the authors proposed well-balanced sequence (binary case) can be a prominent candidate for a stream cipher like applications.

6 Conclusion

The authors have proposed a multi-value sequence (including a binary sequence) which defined over the odd characteristic field. The k-th power residue symbol utilized in this paper which is an extension of the Legendre symbol. Additionally, the proposed sequence also includes the case of the signed binary sequence. Prominent features regarding a sequence for instance, its period, autocorrelation and cross-correlation of the proposed sequence discussed based on experimental results along with a theorem (by which the value of the correlation can be explicitly given). In addition, the authors also introduced the flexibility of their proposed sequence by making a binary sequence from a well balanced multi-value sequence. Furthermore, a comparison result regarding the linear complexity and distribution of bit patterns properties are also included in this paper. According to the comparison results, the authors proposed well balanced sequence holds better properties compared to our previous sequence.

As a future work, the more efficient calculation will be introduced for instance power residue symbol needs exponentiation calculation.

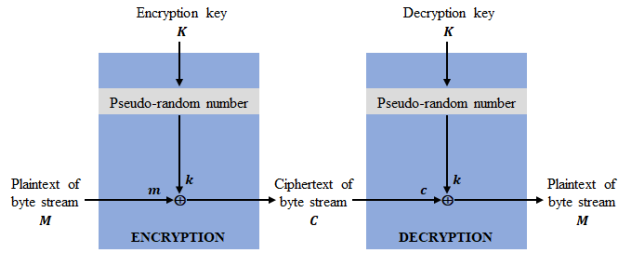


Figure 9: Application of the proposed sequence (binary case) in stream cipher.

Conflict of Interest The authors declare no conflict of interest.

References

[1] M. Luby, Pseudorandomness and Cryptographic Applications, Princeton Univ. Press, 1996. https://doi.org/10.1016/0898-1221(96)87348-2

[2] T. W. Cusick, C. Ding, and A. Renvall, Stream Ciphers and Number Theory, North-Holland Math. Library, North-Holland Publishing Co., Amsterdam, 1998. https://doi.org/10.1016/s0924-6509(98)x8001-3

[3] S. W. Golomb and G. Gong, Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar, Cambridge University Press, NY, 2005. https://doi.org/10.1017/CBO9780511546907

[4] A. Kinga, F. Aline, and E. Christain, "Generation and testing of random numbers for cryptographic applications", Proc. of the Romanian Academy, **13**(4), pp.368–377, 2012.

[5] M. Matsumoto, and T. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator", ACM Transactions on Modeling and Computer Simulation, **8**(1), pp.3–30, 1998. https://doi.org/10.1145/272991.272995

[6] L. Blum, M. Blum, and M. Shub, "A simple unpredictable pseudorandom number generator", SIAM Journal on Computing, **15**(2), pp.364–383, 1986. https://doi.org/10.1137/0215025

[7] J. S. No, H. K. Lee, H. Chung, H.Y. Song, and K. Yang, "Trace representation of Legendre sequences of mersenne prime period", IEEE Transactions on Information Theory, **42**, pp.2254–2255, 1996. https://doi.org/10.1109/18.556617

[8] C. Ding, T. Helleseth, and W. Shan, "On the linear complexity of Legendre sequences", IEEE Transactions on Information Theory, **44**, pp.1276–1278, 1998. https://doi.org/10.1109/18.669398

[9] A. M. Arshad, Y. Nogami, C. Ogawa, H. Ino, S. Uehara, and R. H. Morelos-Zaragoza, "A new approach for generating well balanced pseudo random signed binary sequence over odd characteristics field", International Symposium on Information Theory and Its Applications (ISITA), Monterey, CA, USA, 2016. ISBN 978-4-88552-309-0.

[10] A. M. Arshad, Y. Nogami, H. Ino, and S. Uehara, "Auto and cross correlation of well balanced sequence over odd characteristic field", Fourth International Symposium on Computing and Networking (CANDAR),Hiroshima, Japan, 2016. https://doi.org/10.1109/candar.2016.0109

[11] Y. K. Han and K. Yang, "New M-array sequence families with low correlation and large size", IEEE Transactions on Information Theory, **55**(4), pp.1815–1823, 2009. https://doi.org/10.1109/tit.2009.2013040

[12] N. Y. Yu and G. Gong, "Multiplicative characters, the weil bound, and poly phase sequence families with low correlation", IEEE Transactions on Information Theory, **56**(12), pp.6376–6387, 2010. https://doi.org/10.1109/tit.2010.2079590

[13] R. Lidl and H. Niederreiter, Finite Fields (Encyclopedia of Mathematics and Its Applications), Cambridge Univ. Press, 1996. https://doi.org/10.1017/CBO9780511525926

[14] E.R. Berlekamp, Algebraic Coding Theory, Aegean Park Press, Revised edition, 2014. https://doi.org/10.1142/9407

[15] M. Joye and B. Libert, "Efficient cryptosystems from 2^k-th power residue symbols", in Proceedings of Eurocrypto 2013, Johansson T. et al. Ed., ser. LNCS, vol. 7881. Berlin, Heidelberg: Springer, pp.76–92, 2013. https://doi.org/10.1007/978-3-642-38348-9_5

[16] P. V. Kumar and O. Moreno, "Prime-phase sequences with periodic correlation properties better than binary sequences", IEEE Transactions on Information Theory, **37**(3), pp.603–616, 1991. https://doi.org/10.1109/18.79916

[17] P. V. Kumar, T. Helleseth, A. R. Calderbank, and A. R. Hammons, "Large families of quaternary sequences with low correlation," IEEE Transactions on Information Theory, **42**(2), pp. 579–592, 1996. https://doi.org/10.1109/18.485726

[18] D. Hertel, "Cross-correlation properties of perfect binary sequence", in Sequences and Their Applications SETA 2004, T. Helleseth et al. Ed., ser. LNCS, vol. 3486. Berlin, Heidelberg: Springer, pp.208–219, 2005. https://doi.org/10.1007/11423461_14

[19] D. V. Sarwate and M. B. Pursley, "Cross-correlation properties of pseudorandom and related sequences", Proceedings of IEEE, **68**(5), pp.593–619, 1980. https://doi.org/10.1109/proc.1980.11697

[20] Y. T. Kim, M. K. Song, D. S. Kim, and H. Y. Song, "Properties and cross-correlation of decimated sidelnikov sequences", IEEE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, **E97-A**(12), pp.2562–2566, 2014. https://doi.org/10.1587/transfun.e97.a.2562

[21] G. Gong, "New designs for signal sets with low cross-correlation, balance property and large linear span: GF(p) case", IEEE Transactions on Information Theory, **48**(11), pp.2847–2867, 2002. https://doi.org/10.1109/tit.2002.804044

[22] R. A. Rueppel, "Linear Complexity and Random Sequences", F. Pichler (Ed.): Advances in Cryptology-EUROCRYPT'85, Springer, LNCS 219, pp. 167-188, 1986. https://doi.org/10.1007/3-540-39805-8_21

[23] A. Alecu and A. Salagean, "Modified Berlekamp-Massey Algorithm for Approximating the k-Error Linear Complexity of Binary Sequences", IMA Conference on Cryptography and Coding (S.D. Galbraith, Ed.), Springer, LNCS, vol. 4887, pp. 220–232, 2007. https://doi.org/10.1007/978-3-540-77272-9_14

[24] Y. Nogami, S. Uehara, K. Tsuchiya, N. Begum, H. Ino, and R. H. Morelos-Zaragoza, "A Multi-value Sequence Generated by Power Residue Symbol and Trace Function over Odd Characteristic Field", IEICE Trans. on Fund. of Electronics, Communications and Computer Sciences, **E99.A**(12), pp.2226–2237, 2016. https://doi.org/10.1587/transfun.e99.a.2226

[25] T. Moriuchi and S. Uehara, "Periodic sequence of the maximum linear complexity simply obtained from an m-sequence", IEEE International Symposium on Information Theory (ISIT), 1991. https://doi.org/10.1109/isit.1991.695231

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 190-193 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Photoluminescence Investigation of Inhomogeneous Porous P-type Si

Róbert Brunner[1,*], Pavel Vojtek[2], Zuzana Zábudlá[2], and Emil Pinčík[1]

[1]*Institute of Physics SAS, Dúbravská cesta 9, 845 11 Bratislava, Slovak Republic*

[2]*Department of Experimental Physics, Faculty of Mathematics, Physics, and Informatics, Comenius University, Mlynská dolina F2, 842 28 Bratislava, Slovak Republic*

ARTICLE INFO

Article history:
Received:20 March, 2019
Accepted:22 May, 2019
Online: 18 June, 2019

Keywords:
Photoluminescence
Inhomogeneous film
Porous Si
Electrochemical etching

ABSTRACT

Photoluminescence (PL) of inhomogeneous porous silicon (PS) of p-type is investigated in this contribution. We measured the PL signal at equidistant positions separated by 0.05 mm in area localized between original crystalline Si (c-Si) wafer surface and electrochemically prepared PS layer. Two PL peaks localized at energies 1.8 and 1.9 eV were identified and their parameters were determined. The changes of PL maxima energy ("blue shift") in dependence on position were observed.

1. Introduction

Researching porous silicon (PS) structures is time-intensive. Extensive collection of results is presented e.g. in [1], but the interpretation of the origin of photoluminescence (PL) signals is still a problem.

Structural characteristics and luminescence of PS prepared by chemical etching were studied by Korsunskaya et al. in [2]. The luminescence spectrum of this type of PS was the result of superposition of two PL bands. The first band represents excitonic recombination in amorphous Si nanocluster smaller than 3 nm. The second band, which prevails at room temperature, is generated by the recombination of charge carriers mediated by the defects in silicon oxide.

Silicon nanocrystallines (NCs) are included in the PS in the form of nanowires on monocrystalline silicon surface composed of different phases of crystalline and amorphous Si. They are covered with oxides (SiO_y) and hydrides (SiH_x) (Tynyshtykbaev [3]). Authors concluded that the light emission originates in hydride coverages of nanocrystallines. In order to explain the character of PL centers, various models have been elaborated (see e.g. Lenshin [4], experimental analysis of aging effects). The model, in which the maximum of PL localized at 640 nm is related to defective levels os SiH_x and SiO_y components, is greatly accepted. The presence of groups SiO_xH_y, observed using FTIR spectroscopy in IR region, was reported in the work [5]. Only small influence of PL on surface morphology was found.

The homogeneity of effective PL from PS was studied by Kayahan [6]. Imaging spectroscopy technique was used by author in order to examine both the effect of atmospheric aging and light illumination applied during anodization process on spatial distribution of PL intensity. He concluded that chemical structure of PS surface plays role in PL and that quantum confinement effect related to surface states can be understood to be a source of luminescence.

Chan [7] observed dependence of energy of PL spectrum on size of pores in PS. Their size was controlled by concentration of HNO_3 in etching liquid. Observed "blue shift" of peaks was attributed to quantum confinement effect.

In this work are presented results of our PL experiments on PS. For evaluation of PL signal we used the decomposition procedure based on the fitting of the spectrum by the set of Gaussian peaks. The proper number of the peaks was estimated from the residuum of the fit (Brunner [8]).

2. Experimental Part

The contribution topic is to investigate the PL signal recorded on an inhomogeneous PS sample. The p-type boron doped Si wafers (100) with resistivity of 8 – 10 Ωcm and thickness 600 –

* Róbert Brunner, Email: robert.brunner@savba.sk

650 nm were used in the experiment. Standard PS was prepared by electrochemical etching process in the solution of HF and MeOH without illumination. Used apparatus is schematically depicted in the Figure 1. Electrical voltage 6V – 11V was applied through mask (Fig. 1) between the exposed area of the Si sample (positive potential) and Pt negative electrode. A constant current of 30 $mAcm^{-2}$ was maintained during the few minutes of the etching process. The PS was rinsed in ethanol and annealed in nitrogen at 250°C for 30 minutes.

Optical and electrical parameters of PS layer depend on the type and size of the formed crystallites that are controlled by etching process parameters. The mask affects current density in vicinity of their boundary. This shielding effect leads to generation of transient area between exposed and non-exposed part of surface of silicon wafer.

Measurements were performed in this transient area in 17 points separated by 0.05 mm starting from initial crystalline surface (after polishing and cleaning procedures) to prepared PS layer as shown in the Figure 2.

The PL of the sample was measured at ambient temperature. PL spectra were decomposed into Gaussian peaks (this process is explained in more details in [8, 9, 10]). Usage of Gaussian peaks is suitable for sufficiently homogeneous structures composed of domains (layers). Photon emission is considered as an event which is independent of other emission acts. This independence makes possible to use the central limit theorem, leading directly to the Gaussian profile of the PL peak. The presence of light emitting centers of different types in the sample gives the spectrum composed of a set of Gaussians.

The observed spectra were fitted in order to find parameters of the set of Gaussian functions (energy, intensity and width given as FWHM) and to calculate constant bias. This bias represents both noise and signal offset contribution. Root-mean-square (RMS) value was used in the role of the object function in the fitting process.

The calculated parameters of PL peaks (without bias) were statistically tested using bootstrap method (e.g. [11]) based on measured data set resampling. Confidence intervals of peak parameters were estimated using BCa percentile method ("bias-corrected and accelerated"). This version reduces the requirements on number of the used bootstrap data set. In our calculations we used 500 resampling procedures for each sample. Standard 5% confidence limit was used for confidence intervals.

3. Results and Discussion

Estimation of optimal number of PL peaks during fitting process was based on size of residuum. Figure 3 shows how residuum depends on the number of used peeks. Well visible "knee" indicates the optimal model. This approach is explained in more detail in [8] (but briefly speaking, usage of too small number of peaks omits relevant parts of the PL signal, whereas redundant peaks fit mainly the noise). In this case, two relevant PL peaks were found: ~1.8 eV and ~1.9 eV.

Figures 4 – 6 show dependence of parameters of PL peaks (in successive steps energy, intensity and FWHM). Signal was measured at ambient temperature. The visible dependence of parameters of peaks on the position of the measurement point is attributed to the lateral non-homogeneity of the PS layer in the transition region (Fig. 2). Calculated confidence intervals are slightly non-symmetrical. Possibility to evaluate such data sets is the main advance of the bootstrap method, because the approach based on the normal error distribution would be less effective.

Several trends were observed:

Increase of the PL peak energy in dependence on position of measurement point is visible in Figure 4. We suppose that this may be the consequence of the size effect caused by the etching of nanostructural elements. Namely it is related to the SiO_xH_y compounds covering the structure's surface. The presence of the SiO_xH_y has been confirmed in the works [5, 12] using FTIR spectrometry of the samples of this type in wavenumber range 790-1250 cm^{-1} (0.098 - 0.155 eV).

A noticeablw changes of power of PL signal in dependence on position of measured point can be seen in Figure 5. This is mainly the consequence of increased thickness of light-emitting PS layer in combination with interference of exciting light during PL experiment, but the size effect (controlling light conversion effectivity) can take place in this process also.

Observed changes of FWHM of peaks 1 and 2 on Figure 6 are close to zero and may be non-significant in comparison with the width of confidence interval. Otherwise the increase of FWHM may indicate the inhomogeneity of the sample.

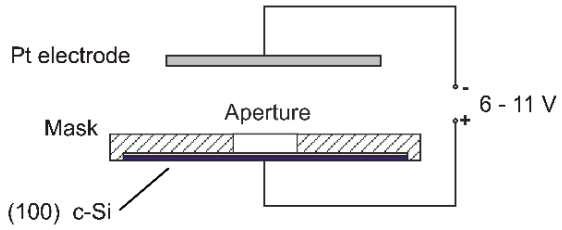


Figure 1: Experimental setup

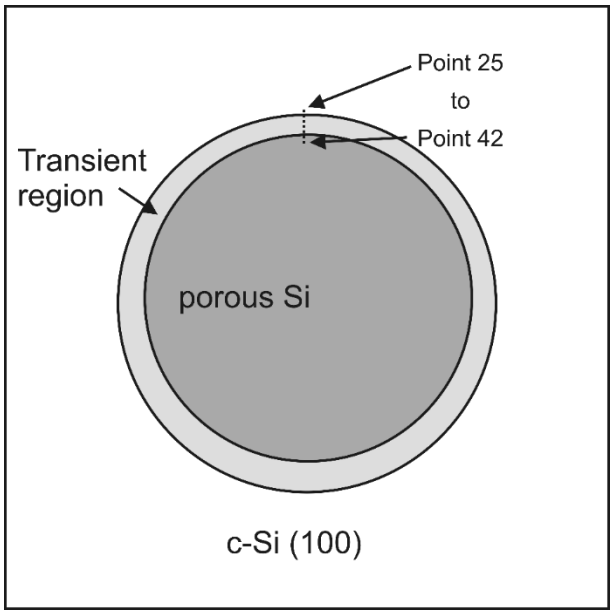


Figure 2: Sample geometry and measured area

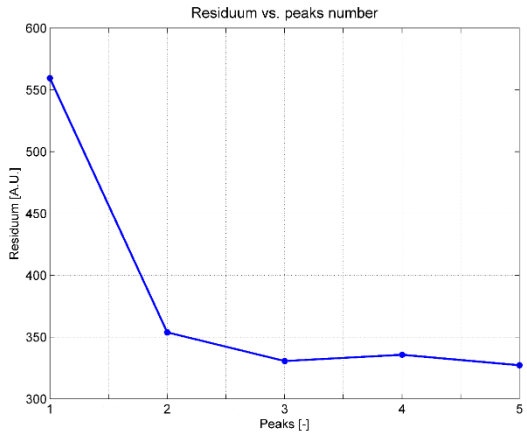


Figure 3: Residuum as a function of number of peaks

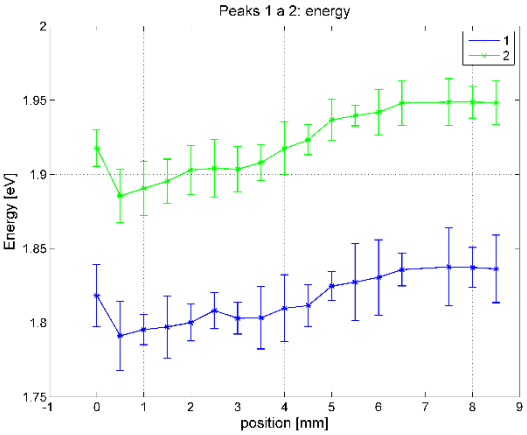


Figure 4: Dependence of energy on the position. Confidence intervals at confidence level α=0.05 are given

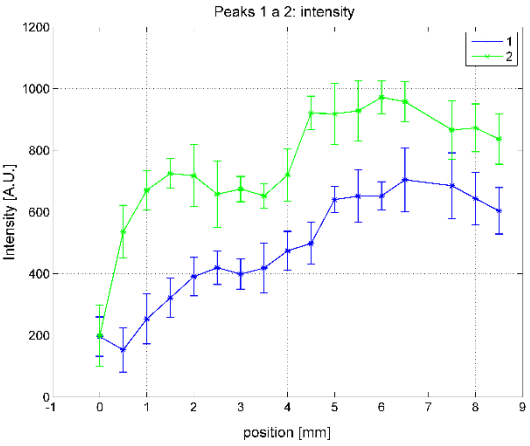


Figure 5: Dependence of intensity on the position. Confidence intervals at confidence level α=0.05 are given

4. Conclusion

Visible increased "blue shift" of both PL maxima (1: at 1.9 eV, 2: at 1.8 eV) occurred in all records in dependence of location of measuring point has been observed. We suppose that this effect is mainly related to the SiO_xH_y compounds that cover the PS structure's surface. This structure was modified during preparation of the sample by the non-homogeneous density of the current in transient region (in the vicinity of the mask boundary, see Fig. 1).

The corresponding FWHM of the peak marked as 1 is decreasing, whereas the value of the peak 2 is increasing in dependence on the position of measuring point. These effects are weak, however, the increase of the FWHM parameter may indicate the degradation of the sample homogeneity.

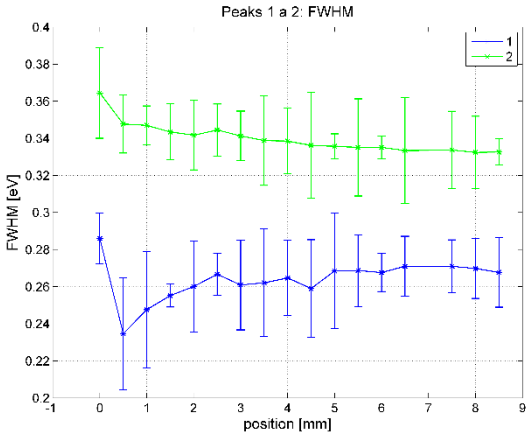


Figure 6: Dependence of width on the position. Confidence intervals at confidence level α=0.05 are given

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The following agencies partly financially supported this contribution: APVV – project No. 15-0152 and VEGA – projects Nos. 1/0900/16, 1/0676/17, 2/1061/18, and 2/0149/18.

References

[1] Porous Silicon – from Formation to Application, Volume One, edited by Ghenadii Korotcenkov, CRC Press, Taylor and Francis Group, Boca Raton, London, New York, 2016.

[2] N.E. Korsunskaya, T.R. Stara, L.Yu. Khomenkova, K.V. Svezhentsova, N.N. Melnichenko, and F.F. Sizov, The nature of emission of porous silicon produced by chemical etching, Semiconductors, 44, 79-83, 2010.

[3] K.B. Tynyshtykbaev, V. B. Glazman, M.A. Yeleuov, A.T. Isova, B.A. Rakymetov, D. Muratov and S.Z. Tokmoldin, Photoluminescence of Por-Si with High-ordered Mosaic Structure Received at Long Anodic Etching p-Si (100) in the Electrolyte with an Internal Current Source, Optics and Photonics Journal, 3(2B), 217-221, 2013. doi: 10.4236/opj.2013.32B051.

[4] A.S. Lenshin, V.M. Kashkarov, S.Y. Turishchev, M. Smirnov, E. Domashevskaya, Influence of Natural Aging on the Photoluminescence of Porous Silicon, Technical Physics, 57(2) , 305–307, 2012.

[5] M. Kopani, M. Mikula, D. Kosnac, P. Vojtek, J. Gregus, E. Vavrinsky, M. Jergel, E. Pincik: Effect of etching time on structure of p-type porous silicon, Appl. Surf. Sci. 461, 44-47, 2018. doi: 10.1016/j.apsusc.2018.04.228.

[6] E. Kayahan, Investigation of Photoluminescence Homogeneity Distribution of Nano-Porous Silicon by Imaging Spectroscopy, Journal of Natural and Applied Sciences, 21(3), 738-742, 2017. doi: 10.19113/sdufbed.19374.

[7] K.S. Chan, Dwight T.J.E, Photoluminescence, morphological and electrical properties of porous silicon formulated with different HNO3 concentrations, Results in Physics 10, 5–9, 2018, https://doi.org/10.1016/j.rinp.2018.04.011.

[8] R. Brunner, E. Pinčík, H. Kobayashi, M. Kučera, M. Takahashi, J. Rusnák, On photoluminescence properties of a-Si:H-based structures, Appl. Surf. Sci. 256, 5596-5601, 2010. http://dx.doi.org/10.1016/j.apsusc.2010.03.036.

[9] E. Pincik, R. Brunner, H. Kobayashi, M. Mikula, P. Vojtek, J. Greguš, Z. Zabudla, K. Imamura, P. Švec Jr., The photoluminescence of multicolor

silicon, Journal of the Chinese Advanced Materials Society, 4(2), 152-171, 2015- http://dx.doi.org/10.1080/22243682.2016.1151374.
[10] M. Takahashi, T. Fukushima, Y. Seino, W.-B. Kim, K. Imamura, and H. Kobayashi, Surface Structure Chemical Transfer Method for Formation of Ultralow Reflectivity Si Surfaces, J. Electrochem. Soc. 160(8), H443-H445, 2013. doi:10.1149/2.044308jes.
[11] B. Efron, R.J. Tibshirani: An Introduction to the Bootstrap. Chapmann&Hall, New York, 1993.
[12] M. Kopani, M. Mikula, D. Kosnac, J. Gregus, and E. Pincik: Morphology and FTIR spectra of porous silicon, Journal of Electrical Engineering 68, 53-57, 2017. doi: 10.1515/jee-2017-0056.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 193-206 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Prediction of Non-Communicable Diseases Using Class Comparison Data Mining

Ammar Al-Dallal[*,1], Amina Al-Moosa[2]

[1]*Ahlia University, Computer Engineering Department, P.O. box 10878, Bahrain*

[2] *Ahlia University, Information Technology Department, P.O. box 10878, Bahrain*

ARTICLE INFO

Article history:
Received: 30 April, 2019
Accepted: 07 September, 2019
Online: 08 October, 2019

Keywords:
Communicable disease
Data mining
Diabetes
Predictive data mining
Heart disease
Class comparison

ABSTRACT

Data mining is recognized as an effective technique for extracting and retrieving valuable information or decision from the vast available data. Because of the nature of the functionality of medical centers and hospitals, their data centers contain a collection of valuable information about their patients. By properly processing these data, different applications can be developed to utilize them. These applications could participate in predicting and diagnosing particular diseases. Two prime diseases realized to impact the overall health of society are heart diseases and diabetes. The presented work intends to develop and test a software application that helps doctors and practitioners predict the emergence of noncommunicable diseases (NCDs) such as diabetes and heart diseases. The application applies the predictive data mining model to the medical records which are collected from the Bahrain Defense Force Hospital (BDFH). The BDFH doctors evaluated the application and executed it on actual patients. The results obtained are accurately matching the expectation of doctors in BDFH. All kinds of risks are categorized appropriately according to the defined categories. As a conclusion, this application can help doctors in making proper decisions toward patient health risks. In addition, data mining is more supportive for the health sector and is essential for exploring the knowledge to be used in the health care sector.

1. Introduction

The term data mining refers to the process of knowledge discovery discipline in databases (KDD). It is associated with various computing fields; such as databases, artificial intelligence, as well as software engineering. It can also be used to overcome various problems that emerge where a large volume of data is involved.

The World Health Organization (WHO) is the world body that monitors international health. It showed that 68% of all mortality around the globe is due to noncommunicable diseases (NCDs), and the major killer NCDs are cardiovascular diseases (CVDs), diabetes, lung diseases, and cancer [1-2]. It added that heart diseases cause 12 million deaths globally. CVD comprises of various heart-related disease and its functioning problems that contribute significantly to the adult mortality in some countries.

The (popular) disease or new-generation disease is diabetes. It has emerged as a consequence of inactivity, fewer movements, sedentary lifestyle, and unhealthy eating habits. Given the low activities and other lifestyle reasons, the pancreas produces insufficient insulin to control blood sugar, resulting in diabetes. In a small country in the population such as Bahrain, these two NCDs present a multifaceted problem that must be addressed [3]. Hospitals and health centers accumulate a massive amount of patient's data. This data can be entered or fed to a data mining engine to help in predicting and diagnosing various types of diseases.

The accurate and efficient prediction is a consequence of medical diagnosis. Therefore, it is an important task at that stage. Unfortunately, not all doctors hold expertise for various specialization. In addition, the number of specialized practitioners is also found to be insufficient in many health centers, especially when it is related to heart diseases. Hence, there is a great need to develop a system that assists in predicting and forecasting the patient's diagnosis. It would certainly bring plenty of relief and help doctors.

[1] Corresponding author: Ammar Al-Dallal, Computer Engineering Department, Ahlia University, Manama, Bahrain. Email: aaldallal@ahlia.edu.bh, as_aldallal@yahoo.com

This research is designed to assess the data of the patient concerning his health which is necessary to forecast NCDs using inferential data mining and KDD to assists the healthcare professionals in evaluating, as well as predicting, the recurrence or occurrence of NCDs. As a result, this research offers a platform for doing such a system to be able to develop all sets of rules and make them available to medical doctors and other practitioners to anticipate with a degree of correctness the prospect of NCDs among patients. To be able to detect the prime elements which are vital for NCD prediction, the development of an application takes place to compile the massive volumes of patient data for forecasting the potential future trends by employing data mining. Afterward, its testing occurs, which allows the physicians and experts to use the information pertaining to NCDs for improving its diagnosis and reducing the diseases. This project utilizes huge amounts of data obtainable from the BDFH to anticipate NCDs by employing data mining technique.

The contribution that the study adds is that it develops a software which predicts the diagnosis and is tested by BDFH actual practitioners. Using the software, the practitioners were gratified with the study results.

1.1. Purpose of the Study

The purpose of the study is to apply KDD and execute the inferential data mining technique to examine health data, which is essential for predicting NCD. This assists doctors in analyzing or even predicting the recurrence or occurrence of NCDs in patients. Currently, BDFH has no tools or techniques to undertake this task. Hence, a Predicting NCD Application (PNCDA) software will be developed by applying the inferential data mining technique on the compiled huge volumes of available vital data. This application will be available for doctors to be able to predict future trends with accuracy, for the NCD potential among the BDFH patients. Following it, the application will be assessed for demonstrating the model which enables the clinicians and other stakeholders to have access to this facility and use it during their diagnosis process to improve patients' health and to reduce such diseases.

Because of the sensitivity of this project as it deals with confidential patient data, two types of data will be utilized: secondary and primary data. Secondary data will be extracted from the records of the BDFH, while primary data are gathered in person where the related data are organized as per the software requirements. The approach of the interview will be used for collecting primary data, which will recruit about 30 percent to 40 percent of doctors, nurses, as well as BDFH paramedics.

1.2. Study Significance

The objective of this project is to experiment as well as assess different algorithms of data mining, which will assist in NCDs prediction, comparison, and contrast of the effective ways of predicting the disease. The experiment is assumed to serve as an instrumental tool for the doctors for predicting in complex medical cases related to NCDs and advice their parents as per the predictions generated through accurate algorithms, that will have a huge advantage for the field of medical science. This research would provide evidence that the technique of data mining is effective for physicians for predicting and forecasting risky medical cases as practiced across developed nations. This research is set to act as a predecessor for accumulating data for patients in the Kingdom of Bahrain in the future.

1.3 Study Limitations

The study has certain limitations such as it includes patients recorded in BDFH only in Kingdom of Bahrain. Once completed, the findings of the research would be communicated and discussed with the healthcare authorities in the kingdom, where additional comprehensive research can be conducted for covering the whole Bahrain population.

This whole research is categorized into five sections; where section two presents a comprehensive review of the data mining methods and its implementation in the medical discipline. Following it, section 3 provides an explanation of the process for data mining, which is used for the prediction system for NCD. It also provides a description of the prediction software, which is developed. The developed software for prediction and its results are reflected and analyzed in section four. Lastly, section five briefly summarize the overall findings of the research and provides a direction to the future researches for expanding the research horizon.

2. Theoretical Background and Literature Review

In the present times, an increased acceptance of the data mining on the international forums is being recognized, across different medicine and life spheres. Considering the dynamics scope of data mining related to its efficacy for enhancing the healthcare outcomes, this section intends to highlight the theoretical basis which assists in NCD determination as well as the application of the data mining in forecasting.

2.1. Non-communicable Disease

At present, the unbridled growth of NCDs is recognized as the primary cause of mortality across the world. It is reasoned that the increase in sedentary lifestyle and lack of exercise has added to its increase. As it is well-recognized that an active lifestyle is vital for proper maintenance and functioning of the human body. According to 2016 WHO statistics, more than 15 million deaths occurred as a result of NCD such as diabetes and cardiovascular disease [1]. The most surprising and alarming finding of this statistics was that almost half of the deaths were of individuals who were less than 70 years of age.

A. Diabetes

The statistical evaluation of World Health Organization (WHO) findings, and its comparison with the outcomes of 2014, it is found that diabetes had an increasing percentage of about 409 percent for 34 years, ranging from 1980 to 2014. The main contributor to the increasing percentage is the changing world dynamics where the living status of the majority of the countries has changed for both developing and underdeveloped countries [1]

As per the healthcare professionals, the disease of diabetes occurs when the gland of the body (pancreases) is able to release an adequate amount of insulin. It affects the human body function as insulin is responsible for carrying sugar from the bloodstream to numerous cells in order to be used as energy. The deficiency of insulin in the body makes its natural functioning difficult. As a

consequence, the body releases high levels of glucose in the urine. Its prevalence for a long-term can cause an organ to fail, cause CVD, and affect the other functions of the body. The deteriorative effects which emerge as a consequence of diabetes are the reason it is ranked as the fourth primary NCD by WHO [2]. It can also lead to the complication of CVDs. The WHO organization [2] documented that diabetes and CVDs were the reasons for more than 11 million fatalities that occurred before the age of 70, as shown in Figure 1.

The primary factor which contributes to the occurrence of diabetes includes the uncontrolled level of blood glucose. Using data mining techniques, the healthcare professionals across the world would be able to forecast illness in an effective manner and will be able to integrate better management technique for patients at high risk. This adds to the significance and needs for the disease analysis and predictions for overcoming it and providing relief to the majority of the patients.

Similar to other regions, Bahrain also lists diabetes as a major health concern. The prevalence of diabetes is found across different ages and populations comprising of different characteristics. The report by WHO for Bahrain documented that diabetes account to 13 percent of the deaths in the region. It also indicates the increasing detrimental outcomes of diabetes, which continue to grow at an increasing speed [3].

Generally, diabetes is categorized into two types; namely, type 1 and type 2. In type 1, the diabetic patient is required to be infused with artificial insulin using medicines and injections. Whereas, in type 2, the body gland (pancreas) produces insulin but is inadequate for the body. Majorly, type 2 diabetes is more prevalent among patients across the world. The population that is generally found to be affected by it comprise of adults, particularly people, in the middle-aged group, however, with the changing lifestyle and dynamics, its prevalence is found in children also.

The low prevalence of Type 1 diabetes is due to its interconnectivity with the external environment, which affects the body insulin-releasing cells. Though, the change in the lifestyle such as regular exercise and maintenance of adequate body weight can help prevent diabetes of type 2.

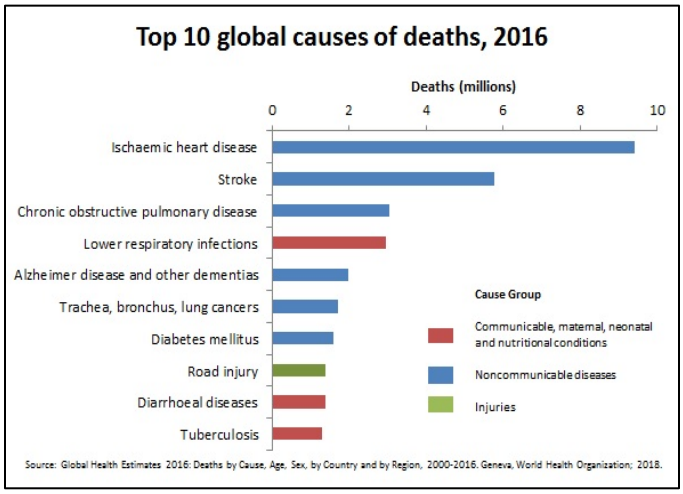


Figure 1: Primary Ten Causes of Death as Reported by WHO [1]

International Diabetes Federation has listed down following practices for preventing against diabetes:

1. It recommends that individuals must start a regime for losing some weight to overcome the effects of diabetes by enhancing insulin resistance and mitigating the prospects of hypertension. So, the people who are overweight are encouraged to sustain adequate body weight.
2. The consistency of physical exercise is integral for sustaining weight loss. It is because the indulgence in the physical activity overcomes arrhythmia, hypertension, and insulin sensitivity; improving the composition of the body; and developing psychological health.
3. Sustaining a healthy diet also overcomes the CVD related risk factors.
4. Smoking, depression, stress, and inadequate sleep can also be considered risky behaviors; therefore, it is preferable to avoid them.

B. Cardiovascular Diseases

Cardiovascular diseases comprise a number of disorders that are related to the heart and blood vessels for example;

Coronary heart disease which is caused by the clogging and narrowing or blockage of the blood vessels which supply the blood to heart muscles.

Cerebrovascular disease is caused by the issues in the arteries, which can affect the supply of blood to the brain. Example of the cerebrovascular disease is a stroke.

The peripheral arterial disease occurs due to plaque (calcium, fats, and other substances) that build up in the arteries supplying the blood to the head, limbs, and the other organs.

Rheumatic heart disease is caused because of the streptococcal bacteria, which attacks the tissues of the body, particularly of the brain and the heart.

Congenital heart disease is caused because of the anomalies of structure and the malformations at the time of birth.

Deep vein thrombosis and pulmonary embolism are caused due to the clotting of blood in the veins of the leg, which might be transferred to the organs like the lungs.

Myocardial infarction (MI) is the term which is used in the medicine for the NCDs in common, which also known as the heart attack. Cerebrovascular accident, which is also known as stroke, is also a kind of NCD and is a serious one. These illnesses or disease occur due to the defects and the faults of the heart and the arteries. The arteries are the vessels for the pumping of the pure blood from the heart into the body. Lack of exercise, bad eating habits, and fat accumulation in the body lead the fat cells to get deposited in the arteries' inner walls. Misuse of alcohol, the use of irregular tobacco, habits of eating, hypertension, and a host of circumstances and conditions lead to the CVA or MI.

Building up of fat inside the blood vessels and the arteries leading to the gradual vessel clogging. Without any treatment, medical care, or the changes in habits of eating or the lifestyles, it can lead to complete blocking of the blood flow. The unmanaged glucose levels in the blood, physical inactivity, and obesity are very common in the population nowadays. Formerly, people had a walking habit, they used to work in the farms, and the field were heavily involved in labor physically, which is rapidly reducing nowadays. Man, in search of comfort and the material gains, has

invented a lot of devices and machines, which has reduced the physical activity of humans. The emergence of the fast-food culture is also a danger to health and body, which most of us fail to understand.

If once diabetes, hypertension or the cholesterol imbalance is diagnosed, it is the high time to make sure that the person has the information regarding the upcoming dangers which might lead to the NCDs. These intermediate-risk factors defined by WHO must be highlighted by the primary health center and the clinics.

It is a general knowledge for the people who are having awareness that to stop or to reduce the dependency on the tobacco or the alcohol, to limit the oily and the greasy food that contains high fat and reducing sugar and the salt intake can help them in reducing the chances and the risks which are linked with the NCDs and the CVDs specifically. When a possible danger is detected, the patient should seek medical help immediately and make his lifestyles better. Further, people must use the proper medication for the control or restrict the damage that the factors can cause.

In most of the cases, there are no clear and visible indications for the MI and the stroke as well. It happens all of a sudden, and are not aware of it. Most of the MIs occur when people are sleeping. When the people feel the discomfort, and the pain in the chest or an extreme pain shoots up in the area of the right shoulder or jaws or the elbow, it can be considered as a warning that it can lead to developing of MI or CVA. The only option in this scenario is immediate medical attention.

The common symptoms which are linked with the CVAs are the insensitivity or the numbness in legs, arms, or face as well and it can be on one complete side of the body. There can be disorientation of speech, and even the difficulty in the vision as well. Severe dizziness or hallucinations or the headaches can happen, and in the intense cases, the patients got fainted too.

The third world is mostly affected by the NCDs; it is populated by middle and the low-income groups. The records of WHO speaks for themselves only. There is the accessibility to the medical care for the rich people while the people with the low or middle-income groups, if we look at the profiles of the country, have no or little access to the primary health care. In the countries across Southern America, Asia, and Africa, to approach the medical center is costly, which causes the late detection of NCD for the poor people.

According to the fact sheet of WHO [1], 26% of deaths in Bahrain are caused due to the CVDs. These statistics highlight the possibility of 13% deaths between the ages of 30 to 70 years and is caused by four NCDs. The physicians should record the factors of risks for the CVDs in order to reduce the strokes or the heart attacks by getting the right system for the storage of medical record and the analysis of the data.

C. Diabetes and cardiovascular disease: double jeopardy

Diabetes mainly contributes to the CVDs as proved by the clinical trials and the situation of the people who have suffered from CVDs. Due to the defects of crucial organs like kidneys and the liver, it becomes more difficult to pump the blood for the heart. The defects or the failure of kidney, liver, and the pancreas cause the threatening amount of toxins to sustain in the bloodstream.

The world of medicine has the challenge to handle this double-edged sword. This is the crucial time for the health centers and the governments to look at these two killers; CVDs and diabetes. To aid in reducing the deaths which are caused due to CVDs, data mining methods must be applied to predict the accurate occurrence or the reoccurrence of the CVDs and diabetes.

2.2. Evolution

Organizations nowadays produce a substantial amount of data specific to the institute. For improving the practicability concerning the use of data, researches have introduced algorithms which allow micro-focus on the data and position it as per the super-specific requirements. This advanced the efforts for the development and creation of the machine language algorithms which are useful in the analysis of the different analysis types and formation of decision without or little human supervision. Thus, the evolution of data mining is based on human needs, which assists in the identification of the relationship patterns and forecasting based on the presented layout of the program rules as well as stipulations.

The data mining is described by the researchers and practitioners using various terms. The concept of discovering knowledge from databases has been evolved from data mining. Earlier research of Fayyad et al.[4] has defined data mining as a procedure in which the data sets are implicitly and which reveal previously unidentified but significant information for effective decision-making. This whole procedure is termed as knowledge discovery in database (KDD).

This procedure can be applied in health care to predict the trends of many kinds of diseases and illnesses. Hence, instead of relying on the knowledge and experience, the data mining technique can be used by the doctors, more precisely for KDD to predict trends that would lead to better diagnoses.

KDD increases the efficiency and effectiveness of doctors by allowing them to treat a large number of patients at a given period. Moreover, such a system increases the opportunity for doctors of the same specialization across multiple firms, locations, as well as countries to e-share medical reports for devising best possible diagnoses in a time-effective manner.

The significance and usefulness of the data are evident from different perspectives. The logical alignment of the irrelevant data can be emphasized on the concealed or undiscovered correlations as well as patterns. This can provide valuable data which is critical for examining the individual as well as health being. The main notion is to briefly summarize the voluminous data and conclude its useful findings and information [5].

Data mining is regarded as a statistical interface, which is inclusive of other interference such as statistics, technology database, pattern recognition, data in machine-readable form as well as intelligent expert system [6].

Several definitions are set for data mining, which are raised according to the area of implication.

According to Krishnaiah et al. [7], the facility of data mining enables the use of data for identifying and using the data set trends. The primary findings of this database are to identify the

mechanical or automatic patterns, which require less input as well as efforts from the user.

Recently, the spectrum of data mining has enhanced, which is inclusive of artificial intelligence concepts which help in effective and fast-paced management and visualization of the data [8]. The other definition of data mining is provided by Han et al. [9], which states that it helps in the effective extraction of useful data.

Generally, the actual task of data mining is linked with the mechanical analysis of the voluminous amount of data, that is used to attain information which is not yet discovered. It assists in the identification of the data patterns, its categorization using the cluster analysis, odd records identification which require anomalies detection as well as its associated mining rule or dependencies.

Different statisticians' information system communities, as well as data analysts, use the data mining term. The procedure of KDD is recognized as complete, which requires the attainment of the data or discovery of new information. KDD core concept is of data mining; hence, it can be defined as an application which uses specific algorithms required for the analysis and extraction of the data patterns. It is also recognized as the KDD focal hub. It is inclusive of the intellectual approaches required for the data patterns extraction. From a healthcare perspective, the traditional methods are considered to integrate into statistical procedures for the process control comprising of numerous functions pertaining to the fundamental probability distribution which can be executed successfully for controlling the infection rate in hospitals [10].

Previously, in the 1960s, data mining was perceived to be an looked down by mathematicians as well as to statisticians alike. It was regarded as an unhealthy practice which improved its recognition using the term data fishing and data dredging, based on the hypothetical analysis of the data.

The terms database mining emerged in 1980 by HNC, a San Diego–based company, for describing the Database Mining Workstation [11]. With increasing years such as 1990, the term data mining was appeared to have been accepted as a legitimate phrase for describing the harvesting methods employed at the available data for forecasting the happenings in the future. Actually, different words were used for data mining, such as information harvesting, data archeology, knowledge extraction, information discovery, and so on [12].

The term "knowledge discovery in databases or KDD" was first initiated by Gregory Piatetsky-Shapiro and Parker [13]. Though, the relevance of data mining attained relevance as well as acceptance from the communities concerning artificial intelligence and machine learning. Furthermore, this term was also recognized within the business domain for the fourth estate communities. At present, both the terms, such as "data mining" and "knowledge discovery" are used interchangeably. Later in 2011, data mining was defined as data science.

2.3. Knowledge Discovery in Databases (KDD) Process

Knowledge discovery in databases (KDD) serves as a useful process that is used to extract important information from the expanded data. The information is gathered and filtered to extract only the required information. Information that has been collected is employed in the data set, which is further interpreted to attain a useful understanding regarding the given outcomes.

According to Maimon and Rokach [14], the target goals serve as important factors to successfully employ the KDD process. The process begins through fir, considering over the objectives, however, the final product is achieved in the form of newly developed information. This marks an end to the loop. Next involves data mining segments, where final output is achieved in the form of changes involved in the application domain. The results of the process are evaluated by employing fresh data sources while putting an end to the given loop. The finalizing of the process restarts the KDD process. Han et al. [9] elaborated the overall process through the given figure below:

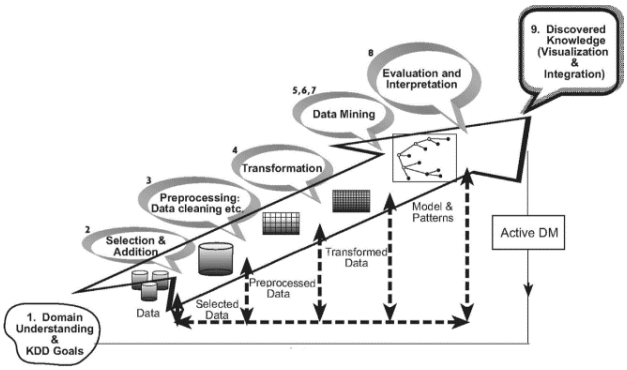


Figure 2. KDD process (Maimon & Rokach 2011)

Step 1. Develop a clear understanding of the application domains

The understanding related to the requirements of the healthcare professionals, and end-users, is of foremost value. This helps in providing a clear-cut vision to data miner in meeting the expectations while illustrating important knowledge regarding things that need explanation. It is further important to provide important knowledge regarding things that may help in achieving the desired objective. However, these requirements could be modified after the first round as the end-user might want to add additional functionalities. After identifying important knowledge, it is important for people contributing to the KDD process to function in accordance with the below-given steps. The steps are important for indicating preprocessing measures in the KDD process.

Step 2. Select data sets.

Selection of data sets is important to elaborate on the idea of the target goal. It further helps in providing important details regarding the type and the amount of data that is needed to achieve the target goal. The stage is of greater importance, since any wrong selection of data may lead to several complexities in preparing data. In such cases, the process may result in providing unimportant inferences that may weaken the overall effectiveness of the process.

Step 3. Perform preprocessing and cleansing.

Before processing the selected set of data, it must be filtered and cleaned. The idea is to improve and augment the reliability of the chosen set of data. The factor is important in providing maximum reliability to the selected data set. The cleansing

process includes removing noise or barriers that may impede processing; hence, eradicating the given barriers are crucial here. This might serve as a time-consuming process; still, it is effective in indicating maximum surety in terms of validity and reliability of data. In certain cases, insufficient data sets are achieved. However, situations, where ineffective data set, is intervened with accurate predictions, conscious efforts are required to attain expected outcomes. Consider an example; where irrelevant data of patient may act as a barrier in the overall process. Therefore, it is important to remove any such data at this stage

Step 4. Complete data transformation.

The next step is to make the data project-specific. This step ensures the provision of accurate data in a format that produces the required outcomes or results that will be utilized by the doctors to provide accurate predictions. The step is of significant value to provide valuable outcomes in any KDD project. However, important provisions in this regard may help in restoring the KDD process. Two significant methods, including; attribute transformation and dimension reduction are suggested by authors. KDD process serves as a useful tool in providing useful transformations. This helps in indicating maximum validity and efficacy to the formulated results.

Step 5. Choose the adequate data mining task.

Following the above procedure, the next task is to select the data mining type as per the nature of the project. This selection is based on the expected outcomes. Moreover, the selection between regression, and classifications or clustering is based on the desired outcome. Normally, the data mining outcome can be either forecasted, where data mining is done under controls of supervision, or description, where data mining occurs under visualization.

Step 6. Choosing the right algorithm for data mining.

This stage is reflected as the one which discusses the tactics to be used for obtaining the strategic objectives. Is it sufficient to use neural networks? Or is it better to use decisions? In fact, the selection is based on the searching pattern type which is found consistent to the given project and the desired end result. Does it relate to the results precision or decipherability or understandability that forms the project objective? The preference is more towards the neural networks when the precision of the results is required, while the appropriateness of the decision tree is found when gaining an understanding of the patterns and trends is required.

Step 7. Employ the algorithm of data mining.

In this step, the algorithm of data mining is implemented. It could be applied several times for obtaining the best result. For instance, the algorithm could be iterated using different controlling parameters, i.e., the low number of occurrences in a certain leaf or probably executed until the desired accuracy is obtained.

Step 8. Evaluate the mined data.

The mined data evaluation is the core of this KDD stage. This is primarily related to the reasonable interpretation and findings with respect to the discussed and defined goals of the project. This evaluation may promote the need to either add or remove a certain feature from the transformational stage of the data. This stage helps in achieving the usefulness of the data and its comprehension by the end user. Upon finalization, the found information is documented. When the found knowledge is being finalized by the KDD team, the whole data mining procedure is documented and assessed against the predetermined outcomes.

Step 9. Use of the discovered knowledge.

At this stage, the developed KDD process is evaluated for its effectiveness and efficiency by the end user, where refinement occurs in case the fine tuning of the data is required. Also, the created knowledge is being assessed for pilot-testing by practitioners and doctors for ascertaining the production of the desired outcomes. When a new product is being tested in a real environment, some conditions might variate from that at the laboratory. Therefore, these must comprise of the built-in ability to change, adapt, modify the derived knowledge in order to satisfy the end user.

2.4. Data Mining Techniques

Currently, an increased inclination of the researchers is concentrated on data mining. Generally, there are two types of data mining; namely, predictive model and the descriptive model. The two models are as follows;

A. Predictive Model

Prediction, as the name indicates constitutes of the correct envision of the future trend be logical computation of the data. The predictive model uses the previously available information for predicting future outcomes. This model is employed by various firms such as organizations who attempt to data mine the worth of an individual.

The predictive model data mining applies several techniques encompassing regression, classification, time series analysis, and prediction. This model is used for identifying the model that effectively matches the identified ideas or data sets. It is also helpful for class prediction of the objects when there is no availability of the class objects. The achieved model is primarily focused on the assessment of the identification classes set. To assess the numerical values and forecast, the statistical model of regression is used.

B. Descriptive model

This data mining model is employed to identify the data patterns to understand the relationship between the data attributes. The fundamental feature of the data is represented and summarized using the descriptive model. For instance, the customer database and identification of different sets can be useful to the marketers. The identified method can be employed for devising effective marketing programs for targeting audience.

The descriptive model applies several techniques such as clustering, summarization, association rules, and sequence discovery. In the clustering method, the analysis of the data object occurs without reference to the detected class label. Whereas, in the summarization, different properties and specialties of data values are to be considered as noise or outliers.

For the association rules, the same recreated rules are used for assessing the patterns and association among the different obtained characteristics. These are used for analyzing the data patterns for determining or forecasting their classification. Other systematic order identification is through sequence discovery which occurs for identifying the events in a timely manner rules and regulations. Figure 3 demonstrates the classifications of data mining.

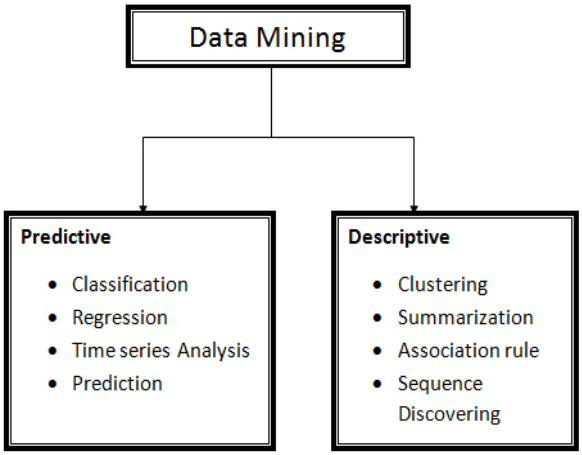


Figure 3. Tasks of Data mining

Among all the models presented in figure 3, the focus of the present study is on one named prediction-mining method.

2.5. Medical Applications Data Mining, Forecasting

Data mining has been widely used in medical applications to solve varies aspects. In this section we highlight some of them.

A. Health care informatics

Medical informatics is termed as dynamic because of the increasingly developing and shifting nature of both medical science and technology. According to Hersh and Hoyt [15], the discipline of health care informatics is related to the resources, materials, tools, and formal ways that are utilized for optimizing the storage, its management and retrieval of the needed biomedical information for effective decision-making, and problem-solving. Likewise, Athina and Konstantinos [16] highlighted that health care informatics employ modern communication, computer applications, as well as IT and computer system in medicine fields in the form of care provided, the delivered education, and research undertaken. Bronzino [17] stated that health care informatics uses different type of tool for using diverse which assists in utilizing and sharing information for improving the delivery of health care. Another term used is medical informatics whereas, it is also sometime referred by practitioners as clinical informatics or bioinformatics. Though, bioinformatics regard it as biological information analysis of the databases that are based on computer and statistical analysis of biological information in a speedy and improved manner. Clinical informatics concerns with maintaining data structure, data organization, and data search to make decisions using relevant data to conduct significant research in the medical discipline.

B. Data mining in medical discipline

Over the years, there is an increase in the computation analysis for deriving computer-based analysis and interference for improving the medical outcomes. The use of software and highly complex computation tools have increased. Previously, healthcare professionals majorly relied on their knowledge, skills, intuition and experience. However, at present, various channels have been developed across firms, regions, as well as states using both formal and informal sharing means for clinical experience which can be employed by the rest of the world. Data mining has provided great benefit to the medical field such as to forecast NCD and major other diseases.

This proliferation of the data mining has increased its efficiency in the medical discipline, where the use of data mining has become a norm where healthcare firms and facilities are being encouraged for meeting the need to predict the trends in disease and their occurrence. The significance of data mining, particularly for information management, is well established. Acharya and Yu [18] the developers of various computer-aided algorithms, have made the health facility more effective and have tremendously improved health care outcomes. Such as, it has aided in informatics, monitoring systems, as well as epidemiology. The difference between the pathological and normal data has facilitated improvement in the diagnosis and decision making.

Recent data mining techniques, as well as the essential statistical tools, can be used to assist doctors in diagnosing diabetes and CVDs. With the use of the statistical analysis, various CVDs encompassing stroke (cerebrovascular disease), cardiac arrests (coronary heart disease), congenital heart diseases, hypertension, peripheral artery disorder, arrhythmia, and heart failure with diabetes can be predicted. An overview of such systems is listed below.

Khaing [19] proposed an efficient approach for assessing and predicting the risk of heart disease. The primary stage is the execution of the k-means clustering algorithm with K=2 for extracting relevant data through clustering the heart disease database. By using k-parameter the number of fragments is mastered. Consequently, the extracted data of heart disease mine the frequent patterns with the use of maximal frequent itemset algorithm (MAFIA). Afterward, the execution of ID3 training algorithm occurs to produce the heart attack level following the decision tree. Although the accuracy of the findings obtained was 74% using k-mean–based MAFIA, there is a need to further increase this accuracy.

Many real-world medical data sets suffer from overlapping information. Overlapping k-means (OKM) is extended from the conventionally used k-means algorithm, and it is recognized as one of the most real-world medical data sets that inherent overlapping information. This clustering method enables one sample to be related to one or more than one cluster. Though, the issue of sensitivity in OKM is prevalent to the initial cluster centroids. Khanmohammadi et al. [20] proposed a hybrid method which assimilates the k-harmonic means and overlapping k-means algorithms (KHM-OKM). In this approach, the initialization of the cluster centers of OKM are obtained from the output of KHM. The performance of KHM-OKM is evaluated in

terms of completeness, homogeneity, and cluster size-quantity tradeoff, and it is found that it outperforms the OKM algorithm.

Feature selection is one of the integral steps in data mining. When selected properly, it improves prediction accuracy. Orthogonal Feature Extraction (OFE) is proposed by Jiang et al. [21], which uses the feature ranking techniques. The technique proposed is applied to improve cancer prediction accuracy. To make sure that the attributes of the selected features are accurate, they developed an algorithm that selects the linearly independent vectors iteratively that belong to top-ranked attributes. The results are compared with the analysis of principal component, neighborhood component, and linear discriminant. These three methods are outperformed by OFE in terms of computational complexity and performance.

Another heart disease prediction system is developed by Suvarna et al. [22]. This system combines optimization techniques and data mining. The optimization technique used is particle swarm optimization, which is modified by applying a constriction factor. In this inherently distributed algorithm, the solution is obtained through the interaction among several simple individuals called particles. The authors used *de facto* standard data sets for the reliability ranking of heart disease prediction. The three techniques used for comparing the result are: ID3, multilayer perceptron with backpropagation training and classification, and regression trees. The proposed approach outperforms these three techniques, producing accuracy of classification equal to 53.1%. But the accuracy is much lower than what is expected from actual practitioners.

Rairikar et al. [23] applied the backpropagation technique in genetic algorithm to develop a system that predicts. This prediction system uses 13 attributes obtained from cardiovascular disease information. The results are compared in terms of time complexity with two well-known classification of data mining techniques, which are KNN and Naive Bayes. The results showed that KNN is much better in its performance. However, this method is lagging measuring the accuracy of heart disease prediction.

Tomczak and Zieba [24] applied classification data mining in machine learning to solve two issues that appear when applying machine learning to medical diagnosis. One issue deal with the implementation of interpretable models such as decision tree and classification rules. The other issue deals with the imbalance between classes with high number of examples such as healthy patients and a class with low number of examples such as ill patients. The proposed model is a probabilistic combination of *soft rules*. These rules are constructed by introducing a new random variable called *conjunctive feature*. In addition, a new estimator is introduced to incorporate the knowledge about imbalanced data. This approach is tested using oncology data set. The results show that soft rules can perform well on data sets with small number of examples. In addition, the outcome is comparable with expert oncologists. Meanwhile, this work needs to be examined against large data sets.

Zhu and Fang [25] noticed that the current existing classification trees might not be able to properly provide classification for the provided predictor variables set, which may be categorized using a high error rate. To solve this, they proposed an algorithm that combines the logistic regression model and the trichotomous classification tree. This tree is used to split and establish the tree recursively until it meets the stopping rule of the tree splitting. The algorithm is tested on two real data sets: Pima Indian data set and Wisconsin Breast Cancer data set. The findings showed that the proposed algorithm is more accurate in predicting some diseases than the classification and regression tree. However, this technique suffers several limitations, including using binary response variable, the cost of variables used in classifications ignores the actual cost and thirdly, the variance-covariance between 2 categories use limited simulated normal distributions.

Several algorithms and methods such as regression, clustering, classification, neural networks, artificial intelligence, genetic algorithm, association rules, nearest neighbor, and decision tree have been executed by researchers to assist healthcare practitioners in the effective diagnosis of heart diseases or diabetes to prevent further diseases.

The above-mentioned research studies provide evidence for the effectiveness of the data mining concerning the field of NCDs prediction and correction. The objective of this research is to devise a practical model which can be executed by the healthcare practitioners in the main hospital of Bahrain, or where individuals belonging to particular demographics which adds to its generalizability across the worldwide health care centers.

What features the developed PNCDA is that it allows realistic and timely mean to forecast the general NCD forms, particularly CVDs and diabetes. Rather than using pre-defined datasets, this approach uses real data obtained form BDFH. Further, the developed PNCDA has been tested by actual practitioners from the mentioned hospital and the results met their expectation obtained from the conducted questioner as will be explained later.

3. Research Methodology

As the major outcome of this research is the prediction software system that aids the doctors and practitioners of the BDFH to effectively predict the NCDs of their patients, the first stage of this project is to identify the most effective factors and classify them. These factors must provide good input data (parameter) using an adequate model for data mining. These factors can be identified by the users of the proposed system. Three main factors are included in this research, such as, prediction methods for data mining, patient database, and software application implementation, which are used to characterize and devise a model to predict the cardiac arrest diagnosis as well as diabetes risk among the patients that are listed in the obtained BDFH data set. This research will use both forms of data (primary and secondary). It will integrate in data mining, its significance in healthcare, and medical field application for developing a software system that will be assessed in real situations for evaluating its performance.

3.1 Data Types

In this research, two types of data are used; secondary data and primary data.

A. Secondary data

Secondary data refers to the data that has already been assessed, used and collected by certain firms. In this research, the record of 23000 patients is obtained from BDFH, which are used for designing the required NCD prediction application.

B. Primary data

The data that is collected for the first time as per the defined objectives of the research are termed as primary data. The primary data for the study is collected using a questionnaire which is distributed across 30 specialized healthcare professionals in the hospital. Since all the questionnaire were completed and received; therefore, the response rate is termed as 100 percent. The objective of the questionnaire is to gather the necessary requirement, NCD diagnosis for medical characteristics, which include CVD and diabetes, which are the commonly found forms of NCD.

3.2 Data Collection Procedure

A survey was conducted using the questionnaire as a primary form of data collection. While, for the secondary form of data collection, the health center record was used following the development of the software application.

A. Primary data collection

The questionnaire-based survey was conducted among the doctors to clearly understand the expectations they hold as well as ideas about the current situation concerning NCDs from one side and to examine the awareness and the level of utilization of having a new software application that will help them in predicting NCDs from the other side. To evaluate the current situation concerning NCDs prediction, the BDFH doctors who are either endocrinologists or cardiologists were requested to fill in a questionnaire survey. It was done to gain an understanding pertaining to their perceptions and opinions about the expected software application.

B. Study Population and Sampling

In BDFH, 13 cardiologists and 17 endocrinologists are present. The sample of the study is the entire population given the relative ease to work around. Therefore, the sample would constitute of all the related doctors. The structured questionnaires were administered by the researcher to assess their perception and opinion.

4. Survey Data Analysis and Findings

This section provides information concerning the questionnaire model, which was used in the survey. The survey aimed to establish the importance of predicting model concerning NCD trends among the end-user, and doctors. The direct interviews were held with the doctors to explain to them the significance of developing the software.

The conducted questionnaire was aimed to examine the present scenario concerning the use of NCDs forecasting tools in the BDFH. The sample is formed by the total endocrinologists and cardiologists' populations who are presently on the rolls of the BDFH. As the interviews are conducted on a one-to-one basis; therefore, no interview was missed. The following results were achieved by the questionnaire;

The answers to the first question show that 70% of patients treated by these 30 doctors suffer CVDs and diabetes. This provides evidence that NCDs, more particularly, diabetes and cardiovascular disease are majorly found. The generic and common disease such as influenza cold, allergies, and sprain make a total of about 30 percent. The BDFH does not provide treatment for critical diseases such as cancer.

The second question examines the primary NCDs causes, where the doctors mentioned unhealthy eating habits, smoking, no exercise, and consumption of high-fat food as the major NCDs reasons.

Following third and fourth questions in the questionnaire are related to each other and are also analyzed together. Though 80 percent of the doctors are aware of the prevalence of some form of NCD prediction mechanism across the globe, they firmly say (i.e., 100%) that BDFH lacks any such system, mechanism or facility.

In the fifth and sixth questions, 24 among the 30 doctors (i.e., 80%) showed agreement that presence of computer-based application would assist in predicting such NCDs, which aids the BDFH in notifying patients on a continuous basis. In addition, 90 percent of them welcomed the idea of offering an application that can "correctly predict" the NCDs occurrence, which is essential.

5. Proposed Application System

The software system developed comprises of two main components; the prediction engine and the software application. The system functions as an intermediator (practitioner) between the prediction system and the user. The explanation of each component is presented in which first describes the data that is applied in the proposed prediction system.

5.1 Data Mining Engine

The application purpose is to assess the data of the patients to predict if the patient is at an NCD risk or not. The following tools are used for application programming;

1. VS.Net: Tool for writing the application.
2. VB.Net: Tool for coding the application.
3. Oracle DB: The database which is used for data storage and manipulation.
4. Toad Oracle: Tool for gathering the code and retrieving the data from the database.

Among various techniques for data mining, the research adopted the mining class comparison. This is used to mine a description which compares or distinguishes one class from another related class. It is used for meeting the project purpose based on stages which are followed for the process of mining comparison [9]. The stages are as follows;

1. *Data collection*. It comprises a set of relevant and useful data that is collected by the use of query processing. The segregation of the data occurs in the target class along with contracted classes set.

2. *Dimension relevance analysis*. This is executed only where there are several dimensions and only a few relevant dimensions are to be selected for analysis.
3. *Synchronous generalization*. This is applicable on the target class concerning the degree of control exercised by the user. The contrasting classes concept is to generalize the same level as those in the main target relation, to form the prime contrasting classes' relation.
4. *Presentation of the derived comparison*. The achieved class comparison can be represented in the form of tables, graph, and rules.

5.2 Patient Data Set

The researches, which tackle data mining for medical application use the available data sets such as heart-Statlog Medical data set [26], liver disorder, Indian liver patients, heart disease (Statlog), heart disease (original), hepatitis Parkinson's, Parkinson's, breast cancer Wisconsin (diagnostic), dermatology, and lung cancer [20].

This work is featured using the actual data set obtained from the BDFH, which contains the details and vitals of 23,000 patients.

Using the developed data set, that was extracted with the use of the discussed algorithm of data mining, the recognized pattern assists to predict the risk of getting heart disease and diabetes.

The data of all the patients were processed through the elimination of the repetitive records and completion of the missing details. The dummy values were used for filling the personal details (name and ID number, etc.). Moreover, the recurring patterns are mined with the use of proposed process of data mining. The attributes of the patients gathered from the database are; sex, age, chest pain (CP), FBS (fasting blood sugar), test BPS, rest ECG (electronic cardiographic), smoking, diet, as well as alcohol. Following it, these are used for classifying the patients, which are modified from the study conducted by Khaing [19] as some of them are unavailable in the data set, we have. Afterward, they are adjusted as per the survey conducted with the actual practitioners in the BDFH.

5.3 Application Process Description

The overall exercise was conducted with doctors who will be the main user of the developed software. This is to ascertain that all the unnecessary data is removed so that the prediction of NCDs is barrier-free. The data of the patients is derived from the three tables by executing the second step of KDD. Figure 4 presents the diagram for the entity relationship (ER). The information record of the patients is presented in the first table, second patients' vitals, such as, diet, blood pressure, chest pain type, smoking, and heart rate. The blood test collected from the lab is collected in table 3, where cholesterol and fasting blood sugar are its attributes.

Reflecting that the target class are the lab results and vitals, the "if condition" is used for the constructing class, along with attributes such as sex, age, chest pain (CP), FBS (fasting blood sugar), test BPS, rest ECG (electronic cardio graphic), smoking, diet, as well as alcohol (Table 1). The data mining query language (DMQL) can be used for the data mining task, as follows;

1. Used Patient_DB
2. Mine comparison as "Patient_NCD_Risk_Level."
3. In relevant to sex, age, CP, chol, FBS, test BPS, rest ECG, smoking, diet, alcohol
4. For "Vitals_View" and "Lab_Results"
5. Versus "if_condition_rule"
6. Evaluate sum
7. Show as "risk level."

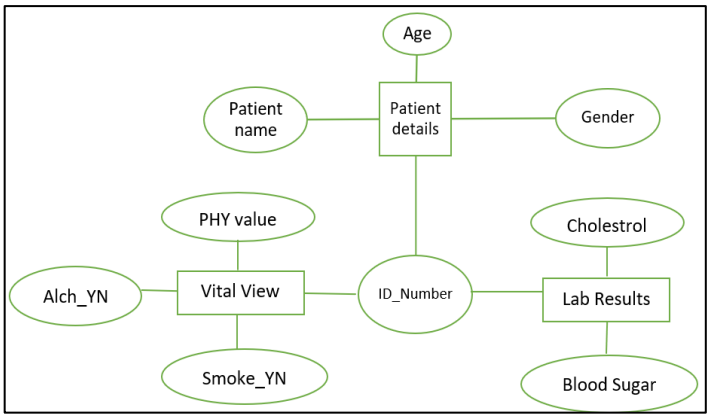


Figure 4. ER diagram of the proposed application

Initially, the change in the query occurs in two forms such as in the two interrelated queries where the two sets of relevant data tasks are accumulating; such as, one is obtained from Vital View table target class and Lab Results table, whereas the other is derived from the contrasting class, i.e., "if condition rule" which is written in the code classes.

Secondly, the two tables, i.e., Vitals View and Lab Results, are analyzed for their dimension relevancy. Furthermore, the weakly relevant dimension are removed, i.e., the removal of the patient information from the found test classes.

Thirdly, the synchronous generalization is applied from the target class to the controlling class level in order to produce the prime relation for the target class. With the use of "if condition" rule, the categorization of the attributes in the form of group is held. Initially, the ages are evaluated. In case, the age is less than 40, it is grouped as young age, whereas, if it lied between 40 to 60, it is grouped as medium age while for 60 or greater than 60, it is regarded as older age. Following it, smokers or alcohol drinkers are categorized. In case the patient is a smoker, he is characterized as group 1, and in group 2 if the situation is vice versa. Similarly, if the patient is an alcohol drinker then he is characterized in group 1, and in group 2 if he does not.

If the blood pressure (BP) of the patient is less than 80, then he is grouped into group 0, and group 1 if it is equal to or more than 80 and group 2 when the blood pressure is higher than 90. Likewise, for the classification of heart rate, if the heart rate is below or equal to 100, it is categorized in group 0 while if it is more than 100 and less then 150, it is categorized as 1. Similarly, in case the heart rate is greater or equal to 150, the patients are categorized in group 2.

Table 1. Attributes and values used in the proposed application

No.	Attributes	Description	Cut-off Value	Type
1	Age	Age of the	Age <=40 Age <=60 and >40	Young age Middle age Old age

		patients in years	Age >60	
2	Gender	Gender	1 0	Male Female
3	CP	Type of Chest Pain	Value 1 Value 2 Value 3 Value 4	Stable Angina Unstable Angina Nonangina Pain Asymptomatic
4	Test BPS	Resting blood pressure (in mmHg)	BP <80 BP <90 BP >90	Normal Normal to high High
5	Chol	Serum Cholesterol (mmg/dl)	Chol <5.2 Chol <5.2 and >6.2 Chol >=6.2	Normal High Severe
6	FBS	Fasting blood sugar	1 0	True False
7	Rest ECG	Resting electro cardiographic results	Val=0 Val=1 Val-2	Normal Abnormal Probable
8	Diet	On a healthy diet	0 1	True False
9	Smoking	Smoker patient	0 1	True False
10	Alcohol	Alcohol drinker	0 1	True False

```
If txtAge.Text.Trim() <= "40" Then
    strAge = "0"
    'MessageBox.Show("Young Age")
ElseIf txtAge.Text.Trim()> "40" And txtAge.Text.Trim() <= "60" Then
    strAge = "1"
    'MessageBox.Show("Middle Age")
ElseIf txtAge.Text.Trim() > "60" Then
    'MessageBox.Show("Old Age")
     strAge = "2"
End If
```

Figure 5. Comparison of condition for the age

Lastly, Table 1 presents a comparison of the resulting classes in the rules form. Such as, in case the classes are equal to 0 when there is no risk for patients. While patients are at high risk when the classes are equal to 2. Likewise, the patient is at lower risk, when the classes are equal to one, where the other classes combination occurs based on the recommendation of the actual practitioners.

Following the preprocessing, the previously mentioned attributes in Table 1 are utilized for the code application for using the IF condition. The classification of these values done in the form of cut-off value, following its comparison.

```
If Trim("" & sqlds.Tables(0).Rows(0).Item("SMOKE_YN")) = "Y" Then
       cmbSmoking.SelectedIndex = 1
ElseIf Trim("" & sqlds.Tables(0).Rows(0).Item("SMOKE_YN")) = "N" Then
       cmbSmoking.SelectedIndex = 2
End If

If Trim("" & sqlds.Tables(0).Rows(0).Item("ALCO_YN")) = "Y" Then
       cmbAlco.SelectedIndex = 1
ElseIf Trim("" & sqlds.Tables(0).Rows(0).Item("ALCO_YN")) = "N" Then
       cmbAlco.SelectedIndex = 2
End If
```

Figure 6. Grouping smoke and alcohol

The process of grouping the attributes are described by explaining the Figures 6–10 as follows:

Figure 6 shows the comparison if patients smoke or drink alcohol. In case the patient is a smoker, he is characterized as group 1, and in group 2 if the situation is vice versa. The same situation is followed if the patients is an alcohol drinker than he is characterized in group 1, and in group 2 if he does not.

Figure 7 presents the BP, for which if the patient BP is less than 80, then he is grouped into group 0, and group 1 if it is equal to or more than 80 and group 2 when the blood pressure is higher than 90.

```
If txtBP.Text.Trim().Substring(4, 2) < "80" Then
       strBp = "0"

ElseIf txtBP.Text.Trim().Substring(4, 2)>= "80"
      And txtBP.Text.Trim().Substring(4, 2) <= "90" Then
       strBp = "1"

ElseIf txtBP.Text.Trim().Substring(4, 2) > "90" Then
       strBp = "2"
End If
```

Figure 7. Grouping blood pressure (BP)

Figure 8 demonstrates the grouping of heart rate. If the heart rate is below or equal to 100, it is categorized in group 0 while if it is more than 100 or less then 150, it is categorized as 1. Similarly, in case the heart rate is greater or equal to 150, the patients are categorized in group 2.

Figure 9 displays the patients grouping as per his cholesterol level. The patient is grouped in 0, when the level of cholesterol is below 5.2, and in group 1 if the value of cholesterol is more or equal to 5.2. Similarly, these are grouped 2, when it is more than or equal to 6.2.

Figure 10 provides a list of conditions for chest pain. As per the program, if the grouping of the chest pain is done as 1, the patient experience chest pain, and 2 in case a risk of cardiac arrest prevails. It is grouped as 3 if a patient is experiencing coronary artery disease. Whereas for being grouped as 4, the patient is at high risk of chest pain.

Lastly, the class comparison results are presented in rules forms (Table 2) and process flow of the system is depicted in Figure 11.

5.4 Testing Results of the Application and its Predictive Performance

As explained above, the resulting conditions provide a comparison of the prime target classes and the executed "if condition rule" (Table 2), where risk results are the outcome of applying the aforementioned rules.

```
If txtHeartRate.Text <= "100" Then
       strHR = "0"

ElseIf txtHeartRate.Text > "100" And txtHeartRate.Text < "150" Then
       strHR = "1"

ElseIf txtHeartRate.Text >= "150" Then
       strHR = "2"

End If
```

Figure 8. Grouping heart rate

```
If txtCholesterol.Text < "5.2" Then
       strTotCh = "0"

ElseIf txtCholesterol.Text >= "5.2" And txtCholesterol.Text < "6.2" Then
       strTotCh = "1"

ElseIf txtCholesterol.Text >= "6.2" Then
       strTotCh = "2"

End If
```

Figure 9. Grouping cholesterol

```
If cmbChestPain.SelectedIndex > 0 Then

   If cmbChestPain.SelectedIndex = 1 Then
       strChPain = "1" 'Predictable Chest Pain

   ElseIf cmbChestPain.SelectedIndex = 2 Then
       strChPain = "2" 'Chest Pain Signal to Heart Attack

   ElseIf cmbChestPain.SelectedIndex = 3 Then
       strChPain = "3" 'Have Coronary Artery Disease

   ElseIf cmbChestPain.SelectedIndex = 4 Then
       strChPain = "4" 'Predictable Chest Pain

   End If

End If
```

Figure 10. Grouping chest pain

The software application is implemented using VB, and the samples of the screenshots are depicted in Figures 12 to 15. These screenshots serve as evidence for the program utility concerning the NCDs prediction for the tested patients. These samples of NCD predictions are shown for different risk levels: no risk, low risk, medium risk, and high risk.

5.5 Findings

The results are examined on the basis of the response gathered through a questionnaire survey, which was held among the doctors and the results achieved from the use of the developed application.

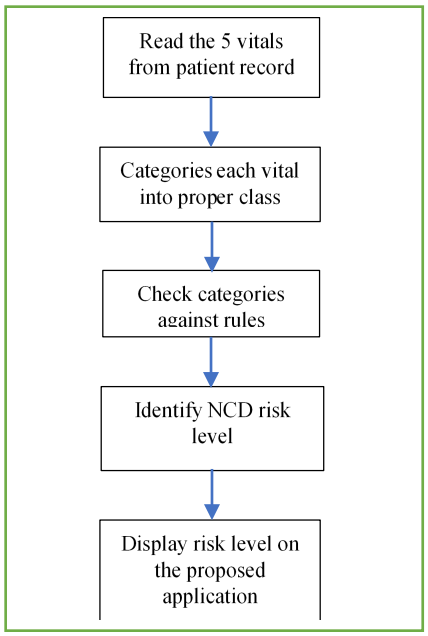


Figure 11. Flow diagram of the proposed application

Table 2. Classes and Risk Results Comparison

BP	Heart Rate	Cholesterol	Chest Pain	Age	Risk Result
0	0	0	0	0	No risk
0	1	0	1	0	No risk
0	2	0	1	0	No risk
1	1	1	1	1	Low risk
1	2	0	1	0	No risk
1	2	0	1	1	No risk
1	2	0	1	2	Low risk
1	2	0	2	2	Medium risk
1	2	1	1	1	Low risk
1	2	1	2	1	Medium risk
2	2	1	1	1	Low risk
2	2	1	1	2	Medium risk
2	2	1	2 or 3	2	High risk
2	2	2	1	1	Medium risk
2	2	2	2 or 3	2	High risk

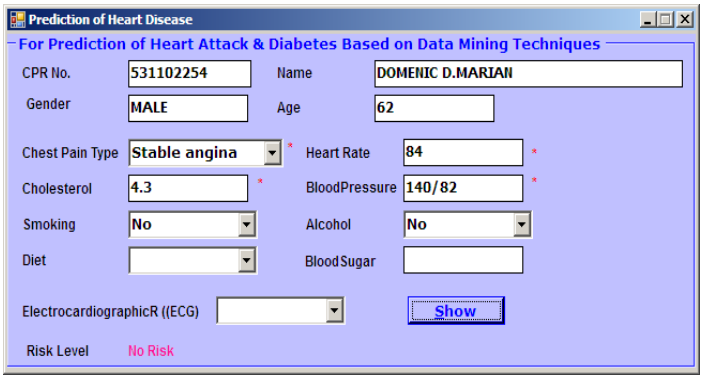


Figure 12. Sample of NCD prediction level: no risk

A. Survey Based Questionnaire Inferences

The questionnaire survey provides evidence that at present, there are no tools or the mechanism available in the BDFH, which is having almost 23,000 registered patients, for NCD forecasting

and prediction. The patient's record provides a massive volume data, yet there is no examination or the analysis of such essential data for the purpose of forecasting. The survey highlighted that most of the doctors are aware regarding the tools which are being used somewhere else and they will like to have applications that can help them better in the prediction of a patient's possibility for the occurrence or the recurrence of NCDs.

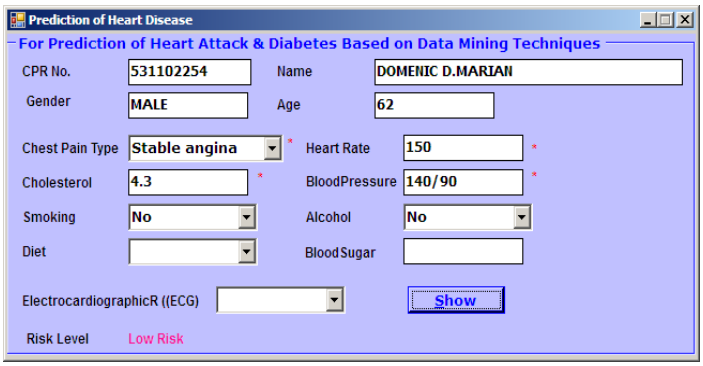


Figure 13. NCD prediction level (sample): low risk

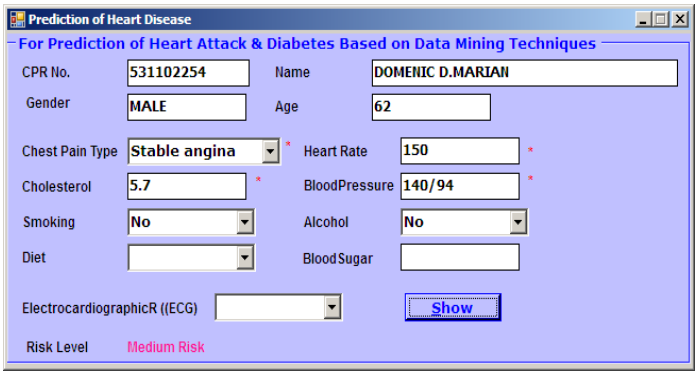


Figure 14. NCD prediction level (sample): medium risk

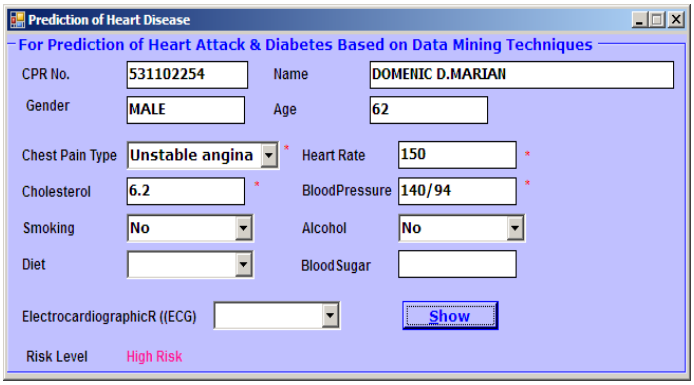


Figure 15. NCD prediction level (sample): high risk

B. Analysis of the developed software Inferences

The results obtained were discussed with the doctors and are considered as the levels of risks. The doctors were highly contented with the results as they accurately matched their expectations regarding the levels of risk. All kinds of risks are categorized appropriately according to the formerly mentioned categories. Additionally, they indicated high interests in the implementation of such applications live in their clinics. The application of the software has a proof for the accuracy and the efficacy of the suggested technique of the data mining for the forecasting of the NCDs for the patients of BDFH. However, the NCDs can lead to the possible mortality among the patients, the doctors who had a trial of the software feel excitement for its application as it can help them in the detection. As a matter of fact, they want to use it as soon as possible because they highly noticed that it supports quickly and gives better forecasting along with the briefing of the patient's record. On the other hand, the statistical analysis is not applicable to this study as the results are evaluated manually by the consulted practitioners.

Using the application, more savings in medical expenditure can occur avoiding loss of the duty timings and maximizing the utilization of the crucial medical facilities. This application will give the one-stop-shop to all the patients to get the accurate and the calculated information regarding their health. This application is considered to be an asset for the doctors so that they make sure that their detection or inferences are professional as well as correct. Further, such information can be utilized globally, for the patients having the health problems and have to travel abroad, the medical history and the forecasting for the NCD can be made online available or can be shared via emails.

6. Conclusions and Future Work

An application which uses the data mining algorithm for the class comparison has been invented to forecast the level of risk of occurrence or recurrence of NCDs such as diabetes and the heart diseases. Additionally, the outcomes of the application highlighted that the forecasting system could forecast NCDs efficiently, instantly, and effectively. This application helps the doctors to make the decisions regarding the health risks of a patient. It also creates the results, which make it closer to the situations of real life. Data mining, therefore, is more supportive for the health sector and is essential for exploring the knowledge to be used in the health care sector.

This research confirms that the health centers should have a software application or the system which can forecast the NCDs. To have raw data in a large volume is a must for the proper categorization rules which lead to the proper forecasting of NCDs. The forecasting is helpful for not only the physicians but for the patients too for their constant health check-ups. The application developed must be trialed on a constant basis, and the coding must be accurate in order to adopt the application across the various health centers and the clinics. Further, in the future, it can also be considered for the development of the mobile application for the patients where they can input their vitals and get the results instantly.

The developed application and the software can be extended for assisting in forecasting the level of risks for the other NCDs, for example, cancer, Alzheimer, chronic lung disease, stroke, and the osteoporosis. This can be achieved more by highlighting the associated vitals and their analysis to set the appropriate rules by the consultation linked with the expert practitioners. Finally, the obtained results can be further analyzed once the approach is applied on predefined data set with previously known level of risk of each NCD disease.

References

[1] Anon., 2016. World Health Day 2016. [Online] Available at: http://www.who.int/diabetes/global-report/WHD16-press-release-EN_3.pdf?ua=1.

[2] W. H. Organization, "The top 10 causes of death," 24 May 2018. [Online]. Available: http://fmrglobalhealth.com/frame/top10.html. [Accessed 28 June 2019].
[3] Anon., 2018. Diabetes country profiles. [Online] Available at: https://www.who.int/diabetes/country-profiles/bhr_en.pdf?ua=1 [Accessed 28 June 2019].
[4] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P., 1996. From data mining to knowledge discovery. In Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, pp. 495–515.
[5] Clifton, C., 2010. Encyclopaedia Britannica: definition of data mining.
[6] Obenshain, M. K., 2004. Application of data mining techniques to healthcare data. Infection Control and Hospital Epidemiology, 25(8), pp. 690–695. DOI: https://doi.org/10.1086/502460
[7] Krishnaiah, V., Narsimha, G. & Chandra, S., 2016. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. International Journal of Computer Applications, 136(2), pp. 43–51. DOI: 10.5120/ijca2016908409
[8] Yoo, I. et al., 2012. Data mining in healthcare and biomedicine: a survey of the Literature. Journal of Medical Systems, 36, pp. 2431–2448. DOI: https://doi.org/10.1007/s10916-011-9710-5
[9] Han, J., Pei, J. & Kamber, M., 2011. Data mining: concepts and techniques. Walthman, MA: Morgan Kaufman Publisher. DOI: 10.1007/978-1-4419-1482-6_3752
[10] Benneyan, J. C., Lloyd, R. & Plsek, P., 2003. Statistical process control as a tool for research and healthcare improvement. Quality & Safety in Health Care, 12, pp. 458–464. DOI: 10.1136/qhc.12.6.458
[11] Mena, J., 2011. Machine learning forensics for law enforcement, security and intelligence. In CRC Press.
[12] Campos, J. et al., 2017. A big data analytical architecture for the asset management, Elsevier, pp. 369–374. https://doi.org/10.1016/j.procir.2017.03.019
[13] Piatetsky-Shapiro, G. & Parker, G., 2011. Lesson: data mining, knowledge discovery: an introduction. In Introduction to Data Mining.
[14] Maimon, O. & Rokach, L., 2011. Data mining and knowledge discovery handbook, Springer.
[15] Hersh, W. R. & Hoyt, R. E., 2018. Health informatics: practical guide seventh edition. lulu.com.
[16] Athina, L. A. & Konstantinos, S. M., 2008. Handbook of research on distributed medical informatics and e-health. Medical Information Science Reference.
[17] Bronzino, D., 2006. Medical devices and systems, CRC Press.
[18] Acharya, R. U. & Yu, W., 2010. Data mining techniques in medical informatics. The Open Medical Informatics Journal, 4, pp. 20–21. DOI: 10.2174/1874431101004020021
[19] Khaing, H. W., 2011. Data mining based fragmentation and prediction of medical data. Shanghai, pp. 480–485. DOI**:** 10.1109/ICCRD.2011.5764179
[20] Khanmohammadi, S., Adibeig, N. & Shanehbandy, S., 2017. An improved overlapping k-means clustering method for medical applications. Expert Systems With Applications, 67, pp. 12–18. DOI: 10.1016/j.eswa.2016.09.025
[21] Jiang, H., Ching, W. K. & Hou, W., 2016. On orthogonal feature extraction model with applications in medical prognosis, Applied Mathematical Modelling. 40, pp. 8766–8776. DOI: 10.1016/j.apm.2016.05.011
[22] Suvarna, C., Sali, A. & Salmani, S., 2017. Efficient heart disease prediction system using optimization technique, Erode, India, pp. 374–279. DOI: 10.1109/ICCMC.2017.8282712
[23] Rairikar, A. et al., 2017. Heart disease prediction using data mining techniques, Coimbatore, India, pp. 1–8. DOI: 10.1109/I2C2.2017.8321771
[24] Tomczak, J. & Zieba, M., 2015. Probabilistic combination of classification rules and its application to medical diagnosis. Machine Learning, 101, p. 105–135. doi: 10.1007/s10994-015-5508-x
[25] Zhu, Y. & Fang, J., 2016. Logistic regression-based trichotomous classification tree and its application in medical diagnosis. Medical Decision Making, 36(8), pp. 973–989. DOI: 10.1177/0272989X15618658
[26] Jaganathan, P. & Kuppuchamy, R., 2013. Threshold fuzzy entropy based feature selection for medical database classification. Computers in Biology and Medicine, 43(12), pp. 2222–2229. DOI: 10.1016/j.compbiomed.2013.10.016

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 197-205 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Dysphoria Detection using EEG Signals

Norhaslinda Kamaruddin*,[1], Mohd Hafiz Mohd Nasir[1], Abdul Wahab[2]

[1]*Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40540 Shah Alam, Selangor, Malaysia*

[2]*Kulliyyah of Information & Communication Technology, International Islamic University Malaysia, P. O. Box 10, 50728 Kuala Lumpur, Malaysia*

ARTICLE INFO

Article history:
Received: 19 May, 2019
Accepted: 14 July, 2019
Online :30 July, 2019

Keywords:
Dysphoria
EEG emotion
Affective Space Model
Multilayer perceptron

ABSTRACT

Dysphoria is a state faced when one experienced disappointment. If it is not handled properly, dysphoria may trigger acute stress, anxiety and depression. Typically, the individual who experienced dysphoria are in-denial because dysphoria is always being associated with negative connotations such as incompetency to handle pressure, weak personality and lack of will power. To date, there is no accurate instrument to measure dysphoria except using questionnaire by psychologists, such as: Depression, Anxiety and Stress Scale (DASS) and Nepean Dysphoria Scale (NDS-24). Participants may suppress or exaggerate their answers resulting in misdiagnosis. In this work, a theoretical Dysphoria Model of Affect (DMoA) is developed for dysphoria detection. Based on the hypothesis that dysphoria is related to negative emotion, the input from brain signal is captured using electroencephalogram (EEG) device to detect negative emotions. The results from analyzing the EEG signals were compared with DASS and NDS questionnaires for correlation analysis. It is observed that the proposed DMoA approach can identify negative emotions ranging from 55% to 77% accuracy. In addition, the NDS questionnaire seems to provide better distinction for dysphoria as compared to DASS and is similar to the result yielded by DMoA in detecting dysphoria. Thus, DMoA approach can be used as an alternative for early dysphoria detection to assist early intervention in identifying the patients' mental states. Subsequently, DMoA approach can be implemented as another possible solution for early detection of dysphoria thus providing an enhancement to the present NDS instruments providing psychologists and psychiatrists with a quantitative tool for better analysis of the patients' state.

1. Introduction

Stress is a condition experienced when an individual face a pressure or strain from adverse or demanding circumstances [1]. The individual life satisfaction is undermined when difficulties and challenges happen; ranging from the daily annoyances of the consequences of overstretched, time-pressured lifestyle such as traffic jams, lateness and work incompetency to the unexpected events faced in workplace, family lives, interpersonal relationship or even interaction with environment, such as road accident, bankruptcy, demise of a loved ones, loss of a pet or even handling natural disasters like flood or drought.

*Corresponding Author: Norhaslinda Kamaruddin, Advanced Analytics Engineering Centre, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia, norhaslinda@tmsk.uitm.edu.my, +603-55211163

Stress can cause positive reinforcement as it can be the source of motivation. For instance, healthy competition in workplace can boost morale of a worker to work harder and produced better results. However, if stress is not handled correctly, it can lead to other medical issues, such as; headaches, upset stomachs, anxiety, depression and other undesirable circumstances. [2]. Moreover, long term stress can contribute to other productivity issues, for example; high rate of absenteeism, high labour turnover, inefficient workforce and even high rate of accidents [3]. Hence, it is important to detect stress before it can be worse.

Generally, the number of individuals in every generation experienced stress level are increasing. This finding has been reported in a comprehensive study by American Psychological Association (APA) [4] with 23% and 37% are recorded for

millennials and older adults respectively from 2014 to 2017. The study segregated participants into four groups based on age, as follows; Millennials (18–38 years), Gen X-ers (39–52 years), Baby Boomers (53–71 years) and Older Adult (72 years and above). The survey was conducted online within the United States with participation of 3361 of adults. Figure 1(a)-(d) illustrate the American stress level by generations from 2014 till 2017.

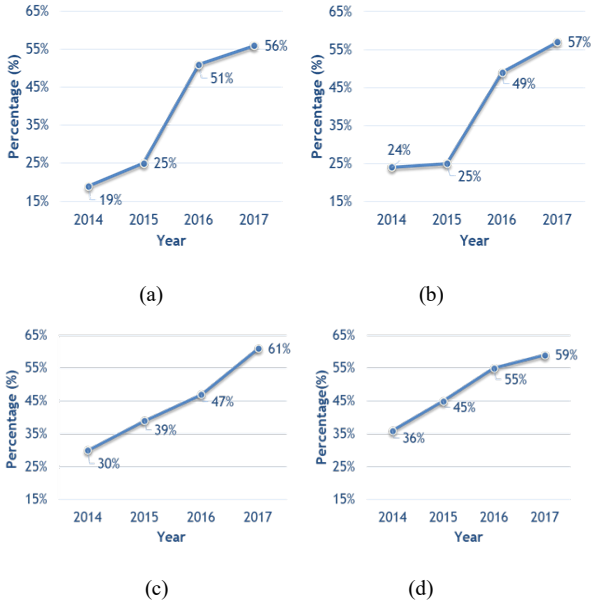


Figure 1: Graph of Stress Level Based on Generation. (a) Older Generation (b) Baby Boomers (c) Gen X-ers and (d) Millennials [4]

The survey results in Figure 1 show that higher number of younger generation experienced stress compared to Older Adult. Older Adults had the least stress levels as presented in Figure 1(a). Such trend is observed to be consistent from 2014 till 2017. This scenario may be linked to work stress as Older Adults typically have retired and are not formally working therefore their stress level is lower than younger generation who are still in employment. However, the highest change is observed in the year of 2015 to 2016 for Older Adults with 26% spike followed by 24% spike for Baby Boomers in the same duration as shown by Figure 1(b). This situation may be related to the United States Presidential Election that happened in 2016. The Baby Boomers and the Older Adults groups are highly affected as they are more in tune with the US political scenario as compared to the younger generations. Additionally, the survey revealed that a critical increment in the number of Americans who had encountered at least one symptom of stress in the previous month is observed from 71% in 2016 to 75% in 2017. Hence, the trend of stress experienced by individual is on the rise.

The finding by American Psychological Association [4] is consistent with the National Health and Morbidity Survey (NHMS) [5, 6] discovery that the trend of stress is rising, and it is reflected in the high leap of patients who suffer from mental illness as presented in Figure 2. In 2011, a total of 24,498 respondents were successfully interviewed and 6.3% reported to have mental health problem in Malaysia [5]. However, the NHMS V (2015-2018) survey [6] reported that 29.2% Malaysian adults aged 16 years and above in Malaysia experienced mental health problem. Such pattern shows a large increase over the last 4 years with almost four folds increment with factors such as younger Bumiputras female adult from low-income family is predicted to be at higher risk to experience mental health problem [5]. The report also mentioned that early recognition/detection and early treatment of depression must be enhanced.

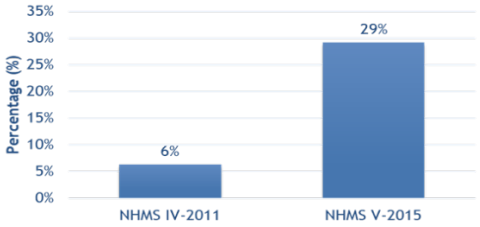


Figure 2: Mental Health Among Adult in Malaysia

Before an individual experienced stress, a state called dysphoria will be endured. Dysphoria can be defined as an emotional state with high degree of complexity and have a close relation to negative mood and in some cases can lead to frustration and dissatisfaction [7]. Dysphoria is usually connotated to anxiety, depression and stress, and can be experienced by any individuals at any time. Typically, dysphoria is viewed as everyday response once an individual face disappointment in his/her life. Figure 3 illustrates the simplified stages of mental cycle in handling disappointment [8]. Dysphoria is viewed as a starting point of either positive or negative adaptation states starting with normal condition to the circumstances when an individual faces a disappointment. When the individual cannot adapt to the dissatisfaction or disappointment, it can lead to dysphoria where the state of frustration is intense thus can lead to clinical depression or stress. On the other hand, the individual who is able to adjust and adapt to the disappointment may be able to carry out his life normally through adaptation. Since emotion indicates the mental state of an individual and is closely related to dysphoria emotion measurement methods is adopted [9]. Hence, the negative emotion that one experienced after disappointment can be empirically measured.

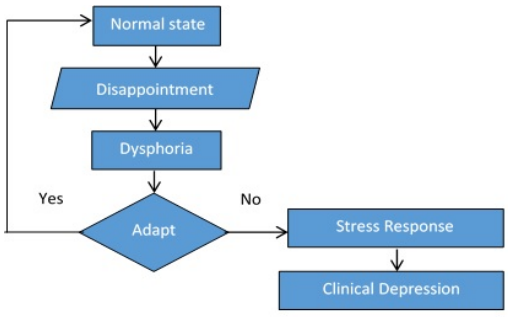


Figure 3: Mental cycle of handling disappointment

In this work, a Dysphoria Model of Affect approach is proposed to measure dysphoria from the brain signals captured using electroencephalogram (EEG) device. Such model can provide a datum for better comparative analysis with the standard psychological instruments to understand dysphoria such as DASS-21 and NDS-24 questionnaires. This is to eliminate the need to

validate the patient's answer by directly using neurophysiological input of the individual because patient have tendency to exaggerate or suppress the answers. Because dysphoria is very much related to negative emotion, emotion recognition methodology is adopted. For simplification, the working definition of dysphoria in this work is the state after disappointment of event that is prior to stress and it is always be associated with negative emotions [8].

This paper is organized in the following manner. Section literature review provides a summary of researchers concentrated efforts to recognize dysphoria and the fundamental concepts of dysphoria study are discussed. The methodology of the proposed approach of Dysphoria Model of Affect (DMoA) is presented in Section 3. The data collection, experimental result and discussion are given in Section 4. Subsequently, this paper concludes with conclusion and future work in Section 5. This paper is an extension of work originally presented in 2018 12th International Symposium on Medical Information and Communication Technology (ISMICT) [8].

2. Literature Review

2.1. Dysphoria and Psychological Instruments

Commonly, dysphoria is associated with depression, stress and anxiety [7,10,11,12]. Dysphoria is associated with depression because it shares the same features of depressive syndrome and dysphoric mood, such as sadness, despondence, despair, and others as suggested by [10]. Such notion is supported by Berner, Musalek & Walter [11] that the term dysphoria should be restricted to a condition of a morose, tense and irritated mood. It can also be recognized from stable and unstable mixed state. Moreover, anxiety may also refer to dysphoria because both have tension and the underlying hyperarousal in common. Dysphoria and anxiety can be experienced at the same time because dysphoria may happen due to long-standing stress. It is not uncommon for anxiety to arise on the background of dysphoric mood. In a certain personality disorder, dysphoria also can be incorporated such as affective instability, irritability, intense anger, and profound unhappiness [12]. These characteristics are shared or overlap with features of dysphoria. Furthermore, Starcevic et al. [7] stated that dysphoria is also associated with depression, interpersonal sensitivity, hostility and stress, reflecting broad but also complex and relatively non-specific relationship.

Dysphoria is assessed using several psychologists' instruments such as the Nepean Dysphoria Scale (NDS) [13], the Depression, Anxiety, Stress Scales (DASS) [14], the Beck Depression Inventory II (BDI II) [15], Perceived Stress Scale (PSS) [16], and the Anxiety Sensitivity Index (ASI) [17]. Typically, the participants are forced to select one answer from the fixed choice response format [18]. The participants need to respond to a series of standard measurement for attitude questions about a pre-determined topic that allow the degree of opinion rather than a simple yes and no answer. Such data enables better quantitative data analysis. The questionnaires can be provided to participants of either online or off-line dissemination method depending on the geographical availability of the participants. In summary, the PSS questionnaire is focusing on measuring the level of stress [16], the BDI-II can measure the level of depression [15] and the level of anxiety of an individual can be measured using the ASI instrument[17]. Based on the dysphoria descriptions that are being discussed earlier, dysphoria is very related to depression, anxiety and stress. All these factors can be measured using DASS [14] and NDS [13] questionnaire. It is expected to observe strong positive correlations between scores on the NDS and scores on the DASS-21. Hence, for the purpose of this work, both questionnaires are used to investigate dysphoria. The summary of the questionnaire and its focus is presented in Table 1.

Table 1: Summary of Questionnaire and Its Focus Study.

	DASS	PSS	BDI-II	ASI	NDS
Depression	√		√		√
Anxiety	√			√	√
Stress	√	√			√

The DASS-21 test is the psychological instrument to measure depression, anxiety and stress of an individual. It consists of 21 questionnaires and is able to measure the degree of depressions, anxiety and stress of the participant taking the test [14]. The participants are needed to specify indication of a symptom over the previous week based on a 4-point Likert scale. Each answer is recorded from 0 (did not apply to me at all over the last week) to 3 ('applied to me very much' or 'most of the time over the past week'). Calculation on the degree of depression (D), anxiety (A) and stress (S) is indicated is provided where every item of the questionnaire is related to either D, A or S score. The cumulative score of the D, A and S are used to indicate the level of severity, namely; normal, mild, moderate, severe and extremely severe. A trained psychologist or psychiatrist is needed to interpret the results and thus providing a more meaningful diagnostic for the patient.

The Nepean Dysphoria Scale (NDS) is a set of questionnaires to measure the severity of dysphoria [13] and expects participants to give evaluation of the frequency on a scale from 0 to 3 ranging from 'not at all', 'always', 'every day' to 'most of the time'. In this questionnaire, there are 4 factors that are being observed, namely; irritability, discontent, surrender and interpersonal resentment. Based on the summation of the score gathered from the itemized questions, the correlation matrix factor of NDS are calculated. The specified weightages are then used to get the maximum and average value of the scale. The value yielded are then compared to get the severity category of either normal, mild, moderate and severe.

2.2. Dysphoria and Negative Emotion

Dysphoria has recently been conceptualized as a complex emotional state that consists of discontent and/or unhappiness and a predominantly externalizing mode of coping with these feelings [7]. Many researchers relate dysphoria to negative emotions [10,11]. Dysphoria is also used as synonym for sadness or depression which describe a mixture of negative and unpleasant emotions [19].

A study by Caseras, Garner, Bradley and Mogg [20] examined the eye movement of the participants on negative images. The participants are grouped into dysphoric and nondysphoric group. In the experimental results, it was observed that dysphoric group have the tendency to look at the negative images longer than the control images as compared to the time taken by the nondysphoric

group. This result is in line with another work by McMakin, Santiago and Shirk [21] that observed the experiential reactivity to emotion in dynamic moment. Experimental results revealed that individual with dysphoric tendency and those who have not are equally responsive to both negative and positive film clips regardless of their dysphoric level. However, dysphoric group shows a lower ability to maintain positive emotion when they are looking at the positive film clips. Hence, it can be said that dysphoria is always associates with negative emotions.

The negative emotions are located in the negative valence axis region of Affective Space Model (ASM) [22]. The ASM consists of three main emotion primitives; namely: valance, arousal and dominance. The emotion primitives are actually the descriptive generic attributes of an emotion. These attributes function as constituents that act as a fully complementary description of the emotion. The affective space model can be divided into quadrants where each quadrant represents positive or negative value of the emotion primitives. The horizontal axis represents the valance value that indicates whether the emotion is perceived as positive or negative whereas the vertical axis representing the arousal which indicates the level of activation from passive to active. In this work, the focus is given on valence since valence is the effect of the emotion to an individual of either positive effect or otherwise. Figure 4 shows the Schlosberg's Model of Affect where dysphoria is labelled in pink rectangle in quadrant 2 and 3. Dysphoria happens at negative valance values regardless of its arousal of either active or passive.

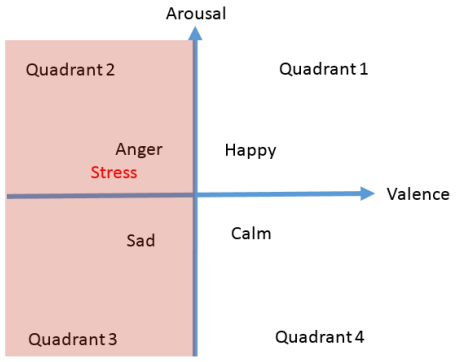


Figure 4: Model of Affect [22]

2.3. Dysphoria and Electroencephalogram (EEG)

Further reading shows that some researchers using EEG in dysphoria understanding. Siegle, Condray, Thase, Keshavan and Steinhauer [23] investigated the Gamma band in EEG signal in relating to emotional states for healthy, depressed and schizophrenia adults. The emotion condition assessment is using 3 methods; which are, never-depressed healthy controls, *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) unipolar major depressive disorder and DSM-IV schizophrenia. During the assessment, the patients need to judge their own emotion of either positive, negative or neutral based on the words appearing on the screen of laptop. It is observed that depressed individual displayed both sustained and increased gamma-band EEG signals activity throughout the task while adults with schizophrenia displayed decreased gamma-band EEG signals activity. It is also observed that schizophrenia subjects displayed more affinity responses to negative word compared to neutral words and not significantly faster to positive than neutral words. Based on this study, it can be summarized that there are different responses towards negative words depending on either the patients suffer from schizophrenia or otherwise. Such responses can be monitored using the EEG signals produced by the depressed and schizophrenia groups.

From previous studies [7,8] it is noted that dysphoria is very much affiliated to negative emotions, thus it is proposed to use the electroencephalogram (EEG) in measuring the negative emotion of the individual and relate it the dysphoria. Lin et. al [24] applied machine learning algorithms to categorize EEG signals on emotional states during music listening. The main purpose of the research is to search emotion-specific features of the EEG signals and test the efficacy of different classifiers. The Valance-Arousal model is used as comparator of emotion. 26 subjects were instructed to keep their eyes closed and remain seated while listening to the sixteen excerpts from Oscar's film soundtracks and the EEG signals were recorded. Features are extracted using the Differential Asymmetry of 12 Electrode (DASM12), Rational Asymmetry of 12 Electrode (RASM12), Power Spectrum Density of 24 electrode (PSD24) and Power Spectrum Density of 30 electrode (PSD30) while Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) are used as classifier. As a result, SVM obtained an improvement classification performance from 81.52% ± 3.71% to 82.29% ± 3.06% compared to MLP classifier. The best result recorded were using DASMI12 using SVM as classifier. This study also concluded that the sensitivity metric for characterizing brain dynamic in response to emotional state (joy, angry, sadness and pleasure) are depending on the spectral power asymmetry across multiple frequency bands using DASM12.

Moreover, Sourina and Liu extended the work to recognize emotion from Fractal Dimension (FD) based algorithm using EEG signals [25]. The FD values are calculated from the EEG signal recorded from the corresponding brain lobes and mapped on the 2D emotion model. The International Affective Digitized Sounds (IADS) and music pieces labelled by other subjects with emotions using a questioner are used as stimuli to generate emotion to the subjects during the assessment. As a result, the FD were able to detect emotion with only using 3 electrodes and the data can be mapped into 2D Arousal–Valance model. The accuracy of 72% to 100% are recorded for arousal and valance level.

Based on the study of other researchers, we combine the understanding yielded from EEG emotion recognition and insights gathered from the DASS and NDS questionnaire to understand dysphoria. Figure 5 summarises the correlation between EEG, questionnaire and dysphoria.

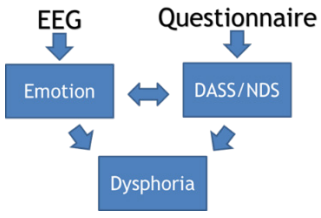


Figure 5: Overview of Correlations between DASS-21/NDS-24, Emotion and Dysphoria

3. Methodology

A theoretical framework of Dysphoria Model of Affect is presented in Figure 6. It consists of five phases, namely; data collection, pre-processing, feature extraction, classification and correlation analysis. Detailed description of each phase is presented in the next sub-sections.

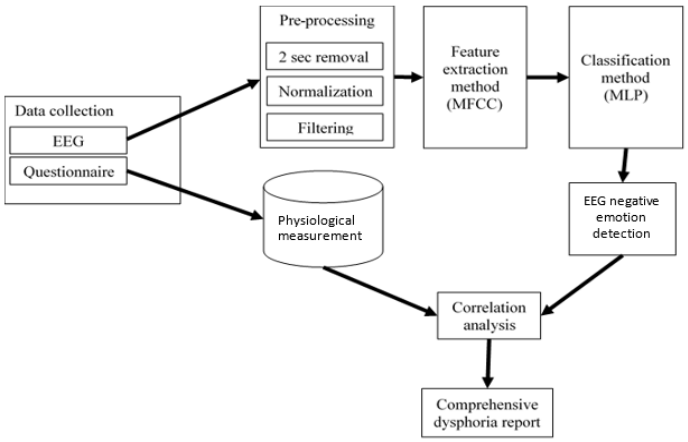


Figure 6: Theoretical Framework of Dysphoria Model of Affect (DMoA)

3.1. Data collection

In this preliminary work only four participants data were collected based on two type of inputs from the psychological instruments (DASS-21) and from the brainwave patterns using the captured EEG signals. The participants are healthy male with no medical record of experiencing acute depression, anxiety and stress with age mean of 27.5 years and variance of ±2.19 years. The participants also indicated that they are under the influence of medicine, drug or alcohol and voluntarily signed the consent form to adhere the IIUM ethic committee procedure.

Prior to taking the brainwave signal using the EEG device, all participants need to fill up two questionnaires namely; the personal information and also the DASS-21 questionnaires. This is to ensure that all four participants do not have any history of acute depression, stress or anxiety. Participants were first briefed about the experiments and were required to sign a consent form as needed by the IIUM ethics committee. To ensure consistency and to relate the study of psychological instruments to the brainwave patterns each individual was required to do the DASS-21 and the NDS-24 questionnaires prior to taking the individual brainwave pattern using the EEG device.

The DASS and NDS questionnaires results are gathered from each of all 4 subjects. The participants are asked to answer the questionnaire truthfully and to take their time to select the most appropriate answer for the questions. The participants can ask the meaning of the word in the questionnaire if they are unsure. This is to minimize misunderstanding that can lead to wrongly selected answer. The DASS questionnaire consists of 21 questions and the NDS questionnaire consists of 24 questions. The participants must answer all questions by selecting their choice in the questionnaire sheet. Once the participants finish answering the task, the questionnaire sheets are collected. Further analysis will be conducted based on the answers given by the participants.

In order to investigate the brainwave patterns signals recorded from the EEG device, the resting state of the participants need to be captured as a measurement basis. It consists of one minute of eyes close and eyes open. Then, it is followed with the reference emotion test based on the International Affective Pictures Systems (IAPS) [26] to analyse the responses of the different parts of the brain to four basic emotions of happiness, sadness, calmness and fear in one minute each respectively. In addition, sadness and fear video stimuli were presented to examine the negative emotion responses of each participant [8]. A final resting state data were collected based on a one minute each of eyes-open and eyes close after the video stimuli. The overall time taken for data collection in this work is 10 minutes without considering the pre-EEG experiment and post-EEG experiment tasks. Also note that all the EEG signals were captured and stored as numeric values to ensure compatibility of data using CSV file format. Table 2 shows the protocol that has been set for the data collection that consists of several condition which is resting state, emotional test and video stimuli.

Table 2: Protocol for EEG Experiment

Resting state (before)		Emotional test				Video stimuli		Resting state (after)	
Eye open 1minute	Eye close 1minute	Happy 1minute	Sad 1minute	Calm 1minute	Fear 1minute	Fear video 1minute	Sad video 1minute	Eye open 1minute	Eye close 1minute
2 minutes		4 minutes				2 minutes		2 minutes	

In the EEG data collection, the raw 19 channels EEG signals are captured using EEG device from BrainMarker. The device allows 19 channels EEG to be captured. However, in this work, only 6 channels are considered; namely, F3, F4, F7, F8, FP1 and FP2 to measure the negative emotion. These six channels are selected because signals from frontal brain are relevant to emotion processing according researchers [27, 28, 29] hence contains a lot of emotion information.

3.2. Pre-processing

Artifact and noise cancellation consist of three steps, namely; removing the first 2 seconds of the EEG signals, normalization and filtering. The pre-processing phase is important to ensure the analyzed data are in the region of interest and not biased or distorted by the effect of artifact and noise. Otherwise the data is regarded as inaccurate and misleading. The first 2 second of the 1-minute data of each protocol typically consists of 500 instances depending on the size of selected window size. The removal of this data is because it may not contain any useful information as the participant just starts to focus their attention to discriminate the emotion based on the presented stimuli. Then, the time vs frequency domain signals are then mapped into 19 channels to be extracted (19 channels x 14,500 instances). For the purpose of this study, only 6 channels data are considered to focus on the data of the brain frontal region.

The data from the 6 channels are normalized to [0,1] and filtered using the elliptical filter. The elliptical filters are employed to ensure to get the cut off frequency/clean signals before going

through feature extraction process. The EEG signals that are higher than 50Hz are also eliminated on the grounds that such data are usually artefacts gathered from various hand, eye and muscle movements.

3.3. Feature Extraction

Once the noise and artefacts are removed, relevant features are extracted for classification purposes. In this work, Mel Frequency Cepstral Coefficient (MFCC) feature extraction method is employed. In MFCC, window size and framing are two main components to observed. Windowing is a technique to cut the large signal into a manageable size for processing and analysis. Windowing is done for minimizing the disruptions at the starting and at the end of the frame, the frame and window function is being multiplied. If the window size selected is too large, the difference maybe too minimal for analysis whereas if the selected window size is too small, the information may not be enough for analysis. Frame is the process of dividing the signal in window into small frames to have reliable information. In the framing process, the raw EEG signals are blocked into small frames of N samples, with next frames separated by M samples (M<N) with this the adjacent frames are overlapped by N-M samples. This procedure of segregating up the signals into frames will proceed until the whole raw EEG signal is separated into small frames. Segregation of frame will make the computation more effective in term of time and processing. In this work, window type used is Hamming window with number of coefficients of 12 resulting to 12 features are extracted for each channel making it 72 features in total.

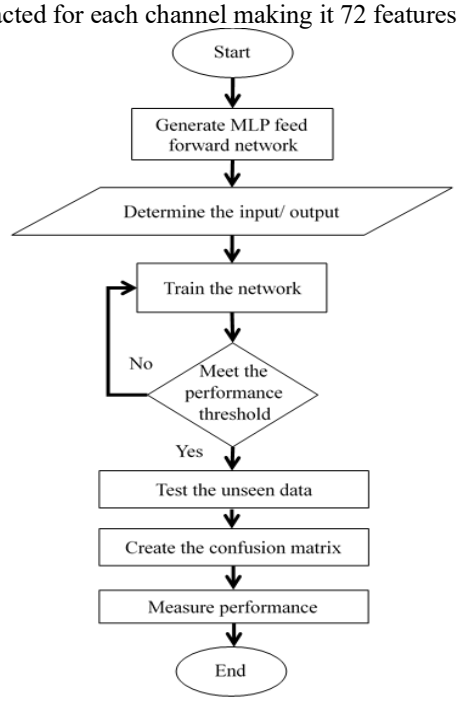


Figure 7: Flow Chart for Emotion Classification

3.4. Classification

Data are arranged according to the experimental set up needs after the relevant features are extracted. In this work, k-fold validation technique is adopted. The data are randomly partitioned into 5 groups where each approximately have an equal size. One group is used for testing while the other four groups are used for prior training. The process is iterated until all fold are used for testing.

In this work, Multi-Layer Perceptron classifier is adopted based on prior work findings [8,29]. The steps of using the MLP are simplified in the flowchart in Figure 7. Input are feed to the 72-neurons input layer before they are passed to the 10-neurons hidden layer. Then, the summation process is conducted in the output layer. The weights of the learning function of each node are iteratively improved until the performance threshold is met.

4. Result and Discussion

There are two types of experiment conducted in this work, namely; emotion recognition and emotion identification. For emotion recognition task, specific emotion is checked whether it is available or not. It is a 2-class classification task. Subsequently, the emotion identification task segregates the emotion individually and compare the intended emotion to the data pool.

4.1. Emotion Detection

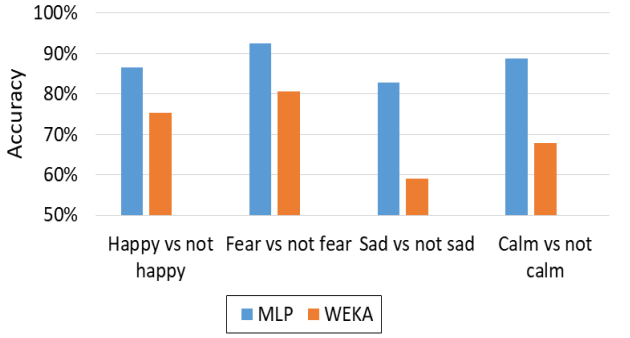
The emotion detection experiment is conducted to discern the existence of specific emotion in a pool of instances. The data are arranged for two-class classification task, for instance; happy versus not happy, fear versus not fear, sad versus not sad and calm versus not calm. The emotion data consists of 67 instances and the data that opposed of the specific emotion are the combination of 1/3 of each other emotions. The total number of instances for the experiment is 134 instances. Features extracted are 72 instances gathered from 12 MFCC coefficients from the six EEG channels of F3, F4, F7, F8, FP1 and FP2. The average of the ten iterations of the experiments are presented in Figure 8(a)-(d). As a comparison, the results gathered from WEKA are also presented. Results from WEKA are used as the base result because the WEKA results are consistent. From the experimental results, it is observed that the result obtained from MLP from Matlab (denoted as MLP) gives higher accuracy as compared to WEKA with increment of 10% to 16% accuracy. The best performance was recorded by MLP for Happy vs Not Happy experiment from Subject4 with performance of 95.5%. The worst performance was recorded by Calm vs Not Calm for Subject3 with performance of 58.2%.

4.1. Emotion Identification

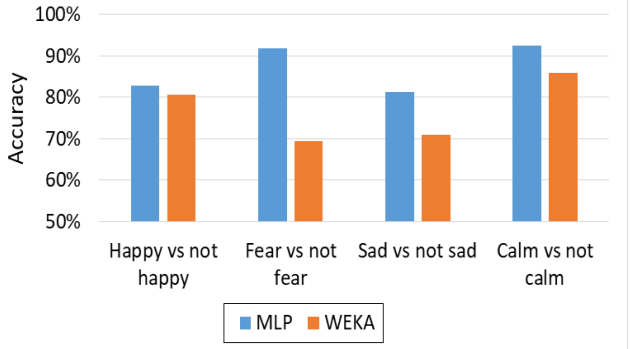
The emotion identification experiment is conducted to observe the ability of the classifier to predict the specific emotion given a pool of different emotion data. The data are arranged in such a way that all emotion has a similar number of instances (67 instances each) gathered from the frontal EEG signals of F3, F4, F7, F8, FP1 and FP2 for all participant. Table 3 shows the detailed MLP experimental result accuracy for emotion identification. Each experiment is repeated 10 times because the MLP performance in Matlab may have difference of ±10%.

From the result in Table 3, it is distinguished that the highest emotion identified is happy by Subject4 which achieved 97.01% accuracy. On the contrary, the lowest emotion identified is fear by Subject3 which yielded 11.94% accuracy. The result then is summarized as in Table 4. Subject4 managed to record the highest average performance for emotion identification whereas the worst

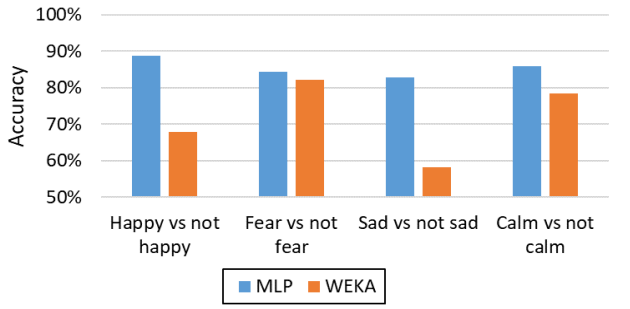
performance is yielded by Subject2 with accuracy of 64.51%. It is also observed that MLP managed to achieve accuracy ranging from 64.51% to 76.72% as compared to WEKA with accuracy ranging from 55.97% to 63.81% that show slight superiority of MLP as compared to WEKA.



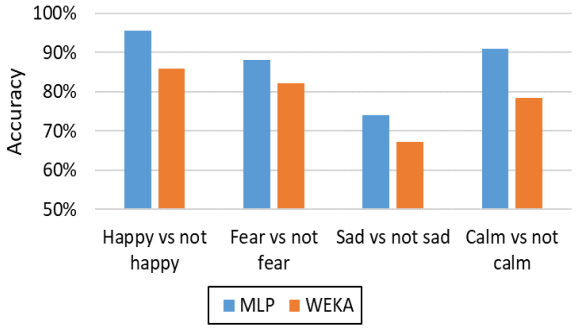
(a) Emotion Detection Result for Subject1



(b) Emotion Detection Result for Subject2



(c) Emotion Detection Result for Subject3



(d) Emotion Detection Result for Subject4

Figure 8: Emotion Detection Result using WEKA and MLP Classification

Table 3: Average Emotion Identification Results accuracy (%)

	Exp#1	Exp#2	Exp#3	Exp#4	Exp#5	Exp#6	Exp#7	Exp#8	Exp#9	Exp#10
Subject1	57.09	65.30	77.99	72.76	69.03	73.13	70.90	70.90	60.82	69.03
Subject2	52.99	77.61	83.58	69.40	82.84	84.33	72.01	79.10	85.45	79.85
Subject3	72.76	83.21	63.43	78.36	22.76	67.16	79.85	82.46	76.49	18.66
Subject4	86.94	61.19	72.76	70.15	73.13	54.10	70.90	68.66	74.63	68.28

Table 4: Summary of Emotion Identification Results

Participant	MLP Accuracy (%)	WEKA Accuracy (%)
Subject 1	68.69	55.97
Subject 2	76.72	63.06
Subject 3	64.51	60.45
Subject 4	70.07	63.81

4.2. Emotion Identification for Video Stimuli

Further analyses are conducted to see the difference in performance of using different media of stimuli, namely; static images and videos. For video stimuli, two videos of sad and fear videos are selected. Figure 9 and Figure 10 show the emotion identification experimental results for fear and sad videos respectively.

Figure 9 shows the bar chart of the experimental result yielded when using fear video. Based on the result, fear is consistently detected for Subject2, Subject3 and Subject4 with accuracy ranging from 33.1% to 37.1%. Subject1 recorded Calm as the highest emotion with difference of 3.5% lower than Fear. However, for the emotion experimental result Figure 10 shows the bar chart result for sad video using emotion identification network. Based on the bar chart, the highest emotion detected is fear in Subject4. Only Subject2 shows the highest sad emotion detected. This result shows inconsistency during video stimuli. This may due to the surrounding when eye open, the information enters through the eye in form of light. This information will react to the brain for visual perception.

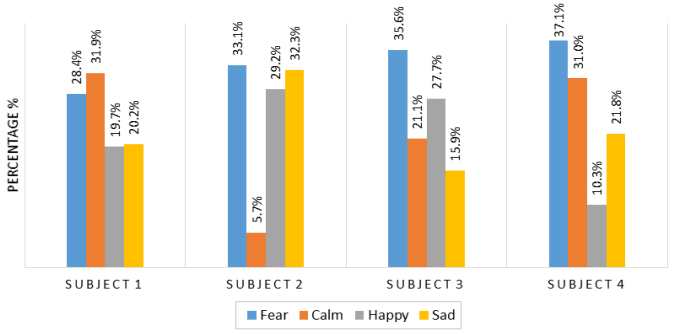


Figure 9: Emotion Identification Result using Fear Video

Extension of the comparison between the effect of emotion elicitation using dynamic emotional contents (video) and static images emotional contents (IAPS) are conducted. Figure 11(a)-(d) show the comparative accuracy of emotion recognition experiments between static images and videos for each subject.

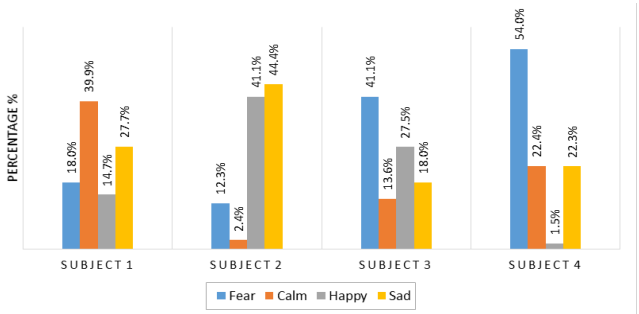
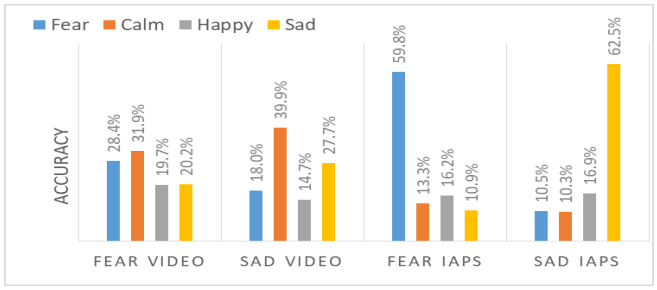
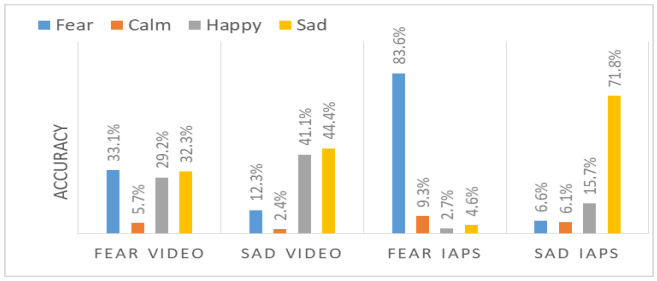


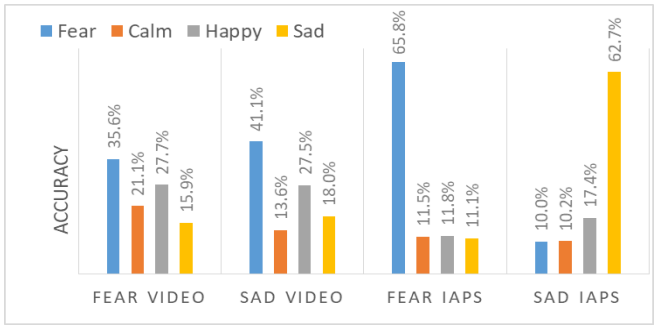
Figure 10: Emotion Identification Result using Sad Video



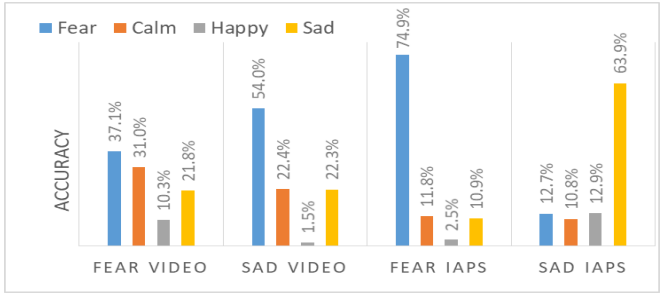
(a)



(b)



(c)



(d)

Figure 11: Result Comparison for Emotion Identification Result using Static Images and Videos for (a) Subject1 (b) Subject2 (c) Subject3 and (d) Subject4.

From the experimental results in Figure 11, the emotion identification result using static emotional contents outperformed the dynamic emotional content in all experiments. The highest accuracy achieved is fear IAPS with 83.6% for Subject2. Meanwhile the lowest accuracy is sad video for Subject4 with 1.5%. This is because the brain needs to process more information when video stimuli is exhibited instead of static images displayed in IAPS stimuli. Hence, lower performance is recorded.

4.3. Questionnaire result: DASS-21 & NDS-24

Table 5 and 6 shows the summary result of the DASS-21 and NDS-24 results respectively for all subjects. According to the depression, anxiety and stress score scale, all participants are in a normal condition. None any of them shows any negative result. However, in Table 6, Subject1 and Subject2 results indicated that they are in mild state as compared to Subject4 that recorded moderate and severe for Subject3 respectively. This result gives better insight as compared to the DASS-21 results that only indicate the participants are normal. Hence, the NDS is able to give more information about the participants' state as compared to the DASS-21 questionnaire and the DASS-21 questionnaire cannot be directly used to determine whether the subject is having dysphoria or otherwise.

Subject	DASS-21 score			Result
	Depression	Anxiety	Stress	
1	2	6	6	Normal
2	2	2	10	Normal
3	2	4	8	Normal
4	2	6	8	Normal

	Total factor Subject1	Total factor Subject2	Total factor Subject 3	Total factor Subject4
Irritability	10.8	8.68	23.2	15.2
Discontent	6.91	5.41	18.3	12.34
Surrender	5.98	5.07	12.9	8.284
Interpersonal resentment	3.41	2.14	7.39	4.55
Result	**Mild**	**Mild**	**Severe**	**Moderate**

4.4. Correlation between DASS-21, NDS-24 & EEG emotion identification

From the emotion identification experiments as in Table 7, fear is highly detected during the eye close which is in line with NDS-24 result that is severe for Subject3. There seems to be correlation between EEG emotion identification and NDS-24. However, it is not highly observable on Subject 1, 2 and 4 because NDS-24 result for each subject is only mild and moderate.

Table 7: Eye Close and Eye Open Before Experiment Training with Emotion Identification MLP Architecture

	Emotion	Before stimuli		After stimuli	
		Eye close	Eye open	Eye close	Eye open
Subject 1	fear	37.5%	43.0%	22.4%	43.0%
	calm	35.1%	25.1%	34.6%	25.1%
	happy	15.7%	24.8%	17.8%	24.8%
	sad	11.8%	7.2%	25.2%	7.2%
Subject 2	fear	41.5%	17.9%	1.2%	17.9%
	calm	10.9%	11.0%	5.4%	11.0%
	happy	31.4%	33.9%	52.0%	33.9%
	sad	16.3%	37.2%	41.5%	37.2%
Subject 3	fear	70.9%	12.2%	75.7%	12.2%
	calm	20.0%	16.4%	9.7%	16.4%
	happy	4.0%	45.7%	5.4%	45.7%
	sad	5.1%	25.7%	9.3%	25.7%
Subject 4	fear	26.4%	49.4%	1.0%	49.4%
	calm	19.6%	18.1%	16.9%	18.1%
	happy	46.3%	27.9%	76.9%	27.9%
	sad	7.8%	4.6%	5.2%	4.6%

5. Conclusion

The Dysphoria Model of Affect (DmoA) uses the neurophysiological data from the EEG devices and by extracting the emotional information from the brainwave patterns different emotions can be detected during the initial state of eyes closed. Such information are very useful in analyzing the mental state of an individual. In this preliminary study, the implementation of DmoA seems to be able to differentiate different classes of emotion fairly well thus resulting in a better tool for early detection of dysphoria as depicted by the NDS-24 results for Subject3. It is also observed that the DASS-21 questionnaire result could not differentiate dysphoria for all the subjects in this work. There are other psychological instruments that can be studied for detail analysis of dyphoria but would be expensive and require experts to interpret score. The proposed DMoA approach provides a new avenue for psychologist and psychiatrist to venture using EEG devices as tools for analyzing dysphoria, depression, anxiety and stress and can also be used in understanding personality traits or behaviour [29].

Acknowledgment

The authors would like to thank MARA University of Technology, Malaysia (UiTM) and Ministry of Education Malaysia (KPM) for providing financial support through the Fundamental Research Grant Scheme FRGS (600 – RMI / FRGS 5/3 (106/2014)) to conduct the work published in this paper.

References

[1] Kranner, I., Minibayeva, F. V., Beckett, R. P. & Seal, C. E. (2010). What is Stress? Concepts, definitions, and applications in seed science. *New Phytologist*. 188: 655–673.

[2] Edwards, J. (2015). *Fundamental facts about mental health 2015*, Mental Health Foundation, United Kingdom

[3] Mallow M. S. (2016). Occupational Stress in Malaysia: Causes, Effects and Possible Solutions, In Proceedings of 3rd International Conference on Education, Social Sciences and Humanities (SOCIOINT 2016), 82-88, Istanbul, Turkey.

[4] Poll H., American Psychological Association (APA) (2017). Stress in America: The state of our nation. Stress in America Survey.

[5] Institute of Public Health (IPH), National Health and Morbidity Survey (2011). NHMS 2011, Vol. II: Non-communicable diseases, Ministry of Health Malaysia, ISBN: 978-967-3887-68-2.

[6] Institute of Public Health (IPH), National Health and Morbidity Survey (2015). NHMS 2015, Vol II: Non communicable diseases, risk factors and other health problems, Ministry of Health Malaysia, ISBN: 978-983-2387-23-7.

[7] Starcevic, V., Berle, D., Viswasam, K., Hannan, A., Milicevic, D., Brakoulias, V., Dale, E. (2015). Specificity of the Relationships between Dysphoria and Related Constructs in an Outpatient Sample, *Psychiatr Q*, Vol 86: 459–469

[8] Kamaruddin, N., Nasir, M. H. M. and Wahab, A. (2018). EEG Affective Modelling for Dysphoria Understanding. *In 2018 12th International Symposium on Medical Information and Communication Technology (ISMICT)*, 1-6.

[9] Kamaruddin, N., Wahab, A. & Quek, C. (2012). Cultural Dependency Analysis for Understanding Speech Emotion, *Expert System with Application (ESWA)*, 39(5): 5115-5133.

[10] Snaith, R. P., & Taylor, C. M. (1985). Irritability: definition, assessment and associated factors, *The British Journal of Psychiatry*, 147(2), 127-136

[11] Berner. P., Musalek. M., Walter H. (1987) Psychopathological Concepts of Dysphoria. *Psychopathology*, 20:93–100

[12] Starcevic, V. (2007). Dysphoric about dysphoria: towards a greater conceptual clarity of the term, *Australasian Psychiatry*, Vol 15(1): 9 -12.

[13] Berle, D. & Starcevic V. (2012). Preliminary validation of the Nepean Dysphoria Scale, *Australasian Psychiatry,* 20(4) 322–326.

[14] Lovibond, P.F. & Lovibond, S.H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33, 335-343.

[15] Beck, A. T., Steer, R. A. & Brown, G. K. (1996). Beck Depression Inventory manual 2nd ed San Antonio: Psychological Corporation.

[16] Cohen D., Kamarck T., & Mermelstein R. (1983). A global measure of perceived stress, *Journal of Health and Social Behavior*, 24: 386-396.

[17] Reiss, S., Peterson, R. A., Gursky, D. M. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behav Res Ther*, 24: 1–8.

[18] Likert, R. (1932). A technique for the measurement of attitudes, *Archives of Psychology*, 140: 1–55.

[19] D'Agostino, A., Manganelli, E., Aportone, A., Monti, M. R., Starcevic V. (2016). Development, cross-cultural adaptation process and preliminary validation of the Italian version of the Nepean Dysphoria Scale, *Journal of Psychopathology*, 22: 149-156

[20] Caseras, X., Garner, M., Bradley, B. P., & Mogg, K. (2007). Biases in visual orienting to negative and positive scenes in dysphoria: An eye movement study. *Journal of Abnormal Psychology*, 116(3), 491.

[21] McMakin, D. L., Santiago, C. D., & Shirk, S. R. (2009). The time course of positive and negative emotion in dysphoria. *The Journal of Positive Psychology*, 4(2), 182-192.

[22] Schlosberg, H. (1954). Three Dimensions of Emotion. *Psychological Review*, 61(2), 81-88.

[23] Siegle G. J., Condray R., Thase M. E., Keshavan M. & Steinhauer S. R. (2010), Sustained gamma-band EEG following negative words in depression and schizophrenia, *International Journal of Psychophysiology*, 75 : 107–118.

[24] Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R. & Chen, J. H. (2010), EEG-based emotion recognition in music listening, *IEEE transactions on biomedical engineering*, Vol. 57(7).

[25] Sourina, O. & Liu, Y. (2014). A fractal-based algorithm of emotion recognition from EEG using arousal-valence model*, IEEE international conference on systems, man, and cybernetics.*

[26] Lang, P. J., Bradley, M. M. & Cuthbert, B. N. (1999). International Affective Picture System (IAPS): Technical Manual and Affective Ratings, University of Florida, Center for Research in Psychophysiology, Gainesville

[27] Cacioppo, J. T. (2004). Feelings and Emotions: Roles for Electrophysiological Markers, *Biological Psychology*, Vol. 67(1): 235-243.

[28] Asakawa, T., Hayashi, T. & Mizuno-Matsumoto, Y. (2014). Coherence Analysis of EEG under Emotional Stimuli Related to Mental States. *Electronics and Communications in Japan*, Vol. 97(8): 14-23.

[29] Wahab, A., Kamaruddin, N., Palaniappan, L. K., Li, M., & Khosrowabadi, R. (2010). EEG signals for emotion recognition. *Journal of Computational Methods in Sciences and Engineering*, 10(s1), 1-11.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 207-212 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Integrating Diacritics Restoration and Question Classification into Vietnamese Question Answering System

Bui Thanh Hung*

Data Analytics & Artificial Intelligence Laboratory, Engineering - Technology Faculty, Thu Dau Mot University, 590000, Vietnam

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Diacritics Restoration
Question Classification
Vietnamese Question Answering System
Deep Learning
Encoder-Decoder LSTM
Bi-LSTM

ABSTRACT

This paper presents a solution for question answering system for Vietnamese language by integrating diacritics restoration and question classification via deep learning approach. It could be said that this will be the first research integrating two phases into Vietnamese question answering system. Question classification has a critical role in the question answering system. However if the question has too many missing diacritics, this will make the classification extremely more difficult. In this paper, both automatic insertion of diacritics and question classification tasks are built to rely on deep learning approach. For diacritics restoration task, we apply the Encoder-Decoder LSTM model. The result of the first step will be the input of question classification. We use pre-train word embeddings in the Bidirectional LSTM model for Vietnamese question classification. The deep learning approach for both tasks is powerful and highly accurate model. By integrating diacritics restoration and question classification into Vietnamese question answering system – ICTbot of Binh Duong Department of Information and Communications Support System – it has produced remarkably positive results; thus proves the practicability of this proposed system.

1. Introduction

Internet allows the researchers to save data and make them available to the public. However, it also makes the search of information in such big data environment more complicated and expensive. Question Answering systems have been developed as a new advanced research tools for this issue.

An automated question answering system has several key steps to follow such as:

- The user inputs the question through query interface. The question is analyzed, pre-processed and phrased into words.
- Question classification identifies the type of the question. This step is rather critical to achieve the most appropriate answering type.
- Finally, based on the result of Question classification, the system will choose the appropriate answer from the answer sets.

In the first step, if there are too many missing diacritics in questions, it will make it challenging for the second step - Question classification. A diacritic is a mark written either above or below a letter to alter its original phonetic or orthographic value. Diacritics are used in several languages' orthography. Nevertheless, in daily conversation, as a matter of convenience, diacritics are usually ignored in many languages. The absence of diacritics in question text makes it extremely challenging for both Question Classification and Question Answering system.

The Vietnamese alphabet uses Latin script with diacritical signs. However, Vietnamese has seven modified letters including ă, â, đ, ê, ô, ơ, and ư and also six tone marks including unmarked tone (ngang), acute (sắc), grave (huyền), hook above (hỏi), tilde (ngã), and underdot (nặng). Due to its diacritical complexity, when typing on phones or computers, most of Vietnamese people prefer to type non-diacritic characters as a matter of convenience. As a consequence, it sometimes causes the confusion or mis-interpretation for the recipients or readers. With that being said, integrating diacritics restoration into question classification is a critical task for Vietnamese question answering system.

We propose the Vietnamese Question Answering model with two phases in this paper. The first phase is automatically recovering the missing diacritics. And the second phase is integrating diacritics restoration into Vietnamese question

*Bui Thanh Hung, Engineering - Technology Faculty, Thu Dau Mot University, hungbt.cntt@tdmu.edu.vn

answering system. We use the Encoder-Decoder LSTM model for automatically recovering the missing diacritics and the Bidirectional LSTM model for Vietnamese question classification. We integrate diacritics restoration into Question classification of Vietnamese question answering system.

This paper consists of five sections. The related works are reviewed in Section 2. Our method is presented in Section 3. Section 4 mentions the procedure of constructing dataset, the experiment settings and discusses the outcomes of the experiments. Our work and possible future research are concluded in Section 5.

2. Related Words

Firstly, diacritics restoration – as the first step – could be run following two typical approaches: character-based [1-2] and word-based [3-5]. For Vietnamese language, many researches propose their methods to restore accents such as [6-9]. In this research we choose the Encoder-Decoder LSTM model that proposed recently by Bui [6] because of the advanced approach, and high accuracy for Vietnamese diacritics restoration.

For question classification, there are three main approaches: Rule-based, Machine Learning and Hybrid ones [10]. Using a few manually handcrafted rules to match the question is the rule-based approach. One of key challenges here is to define too many rules. Machine-learning approach is to train a classifier and use the trained classifier to predict the class label. Combining the two above - the rule-based and machine learning - is the hybrid approach. In terms of question classification, the most successful approaches are Machine Learning and hybrid methods. Nguyen et al. [11] proposed a SVM based method for the same task. Firstly, the question was parsed and tokenized, parts-of-speech were tagged, data removed stop-words and was stemmed. Next they extracted a lot of features and selected features before passing the data into a support vector machine for training. For test questions they used the same pre-processing. Deep Learning approaches have proved that it is significantly beneficial for text classification, summarization, machine translation, etc. as well as for question classification [12-15]. Andreas et al. [12] proposed compose neural networks for question answering. Kim et al. [16] employed Convolutional Neural Networks (CNNs) and treated questions as general sentences to achieve remarkably strong performance in the TREC question classification task.

With regards to our research, what differentiates it from the others is the Vietnamese question answering system. This is the first research integrating two phases - diacritics restoration and question classification into Vietnamese question answering system. The diacritics restoration is the Encoder-Decoder LSTM model and Question classification uses the Bi-LSTM model. The result of the first step will be the input of question classification.

3. Methodology

Our model includes two steps: Diacritics Restoration and Question Classification. In the first step, we use Encoder-Decoder LSTM model proposed by Bui [6] and in the second step, pre-train word embeddings are used in the Bidirectional LSTM model for Vietnamese question classification. We integrate the result of the first step into question classification. An overview of our architecture is illustrated in Figure 1. We will describe more details of each layer in our model as follows:

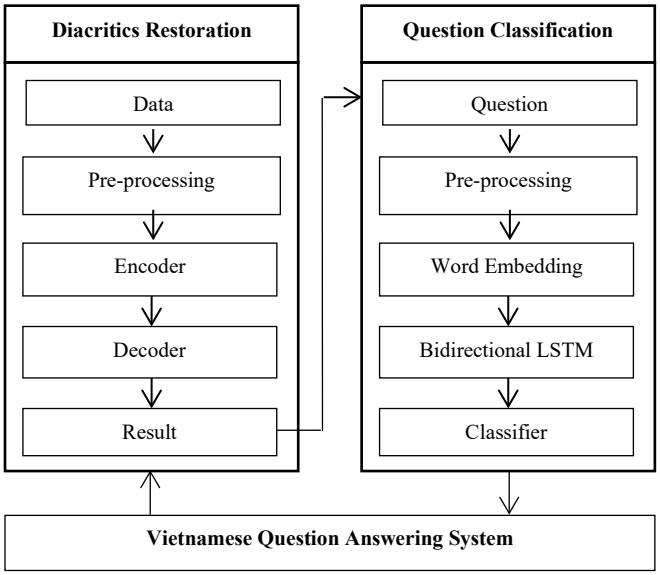


Figure 1: The proposed model

3.1. Word Embeddings

Indeed, it has been discussed for quite a long time to use vector representation for words. And recently there has been more and more interest in word embeddings - a technique which maps the words to vectors. Tomas Mikolov's Word2vec algorithm [17] is one driver for this which takes a big volume of text to create high-dimensional representations of words. By this way, it could identify the relationships between words which are not supported by external annotations. Thus, it proves to succeed in getting lots of linguistic regularities with such representation.

In this research, we use word embeddings - Continuous Bag of Words (CBOW) [17]. In the CBOW model, a variety of words illustrate the context for a defined target word which stands for a modification of neural network architecture. For example, with "trèo" (climbed) as the target word, let's use "mèo" (cat) and "trên" (on) as the words to describe the context. Figure 2 describes this model as below:

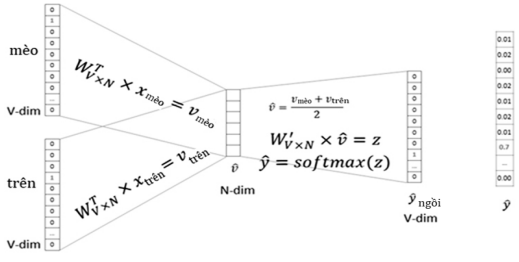


Figure 2: Continuous Bag of Words (CBOW)

3.2. LSTM

Long short-term memory (LSTM) [18] is a modification of the Recurrent neural networks (RNN). Thanks to the feedback loop in its architecture, this model has the ability to store the records of previous outputs.

Figure 3 represents a diagram of a simple LSTM cell. As befits the term deep neural networks, all single cells are combined altogether to shape a huge network. The cell unit represents the memory with five main elements: an input gate *i*, an output gate *o*, a forget gate *f*, hidden state output *h* and a recurring cell state *c*. Given a sequence of vectors (x_1 ,x_2, ..., x_n), σ is the logistic sigmoid function, the hidden state h_t of LSTM at time t is calculated as follows:

$$h_t = o_t * tanh(c_t) \quad (1)$$

$$o_t = tanh(Wx_0x_t + Wh_0h_{t-1} + Wc_0c_t + b_o) \quad (2)$$

$$c_t = f_t * c_{t-1} + i_t * tanh(Wx_cx_t + Wh_ch_{t-1} + b_c) \quad (3)$$

$$f_t = \sigma(Wx_fx_t + Wh_fh_{t-1} + Wc_fc_{t-1} + b_f) \quad (4)$$

$$i_t = \sigma(Wx_ix_t + Wh_ih_{t-1} + Wc_ic_{t-1} + b_i) \quad (5)$$

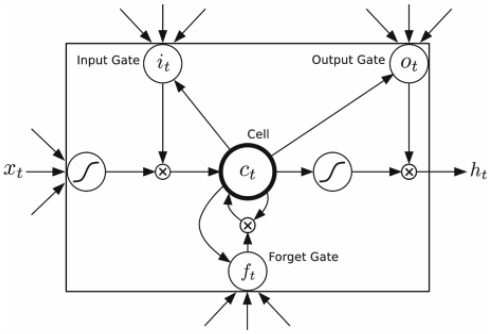


Figure 3: The Long Short Term Memory cell (Source: [19])

3.3. Bidirectional LSTM

Bidirectional LSTM [20] could be seen as a suitable model in many tasks of natural language processing. This model is modified version of LSTM by combining forward and backward LSTMs. The Bi-LSTM model is shown in Figure 4.

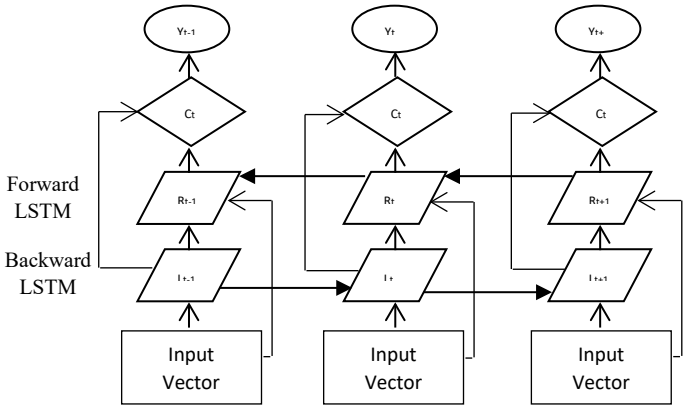


Figure 4: Bidirectional LSTM

Assuming that above equations are abbreviated to

$$h_t = \text{LSTM}(x_t\ h_{t-1}). \quad (6)$$

The forward $\overrightarrow{ht}$ and $\overleftarrow{h}$t the backward LSTM are defined as follows:

$$\overrightarrow{h_t} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}}) \quad (7)$$

$$\overleftarrow{h_t} = \text{LSTM}(x_t, \overleftarrow{h_{t-1}}) \quad (8)$$

Where → denotes the forward and ← denotes backward. By taking the forward and backward LSTM as input, the Bidirectional LSTM $\overleftrightarrow{ht}$ at time *t* leads to further classification procedure:

$$\overleftrightarrow{h_t} = [\overrightarrow{h_t}, \overleftarrow{h_t}] \quad (9)$$

3.4. Diacritics Restoration

We used the Encoder-Decoder LSTM model proposed by Bui [6] for Vietnamese diacritics restoration. The Encoder-Decoder model was introduced by Kyunghyun Cho et al. [14]. This model learns to encode a variable-length sequence into a fixed-length vector representation and to decode a given fixed-length vector representation back into a variable-length sequence. This model is presented in Figure 5.

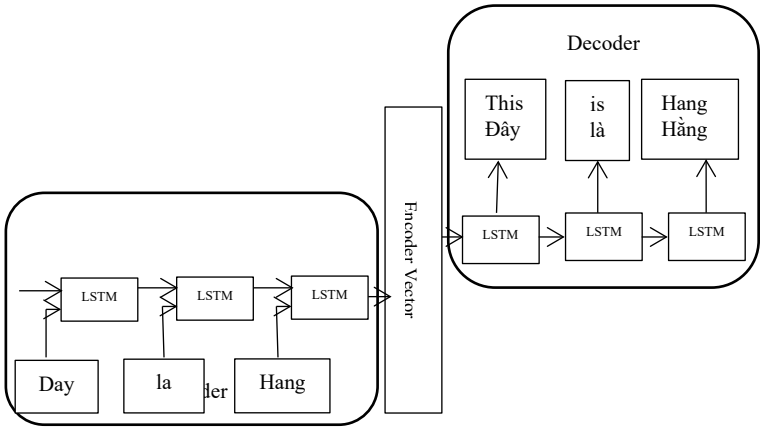


Figure 5: Encoder-Decoder LSTM model

In Figure 5, the input sequence is displayed to the network one encoded character at once. The output produced from this model is a fixed-size vector that shows the internal representation of the input sequence.

3.5. Question Classifier

The output layer is defined as follows based on the text vector studied from the Bidirectional LSTM model:

$$y = (W_d x_t + b_d) \quad (10)$$

where x_t: the text vector studied from the Bidirectional LSTM model; y: the degree of question class of the target text; W_d, b_d: the weight and bias associated with the Bidirectional LSTM model.

The Bidirectional LSTM model is trained by minimizing the mean squared error between the predictive question class and true question class. The loss function is identified as follows:

$$L(X, y) = \frac{1}{2n}\sum_{k=1}^{n}\binom{n}{k}\left\|h(x^i) - y^i\right\|^2 \quad (11)$$

$\mathbf{X} = \{x^1, x^1, x^2, \ldots, x^m\}$: a training set of text matrix

$\mathbf{y} = \{y^1, y^2, \ldots, y^m\}$: a question class ratings set of training sets

4. Experiments

We used dataset in [6] for the first step. This dataset was crawled from the official newspapers website from various categories like sports, world, business and entertainment. The size of dataset is 150 MB. The training dataset is 1500000 sentences and testing dataset is 5000 sentences. We removed all diacritical marks following the rule presented in [6].

To enable a LSTM network to process and remember more effectively, we are going to break dataset down to a lot of n-grams. Also, Vietnamese tone marks for a word can be determined from the small surrounding words in most of the cases; therefore something like a 5-gram or 7-gram model will rather fit for our purpose to some extent. So we have 11 million 5-gram with length varied mostly from 15-25 characters. In addition, as mentioned above, the easiest way to gather a lot of training data to add diacritical marks to Vietnamese text is to get a diacritical text then remove the diacritical marks.

As a machine learning framework, Keras with Tensorflow as a backend has been used. We used same parameters of [6] with 3 LSTM layers, each layer has 256 neurons. We optimized by Adam Optimizer with learning rate 0.002. We used Accuracy score - Accuracy of Diacritics Restoration (ADR) to evaluate the first step as below equation:

$$ADR = \frac{Number\ of\ corrected\ words}{Number\ of\ words} \quad (12)$$

We did experiements with the range of epochs, after about 300 more epochs, or 16 hours we got 97% accuracy and reduced loss to 0.07. Figure 6 shows the Loss of Diacritics Restoration over epochs and Accuracy of Diacritics Restoration per epochs are shown in Figure 7.

Since the used data was extracted from the newspaper websites, it might consist of unpopular words and a couple of errors. Therefore, the diacritic restoration in those cases usually will not achieve highly accurate results.

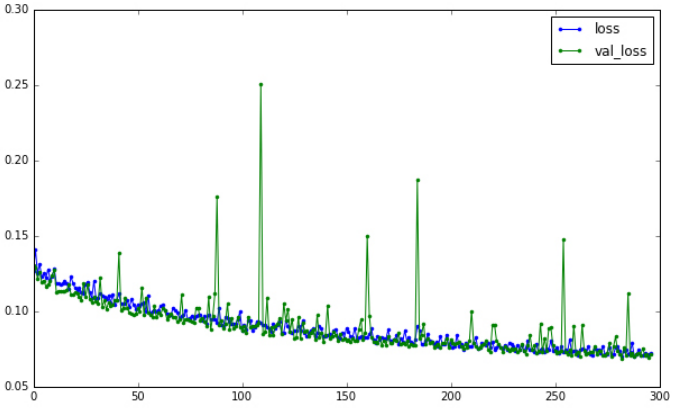


Figure 6: The loss of Diacritics Restoration over epochs

For the second task, we collected question classification dataset from Binh Duong Department of Information and Communications. This data includes 4 areas with 42 types of provincial administrative procedures [21]. The dataset has 560 questions labelled as five primary categories and 270 sub-categories. The main categories and their corresponding number of subcategories are listed in the Table 1. Table 2 shows the details our dataset for training and testing in Question Classification with ratio 8:2.

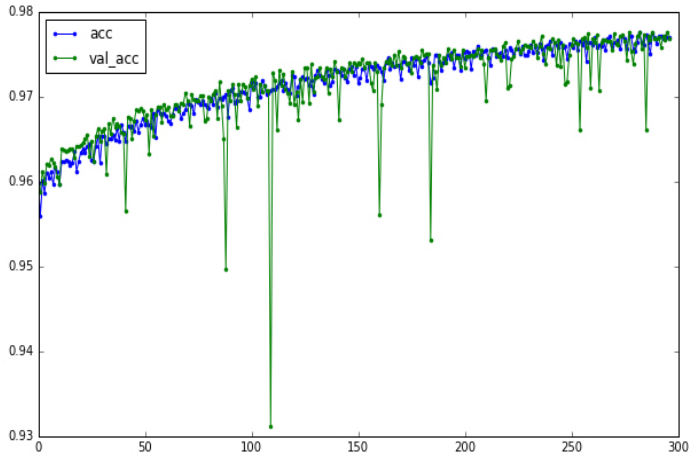


Figure 7: Accuracy of Diacritics Restoration task per epochs

Table 1: Describe of dataset of Question Classification

Categories	Number of Sub-Categories
Journalism	35
Postal	42
Broadcasting	84
Publishing	98
General Information	11

Table 2: Dataset of Question Classification

Name of Dataset	Number of Sentence
The set of questions	570
Sub- Categories	270
Train	456
Test	114

For pre-processing dataset, we used Pyvi (0.0.0.9 - Tran Viet Trung 2016) to tokenize Vietnamese, pre-train Vietnamese word embeddings by streetcodevn (Hung Le 2018). For our training models, we used Tensorflow and Keras libraries. The parameters of our question classification Bi-LSTM model are: Batch size: 500, number of hidden nodes: 128, Drop out: 0.2, epochs: 300, Sigmoid Activate function, Adam Optimization function and Binary cross entropy function. We evaluated performance of question classification by Accuracy of Question Classification (AQC) following below equation:

$$AQC = \frac{\#of\ correct\ predict\ questions}{\#of\ predict\ questions} \quad (13)$$

To evaluate our proposed model, we compared our results with the result of LSTM model separately as illustrated in Table 3. This

result reveals that our proposed model – the Bi-LSTM deep learning approach get the best results. From the result is shown in Table 3, it is obvious that the Bi-LSTM beats the LSTM model; therefore it could be said that the Bi-LSTM is the best-performing score comparing with the LSTM.

Table 3: The results of question classification

Model	AQC
LSTM	92%
Bi-LSTM	**95%**

We made an application by integrating the automated question answering systems into Binh Duong Department of Information and Communications Support System. The name of this application is ICTbot. The ICTbot is used to support Binh Duong Department of Information and Communications in General Information and four management categories: Journalism, Postal, Broadcasting and Publishing. The ICTbot allows users to input new questions and the result is the answers that have the highest accuracy of question classification. Figure 8 shows the ICTbot application.

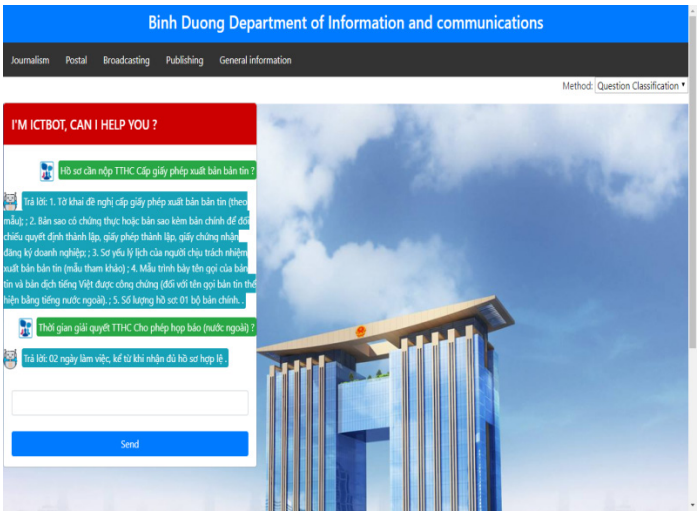


Figure 8: The ICTbot of Binh Duong Department of Information and Communications Support System.

Table 4: The results of question classification in different percentage of Accuracy of Diacritics Restoration

Percentage of Accuracy of Diacritics Restoration	AQC
>= 90%	95%
<= 75 and <90%	70% - 94%
<= 50 and <75%	48% - 69%
<= 25 and <50%	20% -47%
<25%	0-19%

To evaluate the differences in integrating diacritics restoration into question classification, we applied diacritics restoration by percentage into question classification. We counted the number of restoration words in test question dataset and integrated diacritics restoration in test question dataset by percentage of Accuracy of Diacritics Restoration. When percentage of diacritics restoration is equal or larger than 75%, question classification worked well, however when the percentage was so small question classification didn't work well or made a false label of the predict question. The result also depends on the length of question test. Table 4 reveals the result in different percentage of Accuracy of Diacritics Restoration. As a result, it could be seen that integrating diacritics restoration into question classification improves the accuracy of the system.

5. Conclusion

In this paper, we experimented our model by integrating diacritics restoration and question classification into Vietnamese question answering system. The result of the first step will be the input of question classification. It could be said that this is the first research integrating two phases into Vietnamese question answering system. We used the Encoder-Decoder LSTM model for Vietnamese diacritics restoration and the Bidirectional LSTM model for Vietnamese question classification. Our experiments proved that integrating diacritics restoration improves question classification. We also built a useful application based on Question classification – ICTbot and integrated in Binh Duong Department of Information and Communications Support System. In the future, we are looking to apply other deep learning approach and explore more features to improve the results.

Conflict of Interest

I declare no conflict of interest.

References

[1] G. D. Pauw, P. W. Wagacha, and G.-M. de Schryver, "Automatic diacritic restoration for resource-scarce languages," in Proceedings of the Text, Speech and Dialogue, 10th International Conference, TSD, Pilsen,Czech Republic, pp. 170–179, 2007. https://doi.org/10.1007/978-3-540-74628-7_24

[2] M. Simard and A. Deslauriers, "Real-time automatic insertion of accents in French text," Natural Language Engineering , vol. 7, Issue 2, pp. 143–165, 2001. https://doi.org/10.1017/S1351324901002650

[3] Mihalcea, R, "Diacritics restoration: Learning from letters versus learning from words", in Proceedings of CICLing, pp. 339-348. Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, 2002. https://doi.org/10.1007/3-540-45715-1_35

[4] Nikola Santic, Jan Snajder, Bojana Dalbelo Basic, "Automatic diacritics restoration in Croatian texts", The Future of Information Sciences, Digital Resources and Knowledge Sharing, pp. 126–130, Springer, 2009.

[5] R. Nelken and S. M. Shieber, "Arabic diacritization using weighted finite-state transducers", in Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, ser. Semitic '05, Stroudsburg, PA, USA, pp. 79–86, 2005. https://doi.org/10.3115/1621787.1621802

[6] Bui Thanh Hung, "Vietnamese Diacritics Restoration Using Deep Learning Approach", in Proceedings of the 10th International Conference on Knowledge and Systems Engineering – KSE, 2018. https://doi.org/10.1109/KSE.2018.8573427

[7] L.N. Pham, T.V. Hong, V.V. Nguyen, "Vietnamese Text Accent Restoration with Statistical Machine Translation", 27th Pacific Asia Conference on Language, Information, and Computation pp.423 — 429, 2013. https://www.aclweb.org/anthology/Y13-1044

[8] N.M.Trung, N.Q.Nhan, N.H.Phuong, "Vietnamese diacritics restoration as sequential tagging", Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), IEEE RIVF International Conference, 2012. https://doi.org/10.1109/KSE.2018.8573427

[9] T.A. Luu, K. Yamamoto, "A Pointwise approach for Vietnamese automatic diacrics restoration", Proceedings of the International Conference on Asian Language Processing-IALP, 2012. https://doi.org/10.1109/IALP.2012.18

[10] Amit Mishra and Sanjay Kumar Jain, "A survey on question answering systems with classification. Journal of King Saud University - Computer and Information Sciences, 28(3), pp 345 – 361, 2016. https://doi.org/10.1016/j.jksuci.2014.10.007

[11] Nguyen Van-Tu and Le Anh-Cuong, "Improving question classification by feature extraction and selection", Indian Journal of Science and Technology, 9(17), 2016. https://doi.org/10.17485/ijst/2016/v9i17/93160

[12] Andreas, J., Rohrbach, M., Darrell, T., and Klein, "Learning to Compose Neural Networks for Question Answering", Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016. https://doi.org/10.18653/v1/N16-1181

[13] Graves, A., "Supervised Sequence Labelling with Recurrent Neural Networks", Studies in Computational Intelligence, vol. 385, pp. 5-13, Springer, 2012.

[14] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, 2014. https://doi.org/10.3115/v1/d14-1179

[15] Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria, "Recent Trends in Deep Learning Based Natural Language Processing", IEEE Computational Intelligence Magazin, 2018. https://doi.org/10.1109/MCI.2018.2840738

[16] Yoon Kim, "Convolutional neural networks for sentence classification", Conference on Empirical Methods in Natural Language Processing, pages 1746–1751, 2014. https://doi.org/10.3115/v1/D14-1181

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space", in Proceedings of International Conference on Learning Representations (ICLR-13): Workshop Track, 2013. https://arxiv.org/abs/1301.3781

[18] Hochreiter S., Schmidhuber J., "Long Short Term Memory", Neural Computation 9(8), pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Hong J, Fang M, "Sentiment analysis with deeply learned distributed representations of variable length texts", Technical report, Stanford University, CS224d: Deep Learning for Natural Language Processing, 2015. https://cs224d.stanford.edu/reports/HongJames.pdf

[20] Wang P., Qian Y., Soong F. K., He L., Zhao H., "Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Recurrent Neural Network", 2015. https://arxiv.org/abs/1510.06168

[21] Decision No. 1402/QD-UBND dated May 29, 2018 on the announcement of administrative procedures under the jurisdiction of the Department of Information and Communications/District People's Committees of Binh Duong province.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 207-219 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Study of Performance of Bio- Inspired Strategies Applied to Pursuit Evasion Game Under Feedback Laws

Lairenjam Obiroy Singh[1,*], Rajagopalan Devanathan[2]

[1]*Hindustan Institute of Technology and Science, Department of Electronic and Communication Engineering, 603103, India*

[2]*Hindustan Institute of Technology and Science, Department of Electrical and Electronics Engineering, 603103, India*

ARTICLE INFO

Article history:
Received:31 March, 2019
Accepted:07 June, 2019
Online: 18 June, 2019

Keywords:
Bio- Inspired
PEG
Feedback Laws
Closed Loop Control

ABSTRACT

Pursuit Evasion Game (PEG) is an abstract model of various significant problems that appear in both civil and military applications. Bio- Inspired strategies are found to be very useful in studying the PEG. While optimal response to the pursuit strategies are available using geometric control theory, it is shown in this paper that application of linear feedback control laws can further improve the time and tracking response of these strategies in capturing the evader by the pursuer. Empirical results based on computer simulation are used to illustrate the findings. Further, considering the case of sudden turn of the evader, moving at a lower speed, it is shown that both in theory and simulation that the evader can delay the capture by pursuer and in some cases even escape from being captured. These findings are in line with what is found in nature.

1. Introduction

Pursuit evasion game (PEG) is widely prevalent in nature. It is a game that is considered to be between the pursuer and an evader. PEG can be observed among animals when they chase a victim or when they battle for territory and even when they mate. In the context of engineering, PEG finds application in missile guidance and avoidance, aircraft pursuit and evasion, maritime asset protection etc.[1].

Initial study of PEG was mainly from the point of view of game theory [2]. PEG was further studied from the point of view of geometric control theory in [3]. Pursuit manifold is defined in terms of certain criteria involving relative distance and relative velocity between the pursuer and the evader. Using a control law which enables the pursuer to reach the pursuit manifold in an optimum way and keeping the pursuit invariant on the manifold, [3] studied three strategies found in nature, viz., motion camouflage (CM), constant bearing (CB) and classical pursuit (CP). The objective of the work reported in [3] is to find a justification for the prevalent use of CM in PEG in nature. It turned out that, keeping the speed of the pursuer and the evader constant and ensuring that the pursuer moves faster than the evader, [3] has been able to show through evolutionary games that CB and CP strategies always converge to CM strategy under certain fixed assumptions on evader motion thus supporting what is seen in nature.

However, Pais [4] has argued that if the evader control law is reactive (i.e. it is in some way dependent on the baseline vector joining the pursuer and evader locations and its rate of change), then the conclusions in [3] might not hold good. Wei [5] has dropped the constant speed assumption of both the pursuer and evader and allowed the pursuer to change its acceleration as required. Wei also introduced the case of victim turning suddenly left or right when the pursuer comes too close to the evader. A related problem to PEG of confinement and escape is addressed in [6].

Our aim in this paper is this. Considering PEG of a robotic pursuit of an enemy agent, can we consider the problem as a feedback control problem introducing proportional (P),proportional plus integral (PI) and proportional plus integral plus derivative (PID) laws to improve the performance of the pursuer? The performance can be studied under different bio-inspired pursuit strategies, such as CM, CB and CP. The evader may be allowed to follow reactive and non- reactive control laws. A simpler version of this study for the CM case only has been reported earlier in [7]. Further, in view of the increased agility of the evader (i.e. ability to turn at a lower speed compared to the pursuer), we ask what is the outcome of the PEG given the chance

*Corresponding Author: Lairenjam Obiroy Singh, obi3925@gmail.com

that evader makes a sudden 90 degree turn left or right? This paper is a much expanded version of [8] especially in regard to the sudden turn strategy of the evader.

Bopardikar [9] has shown the Bio- inspired co-operative strategies of pursuer for successful confinement of evader. [9] gives the required number of pursuers and speed ratio for guaranteed confinement strategy. X Liang and Y Xiao [10] show the coalition formation of robots for intrusion detection by using the game theory approach. The paper is based on the nature's coalition formation such as predator trying to catch the prey. An analytical method is used to study the tradeoff of coalition or collaboration between the robots.

Initial study of camouflage strategies [11,12] are considered when either pursuer or evader is stationary. Srinivasan [13] analyzed and investigated the motion camouflage and highlighted many ways for future extension. The authors of [14] study the unmanned surface vehicles (USVs) for performing patrolling operations and detecting intruders for harbor protection with a view to reducing the number of humans exposed to threat. An intelligent swarm management unit (SMU) is used for the supervision of all the USVs in operation. A real time motion planner for the USVs in the presence of multiple obstacles is also presented.

[15] addresses the problem associated with the classical pursuit evasion games and mentions that their study is more difficult than the classical one. The author of [15] also explored the geometry of the problem by obtaining sufficient conditions for both pursuer and evader to win. A max- min problem is formulated, from the pursuers point of view, which is solved using outer approximation method. The solution of the max- min formulation is used to synthesize a feedback solution governing the pursuer's behavior in the form of receding horizon control. [16] discusses a harbor defense situation and the author highlights the important features where the problem is different from the classical PEG.

The main contribution of the paper can be summarized as follows: (i) PI and PID feedback laws are shown to enable the pursuer to capture the evader in a shorter time compared to P alone for different conditions of the evader following non-reactive and reactive control laws. (ii) The PI, PID feedback laws followed by the pursuer enable him to capture the evader and not allow him to escape contrary to the the case of P alone. This is termed tracking performance of the pursuer (iii) In the case the evader uses his agility to turn 90 degree left or right suddenly with respect to the baseline, it is shown that the time to capture the evader can be delayed when the pursuer follows any of the CM, CB and CP strategies. (iv) It is also shown that in certain cases of CB followed by pursuer, with the evader following the sudden turn strategy, the evader is able to totally escape from the pursuer inspite of the speed advantage enjoyed by the pursuer.

The rest of the paper is organized as follows: The following section gives the required background on modeling of PEG and the pursuit manifolds under different strategies. Section 3 presents the feedback control system configuration considered together with the derivation of expressions for PI and PID for the three strategies. Also provided in section 3 is an analysis of the effect of sudden turn of the evader in respect of the CM, CB and CP strategies of the pursuer. Section 4 gives the simulation results together with the discussion on the comparative performance of the different control strategies as well as different control modes of the evader. Section 5 concludes the paper.

2. Background

Wei and Krishnaprasad [17], [18] model the interaction in pursuit in terms of gyroscopically interacting particles. We follow a similar approach.

$$\dot{r}_p = x_p\,,\ \dot{x}_p = y_p u_{pr}\,,\ \dot{y}_p = -x_p u_{pr} \qquad (1)$$

where $r_p \in \mathbb{R}^2$is the position of the pursuer, x_p its velocity and y_p is the acceleration of the pursuer. The steering control of the pursuer is given by the scalar u_{pr}. The motion of the evader (with speed υ) is given by

$$\dot{r}_e = \upsilon x_e\,,\ \dot{x}_e = \upsilon y_e u_e\,,\ \dot{y}_e = -\upsilon x_e u_e \qquad (2)$$

where $r_e \in \mathbb{R}^2$ is the position, x_e is the velocity and y_e is the acceleration of the evader. The steering control of the evader, u_e, is a scalar. We also define

$$r = r_p - r_e \qquad (3)$$

which is referred to as the 'baseline' between the pursuer and the evader.

2.1. Pursuit manifolds and cost functions

Two particle pursuer evader system is described in the Euclidean plane of two dimensions. We define the cost functions F on the pursuit manifold $G \in \mathbb{R}^2$ as $F: G \times G \to \mathbb{R}$ associated with different pursuit strategies as in [4] as follows where $(\cdot)$ represents the dot product between two unit vectors and $|\cdot|$ stands for the Euclidean norm.

$$\Gamma = \left(\frac{r}{|r|}.\frac{\dot{r}}{|\dot{r}|}\right) = \frac{\frac{d|r|}{dt}}{\left|\frac{dr}{dt}\right|} \text{ (motion camouflage)} \qquad (4)$$

and

$$\Lambda = \left(\frac{r}{|r|}\ .\ R\ x_p\right) \text{(constant bearing)} \qquad (5)$$

where

$$R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}, \qquad (6)$$

for $\theta \in (-\pi/2, \pi/2)$.

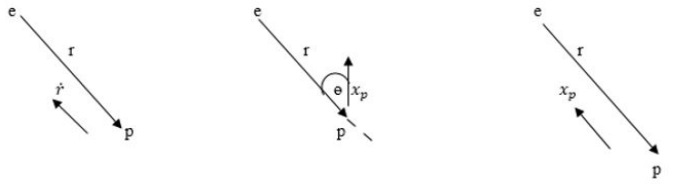
For $R = I_2$ the identity matrix of order 2, we define

$$\Lambda_0 = \left(\frac{r}{|r|}\ .\ x_p\right) \qquad (7)$$

to be the cost function associated with classical pursuit.

All three cost functions Γ, Λ and Λ_0 are well defined and that they take values in the interval [-1,1]. The cost functions Γ, Λ and

Λ_0 define the respective pursuit manifolds. Γ is seen to correspond to the cosine of the angle between r and $\dot{r}$. Camouflage pursuit manifold is defined by the condition $\Gamma = -1$, which corresponds to the case of the angle between r and $\dot{r}$ being π. (see Figure 1(a)).



(a) (b) (c)

Figure. 1 Geometric representation of pursuit manifolds: (a) motion camouflage (CM) pursuit, (b) constant bearing (CB) pursuit and (c) classical pursuit (CP).

The constant bearing pursuit manifold is represented by the condition $\Lambda = -1$. This condition is satisfied when the heading of the pursuer makes an angle θ with the baseline vector (as shown in Figure 1(b)). Similarly, the classical pursuit manifold is defined by the condition $\Lambda_0 = -1$. This condition is satisfied when the heading of the pursuer is aligned (in opposite direction) with the baseline (see Figure 1 (c)).

3. System Modelling

3.1. Pursuit-Evasion system

Representing

$$r_p = [r_{px}\ r_{py}]^t$$

$$r_e = [r_{ex}\ r_{ey}]^t$$

we write (1) and (2) in terms of state equations as follows:

$$\left.\begin{array}{ll} x_1 = r_{px}; & x_7 = r_{ex}; \\ x_2 = r_{py}; & x_8 = r_{ey}; \\ x_3 = x_{px}; & x_9 = x_{ex}; \\ x_4 = x_{py}; & x_{10} = x_{ey}; \\ x_5 = y_{px}; & x_{11} = y_{ex}; \\ x_6 = y_{py}; & x_{12} = y_{ey}; \end{array}\right\} \quad (8)$$

$$\left.\begin{array}{c} \dot{x}_1 = \dot{r}_{px} = x_{px} = x_3 \\ \dot{x}_2 = \dot{r}_{py} = x_{py} = x_4 \\ \dot{x}_3 = \dot{x}_{px} = y_{px}u_{pr} = x_5u_{pr} \\ \dot{x}_4 = \dot{x}_{py} = y_{py}u_{pr} = x_6u_{pr} \\ \dot{x}_5 = -\dot{y}_{px} = -x_{px}u_{pr} = -x_3u_{pr} \\ \dot{x}_6 = -\dot{y}_{py} = -x_{py}u_{pr} = -x_4u_{pr} \\ \dot{x}_7 = \upsilon\dot{r}_{ex} = \upsilon x_{ex} = \upsilon x_9 \\ \dot{x}_8 = \upsilon\dot{r}_{ey} = \upsilon x_{ey} = \upsilon x_{10} \\ \dot{x}_9 = \upsilon\dot{x}_{ex} = \upsilon y_{ex}u_e = \upsilon x_{11}u_e \\ \dot{x}_{10} = \upsilon\dot{x}_{ey} = \upsilon y_{ey}u_e = \upsilon x_{12}u_e \\ \dot{x}_{11} = -\upsilon\dot{y}_{ex} = -\upsilon x_{ex}u_e = -\upsilon x_9u_e \\ \dot{x}_{12} = -\upsilon\dot{y}_{ey} = -\upsilon x_{ey}u_e = -\upsilon x_{10}u_e \end{array}\right\} \quad (9)$$

$$r = r_p - r_e = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} x_1 - x_7 \\ x_2 - x_8 \end{bmatrix} \quad (10)$$

$$\dot{r} = \dot{r}_p - \dot{r}_e = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} - \begin{bmatrix} \dot{x}_7 \\ \dot{x}_8 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} - \upsilon \begin{bmatrix} x_9 \\ x_{10} \end{bmatrix} \quad (11)$$

3.2. Feedback Laws.

In this sub-section, we formulate the pursuit strategies in terms of feedback control laws. The maintenance of the cost function Γ, Λ and Λ_0 associated with a strategy at the reference value of -1 on the respective manifold is represented in the form of a feedback control system as shown in figure 2. u_{pr} will take expresson for proportional control as u_p, for proportional integral control as u_{pi} and for proportional, integral and derivative control as u_{pid}.

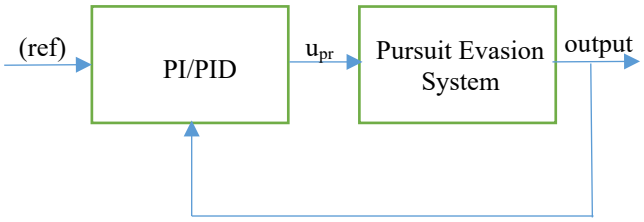


Figure. 2 Manifold Control

3.2.1. Feedback Laws for CM.

Using the results of [4] the feedback control law is defined as

$$u_p = -\mu_1 \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \quad (12)$$

where $\langle\cdot\rangle$ represents the dot product of two vectors and $x^{\perp}$ is defined as the vector x rotated counter- clockwise in the plane by an angle $\frac{\pi}{2}$. In terms of the system model, (12) can be put as

$$u_p = -\mu_1 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle \quad (13)$$

where μ_1 is the proportional (P) setting of the controller.

Proposition 1:

Using (13) under CM, PI and PID control laws can be derived as

$$u_{pi} = -\mu_1 \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle - \mu_2 \int \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \, dt$$

$$= -\mu_1 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle - \mu_2 \int \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle \, dt \quad (14)$$

where μ_2 is the Integral (I) setting of the controller.

$$u_{pid} = -\mu_1 \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle - \mu_2 \int \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \, dt - \mu_3 \frac{d}{dt} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle$$

$$= -\mu_1 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle - \mu_2 \int \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle \, dt$$

$$- \mu_3 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon^2 x_{12}\, u_e, -\upsilon^2 x_{11} u_e)] \rangle \quad (15)$$

where μ_3 is the derivative (D) setting of the controller.

Proof: - see Appendix 1

3.2.2. Feedback Laws for CB.

Considering CB, the expressions for the control law is given as follows.

$$u_p = -\mu_1 \left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right] \quad (16)$$

In terms of system model, (16) can be put as

$$u_p = -\mu_1(\eta \langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.R(x_5,x_6)^t\rangle + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle) \quad (17)$$

Proposition 2:

For CB the control laws for PI and PID can be expressed as follows.

$$u_{pi} = -\mu_1\left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right] - \mu_2 \int \left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right] dt$$

$$= -\mu_1(\eta \langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.R(x_5,x_6)^t\rangle + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle) - \mu_2 \int (\eta \langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.R(x_5,x_6)^t\rangle + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle)dt \quad (18)$$

$$u_{pid} = -\mu_1\left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right] - \mu_2 \int \left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right] dt - \mu_3 \frac{d}{dt}\left(\left[\eta \langle \frac{r}{|r|}.Ry_p\rangle + \frac{1}{|r|}\langle \frac{r}{|r|}.\dot{r}^{\perp}\rangle\right]\right)$$

$$= -\mu_1(\eta \langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.R(x_5,x_6)^t\rangle + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle) - \mu_2 \int (\eta \langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.R(x_5,x_6)^t\rangle + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle)dt - \mu_3\{(((x_3 - \upsilon x_9)\frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}} - (x_1 - x_7)\frac{1}{\left(\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}\right)^3}((x_1-x_7)(x_3-\upsilon x_9) + (x_2-x_8)(x_4 - \upsilon x_{10})), ((x_4 - \upsilon x_{10})\frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}} - (x_2 - x_8)\frac{1}{\left(\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}\right)^3}((x_1-x_7)(x_3-\upsilon x_9) + (x_2-x_8)(x_4 - \upsilon x_{10})).R(x_5,x_6)^t\rangle) + \frac{1}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}, (\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e)\rangle + \left\{-((x_1-x_7)^2+(x_2-x_8)^2)^{\frac{-3}{2}}((x_1-x_7)(x_3-\upsilon x_9) + (x_2-x_8)(x_4-\upsilon x_{10}))\right\}.\langle \frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}.(\upsilon x_{10} - x_4, x_3 - \upsilon x_9)\rangle \quad (19)$$

Proof: - See Appendix 2

3.2.3 Feedback Laws for CP.

Considering CP, the expressions for PI and PID control laws derived under proposition 2 are valid with $R = I_2$.

3.3 Control law for the Evader.

For the evader the steering control common to (CM, CB and CP of pursuer) can be as follows.

$$u_e = \cos t \quad \text{(prescribed)} \quad (20)$$

$$u_e = \langle \frac{r}{|r|}.\dot{r}_e^{\perp}\rangle \quad \text{(reactive)}$$
$$= \langle \left[\frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\right].[(-x_{10}, x_9)]\rangle \quad (21)$$

$$u_e = \langle r.\dot{r}_e\rangle \quad \text{(sudden turn)}$$
$$= \langle \left[\frac{(x_1-x_7, x_2-x_8)}{\sqrt{(x_1-x_7)^2+(x_2-x_8)^2}}\right].[x_3 - \vartheta x_9, x_4 - \vartheta x_{10}]\rangle \quad (22)$$

3.3.1 Sudden Turn of evader

Sudden turn is followed by the evader when the baseline length is shorter than a threshold. The sharp turn is again prevalent in nature where the reduced speed of the evader (victim) compared to that of the predator gives the evader the advantage of agility which enables him to turn sharply trying to escape from the pursuer. The choice of abandoning the other steering laws in favour of the sudden 90-degree turn is taken once the baseline length between the pursuer and evader is perceived to be below a certain threshold. We show below the property that once the evader takes a sudden 90 degree turn with respect to the baseline, he continues to do so in all subsequent moves since the baseline distance between the pursuer and the evader continues to be below the threshold. The property is stated and proved next for the three cases of pursuer strategy, viz, CM, CB and CP.

3.3.1.1 CM

Proposition 3:

Referring the initial baseline between the pursuer and evader as r and after a tangential move for a time dt, the baseline is denoted as r'. Assuming that the pursuer follows the CM strategy it follows that $|r'| \leq |r|$.

Proof: - See Appendix 3.

Table. 1 Time to reach $|r| = 0$ and $\Gamma = -1$ for different combinations of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CM.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	6.6	20.78	21.2	21.1	24.9	5.6	21.8	21.8	27.2	4.5	9.2	1
1,1,0	6.4	13.3	19	19.5	14.7	2.3	3.2	7	19.7	1.8	8.1	0.7
1,2,0	6.3	9.8	18.2	19.7	13.7	1.6	2.9	3.8	18.3	1.4	7.8	0.5
1,3,0	6.3	9.1	17.6	20.1	13.1	1.3	2.8	3.5	17.3	1.1	7.5	0.4
1,1,1	6.4	7.3	19.4	20	13.9	14.6	3.4	4.6	19.7	1.9	8.6	1.3
1,1,2	6.4	7.1	20.1	20.5	13	13.5	3.7	4.5	19.8	1.9	9.1	14.8
1,1,3	6.5	7.1	22.2	22.6	12.5	13.1	4.1	4.1	19.8	1.9	8.4	10.7
1,2,1	6.3	6.9	18.6	21.3	14.1	15.6	3	3.6	18.3	1.4	8.1	0.9
1,2,2	6.4	6.8	19.2	20.7	12.6	12.9	3.2	3.7	18.3	1.4	9	10.4
1,2,3	6.4	6.8	20.5	21.4	12.1	15.3	3.5	3.5	18.3	1.4	8.2	9.4
1,3,1	6.3	6.7	18	20.9	13.5	1.6	2.9	3.4	17.2	1.1	7.8	0.6
1,3,2	6.3	6.6	18.5	20.9	12.4	14.5	3	3.4	17.3	1.1	8.5	1.3
1,3,3	6.3	6.3	19.7	19.8	11.8	14.6	3.2	3.8	17.3	1.1	8	9.2

Table. 2 Time to reach $|r| = 0$ and $\Lambda = -1$ for different combinations of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CB.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	32.3	6.8	32.3	9.8	39.9	6.8	2.9	2.9	NA	6.2	25.7	0.5
1,1,0	17.4	6.6	30.2	9.3	30.6	2.7	9.3	1.7	36.1	2.4	17.8	0.4
1,2,0	17.3	6.6	29.6	9.3	28	2	9	1.5	34.1	1.8	17.8	0.4
1,3,0	16.9	6.6	29.2	9.1	27.6	1.2	8.8	1.2	33.6	2.5	17.7	0.3
1,1,1	17.1	6.5	30	9.4	30.2	2.7	5.4	1.8	36.1	2.2	17.3	0.4
1,1,2	17	6.5	29.8	9.3	30	2.7	4.5	1.9	36.1	2.6	16.9	0.4
1,1,3	16.7	6.5	29.6	9.3	29.6	2.7	3.8	1.9	36	2.7	16.6	0.4
1,2,1	17	6.5	29.5	9.2	28	2	10.2	1.4	34.1	1.8	17.6	0.4
1,2,2	16.8	6.5	29.3	9.2	27.8	2.1	9	1.5	34	1.9	17.3	0.4
1,2,3	16.6	6.5	29	9.1	27.6	2	8.1	1.5	34	2	17.1	0.3
1,3,1	16.7	6.6	29.1	9.1	27.5	1.7	9.7	1.3	33.5	1.6	17.6	0.3
1,3,2	16.5	6.5	28.9	9.1	27.4	1.7	8.7	1.3	33.4	1.6	17.3	0.3
1,3,3	16.4	6.5	28.9	9.1	27.4	1.7	8	1.3	33.4	1.6	16.5	0.3

3.3.1.2 CB

Proposition 4:

For the case of CB, considering the instantaneous case of sudden turn of 90 degrees left or right by the evader, with the pursuer following the CB law under two cases.

where *r* and *r'* are the baseline vectors initially and after the lapse of an infinitesimal time respectively.

and (ii) $\emptyset < \frac{\pi}{2}$, it follows that

$$|r| = |r'|$$

if

$$\left(\frac{\vartheta}{p} + \sin\emptyset\right)^2 + (\cos\emptyset)^2 = \frac{2|r|\cos\emptyset}{p\,dt}$$

where dt is the instantaneous period of the step considered. ϑ and p are the velocities of the evader and the pursuer respectively.

Proof: - See Appendix 4.

Remark 1: To illustrate the use of result of proposition 4(ii)

Assume, $k = \frac{|r|}{p\,dt}$.

When $\vartheta = 0.6, p = 1, |r| = 1, p\,dt = 0.1$,

$k = 10$. Then as per the result of proposition 4 (ii),

$$(0.6 + \sin\emptyset)^2 + (\cos\emptyset)^2 = 20\cos\emptyset$$

$$\emptyset \approx 82.65 \text{ degrees}$$

Table. 3 Time to reach $|r| = 0$ and $\Lambda_0 = -1$ for different combinations of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CP.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{rf}	t_r	t_{rf}	t_r	t_{rf}	t_r	t_{rf}	t_r	t_{rf}	t_r	t_{rf}
1,0,0	6.7	26.5	18.6	NA	12.6	NA	2.7	6.2	17.9	NA	8.3	0.6
1,.03,0	6.7	10.6	18.6	18.8	12.6	13.1	2.6	4.5	17.7	18.3	8.2	0.7
1,.04,0	6.7	9.4	18.5	18.6	12.6	12.7	2.6	4.2	17.7	17.9	8.1	0.6
1,.05,0	6.6	7.6	18.5	18.5	12.5	12.6	2.6	3.9	17.7	17.8	8.1	0.6
1,0,0.1	6.6	8.9										
1,0,.07			18.6	20.6								
1,0,.02					12.6	14.7			17.9	21		
1,0,.03							2.6	18.1				
1,0,.01											8.2	0.6

3.3.1.3 CP

Proposition 5:

Assume *r* and *r'* defined as in Proposition 3. When the evader takes a sudden turn left or right 90 degree, with pursuer following CP law case, it can be shown that

$$|r'| \leq |r|$$

Proof: - See Appendix 5.

4. Simulation Results.

4.1. Time response

4.1.1 Non-reactive case

In this section, we provide the results of simulation of the dynamic equations given in section 3 and discuss the same. We assume u_e given by (20) corresponds to the non-reactive case of the evader. Tables 1-3 provide results obtained through computer, simulation of the pursuit evasion game for CM, CB and CP strategies respectively under P, PI and PID control. The first columns of Tables 1-3 provide the P, I and D gains used corresponding to μ_1, μ_2, μ_3 respectively. The next six columns correspond to different initial starting coordinates for r_p and x_p, r_e and x_e are assumed to be $(0,0)^t$ and $(1,0)^t$ respectively. The two sub- columns in Tables 1-3 correspond to time (t_r) to reach zero for the magnitude of r and the time (t_{mf}) to reach manifold characterized by $\Gamma = -1$, $\Lambda = -1$ and $\Lambda_0 = -1$ corresponding to CM, CB and CP strategies respectively for the first time. In Table 3, the blank cells correspond to the case of no significant results being obtained in the simulation.

It is seen in Table 1, corresponding to CM, that PI and PID laws tend to improve the performance of the pursuer compared to using P alone, which is the existing method. For example, comparing the rows corresponding to (1, 3, 0) and (1, 3, 3) against (1, 0, 0) in the first column, it is seen that t_r and t_{mf} are reduced in the case of PI and PID settings compared to P alone. This holds for all possible starting points considered. Similarly, it is seen in Table 2, corresponding to CB that PI and PID laws tend to improve on the performance of the pursuer compared to using P alone, which can be considered as the existing method. That is, t_r and t_{mf} are much reduced in the case of PI and PID settings compared to P alone. This holds for all possible starting points considered. Similarly, it is seen in Table 3, corresponding to CP, that PI and PID laws tend to improve the performance of the pursuer compared to using P alone, which is the existing method. That is, t_r and t_{mf} are much reduced in the case of PI and PID settings compared to P alone. This holds for all possible starting points considered.

Table. 4 Time to reach $|r| = 0$ and $\Gamma = -1$ for different combination of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CM.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	5.2	5.2	10.1	16.7	10.1	16.2	13.3	21.1	13.7	25.7	8.1	0.6
1,1,0	5.0	6.4	8.8	13.5	9.1	14.1	2.5	6.8	9.4	1.9	7.4	0.5
1,2,0	4.9	5.8	8.4	11.4	7.6	1.7	2.4	6.5	8.8	1.4	7.2	0.4
1,3,0	4.9	5.5	8.2	10.5	7.2	1.3	2.3	34.6	8.6	1.2	7.1	0.3
1,1,1	5	6.3	8.8	13.5	7.6	12.6	2.6	7.4	9.7	2.7	7.3	0.5
1,1,2	4.9	6.1	8.9	13.4	7.3	12.1	2.7	8.1	10.4	14.8	7.3	0.5
1,1,3	4.9	6	9	13.4	7	11.9	2.8	2.7	9.5	14	7.3	0.4
1,2,1	4.9	5.8	8.5	11.4	7.4	10.6	2.4	19.3	8.9	1.5	7.2	0.4
1,2,2	4.9	5.7	8.6	11.4	7.1	10.2	2.5	7.8	8.9	1.5	7.1	0.4
1,2,3	4.8	5.6	8.9	11.5	6.9	9.9	2.6	2.6	9	1.6	7.1	0.4
1,3,1	4.9	5.5	8.3	10.5	7.5	9.8	2.4	11.9	8.5	1.2	7	0.4
1,3,2	4.8	5.4	8.4	10.5	7	9.3	2.4	8.7	8.5	1.2	7	0.4
1,3,3	4.8	5.4	8.7	10.7	6.8	9.1	2.5	14.4	8.6	1.2	7.1	0.3

Table. 5 Time to reach $|r| = 0$ and $\Lambda = -1$ for different combination of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CB.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	5.7(1.6)	5.8	9.7(1.7)	9.7	13.9(1.8)	4.9	2.7	10.9	15.1(1.8)	4.5	9.3(1.8)	0.4
1,1,0	23.1	5.5	25.1	9.3	25.4	2.4	22.9(.52)	1.8	26.4(.51)	2.2	24.5(.52)	0.3
1,2,0	20.3	5.4	23.3	9.1	23.2	1.9	14.2(.53)	1.4	24.5	1.7	22.5	0.3
1,3,0	19.7	5.3	22.5	9.1	22.3	1.7	11.7	1.2	23.6	1.5	21.7	0.2
1,1,1	20.7	5.4	23.3	9.2	23.7	2.4	16.3(.53)	1.8	25	2.2	22.6	0.3
1,1,2	18.8	5.4	21.8	9.1	22.2	2.3	10.7(.51)	1.9	23.8	2.3	21.1	0.3
1,1,3	17.1	5.3	20.5	9	20.8	2.3	8	1.9	22.6	2.3	19.7	0.3
1,2,1	19	5.3	22.4	9.1	22.3	1.9	11.6(.51)	1.4	23.8	1.8	21.6(.51)	0.3
1,2,2	18.5	5.3	21.7	9	21.6	1.9	9.6	1.5	23.2	1.8	20.8	0.3
1,2,3	17.6	5.3	21	8.9	20.8	1.8	8	1.5	22.6	1.8	20.1	0.4
1,3,1	18.9	5.3	22	8.9	21.6	1.6	10.3	1.3	23.1	1.5	21.1	0.2
1,3,2	18.3	5.3	21.5	8.9	21.1	1.6	8.8	1.3	22.7	1.5	20.6	0.2
1,3,3	17.7	5.3	21	8.8	20.6	1.6	8.4	1.3	22.3	1.5	20.1	0.3

Comparing the performance of CM, CB and CP under feedback laws, the initial condition column (1, 3, 5, 6) in Tables 1-3, show that the CM outperforms followed by CP and CB in terms of shorter t_r. In column 2 in Tables 1-3 the CP outperforms followed by CM and CB in terms of shorter t_r, and in column 4 in Tables 1-3 the CP outperforms followed by CM and CB in terms of shorter t_r. Similarly, in terms of shorter t_{mf} columns (1-4) in Tables 1-3, CM outperforms followed by CB and CP. For column 5 in Tables 1-3, CB outperforms followed by CM and CP in terms of shorter t_{mf}. For column 6 in Tables 1-3, CB outperforms followed by CP and CB in terms of shorter t_{mf}.

4.1.2 Reactive case

Simulation results of the dynamic equations given in section 3 by using u_e corresponding to (21) are given below.

It is shown in Tables 4-6, that PI/PID control (proposed) laws correspond to much shorter t_r and t_{mf} compared to P alone for the three strategies CM, CB and CP.

Comparison of the different strategies, under feedback laws, shows that in the column (1, 3, 6) in Tables 4-6 CM outperforms followed by CP and CB. In column (2, 4, 5) of Tables 4-6 CP outperforms followed by CM and CB in terms of shorter t_r. Similarly, in terms of shorter t_{mf} column (1, 4) in Tables 4-6 shows that CB outperforms followed by CM and CP. For column (2, 3) in Tables 4-6, CB outperforms followed by CP and CM in terms of t_r and t_{mf}. For column (5, 6) in Tables 4-6, CM outperforms followed by CB and CP in terms of t_r and t_{mf}. The controller setting of D is not included in Table 6 since no significant improvement is seen through simulation using derivative control.

4.1.3 Sudden Turn

Simulation results of the dynamic equations given in section 3 by using u_e corresponding to (22) are given below.

Tables 7-9 give the data of t_r and t_{mf} of CM, CB and CP strategies under sudden turn evader steering control laws. The empty cells in these tables mean insignificant value within the considered time frame. In CM strategies, we can see from Table 4 and Table 7 that by using sudden turn evader control laws, t_r and t_{mf} are much delayed (compare the first rows of Table 4 and

Table. 6 Time to reach $|r| = 0$ and $\Lambda_0 = -1$ for different combination of $P(\mu_1), I(\mu_2), D(\mu_3)$ for different initial values of r_p and x_p for CP.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	5	8.7	8.1	9	7	9.2	2.2	6	8.5	10.3	7.9	0.5
1,.03,0	5	7.5	8.1	9.1	7	8.4	2.2	10.1	8.5	10.1	7.8	0.5
1,.04,0	5	8	8.1	9.1	7	8.3	2.2	6.3	8.5	10.1	7.8	0.5
1,.05,0	5	8.4	8.1	9.2	7	8.2	2.2	5.1	8.5	10	7.8	0.5

Table 7) from using reactive or non-reactive evader control law for different initial conditions. In case of CB strategy, we can see from Table 5 and Table 8 that by using sudden turn evader control laws, evader escapes from the pursuer for all different initial conditions (compare the first rows of Table 5 and Table 8). This is a very significant result for the evader. For CP strategies, from Table 6 and Table 9, we can see that by using sudden turn evader control laws, capturing of evader is delayed for all initial conditions compared to the case when evader used reactive control law (compare the first rows of Table 6 and Table 9).

Table. 7 Time to reach $|r| = 0$ and $\Gamma = -1$ for P (μ_1) control for different initial values of r_p and x_p for CM.

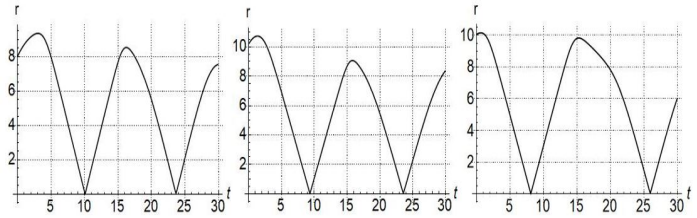
μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	28.6	17.6		20.7	21.4	21.3	21.3	10.9	23.3	4.9	16.1	0.8

Table. 8 Time to reach $|r| = 0$ and $\Gamma = -1$ for P(μ_1) control for different initial values of r_p and x_p for CB.

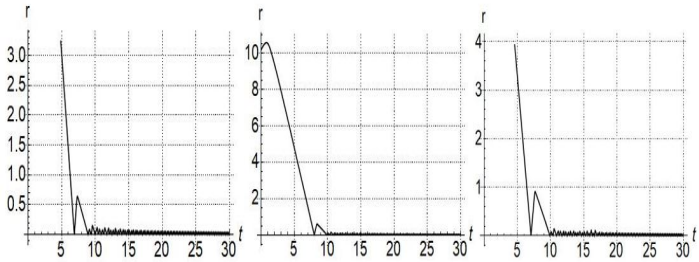
μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0		4.1		7.1		5.4		3.2		4.5		0.5

Table. 9 Time to reach $|r| = 0$ and $\Gamma = -1$ for P(μ_1) control for different initial values of r_p and x_p for CP.

μ_1,μ_2,μ_3	$(r_{px},r_{py}),(x_{px},x_{py})$ (4, -6),(0.735,0.678) (1)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-4, 9),(0.327,0.945) (2)		$(r_{px},r_{py}),(x_{px},x_{py})$ (0, 8),(0.530,0.848) (3)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-1, 1),(0.632,0.775) (4)		$(r_{px},r_{py}),(x_{px},x_{py})$ (-2, 10),(0.335,0.942) (5)		$(r_{px},r_{py}),(x_{px},x_{py})$ (10,-1),(0.9,0.436) (6)	
	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}	t_r	t_{mf}
1,0,0	10.3	14.3	15.1		14.9	18.7	4	8	15.9		15.6	0.6

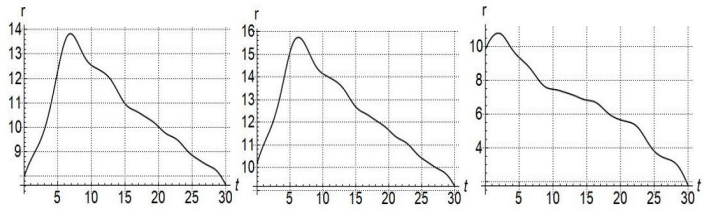


(a)

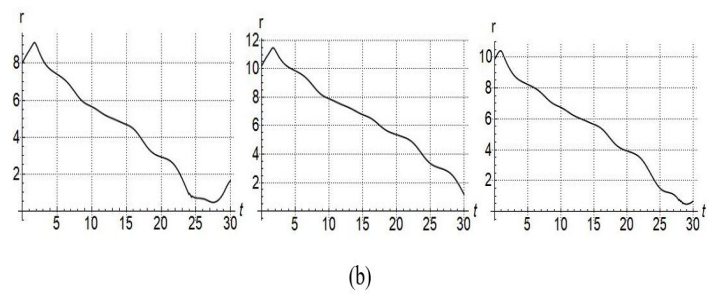


(b)

Figure. 3 r vs t plot for CM using (a) P and (b) PID.

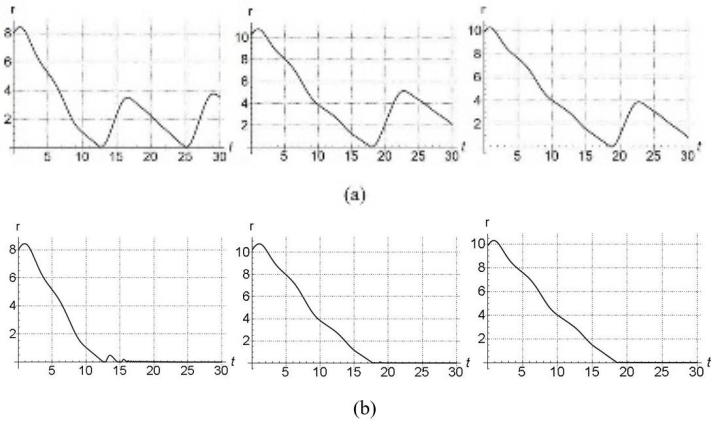


(a)



(b)

Figure. 4 r vs t plot for CB using (a) P and (b) PID.



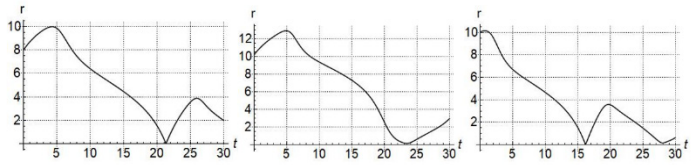
(a)

(b)

Figure. 5 r *vs* t plot for CP using (a) P and (b) PI.

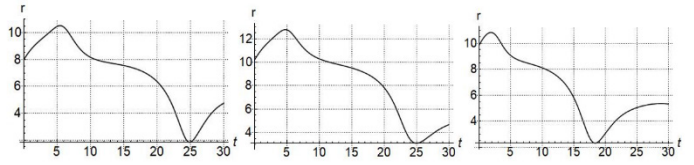
4.2. Tracking response

Figures 3-4 show the response of the baseline magnitude $|r|(t)$ as function of time for different initial settings. When |r| becomes zero, it means the distance between pursuer and evader is zero, and hence the evader is captured. Figure 3(a) and Figure 4(a) correspond to P control while Figure 3(b) and Figure 4(b) corresponds to PID control. Similarly, for Figure 5(a) and Figure 5(b) correspond to P and PI respectively for different initial conditions. It is clearly seen from Figures 3-5 that tracking response of the zero line of the baseline magnitude (i.e. the zero line of the distance between the pursuer and evader) is clearly superior for PID/ PI control, compared to P alone (top rows in Figures 3-5). In fact, P alone is not able to track the zero-baseline magnitude at all. Note that the magnitude of baseline vector going to zero corresponds to evader being captured.



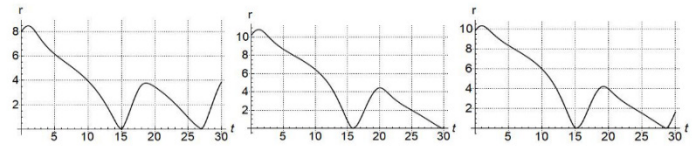
(a) (b) (c)

Figure. 6 r *vs* t plot for CM using sudden turn evader steering law.



(a) (b) (c)

Figure. 7 r *vs* t plot for CB using sudden turn evader steering law.



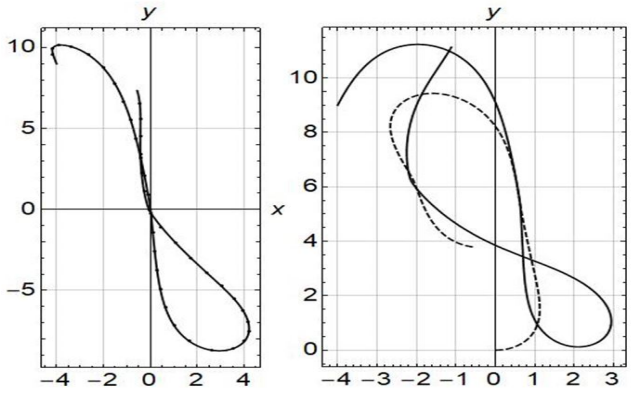
(a) (b) (c)

Figure. 8 r *vs* t plot for CP using sudden turn evader steering law.

Figures 6-8 show the response of the baseline magnitude $|r|(t)$ as function of time for different initial settings for CM, CB and CP strategies respectively. In Figure 6(a) represents the initial starting position of pursuer given as (0, 8) in x- y co-ordinate. 6(b) and 6(c) correspond to starting positions of (-2, 10) and (-4, 9) respectively. For the sake of comparison, (a), (b) and (c) in Figures 7 and 8 also correspond to the same respective starting positions. It is clearly seen from Figures 6-8 that the time to capture the evader by the pursuer is much delayed by using the sudden turn strategy compared to the case when evader used the reactive and non-reactive steering control laws. This can be seen by comparing Figure 3(a) with Figure 6, Figure 4(a) with Figure 7 and Figure 5(a) with Figure 8. In fact in Figure 4, $|r|(t)$ never reaches zero meaning that the evader is never captured in that case.

4.3. Trajectories of pursuer and evader.

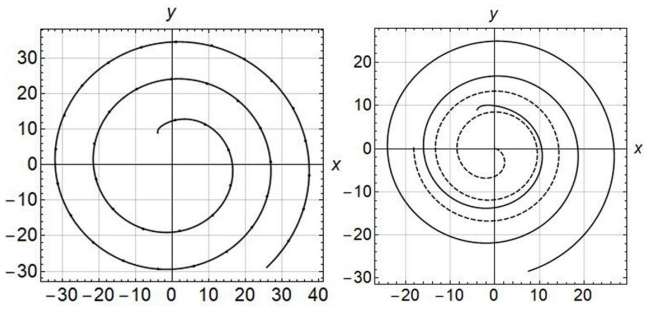
We consider in this section, the actual trajectories followed by evader and pursuer.

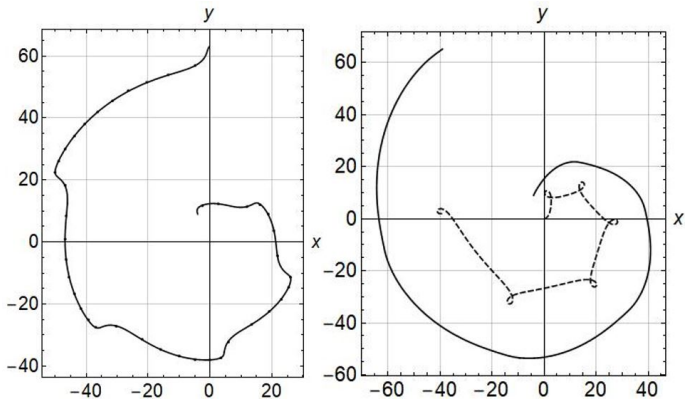


(a) (b)

Figure. 9 For normal pursuer capture evader, (a) magnitude of r, (b) trajectories of pursuer (solid line) and evader (dashed line).

Figure 9 shows evader being captured in a certain time by the pursuer. Figure 9(a) shows the magnitude of r. Figure 9(b) shows the trajectories of pursuer and evader for a 30 sec timeline. The starting point of evader is (0, 0) in x-y coordinate and the starting point for the pursuer is (-4, 9). Figure 10 shows the case when the evader escapes from the pursuer. This happens only in the case of





(a) (b)

Figure. 11 For $\emptyset = 82.65°$, (a) magnitude of r, (b) trajectories of pursuer (solid line) and evader (dashed line).

CB strategies when evader suddenly turn 90 degree left or right as discussed in Proposition 4 case (i). So, in Figure 10 (a) the value of r keeps on increasing. Figure 10 (b) shows the trajectories of pursuer and evader in 30 sec timeline. Figure 11 is for the case when $\emptyset = 82.65°$ and the baseline distance between the pursuer and the evader tend to reach a constant value.

5. Conclusions

Pursuit evasion game (PEG) has been studied in the literatiure using geometric control theory. Control laws for the pursuer have been derived so as to make the pursuer follow a certain manifold according to the different bio-inspired strategies used. The performance of the pursuer control laws is specified in terms of time to capture the evader and tarcking of pursuer path on the manifold given certain fixed evader escape strategies Though the existing pursuer control laws are optimal, it is shown in the paper, through simulations, that adding integral and derivative actions to the pursuer control laws tends to improve the performance of the pursuer further independent of the strategy used. Towards that, the pursuer path tracking on the manifold is specified in terms of a feedback loop with the reference being the cost function of the manifold used. The proportional (P), proportional- integral (PI) and proportional- integral- derivative (PID) control laws are studied on the loop with the existing control law being considered as proportional law. Due to the noninear nature of control laws in a vector setting the derivation of PI and PID control laws are non-trivial.To mimic the nature further, a sudden turn strategy is assumed to be employed by the evader in view of evader's agility advantage while moving at a lower speed compared to the pusuer. The dynamics of the PEG with the sudden turn strategy of the evader has been studied through a theoretical analysis supported by simulation. It is shown that the sudden turn strategy can help the evader delay the capture by the pursuer and in a certain case the evader can even totally excape from the pursuer. As a future work, one could consider the case when the evader moves at a varying speed (instead of the constant speed assumed in the paper) as in commonly seen in nature.

References

[1] P. Svec, A. Thakur, B. C. Shah, and S. K. Gupta, "USV Trajectory Planning for Time Varying Motion Goals in an Environment with obstacles," *Terpconnect.Umd.Edu*, pp. 1–11, 2012.

[2] R. Issacs, "Differential Games," New York: Wiley, 1965.

[3] E. Wei, E. W. Justh, and P. S. Krishnaprasad, "Pursuit and an evolutionary game," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 465, no. 2105, pp. 1539–1559, 2009.

[4] D. Pais and N. Leonard, "Pursuit and Evasion: Evolutionary Dynamics and Collective Motion," *AIAA Guid. Navig. Control Conf.*, pp. 1–14, 2010.

[5] W. Li, "A Dynamics Perspective of Pursuit-Evasion: Capturing and Escaping When the Pursuer Runs Faster Than the Agile Evader," *IEEE Trans. Automat. Contr.*, vol. 62, no. 1, pp. 451–457, Jan. 2017.

[6] W. Li, "Escape Analysis on the Confinement-Escape Problem of a Defender Against an Evader Escaping from a Circular Region," *IEEE Trans. Cybern.*, vol. 46, no. 9, pp. 2166–2172, 2016.

[7] L. Obiroy Singh and R. Devanathan, "Performance improvement of bio-inspired strategies through feedback laws," in *Advances in Intelligent Systems and Computing*, 2018, pp. 143–156.

[8] L. Obiroy Singh and R. Devanathan, "Comparative Study of the Performance of Application of Bio-Inspired Strategies to Pursuit Evasion Game Under Feedback Laws," in *The 18th International Conference on Control, Automation and Systems,South Korea*, 2018, pp. 276–281.

[9] J. H. S.D Bopardikar, F Bullo, "A Cooperative Homicidal Chauffeur Game," *Automatica*, vol. 45, pp. 1771–1777, 2009.

[10] X. Liang and Y. Xiao, "Studying Bio-Inspired Coalition Formation of Robots for Detecting Intrusion Using Game Theory," *IEEE Trans. Syst. Man Cybern.*, vol. 40, no. 3, pp. 683–693, 2010.

[11] H. . Cott, "Adaptive Coloration in animals," *London: Methues*, pp. 141–143, 1966.

[12] K. . Roeder, "Nerve Cells and insect behaviour," *Cambridge, MassachusettsHarvard Univ. Press*, pp. 160--164, 1967.

[13] M. V Srinivasan and M. Davey, "Strategies for Active Camouflage of Motion," in *Biological Sciences*, pp. 19–25.

[14] E. Simetti, A. Turetta, G. Casalino, E. Storti, and M. Cresta, "Protecting Assets within a Civilian Harbour through the use of a Team of USVs: Interception of Possible Menaces," in *IARP workshop on robots for risky interventions and environmental surveillance maintenance (RISE'10)*, 2010.

[15] J. Walrand, E. Polak, and H. Chung, "Harbor Attack: A Pursuit- Evasion Game," in *Forty- Ninth Annual Allerton Conference*, 2011, pp. 1584–1591.

[17] E. W. Justh and P. S. Krishnaprasad, "Steering laws for motion camouflage," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 462, no. 2076, pp. 3629–3643, 2006.

[18] E. W. Justh and P. S. Krishnaprasad, "Natural frames and interacting particles in three dimensions," *Proc. 44th IEEE Conf. Decis. Control. Eur. Control Conf. CDC-ECC '05*, vol. 2005, pp. 2841–2846, 2005.

Appendix 1:

Proof of Proposition 1

P and PI control:

u_p *and* u_{pi} are the control output of P and PI controller derived as given in the proposition in a straightforward way.

PID Control output u_{pid} can be derived as follows.

$$u_{pid} = -\mu_1 \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle - \mu_2 \int \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \, dt - \mu_3 \frac{d}{dt} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle$$

$$= -\mu_1 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle -$$

$$\mu_2 \int \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle \, dt$$

$$-\mu_3 \frac{d}{dt} \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle$$

Using the identity

$$\frac{d}{dt} \langle \alpha, \beta \rangle = \langle \dot{\alpha}, \beta \rangle + \langle \alpha, \dot{\beta} \rangle \tag{A.1}$$

put

$$\alpha = \frac{r}{|r|} \text{ and } \beta = \dot{r}^{\perp}$$

that is,

$$\alpha = \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] \text{ and } \beta = [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)]$$

Considering equation (A.1), since

$\dot{\alpha} = \frac{\dot{r}}{|\dot{r}|}$ is orthogonal to $\beta = \dot{r}^{\perp}$

the first term of R.H.S of equation (A.1) is zero.

Now

$$\beta = [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)]$$

from which

$$\dot{\beta} = [(\upsilon \dot{x}_{10} - \dot{x}_4), (\dot{x}_3 - \upsilon \dot{x}_9)]$$

that is,

$$\dot{\beta} = \left[\left(\upsilon^2 x_{12} u_e - x_6 u_p \right), \left(x_5 u_p - \upsilon^2 x_{11} u_e \right) \right]$$

using (9).

Approximating the term $\dot{\beta}$ as $\hat{\dot{\beta}}$ by neglecting the term associated with u_p,

$$\hat{\dot{\beta}} = [\upsilon^2 x_{12} u_e, -\upsilon^2 x_{11} u_e]$$

Therefore, equation (A.1) becomes

$$\frac{d}{dt} \langle \alpha, \beta \rangle = \langle \alpha, \hat{\dot{\beta}} \rangle$$

$$\frac{d}{dt} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle = \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon^2 x_{12} u_e, -\upsilon^2 x_{11} u_e)] \rangle$$

Therefore

$$u_{pid} = -\mu_1 \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle - \mu_2 \int \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \, dt - \mu_3 \frac{d}{dt} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle$$

$$= -\mu_1 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle -$$

$$\mu_2 \int \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)] \rangle \, dt$$

$$-\mu_3 \langle \left[\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \right] . [(\upsilon^2 x_{12} u_e, -\upsilon^2 x_{11} u_e)] \rangle$$

Hence the result.

Appendix 2:

Proof of Proposition 2:

P and PI control:

u_p *and* u_{pi} are easily derived as given in the proposition.

PID control:

$$u_{pid} = -\mu_1 \left[\eta \langle \frac{r}{|r|} . Ry_p \rangle + \frac{1}{|r|} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \right] - \mu_2 \int \left[\eta \langle \frac{r}{|r|} . Ry_p \rangle + \frac{1}{|r|} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \right] dt - \mu_3 \frac{d}{dt} \left(\left[\eta \langle \frac{r}{|r|} . Ry_p \rangle + \frac{1}{|r|} \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \right] \right) \tag{A.2}$$

Using the identity

$$\frac{d}{dt} \langle \rho, \omega \rangle = \langle \dot{\rho}, \omega \rangle + \langle \rho, \dot{\omega} \rangle \tag{A.3}$$

$$\frac{d}{dt} (\gamma \sigma) = (\dot{\gamma} \sigma) + (\gamma \dot{\sigma}) \tag{A.4}$$

put

$$\rho = \frac{r}{|r|}, \omega = Ry_p, \gamma = \frac{1}{|r|} \text{ and } \sigma = \langle \frac{r}{|r|} . \dot{r}^{\perp} \rangle \tag{A.5}$$

Now

$$\dot{\rho} = \frac{d}{dt} \frac{r}{|r|} = \begin{bmatrix} \frac{d}{dt} \frac{r_x}{|r|} \\ \frac{d}{dt} \frac{r_y}{|r|} \end{bmatrix} = \begin{bmatrix} \frac{|r| \dot{r}_x - r_x \frac{d}{dt} |r|}{|r|^2} \\ \frac{|r| \dot{r}_y - r_y \frac{d}{dt} |r|}{|r|^2} \end{bmatrix}$$

$$\dot{\rho} = \begin{bmatrix} \frac{\dot{r}_x}{|r|} - \frac{r_x \frac{d}{dt} \sqrt{r_x^2 + r_y^2}}{|r|^2} \\ \frac{\dot{r}_y}{|r|} - \frac{r_y \frac{d}{dt} \sqrt{r_x^2 + r_y^2}}{|r|^2} \end{bmatrix} = \begin{bmatrix} \frac{\dot{r}_x}{|r|} - \frac{r_x (r_x \dot{r}_x + r_y \dot{r}_y)}{|r|^3} \\ \frac{\dot{r}_y}{|r|} - \frac{r_y (r_x \dot{r}_x + r_y \dot{r}_y)}{|r|^3} \end{bmatrix} \tag{A.6}$$

$$\dot{\omega} = \frac{d}{dt} Ry_p = R \begin{bmatrix} \frac{d}{dt} y_{px} \\ \frac{d}{dt} y_{py} \end{bmatrix} = R \begin{bmatrix} x_{px} u_p \\ x_{py} u_p \end{bmatrix} \tag{A.7}$$

$\dot{\omega}$ is neglected since it is associated with the term u_p.

Therefore,

$$\frac{d}{dt} \langle \rho, \omega \rangle = \langle \dot{\rho}, \omega \rangle$$

$$= \langle \begin{bmatrix} \frac{\dot{r}_x}{|r|} - \frac{r_x(r_x\dot{r}_x + r_y\dot{r}_y)}{|r|^3} \\ \frac{\dot{r}_y}{|r|} - \frac{r_y(r_x\dot{r}_x + r_y\dot{r}_y)}{|r|^3} \end{bmatrix}, R \begin{bmatrix} y_{px} \\ y_{py} \end{bmatrix} \rangle$$

$$= \langle \left(\frac{x_3 - \upsilon x_9}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} - \frac{(x_1 - x_7)\big((x_1 - x_7)(x_3 - \upsilon x_9) + (x_2 - x_8)(x_4 - \upsilon x_{10})\big)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}^3} \right), R(x_5, x_6) \rangle$$

using equations (10) and (11).

$$\dot{\gamma} = \frac{d}{dt}\frac{1}{|r|} = \frac{d}{dt}\left(\sqrt{r_x{}^2 + r_y{}^2}\right)^{-\frac{1}{2}}$$

$$= -\left(\sqrt{r_x{}^2 + r_y{}^2}\right)^{-3}\left(r_x\dot{r}_x + r_y\dot{r}_y\right) \qquad \text{(A.8)}$$

$$\dot{\sigma} = \frac{d}{dt}\langle\frac{r}{|r|}.\dot{r}^{\perp}\rangle = \langle\frac{\dot{r}}{|\dot{r}|}.\dot{r}^{\perp}\rangle + \langle\frac{r}{|r|}.\ddot{r}^{\perp}\rangle \qquad \text{(A.9)}$$

$\langle\frac{\dot{r}}{|\dot{r}|}.\dot{r}^{\perp}\rangle = 0$, since $\frac{\dot{r}}{|\dot{r}|}$ is orthogonal to $\dot{r}^{\perp}$

So,

$$\dot{\sigma} = \frac{d}{dt}\langle\frac{r}{|r|}.\dot{r}^{\perp}\rangle = \langle\frac{r}{|r|}.\ddot{r}^{\perp}\rangle$$

Now,

$$\ddot{r}^{\perp} = \frac{d}{dt}\dot{r}^{\perp} = \frac{d}{dt}[\upsilon x_{10} - x_4, x_3 - \upsilon x_9]$$

$$= [\upsilon\dot{x}_{10} - \dot{x}_4, \dot{x}_3 - \upsilon\dot{x}_9] = \left[\upsilon^2 x_{12}u_e - x_6 u_p, x_5 u_p - \upsilon^2 x_{11}u_e\right]$$

using (9).

Neglecting the term associated with u_p,

$$\ddot{r}^{\perp} = [\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e] \qquad \text{(A.10)}$$

Therefore,

$$\dot{\sigma} = \langle\frac{r}{|r|}.(\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e)\rangle \qquad \text{(A.11)}$$

Thus,

$$\frac{d}{dt}(\gamma\sigma) = (\dot{\gamma}\sigma) + (\gamma\dot{\sigma})$$

$$= \left\{-\left(\sqrt{r_x{}^2 + r_y{}^2}\right)^{-3}\left(r_x\dot{r}_x + r_y\dot{r}_y\right).\left(\frac{\dot{r}}{|\dot{r}|}.\dot{r}^{\perp}\right)\right\} + \frac{1}{|r|}.\langle\frac{r}{|r|}.(\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e)\rangle$$

$$= \frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} \langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}, (\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e)\rangle$$

$$+ \left\{-((x_1 - x_7)^2 + (x_2 - x_8)^2)^{\frac{-3}{2}}\big((x_1 - x_7)(x_3 - \upsilon x_9) + (x_2 - x_8)(x_4 - \upsilon x_{10})\big)\right\}.\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.(\upsilon x_{10} - x_4, x_3 - \upsilon x_9)\rangle$$

Hence it follows that

$$u_{pid} = -\mu_1(\eta\,\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.R(x_5, x_6)^t\rangle + \frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle) - \mu_2\int(\eta\,\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.R(x_5, x_6)^t\rangle + \frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.(\upsilon x_{10} - x_4), (x_3 - \upsilon x_9)\rangle)dt$$

$$-\mu_3\{(((x_3 - \upsilon x_9)\frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} - (x_1 - x_7)\frac{1}{\left(\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}\right)^3}((x_1 - x_7)(x_3 - \upsilon x_9) + (x_2 - x_8)(x_4 - \upsilon x_{10}), ((x_4 - \upsilon x_{10})\frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}} - (x_2 - x_8)\frac{1}{\left(\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}\right)^3}\big((x_1 - x_7)(x_3 - \upsilon x_9) + (x_2 - x_8)(x_4 - \upsilon x_{10})\big).R(x_5, x_6)^t\rangle) + \frac{1}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}, (\upsilon^2 x_{12}u_e, -\upsilon^2 x_{11}u_e)\rangle + \left\{-((x_1 - x_7)^2 + (x_2 - x_8)^2)^{\frac{-3}{2}}\big((x_1 - x_7)(x_3 - \upsilon x_9) + (x_2 - x_8)(x_4 - \upsilon x_{10})\big)\right\}.\langle\frac{(x_1 - x_7, x_2 - x_8)}{\sqrt{(x_1 - x_7)^2 + (x_2 - x_8)^2}}.(\upsilon x_{10} - x_4, x_3 - \upsilon x_9)\rangle$$

Hence the result.

Appendix 3:

Proof of proposition 3:

Referring to Figure 12, *r* is the initial baseline while *r'* is the baseline in the next step after a lapse of time dt. p and ϑ are the velocities of the pursuer and the evader respectively. The sudden turn with respect to the baseline at every move result in a motion by the evader in a tangential direction resulting in describing the arc 'ab' in Figure 12 by the evader for arbitrarily small dt. We then have

$$|r| = |r'| + p\,dt$$

with $p > 0$, $dt > 0$ and with $dt \to 0$, It follows that $|r'| \leq |r|$. Hence the result.

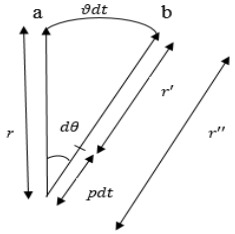


Figure. 12 Case of CM with sudden turn.

Appendix 4

Case (i):

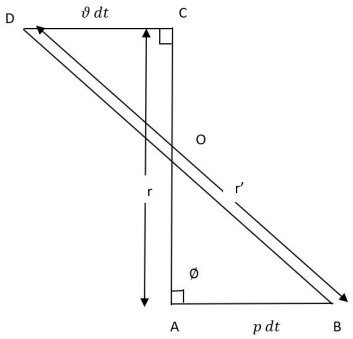


Figure. 13 Case of CB $\emptyset = \frac{\pi}{2}$ with sudden turn.

Figure 13 shows the initial baseline vector r between the pursuer (A) and evader (C). After an infinitesimally small time, with the evader taking the sudden turn of 90 degrees, let the position of evader and the pursuer be D and B respectively.

In right angle triangle AOB

$$OB > OA$$

Similarly, in right angle triangle COD

$$OD > OC$$

So,

$$OB + OD > OA + OC$$

$$|r'| > |r|$$

Case (ii):

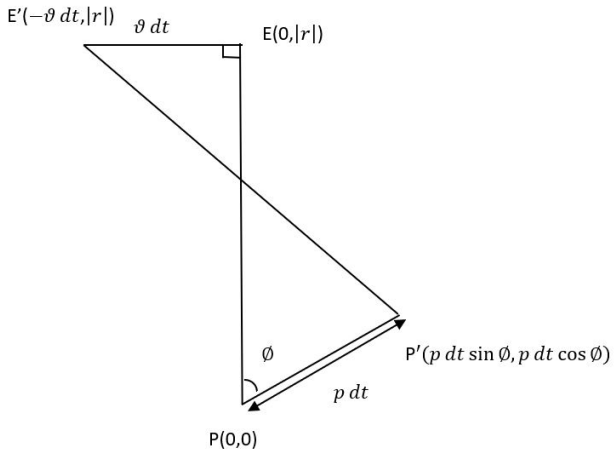


Figure. 14 Case of CB $\emptyset < \frac{\pi}{2}$ with sudden turn.

Figure 14 is redrawn from Figure 13 with $\emptyset < \frac{\pi}{2}$. Considering pursuer (P) as the origin, the coordinates of other points are easily derived. Point E denotes the evader. P' and E' are the positions of pursuer and the evader after a time dt. The baseline after time dt is given by vector

$$r' = (-\vartheta\, dt, |r|) - (pdt \sin \emptyset\,, pdt \cos \emptyset)$$

Now, to impose the condition that $|r'| = |r|$.

$$(-\vartheta\, dt - pdt \sin \emptyset)^2 + (|r| - pdt \cos \emptyset)^2 = |r|^2$$

$$\Rightarrow (\vartheta\, dt)^2 + (p\, dt)^2 (\sin \emptyset)^2 + 2\vartheta\, dt\, p\, dt \sin \emptyset + (p\, dt)^2 (\cos \emptyset)^2 - 2p\, |r| dt \cos \emptyset = 0$$

$$\Rightarrow \vartheta^2 + p^2 (\sin \emptyset)^2 + 2\vartheta p \sin \emptyset + p^2 (\cos \emptyset)^2 - \frac{2|r| p \cos \emptyset}{dt} = 0$$

$$\Rightarrow (\vartheta + p \sin \emptyset)^2 + p^2 (\cos \emptyset)^2 = \frac{2|r| p \cos \emptyset}{dt}$$

$$\Rightarrow \left(\frac{\vartheta}{p} + \sin \emptyset\right)^2 + (\cos \emptyset)^2 = \frac{2|r| \cos \emptyset}{p\, dt}$$

Hence the result.

Appendix 5:

Proof of Proposition 5:

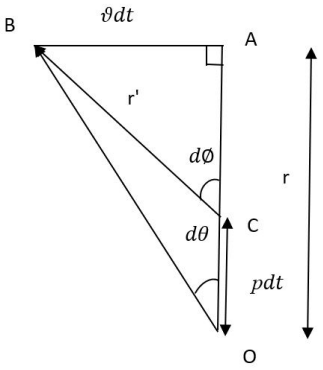


Figure. 15 Case of CP with sudden turn.

Figure 15 is similar to Figures 13 and 14 except that $\emptyset = 0$

Considering Figure 15,

$$\frac{\sin d\theta}{\sin d\emptyset} = \frac{CB}{OB} = \frac{|r'|}{\sqrt{|\vartheta\, dt|^2 + |r|^2}} \leq 1$$

$$\sqrt{|\vartheta\, dt|^2 + |r|^2} \geq |r'| \quad \forall\, dt$$

when $dt \to 0$

$$|r'| \leq |r|$$

Hence the result.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 6, 208-215 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Eye Feature Extraction with Calibration Model using Viola-Jones and Neural Network Algorithms

Farah Nadia Ibrahim*, Zalhan Mohd Zin, Norazlin Ibrahim

Industrial Automation Section, University Kuala Lumpur Malaysia France Institute, 43650 Selangor, Malaysia

ARTICLE INFO

Article history:
Received: 14 June, 2019
Accepted: 28 August, 2019
Online: 05 December, 2019

Keywords:
Eye detection
Calibration
Viola-Jones
Neural network

ABSTRACT

This paper presents the setup of eye tracking calibration methodology and the preliminary test results of the training model from the eye tracking data. Eye tracking requires good accuracy from the calibration process of the human eyes feature extraction from facial region. Viola-Jones algorithm is applied for this purpose by using Haar Basic feature filters based on Adaboost algorithm which extract the facial region from an image. From the extracted region, the eyes feature is selected to find the center coordinate of the iris and be mapped with the calibration point coordinates to create the training model of the eye calibration process. Thus, this paper shows the performance and efficiency of three training functions in Neural Network algorithm to get the best training model with fewer error for more efficient eye tracking calibration process.

1. Introduction

Human features which are once a very selective parameter has been widely applied as the main input for various research of vision. Detection and tracking of human features can improve the performance of system specifically such as security, safety and also human monitory. For eye tracking system, accurate detection of eyes feature extraction is important as it is the first information eye data and calibration process have to be done. Calibration is the process of measuring the eye gaze point calculation which recording the user's gaze before an eye tracking recording is started. For the calibration procedure to be achieve in good accuracy, good eye tracking method has to be applied. There are several types of the eye tracking such as electrography, scleral coil and the most popular method of video tracking. This paper is an extension of work originally presented in 2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR) conference [1].

Electrography is a method that uses electric potentials measured with electrodes placed around the eyes. N. Steinhausen, R. Prance, and H. Prance [2] applied three sensors of electrodes, P. Aqueveque and E. J. Pino [3] aims on developed a low cost electrography system while H. Manabe, M. Fukumoto, and T. Yagi [4] went into automatic drift calibration for electrography. This method requires only very low computational power but relatively poor gaze direction accuracy to locate where a subject is looking compared to a video tracking even though the time of eye movements can be determined.

For video-based eye tracking, calibration has been created by different approaches and presentations. There is an approach of calibration procedure where the user's gaze is being calculated and also an approach of calibration-free where the aims are towards removing the calibration process for a fully automatic system.

D. Model and M. Eizenman [5] proposed general method by extending the tracking range for the remote eye gaze tracking system. By using a stereo pair of cameras, the overlapping field of view is used to estimate the user eye parameters. While, J. Chen and Q. Ji [6] proposed general method by gaze estimation without personal calibration which is the 3D gaze estimation. There is also other research such as [7-9] that also applying the same concept of calibration-free and needed more added inputs while adjusting certain parameters. These research shows that calibration-free system is achievable but the need of calibration for measuring accuracy of user's gaze is still considered as the most important process that cannot be remove in eye tracking.

This paper aims to implement calibration process in order to achieve good accuracy and performance. In order to achieve these objectives, regression technique with good training function is needed. A research from K. Harezlak, P. Kasprowski, and M. Stasch [10] have focus on the process of calibration and analyses the possible steps with ways to simplify this process. The authors compared three regression methods of Classic Polynomial, Artificial Neural Networks (ANN) and Support Vector Regression

*Farah Nadia Ibrahim, Putrajaya, Malaysia, +6014-8271947 & fnadiaibrahim@gmail.com

(SVR) to build the calibration models with presents different sets of calibration points.

Another approach used detection of the user attention by moving calibration target. This technique is presented by K. Pfeuffer, M. Vidal, J. Turner,A. Bulling and H.Gellersen [11] where the work is focused on smooth pursuit eye movement. This system collects sample of gaze points by displaying a smooth moving target to the user. Three realistic applications have been designed with different abilities in each such as Automated Teller Machine (ATM), stargazing and lastly is the waiting screen application. However, this study showed that too slow and too quick effects the accuracy of the system.

Facial region is divided into several features such as eyes including pupil, iris, eye corners, eyelid, nose and mouth. There is a method using template matching in facial image presents by Bhoi, N, & Mohanty, M. N. [12] which use the correlation of the eye template as the eye region and P. Viola and M. Jones [13] that used integral images from Haar Basic filters with Adaboost algorithm to create the region of interest around the area of face features.

After taking consideration of eye tracking types and regression method, this paper adapts the same methodology of [10] in building the calibration models with displaying set of calibration points by using video-based tracking and applied Neural Networks algorithm as the regression method with implementation of Viola-Jones algorithm as the eye feature extraction technique. The extracted eye feature information from Viola-Jones become the input data for training the calibration model. To achieve best training model, three training functions are compared for its performance and processing time.

2. Proposed Method

2.1. Displaying the calibration point as a dot

At the beginning of calibration process, stimuli of calibration point must be display to the user. Thus, we have set up two different calibration points, 4 points and 9 points calibration dots each locate at different position. This process is projected to the computer screen in front of the user to generate the raw data of eye detection. The size of the computer screen is 1920x1080 in pixels. The displayed stimuli are divided evenly on the screen as in Figure 1 and 2.



Figure 1: Screen display of 4 calibration points

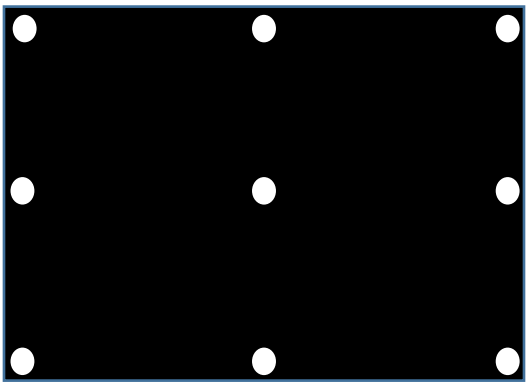


Figure 2: Screen display of 9 calibration points

From Figure 1 and Figure 2, the calibration points are being displayed to the user, one by one point. This process is being conducted by using a Logitech HD Pro Webcam C920 which is a full 1080p high definition. This webcam is equipped with the application of video, and photo capture, face tracking, and motion detection with the size of 640x480 in pixels. The setup of tracking the user's eye features and coordinates are as shown in Figure 3.

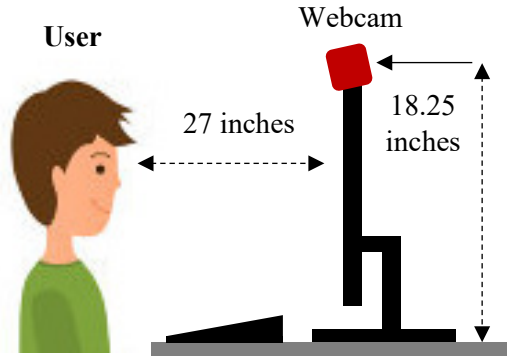


Figure 3: Setup of eye tracking

As seen in the figure above, the user's eye position is being recorded by the web camera. User looked at each point displayed for every 3 seconds at different locations of the calibration point that has been setup. The recorded process is saved in a video format. Thus, the images of user's interaction must be extracted frame-by-frame before the face and eye features detection being carried out. Figure 4 below shown the image of the user looking towards the displayed stimuli of calibration points of 4 points and 9 points on the computer screen.

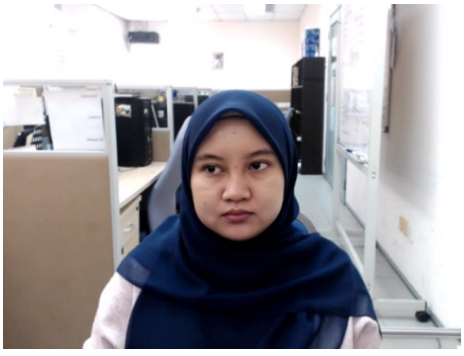


Figure 4: Image of user looking towards computer screen

2.2. Face and eye features detection

A video recording is the first step on capturing the user's eye images for the extraction of the eye center coordinates. As the next step of this research, the face and eye features detection are carried out by using a technique proposed by P. Viola and M. Jones [13], Viola-Jones algorithm. The input data for this algorithm is the image of the user as in Figure 4, that has been extracted frame-by-frame previously. Viola-Jones algorithm is widely known for its ability and performance in detecting the face feature.

P. Viola and M. Jones [13] proposed this method together by using the three main key attributions of integral images, AdaBoost classifier with more complex classifier in a cascade that has been proven to lessen the computational time with high level of detection accuracy. As per mentioned, this research has applied this algorithm which used the Haar-based features that involved the sums of image pixels within the face region in the rectangular areas as shown in Figure 5 and 6.

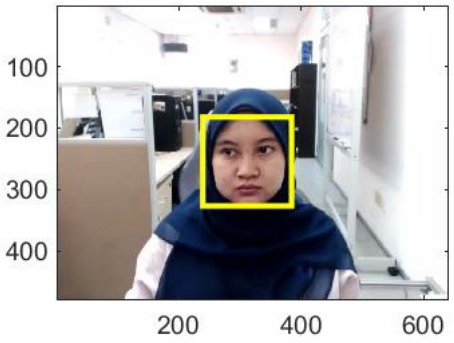


Figure 5: Face feature detection by Haar-like feature

From Figure 6, it shows the rectangular subset of the detection window that will determines whether it looks like a face. Two-rectangle features are shown in (A) and (B) while three-rectangle feature in (C) and (D) a four-rectangle feature. Two-rectangular feature is the different between the sum of pixels within two rectangular regions and both have the same size and shape with horizontally or vertically adjacent. As for three-rectangular feature is the different of the sum within two outside rectangles subtracted from the sum in a center rectangle. While the four-rectangle feature is the difference between the diagonal pairs of rectangles.

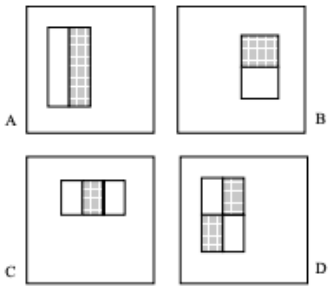


Figure 6: Example rectangle features shown relative to the enclosing detection window

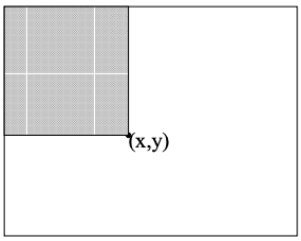


Figure 7: The value of integral image at point (x, y) is the sum of all the pixels above and to the left

Integral image is defined as the summation of all pixels in the image at point (x, y) above and to the left.

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y'),$$

where ii (x, y) is the integral image and i (x, y) is the original image (Figure 7). By using pair of recurrences:

$$s(x,y) = s(x,y-1) + i(x,y) \quad (2)$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \quad (3)$$

where s (x, y) is the cumulative row sum, s (x, -1) = 0, and ii (-1, y) = 0, the integral image can be computed in one pass over the original image. By using integral image, any rectangular sum can be calculated in four array references as in Figure 8.

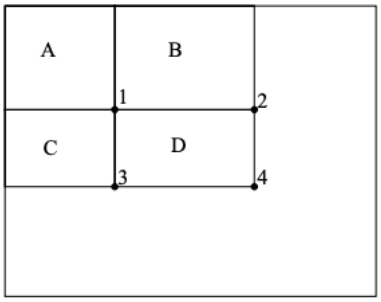


Figure 8: The sum of pixels within rectangle D can be calculate with four array references. The value of integral image at location 1 is the sum of pixels in rectangle A. The value at location 2 is $A + B$, at location 3 is $A + C$, and at location 4 is $A + B + C + D$. The sum within D can be computed as $4 + 1 - (2 + 3)$

Haar feature used the image integral area to compute the value of a feature and its classifier multiply the weight of each rectangle by its area and the results are added together. Since the detection moved across the image of integral, face recognition filter is applied and when the filter gives a positive answer, it will return feedback as face detected in the current rectangle window.

Each face recognition filter contains a set of cascade-connected classifiers that looks at a rectangular subset of the detection window and determines if it looks like a face feature. The classifier will move to the next classifier until all classifiers and filter give a positive answer, the face feature is considered recognized. The Adaboost algorithm is applied to the training of the cascade classifier to minimize the computational time.

The process of feature extraction is continued by filter out the eye pair feature. After the eye pair is detected, the algorithm extracted only half of the eye pair from the user's face feature image. The image of the eye feature is shown in Figure 9 below.

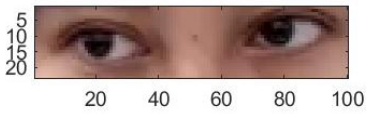


Figure 9: Eye pair feature

From the half of the eye pair feature, a circle detection algorithm is applied to extract the eye center coordinates. This circle detection is a feature extraction technique for detecting circles called Circle Hough Transform (CHT) where the parameters are containing the center of the circle and the radius. A circle can be described by:

$$(x-a)^2 + (y-b)^2 = r^2 \quad (4)$$

where (a, b) is the center of the circle, and r is the radius.

The coordinates will be used as the input data for training the calibration model. When the circle feature has been detected only then the eye center coordinates can be extracted and stored as the raw eye data. Figure 10 shown a circle that are being highlighted in the eye feature image.

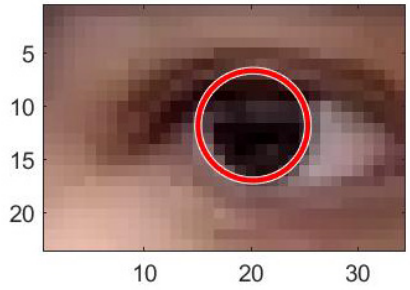


Figure 10: Circle detection on eye feature

2.3. Training the calibration model

After all processes of feature extraction has been done, the raw of eye center coordinates have been produced. These data must be normalized before training the data since the size of the user's eye feature images extracted from the video and the computer screen are different in value.

The normalized eye center coordinates data is stored and be used by Neural Network algorithm for training the model. Neural network can learn and identify images that have been manually trained or labeled for the system to process. A few researches have been conducted using the neural network like [14,15] where they both used neural network for training and identify the parameters of calibration, and input images. Thus, this paper used the approach of neural network to train the calibration model by using the normalized eye center coordinates data from previous process.

There are two sets of calibration points which are 4 points and 9 points that have been distributed at each specific location on the display screen. Thus, two neural networks are been trained for this paper. The type of neural network used is feed-forward back propagation algorithm.

Feed-forward is known for its simplicity of single layer perceptron network which consist of single layer of output nodes where the inputs are fed directly to the outputs via a series of weights. Back propagation is applied with the feed-forward neural network as it is a renowned representative from all iterative gradient descent algorithms that used for supervised learning. This method helps to calculate the gradient descent to looks for the minimum value of the error function in weight space as in Figure 11.

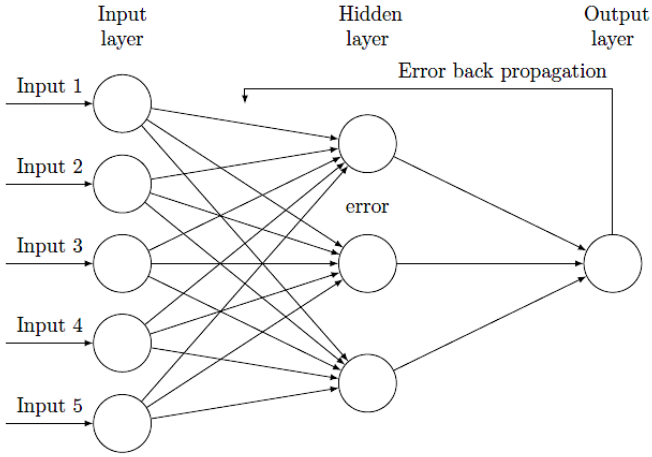


Figure 11: Neural Network architecture with back propagation

As shown in Figure 11, the input given is modeled by using real weights that are usually randomly selected. The output is computed for every neuron from the input layer, to the hidden layer, and to the outer layer. Error is then computed in the output by calculated the differences of the actual output with the desired output. This value will travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased. This process is repeated until the desired output is achieved.

As for the training function, this paper compared the performance and computational time between 3 types of function which are Levenberg-Marquardt known for its fastest, Bayesian Regularization for its efficiency in difficult, small, or noisy datasets, and Scale Conjugate Gradient that requires less memory.

Levenberg-Marquardt (LM) is widely used optimization algorithm for it is outperform simple gradient descent and other conjugate methods. This algorithm provides the nonlinear least squares minimization. Basically, it consists in solving the equation:

$$(J^t J + \lambda I)\, \delta = J^t E \quad (5)$$

where J is the Jacobian matrix for the system, λ is the Levenberg's damping factor, δ is the weight update vector that we want to find, and E is the error vector contained the output errors for each input vector used for training the network. The Levenberg-Marquardt is very sensitive to the initial network weights and does not consider outliers in the data that can lead to overfitting noise. For this situation, we will compare the performance with another technique known as Bayesian Regularization.

Bayesian Regularization can overcome the problem in interpolating noisy data which allows it to estimate the effective number of parameters that actually used by the model such as the number of network weights. It expands the cost function to search for minimal error while using the minimal weights. This function works by introducing two Bayesian hyperparameters, *alpha* and *beta,* to tell which way the learning process must seek. The cost function is as follows:

$$C(k) = \beta^* E_d + \alpha^* E_w, \quad (6)$$

where E_d is the sum of squared error, and E_w is the sum of squared weights. By using Bayesian Regularization, cross validation can be avoided and can reduce the need for testing different number of hidden neurons. However, this technique may be failed to produce robust iterates if there is not much of a training data.

As for Scale Conjugate Gradient, developed by Moller [16], was designed to avoid time-consuming line search. The basic of this algorithm is to combine the model-trust region approach used in Levenberg-Marquardt with the conjugate gradient approach. It requires the network response to all training inputs to be compute several times for each search.

For achieving fast and efficient training model for calibration process, the best training function has to be selected. Thus, the three training functions consists of Levenberg-Marquardt, Bayesian Regularization, and Scale Conjugate Gradient are computed and analyzed the total iterations it takes for achieved minimal error, the processing time and the performance of the algorithm.

The training of 4 points calibration is being carried out first with the total number of neurons for hidden layer of 10, the input data of 200 and the output layer is 150. For each point of the input data, 50 values have been selected. The comparison of those 3 functions are as Table 1 below.

Table 1: Comparison of three training function for 4 points

Training Function	Iterations	Time	Performance
Levenberg-Marquardt	20 Epochs	1m 47sec	1.88e-09 mse
Bayesian Regularization	37 Epochs	5m 29sec	5.39e-08 mse
Scale Conjugate Gradient	76 Epochs	0sec	6.69e-07 mse

From Table 1, Levenberg-Marquardt achieved the minimal training with only 20 epochs of iterations, following with Bayesian Regularization of 37 epochs and lastly with the highest number of iterations of 76 epochs, Scale Conjugate Gradient. For the fastest function, Scale Conjugate Gradient achieved the best computational time with 0 seconds following with Levenberg-Marquardt and Bayesian Regularization. As for the best performance, Levenberg-Marquardt achieved it with only 1.88e-09 mean squared error that is nearest to zero values.

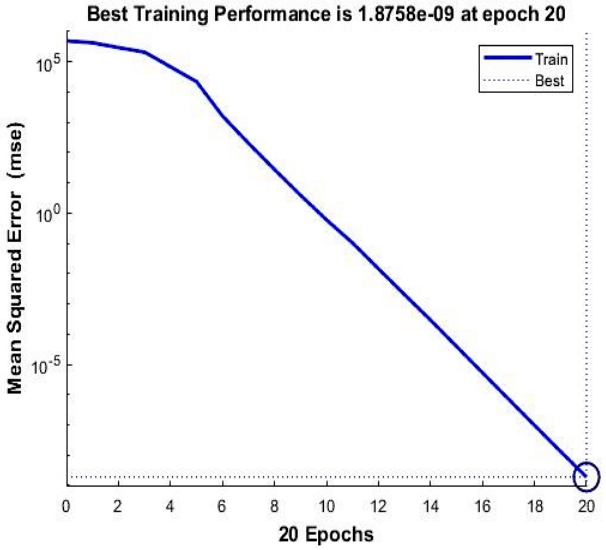


Figure 12: Performance of Levenberg-Marquardt for 4 points calibration model

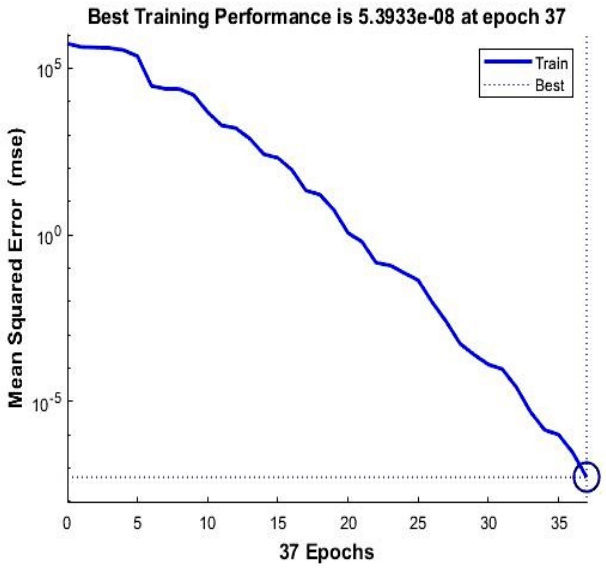


Figure 13: Performance of Bayesian Regularization for 4 points calibration model

From Figure 12,13 and 14, it is shown that the best training performance are from the training function of Levenberg-Marquardt, followed by Bayesian Regularization and Scale Conjugate Gradient. As for the regression values which measure the correlation between outputs and targets data, all three training functions achieved a close relationship with R value of 1 as shown in Figure 15.

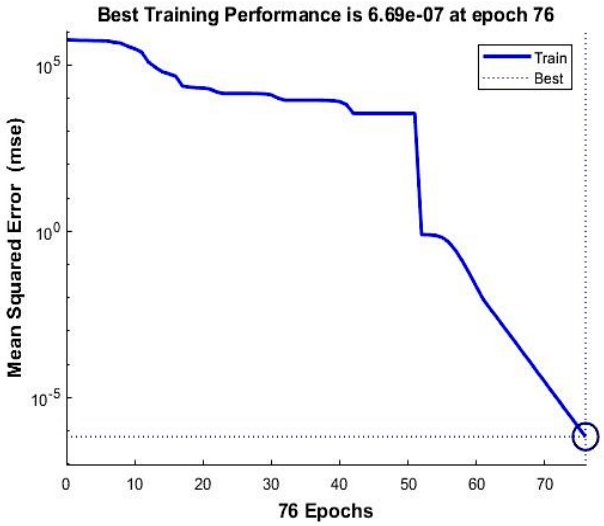


Figure 14: Performance of Scale Conjugate Gradient for 4 points calibration model

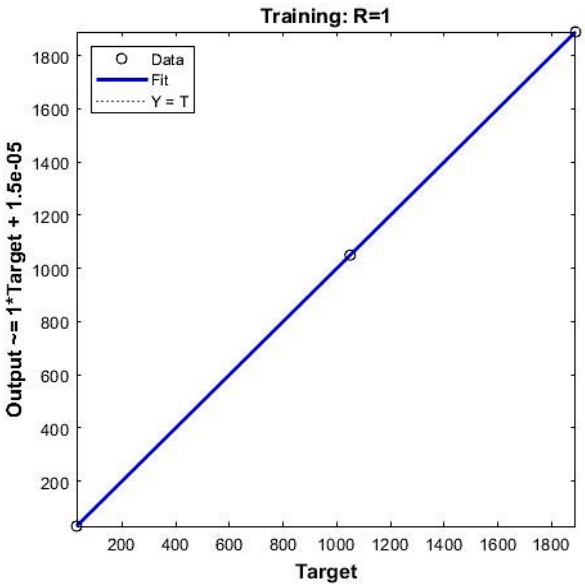


Figure 15: Regression values of all three-training function for 4 points calibration

For the training of 9 points calibration, the total number of neurons for hidden layer is 10, the input data of 455 and the output layer is 404. The total number of input data for each point are also 50 values which make the input data for neural network of 9 points calibration is 455. Table 2 shown the comparison of the three-training function for 9 points calibration.

Table 2: Comparison of three training function for 9 points

Training Function	Iterations	Time	Performance
Levenberg-Marquardt	20 Epochs	24m 55sec	1.97e-09 mse
Bayesian Regularization	43 Epochs	1h 25m 44sec	3.56e-08 mse
Scale Conjugate Gradient	446 Epochs	2sec	8.41e-07 mse

From Table 2, Levenberg-Marquardt achieved the minimum iterations of 20 compared to Scale Conjugate Gradient which achieved the highest iteration of 446 epochs. Even though Levenberg-Marquardt achieved the minimum iterations, it cannot be compared with Scale Conjugate Gradient for its computational time. This function required only 2 seconds to finish compared with the other two functions of Levenberg-Marquardt and Bayesian Regularization each takes around 24 minutes and 1 hour 25 minutes.

Since this modeling is from 9 points calibration which contains 455 inputs data, it's proven that Scale Conjugate Gradient are the fastest training function even by given many datasets. For the best performance of the training model, Levenberg-Marquardt outdone the other two functions by achieved mean squared error of 1.97e-09 as shown in Figure 16 below.

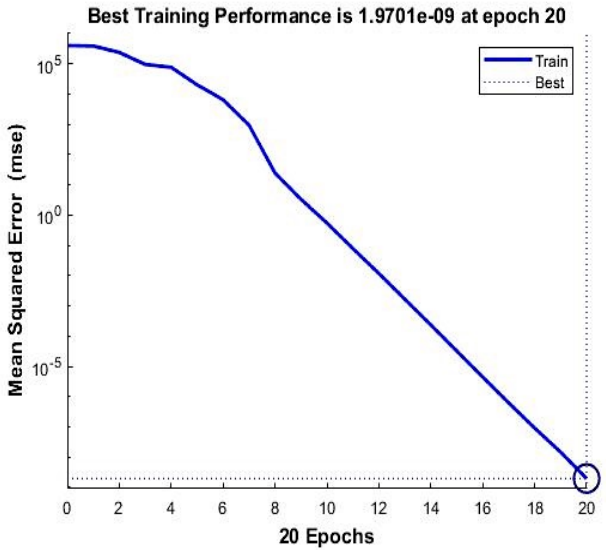


Figure 16: Performance of Levenberg-Marquardt for 9 points calibration model

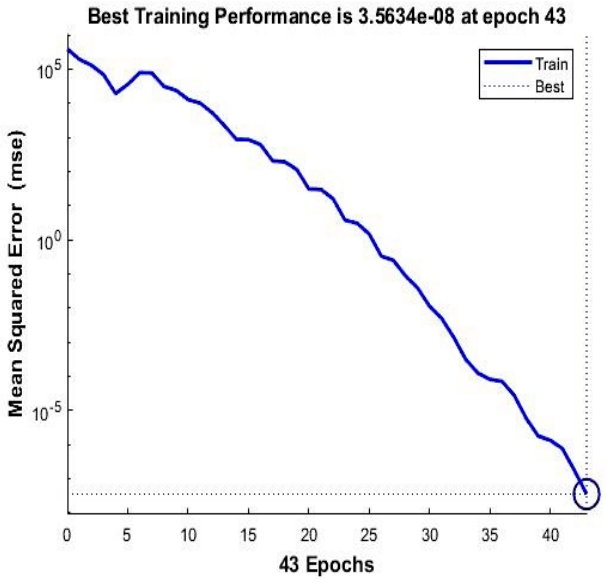


Figure 17: Performance of Bayesian Regularization for 9 points calibration model

Figure 17 and 18 are the performance achieved by Bayesian Regularization and Scale Conjugate Gradient for the 9 points calibration model. Bayesian Regularization achieved its best performance of 3.568-08 mse at the iterations of 43 while Scale Conjugate Gradient achieved its performance at 8.41e-07 mse at the iterations of 446.

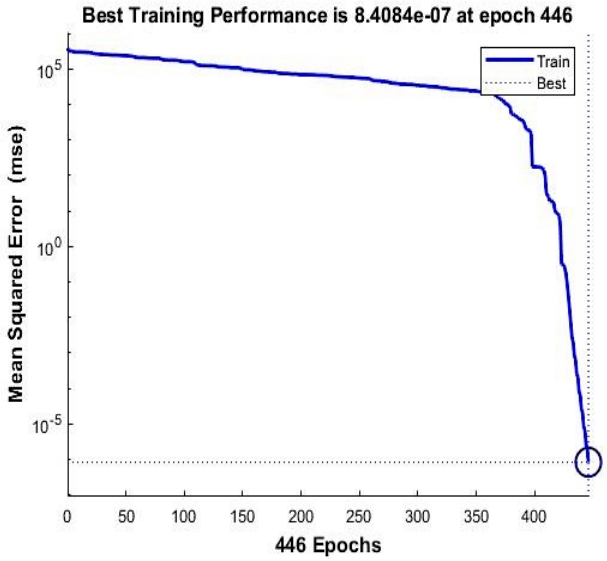


Figure 18: Performance of Scale Conjugate Gradient for 9 points calibration model

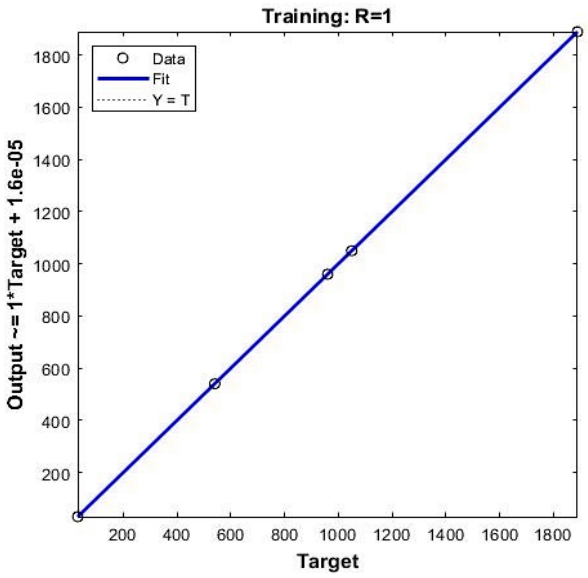


Figure 19: Regression values of all three-training function for 9 points calibration

As for the regression values, all the training functions have gained close relationship with R value of 1 as shown in Figure 19.

3. Discussion

The raw eye center coordinates data required normalization which are crucial process that can affect the performance of the data itself. Thus, this normalization can be improved or automated directly in the algorithm so that the process of the training data can be simplified. As for training models, back propagation is proven to be fast, simple and does not need any special mention of the features of the function to be learned but can be quite sensitive to noisy data and the actual performance dependent on the input data given. From the training functions that we have used, taken from the results of their performance, the best training function is Levenberg-Marquardt that is proven the best for its efficiency and good performance while gaining minimum iterations of error with fast computational time even though Scale Conjugate Gradient is a lot faster, the iterations of the error is too high.

4. Conclusions

In this paper, the eye feature extraction and calibration model were conducted by applying Viola-Jones and Neural Network algorithms. Viola-Jones approach using the Haar-like features that involve sums of pixels with integral image by minimizing computational time with Adaboost algorithm and detected the feature of face. As for the Neural Network, the feed-forward back propagation algorithm is applied together and compared the performance and efficiency of three training functions; Levenberg-Marquardt, Bayesian Regularization, and Scale Conjugate Gradient. The overall result shown that these algorithms can be applied and improve the performance and lessen the computational time for eye data training for calibration model.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] F. N. Ibrahim, Z. M. Zin and N. Ibrahim, "Eye Center Detection Using Combined Viola-Jones and Neural Network Algorithms," 2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR), Putrajaya, 2018, pp. 1-6. doi: 10.1109/ISAMSR.2018.8540543.

[2] N. Steinhausen, R. Prance, and H. Prance, "A three sensor eye tracking system based on electrooculography," IEEE SENSORS 2014 Proc., pp. 1084–1087, 2014.

[3] P. Aqueveque and E. J. Pino, "Eye-Tracking Capabilities of Low-Cost EOG System," 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 610– 613, 2014.

[4] H. Manabe, M. Fukumoto, and T. Yagi, "Automatic drift calibration for EOG-based gaze input interface," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp. 53–56, 2013.

[5] D. Model and M. Eizenman, "User-calibration-free remote eye-gaze tracking system with extended tracking range," Can. Conf. Electr. Comput. Eng., pp. 001268–001271, 2011.

[6] J. Chen and Q. Ji, "Probabilistic gaze estimation without active personal calibration," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 609–616, 2011.

[7] Y. Sugano, Y. Matsushita, and Y. Sato, "Calibration-free gaze sensing using saliency maps," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 2667–2674, 2010.

[8] F. Alnajar, T. Gevers, R. Valenti, and S. Ghebreab, "Calibration-Free Gaze Estimation Using Human Gaze Patterns," 2013 IEEE Int. Conf. Comput. Vis., pp. 137–144, 2013.

[9] Y. Zhang, A. Bulling, and H. Gellersen, "Pupil-canthi-ratio: a calibration-free method for tracking horizontal gaze direction," Proc. AVI, pp. 129– 132, 2014.

[10] K. Harezlak, P. Kasprowski, and M. Stasch, "Towards accurate eye tracker calibration – methods and procedures," Procedia - Procedia Comput. Sci., vol. 35, pp. 1073–1081, 2014.

[11] K. Pfeuffer, M. Vidal, and J. Turner, "Pursuit calibration: making gaze calibration less tedious and more flexible," Proc. 26th …, pp. 261–269, 2013.
[12] Bhoi, N., & Mohanty, M. N.,"Template Matching based Eye Detection in Facial Image," *International Journal of Computer Applications,12*(5), pp. 15-18, 2010 doi:10.5120/1676-2262.
[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I.doi: 10.1109/CVPR.2001.990517
[14] M. Wang, Y. Maeda, K. Naruki, and Y. Takahashi, "Attention prediction system based on eye tracking and saliency map by fuzzy neural network," 2014 Jt. 7th Int. Conf. Soft Comput. Intell. Syst. 15th Int. Symp. Adv. Intell. Syst., pp. 339–342, 2014.
[15] E. Demjén, V. Aboši, and Z. Tomori, "Eye tracking using artificial neural networks for human computer interaction.," Physiol. Res., vol. 60, no. 5, pp. 841–4, 2011.
[16] Møller, M. F., "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks, 6*(4), pp. 525-533, 1993 doi:10.1016/s0893-6080(05)80056-5.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 211-220 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Critical Analysis of Topics in Software Architecture and Design

Janet Bishung, Ooreofe Koyejo, Adaugo Okezie, Boma Edosomwan, Sylvester Ani, Abisola Ibrahim, Austin Olushola, Isaac Odun-Ayo*

Department of Computer and Information Sciences, Covenant University, Ogun State, Nigeria

A R T I C L E I N F O

Article history:
Received: 20 January, 2019
Accepted: 07 March, 2019
Online: 27 March, 2019

Keywords:
Software Architecture
Software Design
Service Oriented Architecture
Design Patterns

A B S T R A C T

Software architecture and design is an important component in the software engineering field. This aspect of software engineering covers the functional and non-functional requirements of any system being proposed to be developed, while software architecture deals with non-functional requirements, software design entails the functional requirements.

The objective of this paper is to critically analyze current topics in Software architecture and design. The method of analysis involved the use of inclusion and exclusion criteria of papers published in journals and conferences. These papers were accessed from digital libraries like ScienceDirect, and IEEE explore*, with a quantitative approach of analysis been imbibed. From the analysis, the result showed that, of 35 papers used in analysis, 34.3% discussed stakeholders' involvement and decisions in software design. 17.1% for design quality, 20% examined software reuse while 11.4% discussed software evaluation and 8.6% of papers reviewed discussed software management, evolution and software development life cycle each which should be more focused as it is the fundamentals of software design and architecture. From the analysis derived, stakeholder's involvement and decision in software design is an integral part in software building for effective use. Thereby making researchers dwell more on the topic. The least discussed topics was due to the expectations of researchers. Expecting readers to have a fore knowledge of the fundamentals of design which includes software management, evolution and software development life cycle.*

1. Introduction

Software Architecture gives the high-level description of a software and the discipline of creating the structures and systems [1]. It gives blueprint for the system, laying out tasks to be executed in a logical manner through the design [2]. It is the fundamental structural choices made vis a vis the business needs of the organization which may be costly to change once implemented [3]. Although there are no standard procedures to follow in software architecture that can address all issues of concern in general software development, certain factors should be of utmost non-negotiable fundamentals in software development, to ensure standardization thereby avoiding incessant collapse of systems witnessed in the early years of software developments [4]. Among the factors that will be enumerated briefly is the factor of proper documentation. This facilitates communication among stakeholders, captures decisions about the structures of the task and the design options – focusing on the decisions that must be right from the onset, otherwise, the imminent collapse of such system will be devastating– [5]. Since software architecture is largely driven by the required or expected functionalities, the current insight to software architecture is that required functionalities should reflect or incorporate all the quality attributes which include fault-tolerance, reliability, backward compatibility, extensibility, availability, maintainability, usability, security amongst others. Stakeholders concerns should reflect these quality attributes at both non-functional and functional stages without recourse to extra cost.

Software design envisions and defines software solutions to problem sets. It involves a sequence of steps that describe all aspects of the software in development [6]. Here, solutions to the problem the software is to solve are expressed in a logical sequence with details of their relationships. It begins with describing the total components to be built, then refine them to every detail. It is

*Isaac Odun-Ayo, Covenant University, +2348028829456
isaac.odunayo@covenantuniversity.edu.ng

the physical expression of all the processes that create solutions according to stakeholders' expectations. The designs are taken within the confines of fundamental principles which ensure designs are traceable to requirement analysis, uniformity, integration, structured for change etc. This ensures standardization while addressing the business objectives and stakeholders' needs [7]. In general, software development is dependent on time and cost and the design option should reflect these critical factors.

Software Design addresses all the expected required functionalities of the business objectives. This includes specifications of services, components, integration, data models and algorithms. Meanwhile, Software Architecture addresses design standard ensuring that it aligns with stated strategy as it pertains to business and technology of an organization. This includes considerations such as compliance, technology standards and operational efficiency. An architecture designed is intended to prevent repetitive mistakes in design or inconsistency with other aspects of the organization. It could be said that architecture is global optimization of software and design is local optimization [8]. In general, software architecture provides standardization upon which software designs are tailored.

Software architecture and software design are extremely important for a software project. So, here are brief points highlighting benefits of software architecture & design; solid foundation for software project, scalability of platform, increase performance, identification of area of cost savings, vision implementation, increase quality, better code maintainability, prioritization of goals, higher adaptability, faster platform, risk management and enable quicker changes among many others [9].

This study is aimed at helping researchers to have an in-depth knowledge of the fundamental as well as critical topics involved in software architecture and design of proposed and existing systems. It would also help in uncovering the critical gaps in which many researchers were not able to explore thereby improving knowledge as regards software architecture and design. The objective of this paper is to conduct a critical analysis on current topics in Software architecture and design. Published papers and articles on the topics discussed in the paper were reviewed with the percentages of each topic in relation to others were calculated.

Other parts of the paper are organized as follows: Section 2 gives review of some related work done, Section 3 gives a highlight on software architectural styles, Section 4 discusses Software Oriented Architecture, Section 5 is a discussion on Design Patterns; section 6 presents the results and discussion of the major topics selected for analysis from literatures. Section 7 gives a conclusion of the paper with recommendations for future work.

2. Related Work

In [10], the importance of software architecture as a vital aspect of software development was examined. The paper explored two important components, software evolution and re-usability. These are very critical component that helps curb the huge expenses involved in the development of software. Software architecture that can be evolved and reused should be in high demand, as software evolution and reuse are more likely to receive higher pay-off. [11] considers the fact that researchers cannot overlook the year's technology as well as the fact that software architecture employs fully detailed explorations of notations, techniques, analysis, tools and creation methods. There exist an intersection and interrelationship of software architecture with the study of software design, domain-specific design, program analysis, software families, specific classes of components and component-based reuse. The comparative analysis of software evolution methods in [12] explains the systematic comparison between architecture models and evolution methods centering its base on the scenario-based approach of software architecture.

The review on the successes and failure for software architecture in [13], gives insight into software architecture development and management process. It assesses previous literature and experiences to identify the factors that cause success and failure for software architecture and classifying these factors into subgroups as indicated by practitioners. [14] proposed a different methodology as a guide for practitioners supporting software architecture and design in an agile environment. Highlights of phases in software design process were covered, tools and techniques were proposed to implement those phases. Architectural design decisions and knowledge in [15], examine the essence of reusable architectural knowledge and the importance of documenting quality attributes along with the decisions captured during architectural design.

In [16], the concept of sustainability was introduced for software design. It was essential in integrating it into the existing catalog of design quality attributes. This was because design is a key factor in software development and this has been noted by many researchers. The information produced during software design tends to evaporate progressively due to certain conditions like software evolution. [17] considered all developed software need to meet required and specified quality standards as requested by users or stakeholders, quality being a major issue in software systems today. To achieve the quality requirements, different analysis approach was explored and a critical evaluation of the software system was carried out. This was to analyze the architecture, thereby verifying that quality requirements have been duly addressed in the design.

From an in-depth review of System Development Methodologies (SDMs) conducted by [18], a list of the important features in each methodology was made. Despite several SDMs, most of them share similar activities which include well known and practiced requirements-analysis, design, codification-test and implementation, all put together in project management. Four stages in the evolution of SDMs were reported which include pre-methodologies, rigor-oriented methodologies, agile-oriented methodologies and emergent service-oriented methodologies. Well-recognized SDMs in software engineering include Rational Unified Process, Microsoft Solutions Framework and Model-based (system) architecting and software engineering.

The software architecture chosen in the development of a software product is dependent on the software requirements and constraints. A Design Association Theory (DAT) to show the causal relationship of why a design should exist was proposed by [19] for associating design concerns, design problems and design solutions. DAT proposes a five-step association-based design

review process, which includes the extraction of requirements, extraction of design, construction of causal relationship between design elements and design solutions, discovery of potential design issues and verification and confirmation of design issues with architects. In documenting design decisions and design reasoning for objective evaluation, architects also utilize DAT. The DAT model helps designers and reviewers in associating architectural knowledge.

3. Software Architectural Styles

Architectural Styles are principles which shapes an application. It is more of an abstract framework of a system in the area of its organization. There are six major types of architectural styles. Namely:

3.1 Dataflow Architecture

In this, all software systems are categorized as lists of shifts on chronological set of input data, where data and operations are independent of each other. When data enters this system, it flows through the modules one at a time until they are assigned to some final destination [20]. Its aim is to achieve the qualities of reuse and immovability and is suitable for applications involving series of independent data computations on orderly defined input and output. There are three execution sequences between modules that the data flow architecture uses; Batch sequential, Pipe and filter or non-sequential pipeline mode and the Process control [20].

3.2 Data-centered Architecture

Data is centralized in this form of architecture and accessed frequently by other components that modify data. Its main purpose is achieving integrality of data. It contains different components that communicate using shared data repositories. The components access a shared data structure and are independent, meaning, they interact only through the data store. The flow of control sums the types of Data-centered architecture into two types; The repository and the blackboard architecture style. This form of architecture is mostly used in information system [20].

3.3 Hierarchical architecture

This views the whole system as a hierarchy structure whereby software systems are decomposed into subsystems at different levels in the hierarchy. It is mostly used in designing system software such as network protocols and operating systems. [20].

3.4 Interaction oriented architecture

The main aim of the interaction-oriented architecture is to separate users' interaction from data abstraction and business data processing. It divides the system into three major partitions: Data module (which provides the data abstraction and all business logic.), Control module (Which identifies the flow of control and system configuration actions) and the View presentation module (This is responsible for the visual or audio presentation of data output. It has two major styles: Model-View-Controller (MVC) and The Presentation-Abstraction-Control (PAC) [20].

3.5 Component based architecture

It is an architecture that decomposes software designs into functional components with their own methods, events and properties. These components become loosely coupled and reusable to provide modular programs that can be tailored to fit any need. [21]

3.6 Distributed architecture

This is a form of architecture that sits in the middle of a system and manages or supports the different components of that distributed system. Its aim is transparency, reliability, and availability. It hides the way in which resources are accessed and the differences in data platform, the resource location, different technologies from users, failures and resource recovery and a host of others. [20]. It is the most widely used form of architecture as it aligns with the technological advancement of the 21st century. Software development has improved greatly with the introduction of the internet. Software is now been distributed, components been reused, as well as introduction of concurrency and simultaneous change in the modification of data. These are the major advantages of the distributed architecture which has made it a common form of architecture in time past. [20]. There are different types of Distributed Architecture;

3.6.1 Broker Architecture: Mostly used to coordinate and enable the communication between registered servers and clients more like a software bus [20].

3.6.2 Client-server Architecture: Is commonly used by search engines, web servers, mail servers, it is mostly based on the functionality of the clients that is, requesting services of other components. The Service Oriented Architecture is a major sub-division of this. It supports business-driven Information Technology (IT) approach in which an application consists of software services and software service consumers. It has the ability to develop new functions rapidly which makes it mostly used along with its basic features which would be explained in the next section. [22]

Service Oriented Architecture would be further discussed due to its relevance in the current IT revolution. It is a known fact that cloud computing has come to stay and due to the integration ability of Service Oriented Architecture it has become a yardstick in the cloud computing revolution as their technologies have become more like bridges to the cloud.

4. Service Oriented Architecture (SOA)

SOA is a software architecture and design styles that entails the use of services as its main building component [23]. A service (as a software component), is a technique that allows access to several capabilities. SOA is now a mainstream software development mechanism. Despite the introduction of new architectural variants like cloud computing or micro-services [24], SOA is still widely used. This is due to its support for fast building applications using assembling of Internet-accessible services, allowing software organizations to hasten the development of distributed applications as well as a result time-to-market. After all this, a service is simply a distinct unit of performance that specifies a business function. This simply means it relies on Web Services for its implementation [24].

SOA carries out two core functions. Which are creating broad architectural models that explain application goals, including the approaches that help meet the goals. The second function is it defines the implementation specifications, which is mostly

integrated to the Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP) specifications [25].

4.1 Major Principles of SOA

The major principles of SOA are [26]:

i. Regulate Service Contract - In this, there must be a form of description that explains what the service is about. This makes it easier for the client applications to understand what the service is meant to do.

ii. Loose Coupling –It entails components having little or no dependency on each other. This is a major characteristic of web services that emphasizes that there should be less dependency between web services and the client initiating this web service. Therefore, if any service functionality changes at any point in time, it should not hinder the client application from working.

iii. Service Abstraction - In this, service is ought to encapsulate its procedures and not expose how it executes its functionality. Explaining to the client application what it does and not how it does it for security purposes.

iv. Service Re-usability - Logic is separated into services with the aim of increasing reuse capability. In any technology, re-usability is a major issue as no one would want to spend time and effort writing the same codes again for multiple applications that require them.

v. Service Autonomy -The service knows everything about the application or system and what functionality it offers so it has complete control over the source code it encompasses.

vi. Service Statelessness - Superlatively, services ought to be stateless. Meaning they should not withhold any form of information from one state to the other.

vii. Service Discoverability - Services can be identified in a service registry. A service registry is a resource that allows controlled access to data for the controlling of SOA.

viii. Service Composability - It splits big issues into little ones. It is to be noted that not all functions should be embedded in an application and moved into one single service. Instead the service should be split into modules each having separate business functionalities.

ix. Service Interoperability - Services should accept and make use of various standards allowing different subscribers to use their various service.

For the implementation of SOA to be a success, you would need a productive SOA method that explains the plans, discoveries, procedures and the selected goals [26].

The integration platform for SOA plays a crucial role in the merging of existing application to cloud services. With SOA, components were split into services that became re-usable and easy to use among several systems. This is similar to the system means used in the automotive industry, where different layouts/systems share the same components e.g. engines. Cloud computing is re-vamping the IT world as we know it. The IT systems are utilized by users and companies. In the automotive industry, owners do not need to buy their own vehicles but can use car-sharing services. These service providers show the cars to several drivers. Comparably, in IT, a cloud service provider acts as a middle man and merges several clouds. Existing IT systems that require combination with new cloud-based solutions or inter-mediated are resolved using cloud service providers [37].

Cloud computing is re-modelling the IT industry and how its services are utilized, just like how petrol is soon going extinct in the automotive industry. Electric cars are now in use even as we use petroleum-based cars. As IT companies continue to make use of more cloud technologies the SOA technologies will continue to serve as bridges to the cloud.

Micro-service-based software architecture is the re-industrialized application of the SOA model. The components are developed as services using Application Programming Interfaces (API), just like the SOA would require. An API broker serves as a mediator to access components, ensuring SOA security and governance practices are followed.

SOA principles have taken us to the cloud and aides the most improved version of cloud software development techniques in use today. [27]

4.2 Advantages of SOA [28]

- Services can be reused in multiple applications independent of their interactions with other services.
- Due to service in-dependency, services can be easily updated or maintained without having to worry about other services.
- SOA-based applications are more reliable since they are small independent services that are easier to test and debug.
- Multiple instances of a single service can run on different servers at the same time.
- It improves Software Quality.

4.3 Disadvantages of SOA [28]

- Every time a service interacts with another service, complete validation of every input parameter takes place. This increases the response time and machine load, and thereby reduces the overall performance.
- There would be high investment cost as implementation of SOA requires a large upfront investment by means of technology, development, and human resource.

5. Design Patterns

Software design could be considered the most important aspect of software development as well as the most difficult process in a software development life cycle. Over the years, based on experience, programmers have embodied and recommended demonstrated results to fulfill the persistent issues that arise during design. Accordingly, the experience-based clarifications are composed and acknowledged as a consistent model for designs patterns [29]. Various design patterns have been presented and classified either as a sanctioned or a variation key to take care of design issues. The current programmed systems for design pattern(s) options help fledgling programmers to choose the more

proper pattern(s) from the rundown of relevant examples, to tackle an issue during the design period of software that is been developed. [29]

In software engineering, a design pattern is a general repeatable result for an ordinarily happening issue in programming structure [30]. Design patterns are used to ensure reuse of software design solutions in the early phases of software development, especially in the requirements engineering phase. Patterns do not provide visible solutions, but they present the concepts from which solutions are derived [31]. Design patterns have been introduced for defining good practices in software design [33]. Design patterns can be used in requirements engineering as patterns exist for core activities of a process. Design patterns are not pure inventions like a light bulb or a car. They are derived patterns that software engineers and architects found, that could be standardized to be used to solve similar problems categorized across three major areas; creational, structural and behavioral. When we determine the proper structure design for a task or issue, it helps us avoid changes that would require budgetary expenses, untenable, multiple and inefficient codes as the system scales up [32].

5.1 Classification and Selection of Design Patterns

The intrigue and involvement of programmers are utilized to present recent arrangement plans for the association of design patterns of specific concerns like object-oriented development and real-time applications [30]. The outline of current efforts describes the number of important categories of design patterns which relies upon the sort and multifaceted nature of target issues. For instance, [33] introduced three important classifications namely behavioral, creational, and structural in the setting of object-oriented advancement to solve recurring issues. In a specific circumstance of working applications, [30] introduced a catalog of thirty-four patterns, which are divided into five categories based on their relevance. In this paper, we would consider the object-oriented advancement categories.

A. Creational Design Patterns

Creational design patterns manage object creation systems attempting to make questions in a way that suits the circumstance. The essential type of object creation could result in design issues or added intricacy to the design. Creational design patterns take care of this issue by controlling this object creation [30]. These patterns can be further divided into class-creation patterns and object-creation patterns. Class-creation patterns use legacy viably in the instantiating of procedures while object-creation designs use assignment adequately to take care of business [35]. Creational design patterns are singleton, abstract factory, prototype, factory method, builder and object pool [31].

B. Structural patterns

Structural Design Patterns are used to ease a design by recognizing a straightforward method to acknowledge relationships [30]. These design patterns are tied in with sorting out various classes and object to frame bigger structures and give new usefulness [35].

C. Behavioral patterns

Behavioral design patterns are design patterns that identify basic correspondence designs among object and understands the patterns. By doing so, these patterns increment adaptability in doing this correspondence [30]. Behavioral patterns are about identifying basic correspondence designs among object and understanding the patterns that exist among them [35].

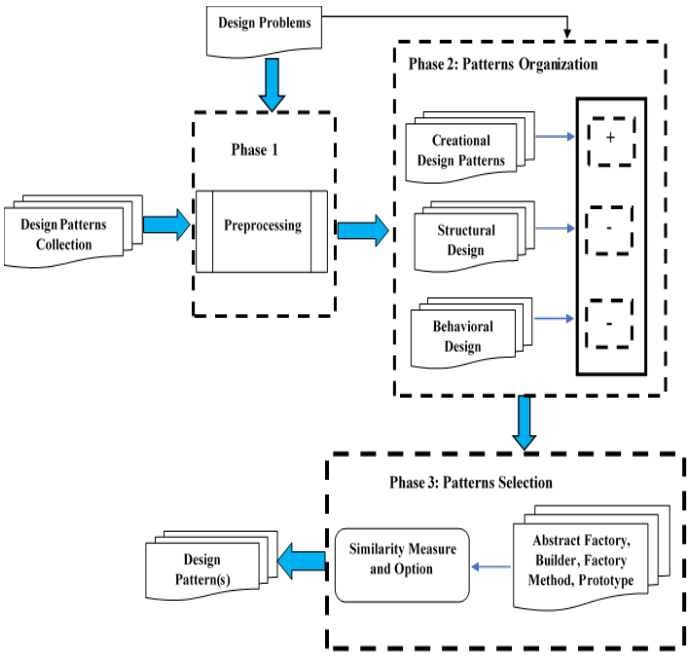


Figure 1: Diagrammatic Representation of Design Patterns Classification and Selection [34].

5.2 Design Pattern Topics

Six research design pattern topics concluded by [36] includes; pattern usage, quality evaluation, pattern mining, pattern specification, pattern development and miscellaneous issues.

i. Pattern development: This involves any advancements in design pattern research such as:
   a. Proposing a new model or new model language
   b. Reviewing model variants, composing models or elaborating a specific model, model evolution.
   c. Arranging current design patterns into distinct areas.
ii. Pattern usage: this relates to the commitment of utilizing patterns in the software development process. They are characterized into two primary groups:
   a. Pattern utilization
   b. Pattern application
iii. Pattern mining: this includes discovering examples of the pattern in the code or design of a system. It can be characterized into two major classes:
   a. Introduction
   b. Evaluation
iv. Quality Evaluation: the quality and effect of a design pattern on a system after applying it, is one important concern developers have. This can be characterized into two major classes:
   a. Pattern evaluation
   b. Application evaluation
v. Pattern specification: it includes utilizing distinctive strategies and notation of representing the patterns. The two main groups are:

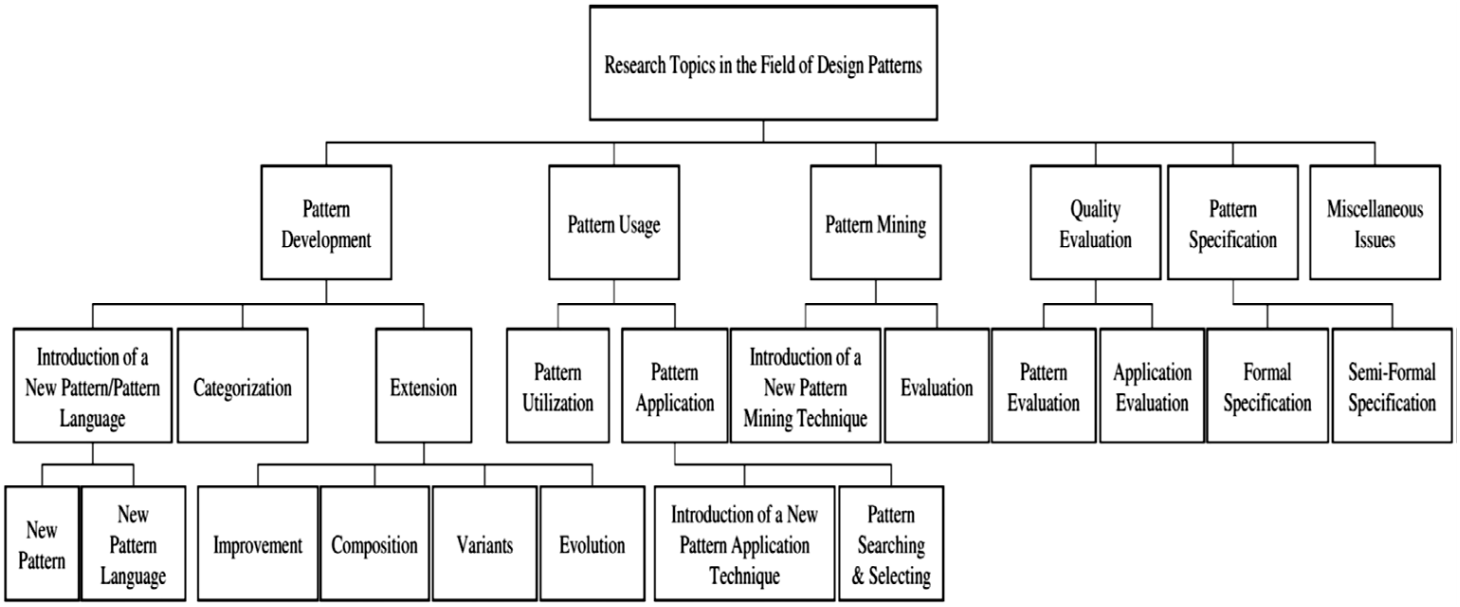


Figure 2: Design Pattern Research Tree [36]

a. Formal specification schemes
b. Semi-formal specification schemes

vi. Miscellaneous issues: this includes other issues that cannot fit into any of the previous classifications' issues such as re-factoring, code smells and anti-patterns. A major challenge of design patterns as discussed in [34,38, 39] is the searching and selection of design patterns before employing a right pattern into the system.

5.3 Pros and Cons of Design Patterns

It is worthy of note that there are very many design patterns available and a lack of understanding of these patterns pose problems for designers in software development especially for the novice [37]. Some of the pros and cons of the design patterns include:

Pros: [40]

- Easy to adapt and very flexible to predictable changes in business needs.
- Easy to test unit and validate individual components.
- Can provide organization and structure when business requirements become very complicated.

Cons: [40]

- Beginner engineers may not understand them, and these can cause a huge delay in development.
- Oftentimes used improperly without a realistic understanding of how the software is likely to change.
- It can add memory and processing overhead, sometimes it is not appropriate for applications such as low-level systems programming or certain embedded systems.

Design pattern, in general, enhances the nature of a product framework by giving a demonstrated solution for repeating design issues [41]. Also, the application of patterns brings about increment in quality and profitability of the software development process [31].

6. Results and Discussion

Table 1 shows several topics in software architecture and design alongside relevant work from several authors. Only one facet was used in this analysis, which are topics that relate to the subject (software architecture and design) discussed by these important authors. These topics were selected using inclusion and exclusion criteria. The essence of the selection criteria was to locate and add all papers that are necessary for the analysis. The inclusion and exclusion criteria were used to eliminate publications that were not significant to the study. It was observed that these topics were the most discussed and they also cut across academics and industry practice. The papers used for this review were access online from digital libraries like ScienceDirect, and IEEE explore. The authors and title of publications are listed in no particular order.

6.1. Software Evolution

Software evolution is the process of developing a software product, employing software engineering principles and methods. It follows the initial development of software and required maintenance. Updates are done till the desired software product is developed, thereby satisfying the (user) expected requirements [42]. This process implements changes to the original software, until the desired software is accomplished. Of the 35 papers used in this analysis, 8.6% of them reported on software evolution. From table 1, the findings show that software evolution though an integral part of software development was not substantially discussed by majority of the literature reviewed during the course of this study. This implies that there may be a decrease in research of software evolution.

6.2. Software Reuse

Software reuse involves creating new software systems from existing software frameworks rather than building software systems from beginning. This simple yet powerful methodology of software development was introduced in 1968 and now widely

Table 1: analysis of software architecture and design topics

Authors	Software Evolution	Software Reuse	Software Management	SDLC	Software Evaluation	Stakeholder Involvement/ Decision Making	Design Quality
Mekni, M., Buddhavarapu, G., Chinthapatla, S. and Gangula, M. (2018) "Software Architectural Design in Agile Environments"	-	-	-	x	x	-	-
Robillard, M.P. (2016) "Sustainable Software development"	-	-	x	-	-	-	x
I. Lytra, G. Engelbrecht, D. Schall and U. Zdun (2015) "Reusable Architectural Decision Models for Quality-driven Decision Support: A Case Study from a Smart Cities Software Ecosystem"	-	x	-	-	-	-	x
A. Ramirez, J. R. Romero and S. Ventura (2018) "Interactive multi-objective evolutionary optimization of software architectures"	-	-	-	-	-	-	-
O. Sievi-Korte, S. Beecham and I. Richardson (2018) "Challenges and recommended practices for software architecting in global software development"	-	-	x	-	-	-	-
N. Hamalainen, J. Markkula, T. Ylimaki and M. Sakkinen (2006) "Success and Failure Factors for Software Architecture"	-	-	x	-	x	-	-
A. Alhar, R. Mazamal and F. Azam (2016) "A Comparative Analysis of Software Architecture Evaluation Methods"	-	-	-	-	x	-	-
David Garlan. (2000) "software architecture"	-	-	-	-	-	-	-
R.Kazman, C.H Lung, (1997) "An approach to software architecture analysis for evolution and reusability"	x	x	-	-	-	-	-
Smrithi Rekha V and Henry Muccini, (2018) "Group decision-making in software architecture: A study on industrial practices"	-	-	-	-	-	x	-
Hans van Vliet and Antony Tang, (2016) "Decision making in software architecture"	-	-	-	-	-	x	-
R.Kazman, C.-H. Lung, s. Bot, and K. kalaichelvan (1997) "An approach to software architecture analysis for evolution and reusability"	x	x	-	-	-	-	-
M.Shaw, and P.Clements(2006) "The golden age of software architecture: a comprehensive survey."	-	-	-	-	-	x	-
H.V. Mohammad, A. Bavar, N.M. Khashayar, and D. Negin(2009)	-	-	-	-	-	-	-
I. Dobrica, and E. Niemela (2002) "A survey on software architecture analysis methods"	-	-	-	-	-	-	-
L. Tan, Y. Lin and H. Ye, (2012). "Quality Oriented Software Product Line Architecture Design,"	-	-	-	-	-	-	x
T. Mens, J. Magee and B. Rumpe(2010). "Evolving Software Architecture Descriptions of Critical System"	x	-	-	-	-	-	-
P. Abrahamsson, M. Ali Babar and P. Kruchen(2010). "Agility and Architecture: Can They Coexist"	-	-	-	x	-	x	-
A.Tang and M. F. Lau (2014) "Software architecture review by association"	-	-	-	-	-	x	-
H.Vlieta and A.Tang(2016) "Decision making in software architecture"	-	-	-	-	-	x	x
P.Y. Reyes-Delgado, M. Mora,H. A. Duran-Limon, L. C. Rodríguez-Martínez,R. V. O'Connorn and R. Mendoza-Gonzalez,(2016)	-	-	-	-	-	-	-
W. Hasselbring (2018) "Software Architecture: Past, Present, Future"	-	x	-	-	-	x	x
S. Orlov and A. Vishnyakov (2017) "Decision Making for the Software Architecture Structure Based on the Criteria Importance Theory"	-	-	-	-	-	x	-
Nitin Upadhyay (2016) "SDMF: Systematic Decision-making Framework for Evaluation of Software Architecture"	-	-	-	-	-	x	-
T. Kim, S. Yeong-Tae, L. Chung and Dung T. Huynh (2015). "Architecture Analysis: A Dynamic Slicing Approach"	-	-	-	-	-	-	-

M. Razavian, B. Paech and A. Tang (2018). "Empirical Research for Software Architecture Decision Making: An Analysis"	-	-	-	-	-	x	-
E. J. Eichwald, E. C. Lustgraaf, & B. Wetzel, (1999). "Transfer of the White Graft Reaction"	-	-	-	-	-	-	-
C. Manteuffel, P. Avgeriou and R. Hamberg (2018) "An exploratory case study on reusing architecture decisions in software-intensive system projects"	-	x	-	-	-	x	-
P.Bengtsson, (1999). "Software Architecture-Design and Evaluation."	-	-	-	-	-	-	-
A. Sharma, M. Kumar and S. Agarwal (2015) "A Complete Survey on Software Architectural Styles and Patterns"	-	-	-	x	-	-	-
M. Ozkaya and M. A. Kose (2018) "SAwUML – UML-based, contractual software architectures and their formal analysis using SPIN"	-	-	-	-	-	-	x
G. Vazquezab, J. Andres, D. Pacec and M Campoab (2014) "Reusing design experiences to materialize software architectures into object-oriented designs"	-	-	-	-	-	-	-
B. Jalendar, A. Govardhan and R. Emchand (2012) "Desiging code level reusable software components"	-	x	-	-	-	-	-
B. Kitchenham, S. Charters (2007) "Guidelines for performing systematic literature reviews in software engineering"	-	x	-	-	-	-	-

adopted all over the world [43]. Only 20% of the papers reviewed, showed software reuse as a major aspect of modern software development**.** More literatures reviewed discussed this subject as indicated in table 1 above, this implies that research still goes on in this area with respect to software architecture and design, as well as in the aspect of software development.

6.3. Software Management

Software management refers to the art and science of leading and planning software projects. From the analysis carried out, software management covers 8.6% of the relationship between software architecture and design and other parameters considered. Like software evolution, the aspect of software management was not widely discussed in literatures reviewed. Substantial research is therefore made in software architecture and design in this area.

6.4. Software Development Life cycle

Software Development Life Cycle (SDLC) gives a description of the development process of a system from the initial study until the time it is updated or replaced. There are six steps that make up the SDLC [44]. The major function of the SDLC is to neatly lay out the process of system development. Despite being a popular and well-discussed topic in practice and theory. 8.6% of the reviewed papers discussed the topic. Unlike other topics, this is a major aspect of software, from the reviewed literature there was no substantial amount of discussion made in this aspect**.**

6.5. Software Evaluation

Comprehensively, non-systematic checklists can be applied to a program in the software evaluation process [45]. In recent times, software assessment using theory-based approaches which incorporates relevant criteria derived from psychological, linguistic and pedagogical models of language learning and teaching has been proposed. 11.4% of the 35 articles reviewed, discussed the importance of software evaluation in developed systems. There has been current research going on in the aspect of software evaluation, hence making it one of the major discussed topics from table 1 above.

6.6. Stakeholder Involvement/ Decision Making

The nature of the design problem also determines the form of decision that will be made. As reported in [46], a structured design problem makes the decision-making process better and easier. In this analysis, 34.3% of the articles reviewed, considered stakeholder involvement and in some cases, decision-making as a factor in software architecture and design, making it the most emphasized topic. This had the highest number of literatures discussing the topic, this implies that currently, more researchers are gearing towards this aspect during the course of their research to emphasize the need of stakeholders' involvement during software development.

6.7. Design Quality

If the quality of a design is not considered properly, it could lead to a negative impact on the product being developed [47]. The quality of any software is dependent on how well it conforms with the design plan of that product, it determines if the product would deliver the requirements desired properly and efficiently. 17.1% of the papers used for analysis discussed this topic, either directly or by evaluation of some design quality factors like quality attributes or design decisions. Also, this was among the topics discussed substantially in the selected literatures, which makes design quality a high recommended aspect of software architecture and design both in terms of academics as regards research and also in the industry.

From the results, the most discussed topics gotten from the analysis of selected literatures were stakeholder involvement and design, software reuse, design quality and software evaluation while the least discussed topics include software management, software evolution and software development life cycle.

7. Conclusion

Software architecture and design is an important component in the software engineering field. For success in the software engineering field both the architecture and design of software must be considered. Hence, various fundamental topics as regards software architecture and design have been analyzed.

The objective of this paper was to critically analyze current topics in software architecture and design. The method of analysis

adopted was the collection of published papers and articles on the topics discussed in the paper and the percentages of each fundamental topic was calculated. From the analysis, the result showed that, of 35 papers used in analysis, 34.3% discussed stakeholders' involvement and decisions, 17.1% for design quality, 20% examined software reuse while 11.4% discussed software evaluation and 8.6% of papers reviewed discussed software management, evolution and software development life cycle each.

From the analysis, it can confidently be concluded that aspects of software architecture and design such as software evolution, management, re-usability and building software which are fault tolerant, reliable, backward compatible, maintainable and secured are under-discussed. Several authors addressed various aspects of software architecture and design, but there are no standard procedures to follow that addresses all issues of concern in general software developments. As earlier stated, some factors should be non-negotiable in software development to ensure standardization thereby reducing incessant collapse of systems witnessed in the early years of software developments. This research stressed the significance and rigorous work involved in the development of software and outlined major factors that should be considered.

Therefore, it is important to note that a critical and rigorous analysis of software architecture and design is required to overcome the overall failure or crash of software in software development process and to also identify relevant gaps in the architecture and design styles or methods. Software architecture and design as an ever-growing field of software engineering, calls for further analysis to test and validate principles as they evolve. This study would help other researchers in the quest of knowing more about software development and the need to research further on the least discussed topics which are software management, software evolution and software development life cycle (SDLC).

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Paul, F. Bachmann, L. Bass, D. Garlan, J. Ivers, R. Little, P. Merson, R. Nord, J. Stafford, (2010). "Documenting Software Architectures: Views and Beyond", Second Edition. Boston: Addison-Wesley. ISBN 978-0-321-55268-6, pp. 592.

[2] D. E. Perry, A. L. Wolf, (1992). "Foundations for the study of software architecture". ACM SIGSOFT Software Engineering Notes. 17 (4): CiteSeerX 10.1.1.40.5174. doi:10.1145/141874.141884, pp. 40.

[3] "SoftwareArchitecture".(2018), https://www.sei.cmu.edu/educationoutreach/courses/course.cfm?coursecode=p34. Retrieved 2018-07-23.

[4] B. Len, P. Clements and R. Kazman, (2012). "Software Architecture in Practice", Third Edition. Boston: Addison-Wesley. ISBN 978-0-321-81573-6.

[5] M. Fowler, (2003). "Design – Who needs an architect?". IEEE Software. 20 (5): 11–44. doi:10.1109/MS.2003.1231144

[6] F. Peter and D. Hart, (2004). "A Science of design for software-intensive systems". Communications of the ACM. 47 (8): 19–21 [20]. doi:10.1145/1012037.1012054.

[7] A. Davis, (2004)" Principles of Software Development", McGraw-Hill, Inc. New York, NY, USA, ISBN:0-07-015840-1.

[8] J. Spacey, (2017) "Software Design Vs Software Architecture.", https://simplicable.com/new/software-design-vs-software-architecture , Retrieved 2018-07-23.

[9] E. Novoseltseva, (2016). "https://apiumtech.com/blog/15-benefits-of-software-architecture/", Retrieved 2018-07-23.

[10] R.Kazman, C.-H. Lung, s. Bot, and k. kalaichelvan, (1997) " an approach to software architecture analysis for evolution and reusability", center for advanced studies conference, pp. 144-154.

[11] M.Shaw, and P. Clements, (2006)" the golden age of software architecture: a comprehensive survey.", institute for software research international school of computer science. Camegie mellon university, Pittsburgh, PA,0740-7459/06/.

[12] A.Alhar, M. Liaqat and F.Azam, (2016). "A Comparative Analysis of Software Architecture Evaluation Methods" Journal of Software Vol 11, No. 9. DOI:10.17706/JSW.11.9.934-942.

[13] N. Hämäläinen, J. Markkula, T. Ylimäki and M. Sakkinen, (2006)"Success and Failure Factors for Software Architecture". Available: https://jyx.jyu.fi/bitstream/handle/123456789/41384/Article_Success_and_Failure_Factors_for_SA.pdf;sequence=1 , Retrieved: 2018-08-21.

[14] M. Mekni, G. Buddhavarapu, S. Chinthapatla and M. Gangula, (2018) "Software Architectural Design in Agile Environments," Journal of Computer and Communications, 6, pp. 171-189. Available: https://doi.org/10.4236/jcc.2018.61018 .

[15] I. Lytra, G. Engelbrecht, D. Schall and U. Zdun, (2015) "Reusable Architectural Decision Models for Quality-driven Decision Support: A Case Study from a Smart Cities Software Ecosystem". Available: https://eprints.cs.univie.ac.at/4330/1/paper.pdf . Retrieved: 2018-08-21.

[16] M. P. Robillard (2016). "Sustainable Software development,". Available: https://www.cs.mcgill.ca/~martin/papers/fse2016.pdf . Retrieved: 2018-08-21.

[17] l. Dobrica, and E. Niemela, (2002) "a survey on software architecture analysis methods", IEEE Transactions on software engineering, vol 28, no. 7. Available: https://www.researchgate.net/publication/3188246. Retrieved: 2018-07-23.

[18] Antony Tang and Man F. Lau, (2014) "Software architecture review by association", Journal of Systems and Software", Volume 88, pp 87-101, ISSN 0164-1212.

[19] P.Y. Reyes-Delgado, M. Mora, Hector A. Duran-Limon, L. C. Rodríguez-Martínez, R. V. O'Connor and R. Mendoza-Gonzalez, (2016) "The strengths and weaknesses of software architecture design in the RUP, MSF, MBASE and RUP-SOA methodologies: A conceptual review, Computer Standards & Interfaces", Volume 47, pp 24-41, ISSN 0920-5489.

[20] "Software architecture and design tutorial," Tutorialpoints, 2019.[Online]. Available: https://www.tutorialspoint.com/software_architecture_design/index.htm. Retrieved:27-Feb-2019.

[21] R. J. de Paula, V. Falvo and E. Y. Nakagawa (2016), "Architectural Patterns and Styles" [Online]. Available:https://edisciplinas.usp.br/pluginfile.php/977101/course/section/268862/S4-%20Architectural%20Patterns%20and%20Styles.pdf.

[22] S. Boyd, M. D'Adamo, C. Horne, N. Kelly, D. Ryan and N. Tsang (2013), "SOFTWARE ARCHITECTURAL STYLES" SENG 403-W2013 Paper Project (Group4) [Online]. Available:http://kremer.cpsc.ucalgary.ca/courses/seng403/W2013/papers/04ArchitectureStyles.pdf

[23] D. Ameller, X. Burgués, D.Costal,C.Farré, X. Franch, (2018) "Non-functional requirements in model-driven development of service-oriented architectures"Science of Computer Programming,Volume 168, Pp 18-37.

[24] G.Rodríguez, J. A. Díaz-Pace and Á.Soria (2018) "Information Systems A case-based reasoning approach to reuse quality-driven designs in service-oriented architectures", Information Systems Volume 77, pp 167-189.

[25] "What is service-oriented architecture (SOA)? - Definition from WhatIs.com," SearchMicroservices, 2014. [Online]. Available: https://searchmicroservices.techtarget.com/definition/service-oriented-architecture-SOA. Retrieved: 13-Dec-2018.

[26] "SOA (Service Oriented Architecture) Principles," Meet Guru99 - Free Training Tutorials & Video for IT Courses, 2018. [Online]. Available: https://www.guru99.com/soa-principles.html. Retrieved: 13-Dec-2018.

[27] S. Pulparambil and Y. Baghdadi (2018) "Service oriented architecture maturity models: A systematic literature review", Computer Standards & Interfaces, Volume 61, pp. 65-76.

[28] "Advantages and Disadvantages of Service-oriented Architecture (SOA)" Techspirited, 2018. [Online]. Available: https://techspirited.com/advantages-disadvantages-of-service-oriented-architecture-soa. Retrieved:27-Feb-2019.

[29] S. Hussain, J. Keung, and A. A. Khan, (2017). "Software design patterns classification and selection using text categorization approach". Applied Soft Computing Journal, vol.58, pp. 225–244. http://doi.org/10.1016/j.asoc.2017.04.043

[30] "Design Patterns and Refactoring", (2018). [Online]. Available: https://sourcemaking.com/design_patterns. Retrieved: 13- Dec- 2018.

[31] J.L. Barros-Justo, B.V.B. Fabiane, L.C. Ania, (2018) "Software patterns and requirements engineering activities in real-world settings: A systematic mapping study", Computer Standards & Interfaces 58, pp. 23-42.

[32] J.O. Coplien, N.B. Harrison, (2005) "Organizational Patterns of Agile Software Development", Pearson Prentice Hall, ISBN-13: 978-0131467408.

[33] E. Gamma, R. Helm, R. Johnson and J.Vlissides , (1995) " Design Patterns: Elements of Reusable Object-oriented Software". Boston: Addison-Wesley Longman Publishing Co., Inc.

[34] S. Hussain, J. Keung, M. K. Sohail, A. A. Khan, and M. Ilahi (2019). "Automated framework for classification and selection of software design patterns". Applied Soft Computing Journal, vol.75, pp.1–20. http://doi.org/10.1016/j.asoc.2018.10.049

[35] "Design Patterns | Set 1 (Introduction) - GeeksforGeeks", GeeksforGeeks, (2018). [Online]. Available: https://www.geeksforgeeks.org/design-patterns-set-1-introduction/. Retrieved: 13- Dec- 2018.

[36] B. Bafandeh Mayvan, A. Rasoolzadegan and Z. G. Yazdi (2017) "The state of the art on design patterns: A systematic mapping of the literature", Journal of Systems and Software, Volume 125, pp. 93-118.

[37] J. Kress, B. Maier, H. Normann, D. Schmeidel, G. Schmutz, B.Trops, C. Utschig-Utschig and T. Winterberg. (2014). "SOA and Cloud Computing", https://www.oracle.com/technetwork/articles/soa/ind-soa-cloud-2190513.html Retrieved: 2018-01-09

[38] A. Birukou, (2010) A survey of existing approaches for pattern search and selection, in: Proceeding of PLoP DISI - University of Trento, Italy, ACM 978-1-4503-0259-3.

[39] P. Velasco-Elizondo, R. Marín-Piña, S. Vazquez-Reyes, A. Mora-Soto and J. Mejia, (2016), Knowledge representation and information extraction for analyzing architectural patterns, Sci. Comput. Program. Vol. 121 pp. 176–189.

[40] 2018. [Online]. Available: https://www.quora.com/What-are-some-pros-and-cons-of-using-Design-Patterns-to-describe-your-business-model. Retrieved: 2018-12- 13.

[41] D. K. Kim, L. Lu, and B. Lee, (2017). "Design pattern-based model transformation supported by QVT". Journal of Systems and Software, vol.125, pp.289–308. http://doi.org/10.1016/j.jss.2016.12.019

[42] Tutorials.point. "Software engineering overview" available:https://www.tutorialspoint.com/software_engineering/software_engineering_overview.html, Retrieved: 2018-12-13

[43] C.W. Krueger, (1992) "Software reuse", School of Computer Science, G'arnegie Mellon University, Pittsburgh, Pennsylvania. Available: https://dl.acm.org/citation.cfm?doid=130844.130856

[44] En.wikibooks.org, (2012) "Computer Revolution". [Online]. Available: https://en.wikibooks.org/wiki/The_Computer_Revolution/MIS/SDLC. Retrieved: 2018-11-2.

[45] IGI Global, (2012) "Software Evaluation". [Online]. Available: https://www.igi-global.com/dictionary/software-evaluation/27677 Retrieved: 2018-11-3.

[46] C. Zannier, M. Chiasson and F. Maurer, (2007) A model of design decision making based on empirical results of interviews with software designers. Inf. Softw. Technol. Vol. 49 (6), pp. 637–653.

[47] L. Tan, Y. Lin and H. Ye, (2012). "Quality Oriented Software Product Line Architecture Design," Journal of Software Engineering and Applications, vol. 5, pp. 472-476, [Online]. Available: http://www.SciRP.org/journal/jsea

[48] A. Ramírez, J. R. Romero and S. Ventura, (2018)" Interactive multi-objective evolutionary optimization of software architectures" Information Sciences, Vol 463–464, pp 92-109 Available: www.elsevier.com/locate/ins

[49] O. Sievi-Korte, S. Beecham and I. Richardson, (2018)"Challenges and recommended practices for software architecting in global software development" Information and Software Technology, In press, Available: www.elsevier.com/locate/infsof

[50] D. Garlan, (2000), "software architecture", School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA.

[51] S. B. chung-Horng Lung, (1997) "An Approach to software Architecture Analysis for Evolution and Reusability," in Centre for Advanced Studies Conference, Toronto, Canada.

[52] V. S. Rekha and H. Muccini, (2018) "Group decision-making in software architecture: A study on industrial practices", Information and Software Technology, Vol. 101, pp. 51-63, ISSN 0950-5849.

[53] M. Unterkalmsteiner, T. Gorschek, R. Feldt and E.riks Klotins, (2015) Assessing requirements engineering and software test alignment—Five case studies, Journal of Systems and Software, Vol. 109, pp. 62-77, ISSN 0164-1212.

[54] H. Vlieta and A.Tang (2016)."Decision making in software architecture", Science Direct-Journal of Systems and Software,Vol 117,pp. 638-644, https://doi.org/10.1016/j.jss.2016.01.017 .

[55] A. Tang and M. F. Lau (2014). "Software architecture review by association", Journal of Systems and Software, Vol 88, pp.87-101, https://doi.org/10.1016/j.jss.2013.09.044 .

[56] T. Mens, J. Magee and B. Rumpe (2010). "Evolving Software Architecture Descriptions of Critical System", IEEE Computer Society, vol 43, pp. 42-48. DOI: 10.1109/mc.2010.136

[57] P. Abrahamsson, M. Ali Babar and P. Kruchen (2010). "Agility and Architecture: Can They Coexist", IEEE Computer Society, pp16-22.

[58] S. Orlov and A. Vishnyakov (2017). "Decision Making for the Software Architecture Structure Based on the Criteria Importance Theory", Procedia Computer Science, Vol 104, pp. 27-34, ISSN 1877-0509.

[59] Nitin Upadhyay (2016). "SDMF: Systematic Decision-making Framework for Evaluation of Software Architecture", Procedia Computer Science, Vol 91, pp. 599-608, ISSN 1877-0509.

[60] T. Kim, S. Yeong-Tae, L. Chung and D. T. Huynh (2015). Architecture Analysis: A Dynamic Slicing Approach. Dept. of Computer Science University of Texas at Dallas https://doi.org/10.1017/CBO9781107415324.004

[61] M. Razavian, B. Paech and A. Tang (2018). "Empirical Research for Software Architecture Decision Making: An Analysis", Journal of Systems and Software, ISSN 0164-1212.

[62] C. Manteuffel, P. Avgeriou and R. Hamberg (2018). "An exploratory case study on reusing architecture decisions in software-intensive system projects", Journal of Systems and Software, Vol 144, pp. 60-83, ISSN 0164-1212

[63] E. J. Eichwald, E. C. Lustgraaf, & B. Wetzel, (1999). Transfer of the White Graft Reaction. Proceedings of the Society for Experimental Biology and Medicine, vol. 126(3), pp. 619–620. https://doi.org/10.3181/00379727-126-32521

[64] P.Bengtsson, (1999). Software Architecture-Design and Evaluation.University of Karlskrona PerOlof Bengtsson Department of Software Engineering and Computer Science. https://pdfs.semanticscholar.org/6c1c/a7056fbb2ee1f82f04c3ae2b7b7e16f41c6c.pdf

[65] A. Sharma, M. Kumar and S. Agarwal,(2015) "A Complete Survey on Software Architectural Styles and Patterns," Procedia Computer Science, Vol. 70, pp. 16-28 https://doi.org/10.1016/j.procs.2015.10.019

[66] M. Ozkaya and M. A. Kose, (2018) "SAwUML – UML-based, contractual software architectures and their formal analysis using SPIN,"Computer Languages, Systems & Structures,Vol 54, pp. 71-94, https://doi.org/10.1016/j.cl.2018.04.005

[67] G. Vazquezab, J. Andres, D. Pacec and M. Campoab, (2014), "Reusing design experiences to materialize software architectures into object-oriented designs,"Information Sciences Vol. 259, pp. 396-411 https://doi.org/10.1016/j.ins.2010.03.013

[68] B. Jalendar, A. Govardhan and R. Emchand, (2012) "Desiging code level reusable software components", International Journal of Software Engineering & Applications, Vol. 3, n. 1, pp. 219-229.

[69] B. Kitchenham and S. Charters, (2007) "Guidelines for performing systematic literature reviews in software engineering", EBSE,Software Engineering Group, School of Computer Science and Mathematics, Keele University, Keele, Staffs, ST5 5BG, UK and Department of Computer Science, University of Durham, Durham, UK .

[70] W. Hasselbring (2018). "Software Architecture: Past, Present, Future in: Gruhn V., Striemer R. (eds) The Essence of Software Engineering". Springer, Cham. https://doi.org/10.1007/978-3-319-73897-0_1]

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 212-220 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Providing Underlying Process Mining in Gamified Applications – An Intelligent Knowledge Tool for Analyzing Game Player's Actions

Anna Tatsiopoulou*, Christos Tatsiopoulos, Basillis Boutsinas

Management Information Systems & Business Intelligence Laboratory, Dept. of Business Administration, University of Patras, 26504, Greece

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 17 July, 2019
Online :30 July, 2019

Keywords:
Semantics
Knowledge mining
Extraction
Concepts
Gamification
Education
Sequential pattern mining
Process extraction
Adoptable knowledge
Management
Behaviour prediction
User-centered prediction

ABSTRACT

This work deals with the issue of understanding a user's behaviour as this is expressed via a gamified application. The notion of ontologies and the association of concepts in relevance to decisions that have to be made is used. The current work introduces a new process-based approach, based on collected large log files and associations of underlying decisions based on them. Both of them deal with work in extracting information for intelligent use, the main difference being that the first discovers but stops on a concept relation basis, while the other based on processes, as knowledge transactions, further to the associations on a 1:1 level maybe applied on a multi associative model. The objective of the current work is to introduce the methodology into gamified environments (such as but not limited to) games, for semi-automated understanding of user behaviour and furthermore, prediction and in instances, guidance via optimal paths of decision making activities, that are useful in gamified applications in various areas like the education. Both the initial ontology based, and the extension work on it, are based on mining association rules, in one instance treated as knowledge nodes (concepts) and in the second as underlying knowledge processes, based on big log files. This may be applicable to online games, that generate big log files of user selections, that are available for study and examination for extracting user behavioural patterns. As a result, maximum length of sequential patterns and items in them, are discovered in an algorithmic based methodological approach, providing in this way a set of guidelines for designing gamified applications.

1. Introduction

The work in this paper is based on the work presented in the 9th International Conference on Information, Intelligence, Systems and Applications (IISA 2018) under the title "On Ontologies and Knowledge Associations in Gamified Environments" [1]. Its contribution is bifold (a) it extends on a more generalized manner the work so far in the above and (b) it is getting additional research by providing an additional mechanism for knowledge associations. More specifically, the case (a) above is mainly based on single concepts relations – associations and its application in mobile devices in the tourist domain, while in the part (b) contribution of additional research the notion of knowledge transaction is introduces as a set of serially associated concepts – nodes, with the presentation of its application in a gamified environment (on line game). For clarity purposes and the distinction of concepts within the domain of gamification vs. the concepts in the game domain a brief introduction is provided below.

2. Gamification – Modeling User Behaviour

Gamification is the process of applying game techniques (e.g. dynamics, feedback, badges, leveling, etc.) to non-gaming contexts [2]. It is a new methodology that flourished just after 2008 [3] and refers to understanding a user's behaviours on making decisions, made in a context of gaming and hence drive them to expected outcomes. As a process, gamification is under study within the scope of computer science, mathematics, artificial intelligence as well as psychology disciplines in the context of persuasive technology [4]. It is important to be noted here the difference between gamification and gaming itself [5]. The evolving processes in gamification are used today in various domains like education, training, marketing, health human

*Corresponding Author Anna Tatsiopoulou, & email: atatsio@upatras.gr

resources in enterprises, e-commerce and so on [6]. Understanding thus the above and applying them on involved processes on the above sectors, is of great importance. In the current work we will go through the gamification-based implications in the context of the designing educational content for computer-based learning and applications.

Designing and implementing gamification applications is an iterative process which consists of six main steps: 1) define learning objectives, 2) delineate target behaviours, 3) describe the players (either students and/or teachers), 4) device activity cycles, 5) implement fun and involve psychological factors that affect learning in it and finally, 6) deploy the appropriate tools. This process is performed so far by experience, by playtesting the designed gamification schemas, examine the corresponding KPIs, check what is working or not and in case that failures of the design and implementation are identified, return to the design phase and start the process all over [7].

Here, we attempt to formally introduce, as gamification is an emergent technology, the design of gamification applications, by analysing game player's actions stored in log files, to the domain of education as a process-based methodology. We try in other words, to study the user engagement into the knowledge acquisition cycle, in a domain agnostic way, based on a semi-formal approach, where, beneath it, underlies the concept of positive and negative reinforcement via the semantics, the mechanics and dynamics that the gamification-based applications are characterized of.

Education and gaming have many in common. Some of them are learning on achieving objectives, introduction and use as much as possible of engagement (interactivity in learning process) concept, motivation in achieving the previous, game elements, like positive and negative reinforcements, awarding and/or discouraging in specific cases and so on.

The above provided, our research was driven by the interest in studying what is a very important issue in the above context, i.e., what are the implications and the impact of the possible social and psychological factors that are being introduced by the gamification methodology to the achieving of the learning objectives. There are certain questions that have to examined. For example, is the educational content being acquired and understood better if the subject is working in such an application alone, or in an antagonistic environment (one to one or one to computer), or finally in a cooperative team-based mode? Based on this, how social media and collaboration-based environment (electronic) assist this process? Should the subject "discover" the knowledge himself, in a role-based game manner, or should be driven/directed to it? If yes, in a "strict" or a more "flexible" way?

To study the above we introduce the concept of underlying processes in the gamification application that supports the learning process. We are going to examine the model of an educational application that underneath it exists a mechanism that discovers patterns in learning and at the same time tries to understand the individual's behavioural patterns and foresee future behaviour (knowledge extraction). This mechanism gets its input from a set of log files of collected raw data, that is generated from playing a respective game.

Log files in most cases, contain structured raw data that are coming and are registered as text files, coming directly from systems' interaction activities. This either comes from the user's or from other systems' components themselves activity or both them. In most cases, this raw data, comes in a structured and in a predefined format, while in other cases, due mainly to some external reasons and abnormalities, comes as incomplete data, introducing thus "noise" to the system. Out of these log files, the important aspect and the objective of the current work is to mine hidden knowledge, that can be characterized as knowledge produced during the game play (knowledge in action, while learning), in terms of underlying processes and patterns. Such concept of getting an advanced level of knowledge out of raw data, is also currently under both research and application, known as Big Data Analysis. To have such a repository of real time produced raw data in terms of log files that registers user behaviour in a game is in this step our objective and part of presenting our design methodology, in terms of underlying techniques.

More specifically, for experimentation purposes, we have selected and dealt with log files, that are being continuously generated out of the context of a game, played worldwide on the WEB 24/7/365, by thousands of players, in real time. Therefore, the generated log files are updated, using a generalized structured data model, that in many cases has the ability to expand or to reduce itself, in a dynamic way, as it will be presented in the following sections. Due to these idiomatic data type characteristics, initial and advanced parsing techniques had to be developed for two major reasons (a) to generate clear, coherent and consistent data sets and (b) to transform them, so to be persistent and hence to be stored and aggregated in a database for further exploration and processing. What we will concentrate on also here, is the discovery of interesting and meaningful patterns that signify the underlying existence of processes, towards a predictable and modifiable behaviour, a major research topic in the field of artificial intelligence [8].

3. Ontology Based Knowledge Discovery - The ONARM+ Methodology

In the previous presented work [1], we have introduced both the theoretical as well as the technical background of the ONARM+ methodology. Based on this, we have also presented in detail an application, in the domain of tourism. For this, we have considered the case of independed users that by potentially using our platform would exchange parts of their personal concepts, based on a centralized tourism [9] concept ontology, in a semi-automated and intelligent way. The scenario to be implemented based on our methodology, required 2 major components: (a) a client component named Concept-Net, that could be able to run either on mobile and/or desktop environments and (b) the Concept-Net Back End set of modules. Both of them and their subcomponents have been implemented as the following conceptual diagrams represent them (Figure 1):

Our work dealt on the concept relevance aspects with to the transformation of raw data obtained, via user selections and decision-making actions, in a specific domain (being domain agnostic), its comparisons with others on a concept basis, represented as an ontology of a specific domain.

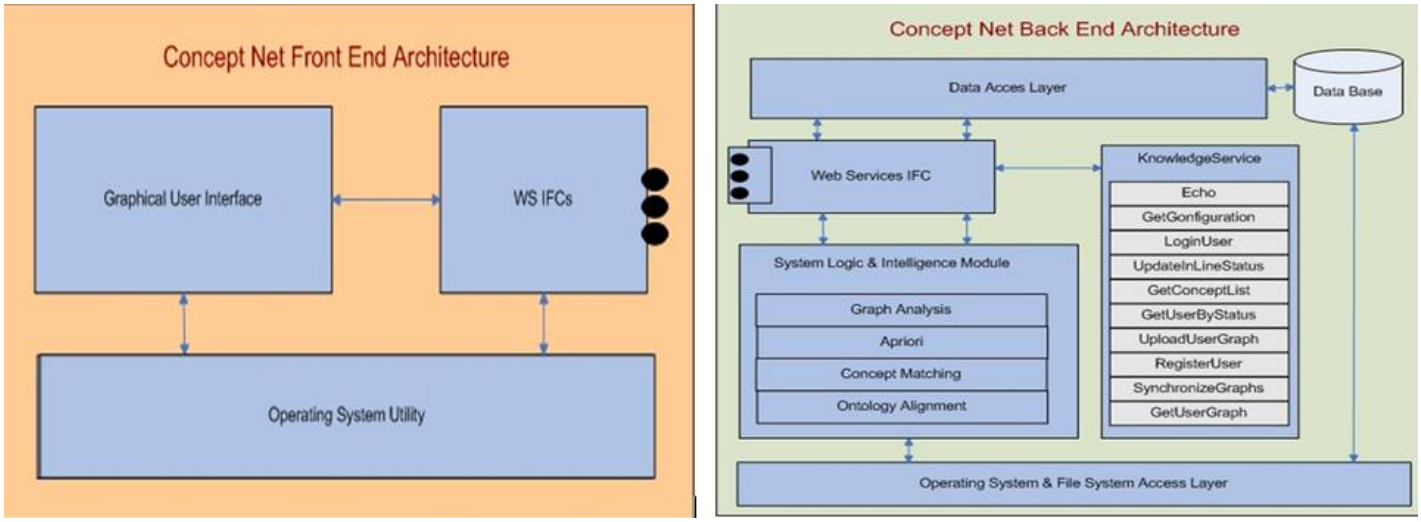


Figure 1: Left: (a) Client logical diagram, on the right (b) server-side components of the Concept-Net application

In short, during the process of knowledge extraction between the 2 users, firstly the discovery of concepts that are of common interest has to take place and then, the comparison via measures while on the final step, the merging of the parts of the specific ontology parts of each user to specific place of the ontology (aligning of them [10], [11]) takes place.

The methodology works under the assumption of existence of at least 2 user profiles, in terms of ontologies of interest. These represent the user's preferences and interests in a specific area and the categorization in a conceptual way that is depicted in a graphical way via the Protégé tool (The Protégé Ontology Editor, http://protege.stanford.edu /, last accessed on June 2, 2018). Also, there exists a central ontology that holds all concepts in the specific domain. As described in the work [1], the methodology is based on the Apriori algorithm [12], to identify. In general, the three major phases of the ONARM+ methodology are: (a) knowledge acquisition, (b) knowledge representation and (c) knowledge exchange.

While the association of concepts in terms of independent knowledge nodes, is of major significance in the discovery and exchanging knowledge processes, when we strive for understanding decision making in terms of a series of choices, then the above provide only partial information. To add value to the previous and essentially to introduce similarities and therefore predict more accurate user choices, we have introduced as an extension to this, the methodology below, which identifies similarities among processes for decision-making, in terms of generalized knowledge transaction processes. This is particularly useful, when not only we need to extract such information so to find similarities among users' profiles, but also equally significantly, to predict such patterns and in a second step, to be able to guide them, so that to trigger dynamic choices for desired user behaviour driving into.

4. Process based Knowledge Transactions Pattern Discovery

4.1. General

In this part, we go through our methodology and we provide an example for supporting the process of designing and development of educational applications based on gamification by using it. For this, we use knowledge extraction methodologies for extracting users' behavioural sequential patterns from large on-line game log files and propose to learners an optimal path or set of actions in a specific order, towards knowledge acquisition. Large game log files stored on line, are collected, processed, stored in an event data base and then transformed to appropriately format, for applying Apriori algorithm sequential pattern mining technique. Extracted patterns then are translated to set of proposed actions. The future path of user preferences in terms of decisions, can be (a) reinforced and (b) predicted by our methodology, while step by step educational/game objectives are achieved. Therefore, designers of educational gamification applications are provided with a set of means to select appropriate type of game mechanics/dynamics and offer selected action types in a user-cantered approach tailored to any individual's requirements. Following advancements in games based on electronic media and the advanced level of new generation familiarization with such environments, in conjunction with the fact that gamification has a wide application in education domain, as it is a useful mean for maintaining student participation and enhance engagement in the educational process, we focus on frequent sequential patterns extraction from game log files and transform them to powerful tools for implementing gamified educational applications.

4.2. Background

Knowledge is the ultimate resource in every enterprise's routine operations and signifies the holy grail to their success. Knowledge in human beings is better understood but not equally explained. Recent developments and technological advancement in all fields of IT technologies (hardware and software) generate on a continuous manner, on real time basis to all kind of unrelated at a first glance data, coming from simple to very complex applications and transactions. The rising need is therefore to discover underlying processes and provide meaning to them as upper level knowledge.

On the other hand, following this trend of advancements, games based on electronic media have evolved from simple applications in the very beginning, to very complex environments, many of them simulating real life experiences and scenaria. In the

domain of education, gamification has a wide application since it has been proved to be a useful mean for maintaining student participation and enhance their engagement in the educational process. Gamification is a mean that is widely used in the marketing domain targeting to foster both brand loyalty and awareness as well as to enhance customer engagement. As mentioned previously, these three key marketing concepts are relevant in the gamification context: engagement – "high relevance of brands to consumers and the development of an emotional connection between consumers and brands", brand loyalty – "the relationship between relative attitude and repeat patronage" and brand awareness, "the rudimentary level of brand knowledge involving, at the least, recognition of the brand name" [13]. Transforming these to learning processes in education, we have the students that are the consumers and the emotional connection to branding in our case the sources of educational context (humans, books, eBooks, web, Internet in general, etc). Therefore, the students can be seen as "highly engaged, emotionally driven players", in this gamified learning approach, where strategic decisions have to be made by them (intermediate learning objectives' achievements, towards the final overall knowledge acquisition, within the provided domain).

These complex gamified applications, therefore, require the track of players' actions and the estimation – calculation of potential responses to their decisions in an intelligent way. In recent years, the tracking and logging of this information and monitoring of the way that game engines and applications perform – has become widespread in the digital entertainment industry, providing sometimes very detailed and specific information on the performance of popular commercial titles with millions of players and installations in digital devices [14],[15]. It has been widely accepted in the academic and business fields that the so-called game telemetry can be a very powerful tool not only for game development but also to assist a variety of stakeholders via user modeling, behavior analysis, matchmaking and playtesting, [16] under the condition that it is processed in the right way. As a consequence, large amounts of game telemetry data are tracked, logged and stored, but not always is being correctly analyzed and used for drawing results and decisions useful to other applications. The challenge quickly became to deal with working with large-scale data, in a process based meaningful way, for extracting behavioral decisions [17]. Quickly, new methods have emerged [15] to assist analysts and decision makers to obtain the information they need to make better decisions. These included: automatic summarization of data, the extraction of the essence of the stored information, and the discovery of patterns in raw data. When datasets became very large (any dataset that does not fit into the memory of a high-end PC as large-scale, i.e. several GB and beyond) and complicated, the breakdown method to smaller sets of data was introduced. Applying on these raw data knowledge discovery processes will help identify and uncover patterns of behavior in it, whether user-derived or learning based-derived, and these, in turn, can be highly valuable.

In this work, we use sequential pattern mining to analyse game player's actions and behaviours, stored in log files.

The objective of the sequential pattern matching process is to discover patterns that take place in sequence(s), so to be used in prediction of future events that generate similar data. It has a great application domain, where human behaviour analysis plays a key role in making strategic role, that varies from students' behaviour and its impact in achieving learning objectives, customer analysis behaviour (and therefore prediction) up to security and alarm analysis, manufacturing engineering, WEB pages' prefetching and so on. In the sequential pattern matching a sequential database of records, expressed as patterns of events is provided to be analysed. The input must be sorted in absolute lexicographical order for the algorithm. The output is a subset of patterns that frequent occur within it, provided that a minimum support (at least), is entered as an expected indicator of the least desired frequency of occurrences of the patterns to be discovered within the total records of the database.

Historically, AprioriAll is the first sequential pattern mining algorithm. The authors of AprioriAll then proposed an improved version called GSP. The AprioriAll and GSP algorithms are inspired by the Apriori algorithm for frequent itemset mining [16].

4.3. Methodology Definitions and Assumptions

There are prerequisites that log files have to support to be useful for the pattern mining extraction. A non-exhaustive list of them is the "case-id", the "supporting activity", "event timestamping", "connectivity with other events, triggering or triggered by", information on resources, such the "user name", "system cost on event" and so on. These, so called event attributes, will allow the extraction of underlying processes, by associating such events and using a generalized model of event logs, consisting of cases, that in turn consist of events, consisting in turn of attributes. That introduces [18]:

1. A set of case(s) based process(es)
2. A set of events associated with processes via respective cases
3. An ordered set (tuple) of such events on a per case basis
4. A set of attributes per event (empty set allowed

As a final point to be noted, is the difficulty and the rareness of availability of user data in terms of raw data (and not in aggregated form), from the game providers, that generate and own these big data files.

At this point, the current work proposes a new methodology for discovering these meaningful patterns, using minimal human intervention (in terms of pre-processing, association and conclusions' analysis). In this section, the main steps of it are going to be presented below:

1. Data Source (Game) Selection. That was a very important step, so the proper game to be selected that conforms with the above requirements and restrictions. At the same time, the game had to offer in an open and continuous way the data being produced, in terms of log files.

2. Data Collection. Based on the fact that the previous criterion would have been met, to develop a method, that in an automatic way would be in a position to collect on a timely basis the generated data.

3. Data Maintenance. Provided the following were met, to further elaborate and develop an automatic way of synchronizing the data collection, so to avoid repeated and unnecessary steps, that

would lead in duplication of data collection. This in turn, most probably, would lead the game data provider to ban the data collection process and stop further experimentation. Unnecessary hits therefore for data collection purposes, had not only to be completely avoided, but at the same time to be carefully scheduled when to occur, avoiding Denial of Service as previously mentioned.

4. Data Transformation. This step was necessary so to introduce unique event game ids, that could be easily backwards traversed, in a meaningful way, reflecting all their attributes, that would lead to processes.

5. Data Persistence. Log files as previously explained are continuously updated in real time, as a stream of data. Out of this, chunks are not useful and have to be thrown away from mining, yet though to be identifiable in case of suspect of usefulness of it. Space and storage limitations had to be considered in the next step, being the storage in a local database.

6. Data Association. In sequence, the asynchronously generated data, has been through data cleansing and in this step, is provided in an automatic way as input to algorithms for association of it. In this step, any type of associative – rule-based algorithm could be used. It is important to be noted that since data can be easily transformed during input process to any given and expected requirement of the algorithm to be used can be applied to any type of these. For the purposes of this work the Apriori algorithm has been used [18].

7. Process Extraction via association rules. For this, the Apriori based algorithm GSP is used, with the objective to speed up the generation of interesting association rules, that will in turn lead to the mining of interesting process as proposed by [8]. The input of GSP is a sequence database and a user-specified threshold named minimum support (minsup, a value in [0,1] representing a percentage).

4.4. Technical Framework and Architecture

Based on the previous criteria, the selected game was the Alter Aeon. According to Wikipedia (https://en.wikipedia.org/wiki/Alter_Aeon), Alter Aeon was built on 1995, continuously and free provided since to thousands of people worldwide. It generates everyday thousands of lines of players' transactions, that are publicly available as streamed log files from the game developers. More details on the game and rules of playing can be found on its web site http://www.alteraeon.com.

For the specific log files, in short, the process was to download the HTML files and write them in local log files. The HTML page was created by the provider to provide the information and had to be parsed until at every entry in formation of value to be found regarding transactions. Then, the files were broken in different ways, so to match the mysqlite data base schema that was adapted. After that, the sequential files were generated and fed into the Apriori algorithm for sequential process associations.

The overall procedure from scheduled data collection from the WEB Data Source, storage of unstructured log files, transformation to structured local log files, database storage, Sequential data sets generation, Apriori as described in (Figure 2) have been implemented using Python 2.7 programming language and were tested in a distributed manner of 2 machines, connected via cloud. The first was used (a) for connection and data collection from game web site (server) of relevant data, (b) the storage of it as raw unstructured data to log files and (c) the transformation of it to local structured log files. These were implemented as a daemon set of processes, on an Ubuntu Linux server-based machine. The second part of the task was performed on a Mac Book Pro OsX with 16 GB RAM and 1 TB solid state HD.

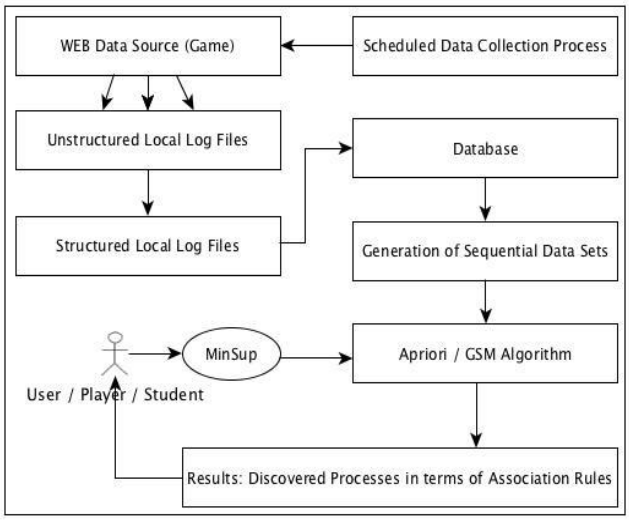


Figure 2: Overall Methodology Architecture

The separation of the overall task into two different sub tasks, in independent hardware platforms (machines), was a critical issue, since the generation of the sequential databases, as the log files were increased in size, was an intense and resources demanded issue. Additionally, the first machine has been configured to update its log files on an hourly basis (dynamically reconfigured). After this repetitive process (as many times as proper data required, here the system was collecting data for 3 consecutive months), the Apriori algorithm was applied, implemented in a Java application environment.

4.5. Results

4.5.1. Initial Results Presentation

In the diagram below (Figure 3), is shortly depicted the process for running the cleared and synthesized log files through the GSP Algorithm.

On the left-hand side of the above image, is a snapshot of the unstructured game log files, collected for a day transactional activity, from the game Web page and on the right is presented a sample of GSP process discovery outputs.

It is worth mentioned that such files were also aggregated to monthly basis and can be further aggregated to any time interval, depending on what time would like to be examined for process extraction. Initially the date / time has been registered and then event type has been recorded. Finally, on the last column, the detailed event description can be found, in a syntactically dynamic format, according the circumstances that the specific event took place.

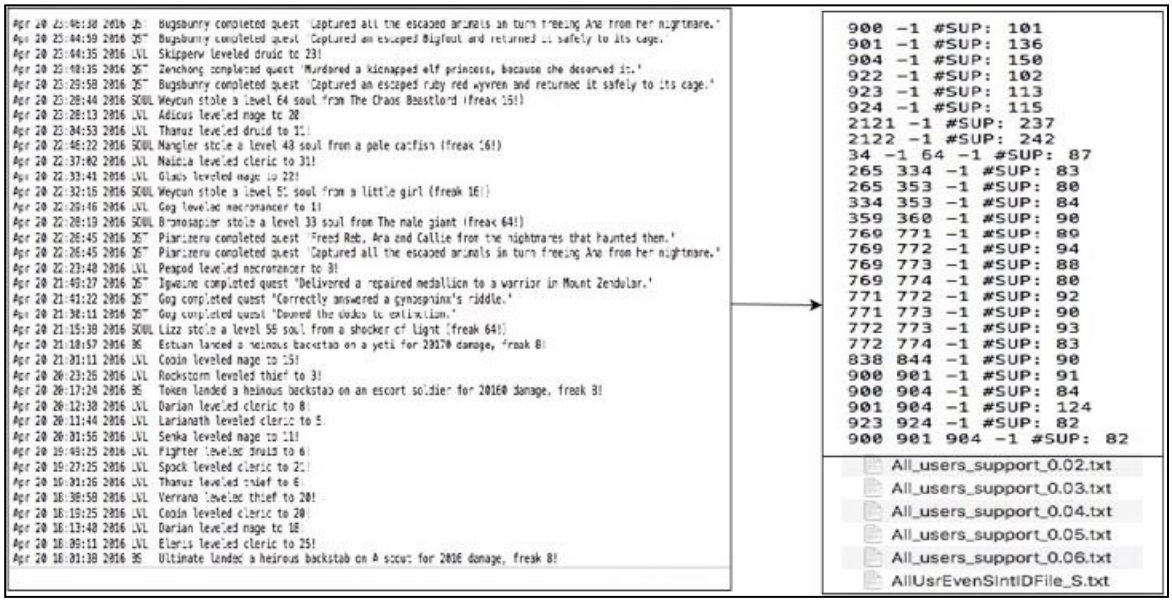


Figure 3: Initial Results Discovered by GSP algorithm.

On the right-hand side, there is (a) a sample of the process discovery after the GSP application; (b) on the lower right-hand side, the format of the filenames of the produced log files are presented, that contain the interesting associations, for processes found. It can be noted at the end of each filename, the respective minsup measure, as previously introduced, under which the process associations took place. (c) finally, on the right top, a sample of the output is presented. The (-1) stands as a separator of the frequently associated process id, that in previous steps has been uniquely composed for a player and player's activity in general to be able to be reversed tracked. As it can be seen in the sample output, there are interesting associations of frequency of 1-member, 2-member and 3-member activities. At the end, the number represents the measure of interestingness, as a measure of frequency of associated processes.

The application of the GSP Apriori implementation algorithm, provides all possible associations that could be found [12], by applying the specific minsup measure. Any sequential pattern mining algorithm could be used instead, as the VMSP [19], that works in a similar way taking the same data format as input and producing a similar file format as output, being different on the fact that provides the maximal model for the associations discovered by the sequential data files.

4.5.2. Processing of Initial Results – Extracting Behavioural Process Patterns

Sequential pattern mining is a research field studied for more than two decades, focusing now in two main areas a) Applications in the aspect of utilizing sequential pattern mining either in new applications or in new ways for existing applications and b) Algorithms either for designing more efficient algorithms or designing algorithm for handling more complex data or for finding more complex and meaningful patterns [20].

The present work contributes to the first area providing results that can be used in the design and development of new gamified applications and in the improvement of existing ones in any domain. The work performed during the experimental phase lead to the following main results:

1. The maximum length of sequential patterns varies inversely with the support value used. The finding was in line with the results stated by R. Agrawal and R. Srikant [8]

2. Sequential patterns discovered only for the events of the type "QUEST". This can be seen as the "repeated" choice of users at different "levels" to perform tasks in the game play in order to achieve certain objectives from simple to more complicated ones.

3. Sequential patterns mined were of maximum length three events for 0.02 support, of two events for support 0.03 when for support 0.04, 0.05 and 0.06 only single items were discovered, which lead to the conclusion that users for various reasons either internal (interest, resources etc.) or external (time availability etc.) are more familiar with this type of pattern.

4. Item members of two or three component sequential patterns were not present as single items in the higher support level sequential patterns mined. This fact may be interpreted and used to the design phase of the potential application(s), as a valuable hint, that helps designers combine "QUEST" elements, being at present not very popular, but of high interest, especially when associated with interaction with other facts / achievements, thus providing added value to them.

5. Single items discovered for support 0.04, 0.05 and 0.06, where not part of the maximum sequential patterns discovered with either two or three event sequence lengths, which can be seen as a cross checking point for implementing gamification strategy, where the importance is the relevance of the "QUEST" with the desired task to be performed.

The above findings are of great significance for designing gamification-based applications for behavioural prediction of users. This is because "QUEST"s can be seen as user behavioural

transactions that include events of high importance, for achieving the maximum objectives set in such applications, in any type of environments supporting well defined processes like, e.g. market(s), sales, e-learning regardless the type of gamification type applied (intrinsic or extrinsic) and so on. The current methodology overcomes the obstacle of the manual observation, association, extraction and in the end generalization, in a (semi-) manual way, based on intuitive approaches. On the contrary, the current work provides an automated algorithmic based methodological approach, for provision of a set of guidelines, to be used in the design process during the application development phase, which so far is based only on empirical findings. In the specific application selected in the current game, this is focused on using "QUEST"s, in the content of gamified applications, with the objective to drive the game user to maximal results, [6] in terms of "QUEST" usage, in a way that is familiar and transparent to him/her. Therefore, based on the fact that the maximum sequence length is 3, provides designers of gamification strategies, for the specific domain, means to select each time the appropriate "QUESTS" for achieving specific goals and drive via behavioural changes in terms of decision making in presented challenges, to the desired performance for selecting either desired actions, or products or services or education-based objectives available in a respective gamified application. On the other hand, the findings that single items gain added value when associated among other, is of great importance and can be used as guide for decision making vis-a-vis to the way that "QUESTS" are presented to the users of these specific gamified application(s).

Getting back to our case in the game selected for experimentation, as mentioned, a sample of 36.000 approximately user interactions have been collected into a respective SQLite database, categorized under 10 relevant tables of specific events types supported by the game. This data was collected on a periodic per hour and per day basis and then aggregated to a monthly based per user model. Running then the specific Apriori sequence algorithm, our methodology provided as an outcome, the conclusion that there was an interesting path of predefined processes expressed in terms of game terminology as "QUEST"s that was followed by the user as a sequence towards the game objectives. These (as all game events), have been coded as follows (processID, questID, questName):

900: q678 -> "Searched the fetid Caves of Pestilence and destroyed the weather machine"

901: q679 -> "Ruined the plans of Malkon the Plague Vampire and saved Archais from his evil machinations"

904: q682 -> "Breached the Fetid Caves of Pestilence to rescue a dwarven child"

It is important to be mentioned here, that the associations of the type (900 -> 901 -> 904) were discovered under the support measure of 0,02 and with frequency of 82 times among all possible processes interactions discovered. The number of the three (3) associated processes, is directly relevant to the sample data collected from the large log files produced by the gamified applications and they are also language/term agnostic. Finally, in the tuples discovered, the order is of significance, providing the sequence of these processes.

What this means, is that there is a significance in users' interestingness, when they use the gamified application, in getting, in a sequential way, from process 900 to 901 and to 904 by achieving their respective objectives. In other words, from the point of view of the behavioural analysis, the future path of potential user preferences getting from a point A to a point B, in terms of decisions, that can be achieved through multiple ways, can be (a) reinforced and (b) predicted by our methodology, while step by step game objectives are achieved.

This is a very important analysis and conclusion for applications that deal and are composed by decision making, based on a process-oriented approach.

5. Comparisons and Advancements – Generalizations & Potentials

The above findings, if applied in the education domain, will be of great significance and will provide initial guidance for developing gamified applications for the educational industry (eProducts, eServices and so on). Such applications need to be compliant and respect the overall model of "Input, Process and Output" (Heis, 2008) in the education domain, where input are the educational objectives that by the appropriate learning process are transformed to learning outputs that in turn are evaluated against the initial objectives (Figure 4).

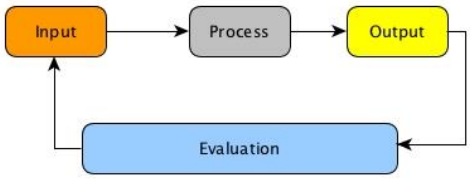


Figure 4: Input Process Output Model (Heis 2008, [220])

As a consequence, for drawing and implementing interesting, effective and challenging input for an educational application, being aware of the patterns that users are familiar with, through the "Evaluation" of their behaviour in such applications, would allow to provide to them more tailored to their needs "Input" for the educational "Process" that will allow the maximum "Output".

In terms of gamification resources, our methodology will enrich the decision process for creating and applying more personalised experience to end users and suitable quests, challenges, rewards etc. for being also compliant with the fact that designing user experience for educational applications that incorporate gamification strategies needs to facilitate the "ABCD" model educators use for writing and using learning objectives, whereas "A" means audience, "B" behaviour, "C" conditions under which the objectives have to be completed and "D" is the degree that learners have to achieve [21].

It is known that the type of proper motivation varies depending on the audience characteristics therefore for succeeding in adopting the above-mentioned model, given that learners are willing to participate in educational processes if they are subjects to appropriate education methods and resources, informed on the foreseen behavioural change and triggered enough for completing the set of achievements they are obliged to.

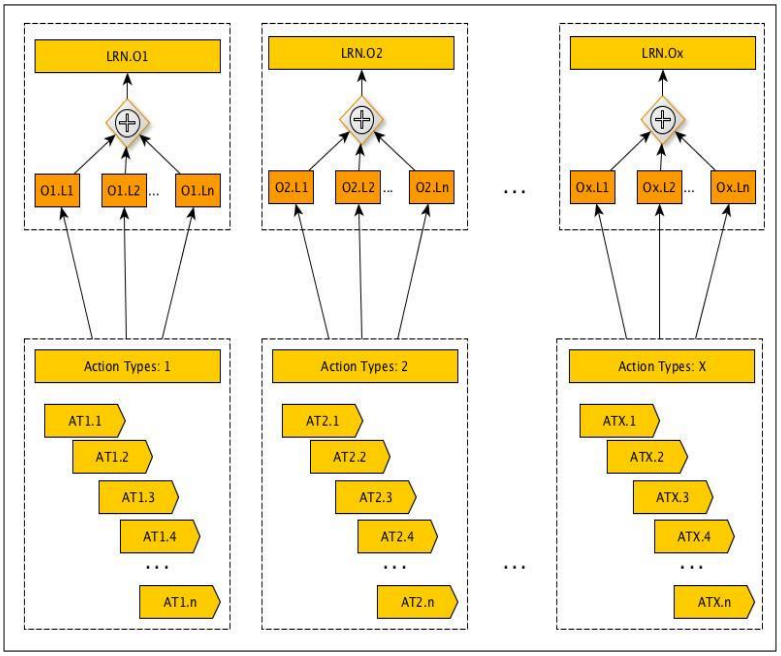


Figure 5: Design of an Educational Gamified Application

Even in Fogg's Behavior Model [22] is clearly stated that providing sufficient motivation, adequate ability and efficient activation to learners, results to the desired from the educator point of view "behaviour" achievement. As gamification contributes towards the provision of both motivation and activation of learners, it is important to employ suitable type and "portion" of them for resulting to successful applications and learning outcomes.

We select to present the potential application of our methodology to the education domain. For this purpose, we consider the education related processes, such that to be related in an input-process-output schema of relations, composing thus a system-oriented view point to it (education).

Our methodology will allow gamification applications in the above-mentioned education model, to use the "right" pattern each time, for supporting the desired behaviour for the learners, overcoming the theoretical obstacle that "the use of challenges in Information Systems though, may be limited because of a high number of monotone and standard tasks" [23]. Such strategy, hence, based on optimal path reinforcement of educational processes, will lead both teachers and students, to be informed and getting up to their educational objectives with a high degree of performance, understanding and productivity in the area.

Based on the above findings of our methodology, we attempt to design a gamified application. We consider again the educational domain for the target group of the application, that is independent of the subject that has to be introduced – taught.

The main course objectives are for the paradigm below (Figure 5), introduced as LRN.O1, LRN.O2, ..., LRNOX. It shows how action types will be connected to Learning Objectives for the design of an Educational Gamified Application. These are the necessary objectives to be achieved so that the course to be successful for the learner (or the learner group).

Every such Learning Objective Ox, in turn, can be achieved in the gamified application to be designed, by the achievement of respective objective game levels (Ox.Ln, meaning for the objective LRN.Ox, the Level n and so on). We assume that the number of levels per learning objective may vary, (but at least may have a value of 2).

On the other side, in our game design methodology we offer, the ability to the developer but at the same time to the player – learner, to design a variety of action types for the achievement of such levels, as described above and play freely by navigating and making actions – decisions through the game scenario. Such action types that are introduced by the game designer and implemented by the game developer, may be represented as a virtual library of actions, containing – being categorized further in sub-action types. For example, somebody may select to achieve the respective level, to read material, then to play a local educational game and then take a quiz to collect points towards next level. On the other hand, somebody else, may select to study much more material in sequence, so that immediately as soon as will take the quiz, to be sure that will also be awarded the next level. This model introduces the independent process in sequence selection towards one, the same objective achievement, allowing edutainment weight to be balanced by the user, on real time basis, i.e. how much game, how much, in app study, out app (text – notes) study, cooperative action types, info from social networks and so on will on demand use.

Collecting this information in large log files, as our application will run, will in turn produce frequent patterns of selected actions, where players achieved the objective much faster, as a function of selected action types. Our methodology then, when enough knowledge and processes will be collected, will propose to next player i, a proposed optimal path towards knowledge acquisition of the respective level, i.e., a set of Action Types, in a specific order, so the player – learner will be more productive. Of course,

this will be only a suggestion and the player will be always free to reject it and hence select his own path, towards the level achievement.

6. Conclusions and Future Work

As an extension to the ONARM+ previously presented work, we developed a formal methodology to support the design of gamified applications. It is based on analyzing game player's actions stored in log files to extract frequent patterns of actions. The extracted patterns are processed and classified according to the expected outcomes which can be supported in a gamification application.

This knowledge base for gamification of frequent sequential patterns of actions, along with the outcomes, they are considered as knowledge process-based transactions. Thus, given a specific domain and the expected outcomes for a gamification application, the designer could build it by selecting the proper patterns of the knowledge base above, so to drive via optimal decision making paths user behaviour changing, which is the basis for a variety of diverse domains, like marketing, learning, consumer/branding loyalty and so on. Apart from patterns, the knowledge base above. Based on this, we plan to extend our work so that the knowledge base above, to contain also meta-rules for controlling game techniques (e.g. dynamics, feedback, badges, levelling, etc.) and their relative association on a per user profile basis.

Finally, we have presented an educational game design concept, based on our proposed methodology, where the proposed application, introduces to new players – learners, accumulated optimal knowledge experience, towards game level achievements and their knowledge – learning objectives. In this game we tried to examine, using our methodology associations in user decisions and choices, as a set of concept-based processes, so based on an algorithmic approach to predict and drive choices of similar user profiles, on the same domain. While availability by developers of such real time on line games have been proven a task difficult to be achieved, future work will be focused in collecting more than an additional year's supplementary data to verify our findings vis-à-vis user behaviour.

Overall, initial interesting results of such associations were found, for specific players. Further work will concentrate also in five main directions (a) application of the maximal VMSP mining algorithm and comparison of results generate; (b) generalization of results found to a broader set of players; (c) application of the proposed methodology to other gamified domains, for mining underlying processes; (d) analysis of more other computer games (types) to extract more patterns and (e) the generalization of our work for groups of users instead of focusing to specific ones.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Tatsiopoulos, D.M. Katsanta, "On ontologies and knowledge associations in gamified Environments", 2018, 9th International Conference on Information, Intelligence, Systems and Applications (IISA), July 23-25, 2018, IEEE, Zakynthos, Greece, DOI: https://doi.org/10.1109/IISA.2018.8633589, e-ISBN: 978-1-5386-8161-9

[2] S. Deterding, Dixon, D. R. Khaled, L. Nacke, (2011). From game design elements to gamefulness. In Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11. http://doi.org/10.1145/2181037.2181040

[3] Z. Epstein, (2012). Enterprise Gamification for Employee Engagement. Marioherger.At.

[4] H. Oinas-Kukkonen, M. Harjumaa, (2009). Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems.

[5] T. Xu, F. Feng, D. Buhalis, J. Weber, H. Zhang, (2015). Tourists as Mobile Gamers: Gamification for Tourism Marketing. Journal of Travel and Tourism Marketing, 33(8), 1124–1142. http://doi.org/10.1080/10548408.2015.1093999

[6] E. B. Passos, D. B. Medeiros, P. A. S. Neto, E. W. G. Clua, (2011). Turning real-world software development into a game. In Brazilian Symposium on Games and Digital Entertainment, SBGAMES. http://doi.org/10.1109/SBGAMES.2011.32

[7] K. Werbach, D. Hunter, (2012). The Gamification Toolkit: Game Elements. For the Win: How Game Thinking Can Revolutionize Your Business. http://doi.org/10.1017/CBO9781107415324.004

[8] R., Srikant, R. Agrawal, (1996a). Mining sequential patterns: Generalizations and performance improvements. Advances in Database Technology — EDBT '96. http://doi.org/10.1109/ICDE.1995.380415

[9] C. Tatsiopoulos, B. Boutsinas, "Automatic knowledge exchanging between tourists via mobile devices," in Journal of Hospitality and Tourism Technology, vol. 1, No2, pp.163-173, 2010.

[10] C. Tatsiopoulos and B. Boutsinas, "Ontology mapping based on association rule mining", in Proceedings of 11th International Conference on Enterprise Information Systems, vol.3, pp.33, Milan, Italy, 2009.

[11] C. Tatsiopoulos, B. Boutsinas, K. Sidiropoulos, "On aligning interesting parts of ontologies", in Proceedings of International Joint Conference on Knowledge Engineering and Ontology Development (KEOD2009), pp. 363-366, 2009, Madeira, Portugal.

[12] R. Srikant, R. Agrawal, (1996b). Mining Sequential Patterns: Generalizations and Performance Improvements. In Proc. 5th Int. Conf. Extending Database Technology, EDBT (Vol. 1057, pp. 3–17). http://doi.org/10.1109/ICDE.1995.380415

[13] A. K. B. Bharathi Gopinath, A. Singh, C. S. Tucker, H. B. Nembhard, (2016). Knowledge discovery of game design features by mining user-generated feedback. Computers in Human Behavior. http://doi.org/10.1016/j.chb.2016.02.076

[14] L. Mellon, (2009). Applying Metrics Driven Development To Mmo Costs and Risks. Versant Corporation.

[15] A. Drachen, A. Canossa, (2011). Evaluating motion: spatial user behaviour in virtual environments. International Journal of Arts and Technology. http://doi.org/10.1504/IJART.2011.041483

[16] D. Kennerly, (2003). Better Game Design Through Data Mining. In Gamasutra.

[17] A. Drachen, M. Seif El-Nasr, A. Canossa, (2013). Game Analytics – The Basics. In Game Analytics. http://doi.org/10.1007/978-1-4471-4769-5_2

[18] W.M.P. Van der Aalst, (2011). Process Mining. Process Mining. http://doi.org/10.1007/978-3-642-19345-3

[19] P. Fournier-Viger, C. W. Wu, A. Gomariz, V. S. Tseng, (2014). VMSP: Efficient vertical mining of maximal sequential patterns. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). http://doi.org/10.1007/978-3-319-06483-3_8

[20] P. Fournier-Viger, J. Chun, W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas, (2017). A Survey of Sequential Pattern Mining. Ubiquitous International.

[21] A. Salam, (2015). Input, process and output: System approach in education to assure the quality and excellence in performance. Bangladesh Journal of Medical Science, 14(1), 1–2. http://doi.org/10.3329/bjms.v14i1.21553

[22] B. Fogg, (2009). A behavior model for persuasive design. In Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09. http://doi.org/10.1145/1541948.1541999

[23] S. Thiebes, S. Lins, D. Basten, (2014). Gamifying information systems A synthesis of gamification mechanics and dynamics. Twenty Second European Conference on Information Systems, (August 2016), 1–17. Retrieved from http://aisel.aisnet.org/ecis2014/proceedings/track01/4/

[24] U. K. Heis, (2008). Input Process Output, 1–3. http://doi.org/10.1007/1-4020-0611-X_462

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 213-223 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Improvement opportunities of a Simulation/Expert System Approach for Manufacturing System Sizing: A review and proposal

Lajmi Imen[1,*], Masmoudi Wassim [2], Elleuch Mounir[2], Chtourou Hedi [3]

[1]*University of Sfax, Research Unit of Industrial Management and Logistics Quality (LOGIQ) Higher Institute of Industrial Management Sfax, Tunisia*

[2]*University of Sfax, Laboratoire des Systèmes Electro-Mécaniques (LASEM), Ecole Nationale d'Ingénieurs de Sfax, Sfax, Tunisia*

[3]*University of sfax, Institut Préparatoire aux Etudes d'Ingénieurs de Sfax, département de Technologie, Sfax, Tunisia*

ARTICLE INFO

Article history:
Received: 03 June, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Manufacturing system sizing
Handling Resources
Production resources
Simulation
Expert System
Performance

ABSTRACT

Manufacturing System (MS) sizing is a crucial task to complete in order to obtain the desired MS performance and efficiency. It involves selecting the required number of resources from each used type in a given planning horizon. In fact, different approaches coupling simulation/optimization tools have been developed to solve this issue and evaluate the MS performance. One of these approaches is the Simulation Expert System Approach (SESA). Unfortunately, the application domain of this approach is limited in sizing only the production resources (machines and labor) but neglects the material handling system (MHS) components. Moreover, omitting the transferring problem is not viable in the real world due to its importance in each shop floor. Thus, the aim of this paper is to describe the evolution of SESA, then, to check if the simulation optimization tools used in SESA are still relevant. This paper also investigates the importance of incorporating MHS in this approach and finally proposes some improvement opportunities for SESA including the tackling of the MHS fleet sizing problem. In fact, the wide literature review performed in this research indicates that SESA is still a pertinent approach but it must be improved. Therefore, it is expected that SESA improvement opportunities proposed in this work will greatly assist industrialists in enhancing the overall MS performance, providing a significant productivity increase and a minimization of the total production costs.

1. Introduction

This paper is an extension of a work presented in the 6th IEEE International Conference on Advanced Logistics and Transport (ICALT) [1].

Due to intense competitivity in the industrial sector and to demand fluctuation, companies need to ensure better performance of their Manufacturing Systems (MS). An appropriately selected number of resources can improve the overall system performance. On the other hand, too much chosen resources means needless investment, whereas an insufficient number of chosen resources can make the system unable to produce the required quantities in the desired due date, which constitutes a financial loss (ie. penalties, loss of customers, etc.). Therefore, MS sizing is an important issue that must be considered to obtain the desired system performance. In fact, this issue has attracted the attention of several researchers. Hence, many types of approaches have been used to solve machines, labor and MHS fleet sizing problems. The Simulation/Expert System Approach (SESA), is one of these approaches used to solve MS sizing. It was proposed by [2] and enhanced by other researchers either by considering other resources or by altering the used tools. However, SESA still needs improvements to be applicable in real cases. However, one of its shortcomings is the lack of consideration for the material handling system (MHS). In fact, appropriate MHS choice and sizing is crucial for the MS design task [3]. Its main role is transferring parts between the various components of a MS in the most economical way, to ensure an efficient material flow. The interest of configuring and sizing the MHS is continually increasing in

*Corresponding Author Imen Lajmi, Sfax 3062, Tunisia, +21652486285, Email: lajmi.mezghani.imen@gmail.com

manufacturing facilities. Therefore, this research proposes some opportunities to improve SESA framework considering the MHS fleet sizing problem to make this approach more pertinent and applicable in real-world cases.

The remainder of this paper is as follows: Section 2 presents a brief literature review on the MS sizing problems. The originality of SESA and its evolution are presented in Section 3. Section 4 presents a survey of recent research studies to show the promising potential of SESA tools. Before concluding, section 5 suggests an improvement proposal for SESA.

2. MS Sizing problem

MS sizing problem is defined as the selection of the required number of resources from each type to be used in a manufacturing process for a given horizon.

2.1 Resources studied in literature

Several research studies dealt with the MS sizing problem, more precisely with the main manufacturing resources used in MS, such as (1) machines, (2) labor and (3) material handling system (MHS).

2.2 Machine selection problem

Machines represent a fundamental industrial equipment that allocates a part, for a period, to generate changes by performing certain tasks (generating the Added Value). Finding the required number of machines is an important task in the MS to achieve the target performances. Among the first research studies discussing the machine selection problem, we can cite [4] who developed neural networks to minimize the number of machines in each work center in a flexible manufacturing system. In addition, [5] proposed a mathematical formulation and a heuristic algorithm to find out the required number of types of machines in a group technology system that reduce the manufacturing cost. More recently, [6] developed a new modelling approach combining simulation/optimization methods to find the minimum number of machines required in each MS type. Besides,[7] used an artificial neural network approach coupled with a simulation tool to estimate the minimum number and the best scheduling of machines that enhance the mean tardiness and the mean flow time. In fact, [8] used two analytical methods to firstly, determine the optimal number of machines and the quantities produced in each period and secondly, determine the production plan and the periodic preventive maintenance scheduling to reduce the failures and minimize the total costs. Moreover, [9] used Discrete-Event Simulation (DES) and OptQuest to find out the optimal machine number. Two objective functions were used: maximizing the machine utilization rate and minimizing the waiting time in the queue.

The continuous dealing with the machine selection problem proves the importance of this issue in the MS improvement.

2.3 Labor selection problem

The human factor plays a crucial role in the creation, the responsiveness, the flexibility and the performance of industrial processes. The determination of the optimal number and quality of labor to be allocated is necessary in each MS. In fact, this issue has been the object of numerous research studies. For instance, [10] used Fuzzy AHP, TOPSIS and Simulation tools to determine the optimal number and assignment of labor in the system. Many performance measures have been used to achieve some objectives such as the average lead-time, the average labor utilization and the average waiting time. [11] used simulation and genetic algorithm to determine the optimum number and allocation of labor and the measure of the efficiency of their operation in an assembly shop. His aim is to maximize the throughput system. Similarly, [12]developed a simulation /optimization approach based on the genetic algorithm (NSGAII) to determine the required number of labor. [13] used CPLEX optimization software and DES (ARENA software) to determine the required number of operators in automobile chassis manufacturer while considering the labor fatigue in their attempt to satisfy the customer demand.

2.4 MHS fleet sizing

MHSs are used to transport goods and materials between workstations of a manufacturing system. Therefore, the number of vehicles greatly influences the performance of the MS. MHS are usually expensive, thus determining the appropriate number of vehicles is very a very important task. Hence, due to the importance of this problem, many recent research works dealt with this problem using different approaches. For instance, [3] developed an analytical model to estimate the minimum number of automated vehicles required by a flexible MS. The results of the regression analysis were compared with simulation results to show that the regression technique is a promising approach to resolve the fleet sizing problem. [14] used a simulation tool to determine the optimal vehicle number and material flow rate in a semiconductor manufacturing systems. [15] applied a queuing model to determine the required number of automated vehicles in a flexible MS containing multiple pickup and delivery points. The objective was to minimize the waiting times. As for, [16] applied the queuing theory to specify the required number of MHSs in a MS in order to maximize their utilization rate and minimize the investment costs. Besides, [17] used the bees algorithm to estimate the optimal number of MHS operated at the MS to maximize the profit. This algorithm is inspired by the foraging behavior of honeybees. On the other hand, [18] used multi-class Closed Queuing Networks (CQN) based on a linear programming to model the movement of MHS in a MS and determine the minimum MHS number. More recently, [19] has developed two analytical methods to determine the AGV fleet size in different layouts of flexible MS. In fact, different factors, such as the processing time, job sequence, job mix, loading/unloading stations and the number of work centers, are included in the model.

2.5 Synthesis

Table 1 presents a literature survey on the MS sizing problem the detailed analysis of which shows that the number of studies about the production resources selection area has grown considerably in recent years. In fact, the growing interest in this

subject shows the importance of the resolution of the resource selection problem in improving MS performance. However, one of the gaps in these studies is the absence of a correlation between the three types problems related to the selection of machines, of labor (operators) and of MHS components, which affect one another.

Table 1. Literature review on MS sizing

Types of resources	Research studies
Machine	[4-9]
Employees	[10-13]
MHS	[14-19]
Machine and employees	[20, 21]

Recently, some studies have combined machine and labor selection problems but without considering the MHS fleet sizing problem. For example, [20] used a simulation tool to size an effective automobile manufacturing process. The aim is to enhance the system throughput by determining the required number of repair stations, labor, machines and the best plan configuration. More recently, [21] has sized a functional MS to determine the required number of machines and employees that enhance the overall system performances for the purpose of respecting the Due Date (DD).

2.6 Used methods

Due to the variety of resources used in the production systems and the influence of one over the other, the search for the optimal size becomes more and more difficult. To overcome this difficulty, many methods are employed in the literature to tackle the sizing issue, which are classified into two main categories:

- Analytical methods;
- Simulation-based methods.

The first category can be classified into two main types: exact and approximate methods.

In fact, the exact methods like Branch-and-Bound, linear programming and others provide an optimal solution, which is suitable only for small sized problems. Therefore, the major weaknesses of these methods are their static and deterministic character, their mis-consideration of the dynamic and stochastic aspects of the system and their inconsistency with the MS sizing problem. Since the number of feasible solutions rises exponentially when changing the total number of resources, it is impossible to use complete enumeration for large-sized problems. As a consequence, heuristic and meta-heuristic methods are widely approved by the researchers to efficiency deal with the MS sizing problems.

Approximate methods, which are heuristic and artificial intelligence methods, as the expert system, artificial neural networks, genetic algorithm, Tabu search, particle swarm optimization, simulated annealing and others do not always guarantee the optimal solution.

On the other hand, the simulation-based methods, which fall into the second category, are widely used to solve MS sizing problems. In fact, simulation can provide a detailed analysis of the system performance and help managers to take the right decision.

Moreover, some studies developed approaches combining simulation and analytical methods to find out satisfying results (not optimal ones). Due to the effectiveness of this combination, [21] used an expert system simulation approach called “SESA” to look for the optimal number of two types of resources simultaneously: machines and labor. In fact, this work considers the stochastic and dynamic aspects. However, the major gap in this work is the neglect of transfer problem when sizing MS.

In the next section, the evolution of SESA and its originality are presented.

3. SESA Evolution

3.1. Originality of SESA

SESA is a prescriptive simulated approach developed by [2] to solve the MS sizing problem and more precisely, to identify the optimal number of machines in a simple static and deterministic MS. The framework of SESA is presented in Figure 1. The two principal tools used in SESA method are the simulation and the expert system, which work in a closed loop and exchange data and results.

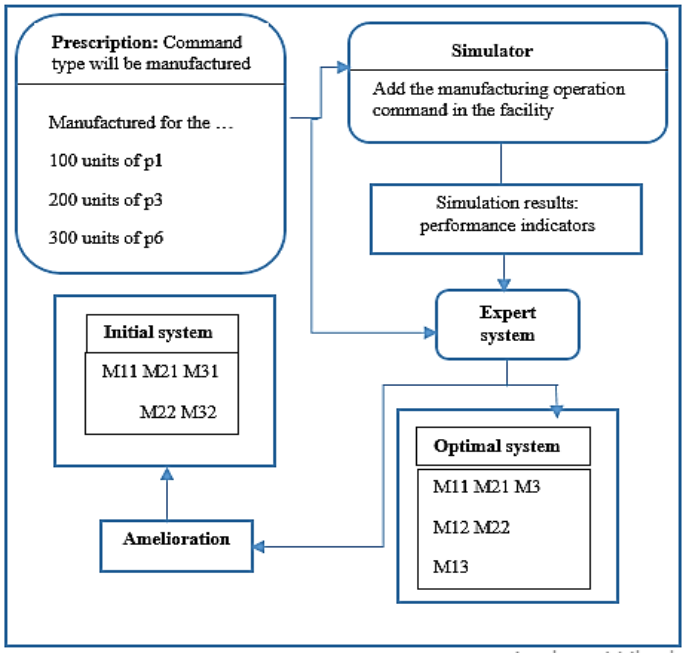


Figure 1. The framework of SESA [2]

The data of the studied MS and the typical order to be manufactured are the required inputs for the simulation step in order to establish the simulation procedure according to which the command type will be executed. This first step provides performance indicators such as the total operation time. For each scenario, the simulation model is run and the system performance measures are recorded. Then, the second step, it involves the expert system, which has as inputs the performance measures (provided

by simulation) and the performance limits. The main role of this step is to analyze the results provided by the simulator and establish one of these two cases:

- The system performance, which is optimized according to the actual context, or;
- The system performance, which needs to be improved according to the actual context. In fact, the expert system offers a problem diagnosis, according to the MS resources, in order to check if there is a lack or an excess of resources. Then, the ES provides a set of recommendations to adjust the resource number, which improves the system performance.

The simulator used in this original version is a deterministic tool named "SICOP". It is a simple program developed by "C language", which provides statistical results (performance measures). The expert system is "SEORAM", which is intended to minimize the Mean Flow Time (MFT). The results of this study indicate that combining the simulation and optimization tool is quite competitive, in terms of precision, when compared to the simulation itself. This approach helps to avoid the "try-error" aspect. Enhancement of the system performance proves the efficiency of this approach. Accordingly, other researches were encouraged to extend the application of this approach to be more applicable in real MS.
The weaknesses of this study are as follows:

- Misconsideration of the stochastic and dynamic aspects;
- Used tools and the total indices of performance chosen in this initiated approach are not used any more in research since they do not reflect any more the current tendencies of the industrialists;
- Misconsideration of the labor and MHS selection problems.

3.2. SESA improvement

Improvement of the Expert System used tool "ESMRS"

In the follow-up research work, [22] developed an improved version of SESA and applied it for Job Shop machine selection problem. The studied MS grouped the same machine type on departments and supposed that the requirement of labor and material handling system are neglected. The main focus was on the development of a new ES prototype named expert system for manufacturing resource sizing (ESMRS developed by Kappa. PC) (see Figure 2). A set of generic diagnosis production rules is developed. The main objectives were to improve the structure of research mechanism, recommend resource modification according to performance indicators provided by the simulation and overcome the "try-error" aspect. The results provide significant insights into the effective application of the enhanced ES.

Improvement of the simulation used tool

After that, [23]enhanced the framework of SESA, previously described, by using the simulation software "ARENA". This tool considered the stochastic and dynamic aspects, governs Manufacturing Orders launching (MO) as well as machine operating parameters. Thus, all entities of the same product type enter the MS according to a stochastic inter-arrival time (random distribution). Then, the entities stay on the machine queue until the machine becomes available. After that, the entity will be treated in a randomly chosen processing time (see Figure 3). The same optimization loop presented previously is used in [23] but the differences are:

- Use of a stochastic simulator tool;
- Use of more relevant performance indicators and;
- Considering the stochastic and dynamic aspect.

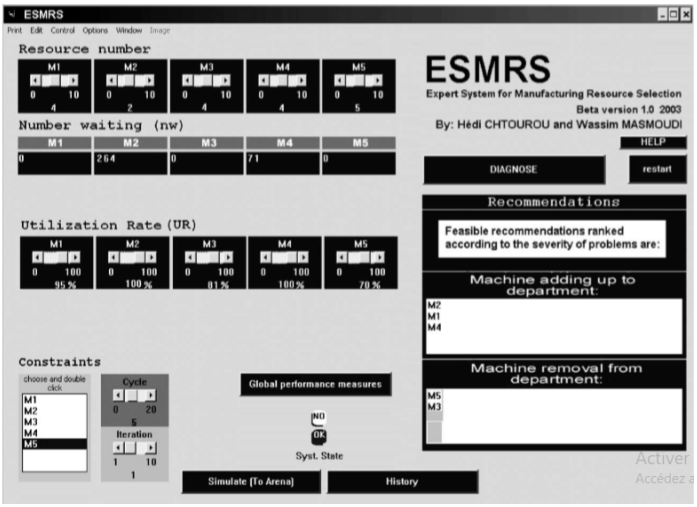


Figure 2. Enhanced ES interface [22]

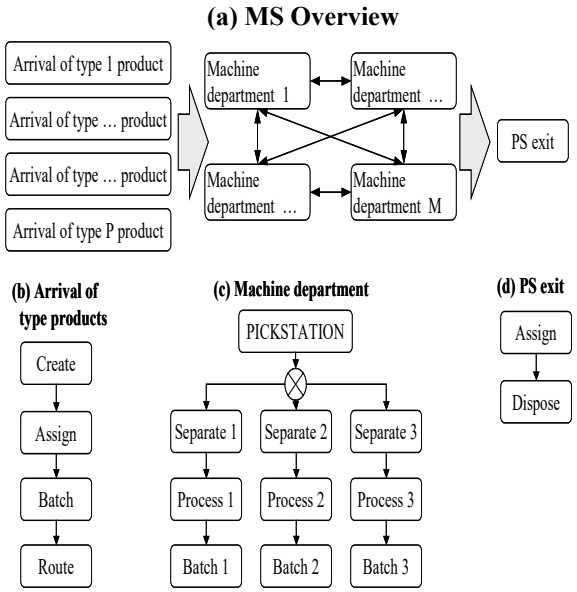


Figure 3. MS "ARENA" simulation model[23]

Despite improvements provided by [22, 23] to enhance the efficiency of SESA tools, this approach still limited in its application domain. It resolve only the machine selection problem.

3.3. Enlargement of SESA application domain

In an earlier work, [21] extended the application domain of SESA to select the optimal number of both machines and labor in a functional manufacturing system. This extension has already involved changes both in simulation model (see Figure 4) and ES interface (see Figure 5).

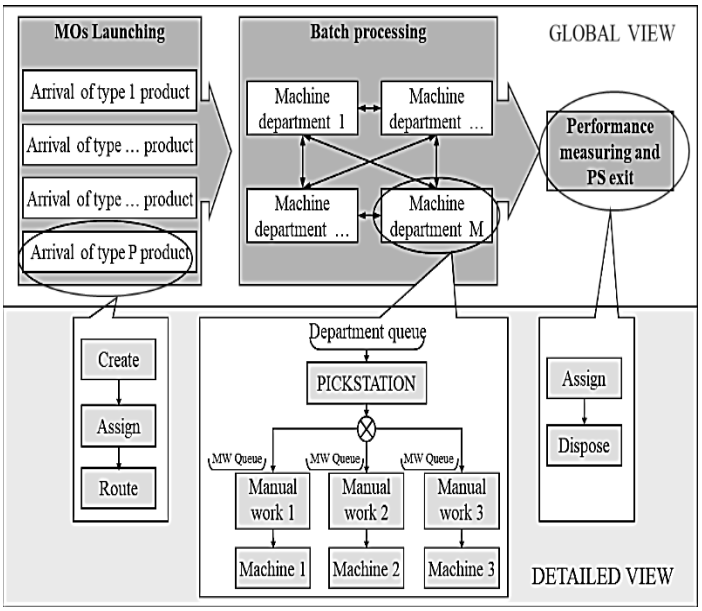


Figure 4. MS "ARENA" simulation model [21]

To build the simulation model, two inputs are required:
The MS data which represents the number of work center, Processing Time (PT) and Setup Time (ST) and;
The demand patterns, which represent the inter-arrival times, product type, load size and sequence.

Different performance indicators have been used to enhance the applicability of SESA. These indicators are: the rate of resource utilization, the average waiting time and waiting number of batches in resource queues, throughput, mean tardiness and mean earliness.

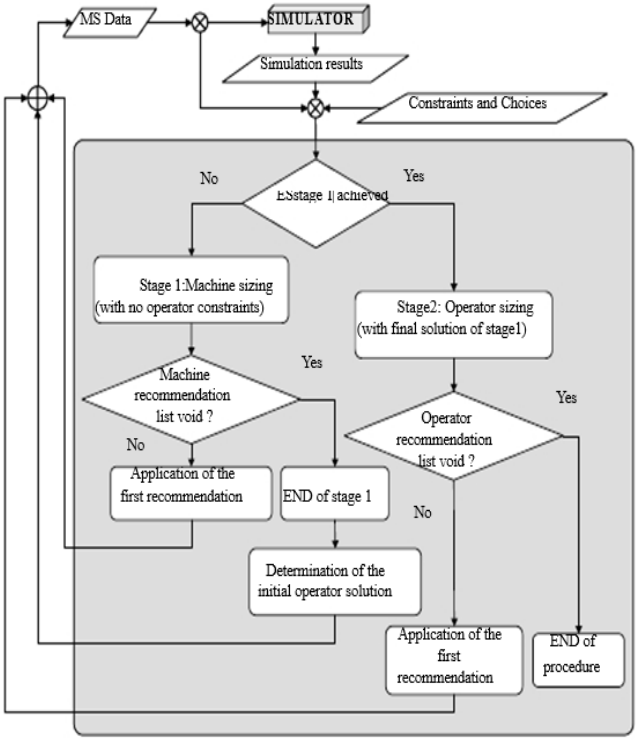


Figure 5. Enhanced ES interface [21]

The main target is to respect the DD in order to minimize the delay penalties and the storage costs. The DD indicator is more adequate for functional manufacturing context than MFT. The first objective is to minimize the Mean Tardiness (MT) and then minimize the Mean Earliness (ME), which is considered as the second objective. Besides, in order to avoid superfluous investment costs, all the resources should be fully utilized. The experiments with different levels of DD closeness indicate the important benefits of the approach improvements through the new performance measures and the enhanced ES interface.

The enhanced ES interface is divided into two main stages. The first stage is the machine selection problem without labor constraints and the second is the labor selection problem.

In the first stage, the ES starts with the study of the MS machine situation (without labor constraints). Thus, this stage supposes that the number of employees is equal to that of machines (assigning a worker to each machine). If the simulated system performance needs to be improved, the ES will recommend modifications regarding machine number. These changes overcome the problem and enhance the system performance. Thus, a new cycle is run with the modified machine number until the ES can not suggest other modifications. This situation presents the end of this stage. In the second stage, the ES resolves the labor selection problem with a machine balanced MS (according to the first stage results). The same optimization cycles will be run until they reach a non-improvable MS.

The confines of SESA developed by [21] is the neglect of the transfer network and MHS constraints when sizing MS component resources. In fact, MHS is significant in MS that has an effect on the system performance [24]. It can account for 30 –75% of the total cost, and efficient material handling can be primarily responsible for reducing a plant operating cost by 15–30% [25]. Therefore, MHS is very important in reducing the total manufacturing costs and increasing profits. Consequently, it is necessary to extend the application domain of SESA by incorporating material handling system constraints when sizing production resources. However, it is necessary to demonstrate the relevance of SESA tools. Therefore, the next section will provide a detailed survey of recent research studies using simulation and/or expert system for solving resource selection problem.

4. Is SESA a promising approach?

4.1. Is simulation always promising?

Simulation has been developing for a long time to became an essential tool for many disciplines. Its application fields may go from industry [26] healthcare [27, 28], urban design [29] military [30], among others. It has been widely proven by the scientific community and practitioners as a flexible technique by allowing modelling, testing and analyzing complex dynamic systems, namely for models where structure and behavior change over time. The simulation model is a representation of the real system. It may reduce the risks related to the implementation of new system design, production cost, production time, etc. It is used to analyze

and process huge quantities of data. As shown in Figure 6, simulation is a solid approach with many years of application and is a rapidly evolving research field [31].

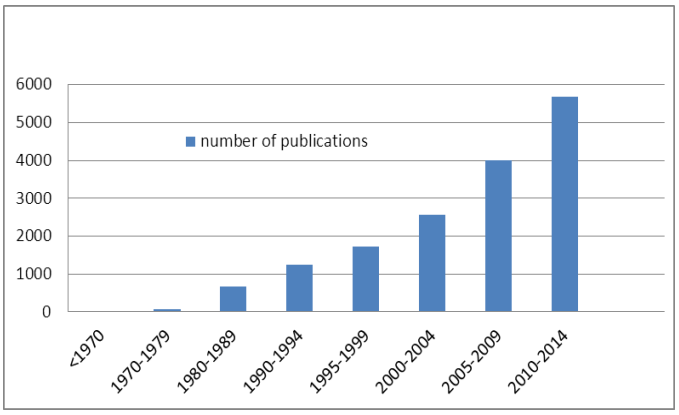


Figure 6. Evolution of simulation application

Discrete-event-simulation (DES) is mostly used to solve MS problems. Numerous benefits to use DES defined by [32]include: simulate a real system in a short time, study the interaction between different components, reduce cost and risk, monitor and control the system and others.

This approach has tackled different problems. A brief summary of DES applications to solve manufacturing systems is presented in Table 2. [33] revealed abundant references in more details about the simulation utilisation on other MS problems.

Table 2. Publications related to simulation approach

Authors	Years	Simulation application
[34]	2008	System and facility design
[35, 36]	2015, 2017	Planning and scheduling
[37]	2016	Production planning and control
[38]	2019	Supply chain design
[39]	2016	Inventory management
[40, 41]	2019, 2018	Maintenance
[42]	2016	Controlling and analyzing energy and power consumption
[43]	2018	Resource allocation

In all the above discussed studies, DES models have been proven as powerful decision support tools, which enables practitioners to accurately test, model, analyze and plan complex systems. Moreover, it is an efficient tool to search for optimal solution and evaluate the obtained results. Different DES tools have been used in literature.in this context, [44] compared these tools to identify the most popular and used ones. In fact, the ARENA commercial software has been the most popular and used tool for modeling and analyzing various complex MSs.

Nevertheless, the simulation tool may not be the first preferred in solving real-life system due to the "trial-and-error" nature of the development process and its lengthy computational requirements. This issue is particularly remarked when the solution space is mostly large and the time available to make decision and analysis is too limited. Thus, it is necessary to add an analytical method to the simulation tool to do an optimal decision-making. Therefore, the simulation tool has been coupled with an expert system approach.

4.2. Is the Expert system still promising?

The ES is an Artificial intelligence (AI) software application that attempts to simulate the knowledge and experience of human experts in a specific domain and make such expertise accessible on demand to the program user. This application was introduced by [45]for the first time in the 1980s. A growing interest in the use of ES has been seen for solving optimization problems in many domains such as Healthcare [46, 47], supply chain [48],transport [49], risk management [50] and others.

According to [51], the expert system includes three principal parts (Figure 7):

- User's interface that enables the user to ask questions and provide advice corresponding to it;
- Knowledge base that consists of facts and rules, which are created from information which are provided to it;
- Inference engine that helps to match the user's query with knowledge base and provides the result.

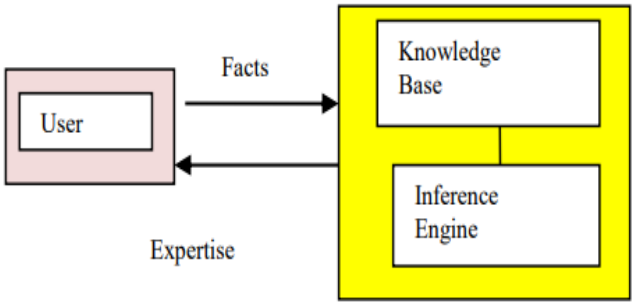


Figure 7. An Expert system environment

Table 3. Recent research studies used ES

Authors	Years	ES applications
[52]	2016	Evaluation of the performance of sustainable manufacturing
[53]	2016	Material handling equipment selection
[54]	2019	
[55]	2019	Layout designs
[56]	2018	Optimization of Design and Manufacturing Process
[57]	2019	Shift work
[58]	2015	Machine tool selection
[59]	2015	Material Selection
[53]	2016	
[60]	2017	
[61]	2018	Risk management

The main advantages of ES consist in reducing human error, always available, providing expertise at a minimum cost, which is used in any risky environment to solve complex decision problems

in different specific fields, obtain flexible and practical solutions, and respond at great speed, etc.

The capability of ESs to acquire performance similar to the human ones makes current researchers used it to solve various types of problem in MS (see Table 3).

4.3. Is coupling Simulation/Expert System promising?

Simulator systems generate large and a high variety of data that make decision making so difficult. Therefore, novel approaches combined simulation and optimization have been developed in literature to eliminate this gap. In fact, simulation based optimization approaches have received considerable attention from many recent researchers to solve MS problems. [62] proposed an improvement framework using genetic algorithm and DES to investigate the buffer size and job sequencing optimization problems. [63] combined the DES models and artificial neural network models of the production system. This combination was realized through an ES in order to evaluate the stability of a production system and implement production plans. [64] developed a practical approach using bee colony to improve throughput while minimizing waste energy and environmental impact. [65] proposed a novel methodology using ant colony and DES to select better facility allocations that minimize fixed and transportation costs.

Recent research studies applied the DES with ES in different application including industrial systems [66], food process [67], supply chain management [68], healthcare systems [69], risk management [70]and others.

The main advantage of this combination is that the DES can integrate compound system dynamics and uncertainties. Then, ES helps in efficiently getting optimal parameter values. The methodology is that the ES accesses the output of a simulation cycle and, depending on the user's objective, selects and retrieves the pertinent data and provides a mechanism for exception reporting. A greater degree of integration would be for the ES to explain the recommendations. Thus, the incorporation of DES with ES improves the decision-making process.

The research studies, cited in the previous three sub-sections, have been conducted to assess the trends of SESA approach and its capability to solve highly complex systems with stochastic variables. We can conclude that SESA is still promising but can be improved.

4.4. Limits of SESA

The application domain of SESA is limited in finding the optimal number of machines and employees without taking into account the MHS, despite its importance. A MHS can be defined as a vehicle that moves work-parts at short distances and usually inside the building. [71] affirms that the MHS represents a major portion of the total manufactured product cost. An efficient MHS can have a main impact on the operating cost, which reduces and increases the total throughput and department utilization. Moreover, a well selection of the required number of MHS is very important to significantly improve the efficiency of the MS. In fact, if the chosen number of MHS is less than the required number, the work-parts will saturate the storage areas between departments and their waiting time increases. Moreover, if the chosen number is higher than the required one, the MHS waiting time increases. This will cause an accumulation of vehicles, a saturation of the traffic lines and a system blockage. Besides, the MHS purchase cost is very high, so an accurate determination of their optimum number is required.

Consequently, considering MHS when selecting machines and labor, it is very important to significantly improve the efficiency of MS.

5. SESA improvement proposal

From previous sections, it is clear that SESA is always promising seen simulation and ES benefits. Nevertheless, it is a priority to take into consideration the MHS. Our contribution in this section is to highlight the opportunities of SESA improvements.

5.1. Proposal to enlarge the application domain

To enlarge the application domain of SESA, we will propose some improvement opportunities and develop a new strategy of SESA by considering the MHS fleet sizing problem (see Figure 8).

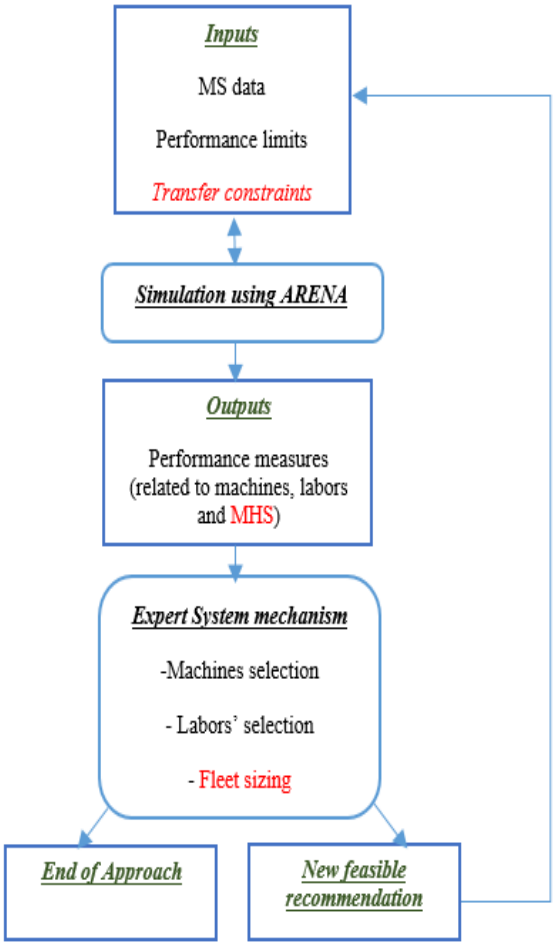


Figure 8. The new SESA framework

In the beginning, different characteristics related to the MHS will be integrated in simulation inputs. Then, new performance indicators related to the MHS will be considered to analyze system performance (see section 5.4). Finally, a new stage related to the MHS fleet sizing will be added to extend the expert system mechanism.

5.2. Improvement proposal of simulation model

To improve the simulation model developed by [21] and make the first investigation of taking into account the MHS when sizing MS, it is necessary to add different data related to the MHS within the MS data, which represent the simulation input. A representation of the improved model is presented in Figure 9.

The MHS data requirements can be:

- Type of material handling;
- Direction of movement (orientation);
- Transport distance;
- Load size;
- Loading and unloading time;
- Speed of material handling;
- Dispatching rule.

The components of this proposed model are as follows:

- Materials warehouse;
- Five departments (each one group the machinery with the same process capability) ;
- Parking (group the MHS);
- Finished Pieces (FP) warehouse.

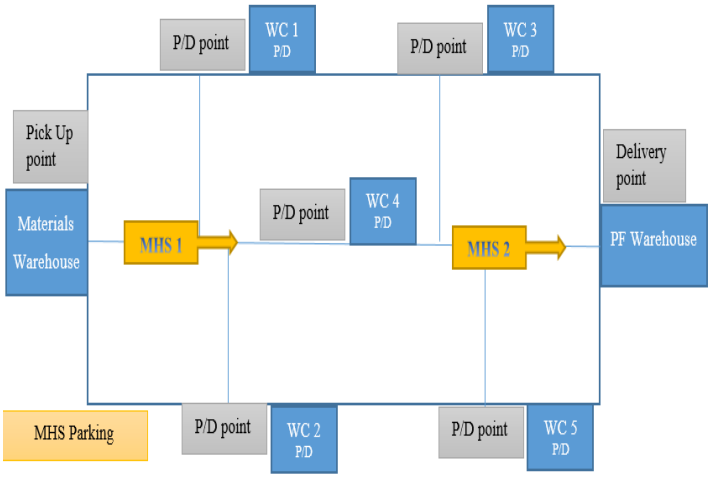


Figure 9. MS model including MHS

The MHS process, in this system, consist of four main steps:

- The unloaded MHS starts from the parking and moves to the pickup point to pick the load and goes to its destination (P/D point) related to the work center;
- The MHS stops to deliver the parts, picks up the new load if it exists;
- The loaded MHS moves to the next point whatever its storage area or work center, then the same operations will be repeated until the end of the procedure;
- MH returns to the parking.

The traveling time depends on: (1) the type of transporter (speed, loading time, unloading time, etc.); (2) the transfer network constraints like dispatching rule, orientation, distance between the departments etc.

5.3. Use of better Performance indicators

The performance indicators used by [21] are divided in two types:

- The main performance indicators are used to evaluate the global MS performance, in terms of tardiness, earliness and DD;
- Diagnostic performance indicators: they are used to resolve resource lack/surplus problems in each work center, which are represented as follows (Table 4):

Table 4.Indicators used by [21]

Indicators	Definition	Objective
Machine utilization rate (RUm)	Machine availability within a period of time	Maximize machine utilization
Labor utilization rate (RUop)	Labor availability within a period of time	Maximize labor utilization
Number of units in queue (WIP)	Total number of parts in queue and under process	Minimize WIP Parts in queue and parts in machine
Waiting time of units in queue	Total waiting time of parts in queue and under process	Minimize the waiting time of Parts in queue and parts in machines
Tardiness (Tr)	Time at which the task is completed after its date.	Minimize tardiness
Earliness (Er)	Time at which the task is completed before its date.	Minimize earliness

We can say that the performance indicators used by[21] are relevant but not sufficient to evaluate the performance of a MS including the MHS. The major shortcoming consists in neglecting the performance measure related to the MHS.

On the other hand, to improve SESA, it is necessary to measure the efficiency of the MHS, which can be characterized by several performance indicators.
In fact, previous research studies, in the field of fleet size, used different performance measures to analyze and evaluate the system performance. [18, 72-77].

Most of the indicators, related to MHS performance are:

- Vehicle utilization rate;
- Throughput rate;
- Vehicle waiting time;
- Vehicle blockage time;
- Empty vehicle traveling time;
- Total number of delivered loads;
- Number of units waiting to be transported;
- Waiting time of units to be transported;
- Availability/ unavailability of MHS.

5.4. Improvement proposal of the ES reasoning mechanism

The ES reasoning mechanism used by SESA is divided into two stages. The first stage aims at resolving machine selection problem without labor constraints (with a number of workers equal to the number of machines in each department). Then, the labor requirement problem resolution starts with a machine balanced MS.

According to our literature review, there has not been yet any expert system developed to select machines, labor and MHS simultaneously. In this work, we try to improve further ES reasoning mechanism in order to select not only the production resources (machines and labor), but also the material handling resources (trucks, pallets, etc.).To attain this objective, it is important to add another stage, in SESA, to resolve MHS fleet sizing problem (see Figure 10).

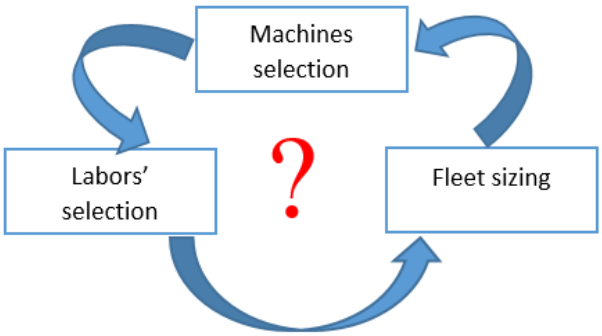


Figure 10. ES reasoning mechanism

The resolution of MHS fleet sizing problem requires new knowledge based rules related to the MHS in order to make the right decision of adding or deleting a MHS. The new stage must be in coherence with the other stages. We can take the example of the Lack and surplus of MHS resources. In fact, If the MHS utilization rate (ur) is less than the lower performance limits (supposed urmin=40%) there is a surplus of vehicles (the number of units waiting to be transported decreases but the MHS cost increases). In addition, If the MHS utilization rate (ur) is more than the higher performance limits (urmax=90%) there is lack of vehicles (the number of units waiting to be transported increases).

Example of the generic diagnosis production rule related to the lack of resources (ur>urmax):

MSS: Manufacturing system state.

If [GlobalPerformance: MSS= = OK] AND [D: ur>= D: urmax] Then D: problem=lack

The ES is a promising tool, as proved in previous section. It can be useful to various types of resource sizing problems. However, integrating an additional stage and finding an optimal order of three stages in the strategy structure is not an easy task. The feasible way to enhance SESA is to start with one problem, by giving a fixed number of resources for the others, then, using the solution of the first stage by solving the next problem, then, using the new information obtained from the second for solving the third problem. Each stage may have several iterations until satisfying results are obtained. In fact, it is necessary to identify the right ordering of selection problems because different ordering can provide different solutions. Unfortunately, the chained stages can address the complexity of the problem when dealing with a high-size problem. This approach becomes challenging to be applied.

Note that it is possible to combine other useful optimization approaches with simulation, such as the Artificial Neural Networks (ANN) Tabu algorithm, Genetic Algorithm, Data Envelopment Analysis (DEA), Regression Metamodeling etc. to identify the required number of resources.

6. Conclusion and Future Work

This paper discusses the originality and the evolution of SESA for sizing MS. The original version was created using static and deterministic tools to solve only the machine selection problems. Later research studies enhanced it by developing new stochastic and dynamic tools and enlarging its application domain by incorporating the labor selection problem. However, SESA is still limited, it neglects the MHS fleet sizing problem despite its influence on the machines and labor quantities. The provided literature survey indicates firstly that the MS sizing problem is continuously relevant due to its importance in ensuring the desired system performance. Secondly, the survey indicates that SESA is a promising approach for the MS sizing problem. The purpose of this study was to extend the application domain of SESA by integrating the MHS fleet sizing problem in the overall MS sizing task. The new SESA framework is thoroughly explained and its application and validation on actual case are proposed as future research.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors gratefully appreciate the reviewers for their valuable remarks and suggestions, which have led to enhance this paper.

References

[1] I. Lajmi, W. Masmoudi, and M. Elleuch, "Importance of Handling Resources' Incorporation Into Simulation Expert System Approach," in 6th IEEE International Conference on Advanced Logistics and Transport (ICALT), pp. 139–143, 2017.

[2] Chtourou Hedi and Guillot Michel, "Une approche prescriptive simulée utilisant un système expert pour l'optimisation des ressources de systèmes

manufacturiers," Actes du 4éme congrés international de génie industriel, pp. 217–26, 1993.
[3] R. Arifin and P. J. Egbelu, "Determination of vehicle requirements in automated guided vehicle systems: A statistical approach," Production Planning & Control, vol. 11, no. 3, pp. 258–270, Jan. 2000.
[4] Chryssolouris, G, Lee, M, Pierce, J, and Domroese, M, "Use of neural networks for the design of manufacturing systems," Manufacturing Review, pp. 187–194, 1990.
[5] R. G. Askin and K. S. Chiu, "A graph partitioning procedure for machine assignment and cell formation in group technology" International Journal of Production Research, vol. 28, no. 8, pp. 1555–1572, Aug. 1990.
[6] F. O., H. Pierreval *, and D. Deflandre, "A simulation-based optimization approach to size manufacturing systems," International Journal of Production Research, vol. 43, no. 2, pp. 247–266, Jan. 2005.
[7] M. B. Yildirim, T. Cakar, U. Doguc, and J. C. Meza, "Machine number, priority rule, and due date determination in flexible manufacturing systems using artificial neural networks," Computers & Industrial Engineering, vol. 50, no. 1–2, pp. 185–194, May 2006.
[8] Z. Hajej, N. Rezg, and T. Askri, "Joint optimization of capacity, production and maintenance planning of leased machines," Journal of Intelligent Manufacturing, Oct. 2018.
[9] A. Rahman, S. Sarker, and M. T. Islam, "Simulating Cutting Line of a Furniture Industry," in 2018 International Conference on Production and Operations Management Society (POMS), pp. 1–7, 2018.
[10] A. Azadeh, S. Nazari-Shirkouhi, L. Hatami-Shirkouhi, and A. Ansarinejad, "A unique fuzzy multi-criteria decision making: computer simulation approach for productive operators' assignment in cellular manufacturing systems with uncertainty and vagueness," The International Journal of Advanced Manufacturing Technology, vol. 56, no. 1–4, pp. 329–343, Sep. 2011.
[11] A. Azadeh, S. M. Asadzadeh, and S. Tadayoun, "Optimization of operator allocation in a large multi product assembly shop through unique integration of simulation and genetic algorithm," The International Journal of Advanced Manufacturing Technology, vol. 76, no. 1–4, pp. 471–486, Jan. 2015.
[12] A. Ammar, H. Pierreval, and S. Elkosantini, "A multiobjective simulation optimization approach to define teams of workers in stochastic production systems," in 2015 International Conference on Industrial Engineering and Systems Management (IESM), Seville, pp. 977–986, Spain, 2015.
[13] S. Samant, V. K. Mittal, and R. Prakash, "Resource optimisation for an automobile chassis manufacturer through value stream mapping enhanced with simulation technique and constraint programming," International Journal of Industrial and Systems Engineering, vol. 28, no. 3, pp. 379–401, Jan. 2018.
[14] J. T. Lin, F.-K. Wang, and P.-Y. Yen, "Simulation analysis of dispatching rules for an automated interbay material handling system in wafer fab," International Journal of Production Research, vol. 39, no. 6, pp. 1221–1238, Jan. 2001.
[15] P.-H. Koo, J. Jang, and J. Suh, "Estimation of Part Waiting Time and Fleet Sizing in AGV Systems," International Journal of Flexible Manufacturing Systems, vol. 16, no. 3, pp. 211–228, Jul. 2004.
[16] D. Raman, S. V. Nagalingam, B. W. Gurd, and G. C. I. Lin, "Quantity of material handling equipment—A queuing theory based approach," Robotics and Computer-Integrated Manufacturing, vol. 25, no. 2, pp. 348–357, Apr. 2009.
[17] H. R. Sayarshad, "Using bees algorithm for material handling equipment planning in manufacturing systems," Int J Adv Manuf Technol, vol. 48, no. 9, pp. 1009–1018, Jun. 2010.
[18] F. F. Choobineh, A. Asef-Vaziri, and X. Huang, "Fleet sizing of automated guided vehicles: a linear programming approach based on closed queuing networks," International Journal of Production Research, vol. 50, no. 12, pp. 3222–3235, Jun. 2012.
[19] V. K. Chawla, A. K. Chanda, and S. Angra, "Automatic guided vehicles fleet size optimization for flexible manufacturing system by grey wolf optimization algorithm," Management Science Letters, pp. 79–90, 2018.
[20] V. Patel, J. Ashby, and J. Ma, "Discrete event simulation in automotive final process system," in Proceedings of the Winter Simulation Conference, vol. 1, pp. 1030–1034 vol.1, 2002.
[21] W. Masmoudi, H. Chtourou, and A. Y. Maalej, "Labor and machine sizing through a Simulation-Expert-System-based Approach," Simulation Modelling Practice and Theory, vol. 15, no. 1, pp. 98–110, Jan. 2007.
[22] H. Chtourou, W. Masmoudi, and A. Maalej, "An expert system for manufacturing systems machine selection," Expert Systems with Applications, vol. 28, no. 3, pp. 461–467, Apr. 2005.
[23] Wassim Masmoudi, H. Chtourou, and Aref Y. Maalej, "A simulation-expert-system-based approach for machine sizing of production systems," Jnl of Manu Tech Mnagmnt, vol. 17, no. 2, pp. 187–198, Feb. 2006.
[24] AshkanNegahban and Jeffrey S.Smith, "Simulation for manufacturing system design and operation: Literature review and analysis - ScienceDirect," Journal of Manufacturing Systems, vol. Volume 33, no. 2, pp. 241-261, Apr. 2014.
[25] D. R. Sule, Manufacturing Facilities : Location, Planning, and Design, Third Edition. CRC Press, 2008.
[26] S. M. Boorla, K. Bjarklev, T. Eifler, T. J. Howard, and C. A. McMahon, "Industry 4.0 – A challenge for variation simulation tools for mechanical assemblies," Advances in Computational Design, vol. 4, no 1, pp. 43-52, Apr. 2019.
[27] C. M. DeRienzo, R. J. Shaw, P. Meanor, E. Lada, J. Ferranti, and D. Tanaka, "A discrete event simulation tool to support and predict hospital and clinic staffing," Health Informatics J, vol. 23, no. 2, pp. 124–133, Jun. 2017.
[28] S. Almagooshi, "Simulation Modelling in Healthcare: Challenges and Trends," Procedia Manufacturing, vol. 3, pp. 301–307, Jan. 2015.
[29] L. Alonso et al., "CityScope: A Data-Driven Interactive Simulation Tool for Urban Design. Use Case Volpe," in Unifying Themes in Complex Systems IX, pp. 253–261, 2018.
[30] X. Song, Y. Wu, Y. Ma, Y. Cui, and G. Gong, "Military Simulation Big Data: Background, State of the Art, and Challenges," Mathematical Problems in Engineering, 2015.
[31] D. Mourtzis, M. Doukas, and D. Bernidaki, "Simulation in Manufacturing: Review and Challenges," Procedia CIRP, vol. 25, pp. 213–229, Jan. 2014.
[32] G. Popovics, A. Pfeiffer, and L. Monostori, "Generic data structure and validation methodology for simulation of manufacturing systems," International Journal of Computer Integrated Manufacturing, vol. 29, no. 12, pp. 1272–1286, Dec. 2016.
[33] A. Polenghi, L. Fumagalli, and I. Roda, "Role of simulation in industrial engineering: focus on manufacturing systems," IFAC-PapersOnLine, vol. 51, no. 11, pp. 496–501, Jan. 2018.
[34] A. Greasley, "Using simulation for facility design: A case study," Simulation Modelling Practice and Theory, vol. 16, no. 6, pp. 670–677, Jul. 2008.
[35] R. Zhang, S. K. Ong, and A. Y. C. Nee, "A simulation-based genetic algorithm approach for remanufacturing process planning and scheduling," Applied Soft Computing, vol. 37, pp. 521–532, Dec. 2015.
[36] D. Krenczyk, W. M. Kempa, K. Kalinowski, C. Grabowik, and I. Paprocka, "Production planning and scheduling with material handling using modelling and simulation," MATEC Web Conf., vol. 112, p. 09015, 2017.
[37] S. Yang, T. Arndt, and G. Lanza, "A Flexible Simulation Support for Production Planning and Control in Small and Medium Enterprises," Procedia CIRP, vol. 56, pp. 389–394, Jan. 2016.
[38] D. Ivanov, "Disruption tails and revival policies: A simulation analysis of supply chain design and production-ordering systems in the recovery and post-disruption periods," Computers & Industrial Engineering, vol. 127, pp. 558–570, Jan. 2019.
[39] W. Ye and F. You, "A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures," Computers & Chemical Engineering, vol. 87, pp. 164–179, Apr. 2016.
[40] Sławomir Kłos and Justyna Patalas-Maliszewska, "The Use of the Simulation Method in Analysing the Performance of a Predictive Maintenance System," in Distributed Computing and Artificial Intelligence, Special Sessions, 15th International Conference, pp. 42–49, 2019.
[41] S. Gallego García and M. García García, "Design and Simulation of Production and Maintenance Management Applying the Viable System Model: The Case of an OEM Plant," Materials, vol. 11, no. 8, p. 1346, Aug. 2018.
[42] J. Stoldt, A. Schlegel, and M. Putz, "Enhanced integration of energy-related considerations in discrete event simulation for manufacturing applications," Journal of Simulation, vol. 10, no. 2, pp. 113–122, May 2016.
[43] E. Romano and D. Iuliano, "A simulation/optimization approach to support resource allocation in a service firms," International Journal of Procurement Management, vol. 11, no. 1, p. 1, 2018.
[44] L. M. S. Dias, A. A. C. Vieira, G. A. B. Pereira, and J. A. Oliveira, "Discrete Simulation Software Ranking: A Top List of the Worldwide Most Popular and Used Tools," in Proceedings of the 2016 Winter Simulation Conference, Piscataway, NJ, USA, pp. 1060–1071, 2016.
[45] F. Hayes-Roth, D. Waterman, and D. Lenat, Building expert systems. 1984.
[46] M. H. Fazel Zarandi, S. Soltanzadeh, A. Mohammadi, and O. Castillo, "Designing a general type-2 fuzzy expert system for diagnosis of depression," Applied Soft Computing, Apr. 2019.
[47] O. T. Nkamgang, D. Tchiotsop, H. B. Fotsin, P. K. Talla, V. Louis Dorr, and D. Wolf, "Automating the clinical stools exam using image processing integrated in an expert system," Informatics in Medicine Unlocked, vol. 15, p. 100165, Jan. 2019.
[48] E. K. Tekez, "Development of an expert system for demand management process," International Journal of Computer Integrated Manufacturing, vol. 31, no. 10, pp. 970–977, Oct. 2018.
[49] Ali Ahmed Mohammed, K. Ambak, A. M. Mosa, and D. Syamsunur, "EXPERT SYSTEM IN ENGINEERING TRANSPORTATION: A REVIEW," vol. 14, p. 25, 2019.
[50] G. Baryannis, S. Dani, S. Validi, and G. Antoniou, "Decision Support Systems and Artificial Intelligence in Supply Chain Risk Management," in

Revisiting Supply Chain Risk, G. A. Zsidisin and M. Henke, Eds. Cham: Springer International Publishing, pp. 53–71, 2019.
[51] Jaskiranjit Kaur and Anjali Walia, "Dimensions of Expert System in Manufacturing Industry," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. Volume 3, p. 7, 2018.
[52] S. Singh, E. U. Olugu, and S. N. Musa, "Development of Sustainable Manufacturing Performance Evaluation Expert System for Small and Medium Enterprises," Procedia CIRP, vol. 40, pp. 608–613, Jan. 2016.
[53] K. Prasad, S. R. Maity, and S. Chakraborty, "Quality Function Deployment-based Expert System for Materials Selection," 2016.
[54] W. Hellmann, D. Marino, M. Megahed, M. Suggs, J. Borowski, and A. Negahban, "Human, AGV or AIV? An integrated framework for material handling system selection with real-world application in an injection molding facility," Int J Adv Manuf Technol, vol. 101, no. 1, pp. 815–824, Mar. 2019.
[55] R. Carreño, V. Aguilar, D. Pacheco, M. A. Acevedo, W. Yu, and M. E. Acevedo, "An IoT Expert System Shell in Block-Chain Technology with ELM as Inference Engine," International Journal of Information Technology & Decision Making, Jan. 2019.
[56] M. S. Amalnik, "Expert System Approach for Optimization of Design and Manufacturing Process for Rotary Ultrasonic Machining," vol. 11, no. 1, p. 13, 2018.
[57] H. Esen, T. Hatipoğlu, A. Cihan, and N. Fiğlali, "Expert system application for prioritizing preventive actions for shift work: shift expert," International Journal of Occupational Safety and Ergonomics, vol. 25, no. 1, pp. 123–137, Jan. 2019.
[58] C. F. Tan, S. N. Khalil, J. Karjanto, L. S. Wahidin, W. Chen, and G. W. M. Rauterberg, "An Expert Machine Tools Selection System for Turning Operation," IOP Conf. Ser.: Mater. Sci. Eng., vol. 88, p. 012044, Sep. 2015.
[59] C. Urrea, G. Henríquez, and M. Jamett, "Development of an expert system to select materials for the main structure of a transfer crane designed for disabled people," Expert Systems with Applications, vol. 42, no. 1, pp. 691–697, Jan. 2015.
[60] M. Ristić, M. Manić, D. Mišić, M. Kosanović, and M. Mitković, "IMPLANT MATERIAL SELECTION USING EXPERT SYSTEM," Facta Universitatis, Series: Mechanical Engineering, vol. 15, no. 1, pp. 133–144, Apr. 2017.
[61] M. Soleymani and M. O. Nejad, "Supply Chain Risk Management using Expert Systems," International Journal of Current Engineering and Technology, vol. 8, no. 04, 2018.
[62] P. S. Kang and R. S. Bhatti, "Continuous process improvement implementation framework using multi-objective genetic algorithms and discrete event simulation," Business Process Mgmt Journal, Sep. 2018.
[63] A. Burduk, K. Grzybowska, and G. Kovács, "Hybrid model of an expert system for assessing the stability of a production system," LogForum, vol. Vol. 14, no. 4, 2018.
[64] A. Singh, D. Philip, J. Ramkumar, and M. Das, "A simulation based approach to realize green factory from unit green manufacturing processes," Journal of Cleaner Production, vol. 182, pp. 67–81, May 2018.
[65] J.-P. Arnaout, G. Arnaout, and J. E. Khoury, "Simulation and Optimization Of Ant Colony Optimization Algorithm For The Stochiastic Uncapacitated Location-Allocation Problem," Journal of Industrial and Management Optimization, pp. 1215–1225, Oct. 2016.
[66] K. Nagadi, L. Rabelo, M. Basingab, A. T. Sarmiento, A. Jones, and A. Rahal, "A hybrid simulation-based assessment framework of smart manufacturing systems," International Journal of Computer Integrated Manufacturing, vol. 31, no. 2, pp. 115–128, Feb. 2018.
[67] D. Stefanovic, N. Stefanovic, and B. Radenkovic, "Supply network modelling and simulation methodology," Simulation Modelling Practice and Theory, vol. 17, no. 4, pp. 743–766, Apr. 2009.
[68] B. Hernández-Vera, A. A. A. Lasserre, M. G. Cedillo-Campos, L. E. Herrera-Franco, and J. Ochoa-Robles, "Expert System Based on Fuzzy Logic to Define the Production Process in the Coffee Industry," Journal of Food Process Engineering, vol. 40, no. 2, p. e12389, 2017.
[69] A. Azadeh, H. Tohidi, M. Zarrin, S. Pashapour, and M. Moghaddam, "An integrated algorithm for performance optimization of neurosurgical ICUs," Expert Systems with Applications, vol. 43, pp. 142–153, Jan. 2016.
[70] H. Behret, B. Öztayşi, and C. Kahraman, "A Fuzzy Inference System for Supply Chain Risk Management," in Practical Applications of Intelligent Systems, pp. 429–438, 2012.
[71] A. Hadi-Vencheh and A. Mohamadghasemi, "A new hybrid fuzzy multi-criteria decision making model for solving the material handling equipment selection problem," International Journal of Computer Integrated Manufacturing, vol. 28, no. 5, pp. 534–550, May 2015.
[72] B. Mahadevan and T. T. Narendran, "A hybrid modelling approach to the design of an AGV-based material handling system for an FMS," International Journal of Production Research, vol. 32, no. 9, pp. 2015–2030, Sep. 1994.
[73] B. M. Beamon and V. C. P. Chen, "Performability-based fleet sizing in a material handling system," The International Journal of Advanced Manufacturing Technology, vol. 14, no. 6, pp. 441–449, Jun. 1998.
[74] Hung, P. C. and Liu, F. H, "Estimation of the fleet size for a multi-load automated guided vehicle system," Proceedings of the National Science Council, Part A: Physical Science and Engineering, pp. 244–253, 2001.
[75] S. Rajotia, K. Shanker, and J. L. Batra, "Determination of optimal AGV fleet size for an FMS," International Journal of Production Research, vol. 36, no. 5, pp. 1177–1198, May 1998.
[76] K.-H. Chang, A.-L. Chang, and C.-Y. Kuo, "A simulation-based framework for multi-objective vehicle fleet sizing of automated material handling systems: an empirical study," Journal of Simulation, vol. 8, no. 4, pp. 271–280, Nov. 2014.
[77] R. Wibisono, T. J. Ai, and D. R. Yuniartha, "Fleet Sizing of Automated Material Handling Using Simulation Approach," IOP Conf. Ser.: Mater. Sci. Eng., vol. 319, p. 012030, Mar. 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 220-226 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Brain-inspired IoT Controlled Walking Robot - Big-Foot

Anna Lekova*[,1], Ivan Chavdarov[1,2], Bozhidar Naydenov[1], Aleksandar Krastev[1], Snezhanka Kostova[1]

[1]*Bulgarian Academy of Sciences, Institute of Robotics, Str. Acad. G. Bonchev, Blk. 2, 1113, Sofia, Bulgaria*

[2]*University of Sofia, "St. Kliment Ohridski", Faculty of Mathematics and Informatics, 5 James Bourchier Blvd. 116, Sofia, Bulgaria*

ARTICLE INFO

Article history:
Received:14 April, 2019
Accepted:04 June, 2019
Online: 18 June, 2019

Keywords:
Brain-robot interaction
Walking robot
3D printing

ABSTRACT

This work presents the development of an original idea for a walking robot with a minimum number of motors, simple construction and a control system based on the brain bioelectrical activities. Described are geometric and kinematic dependencies related to the robot movement, as well as brain-inspired IoT control method. Various aspects are discussed for improving the robot's qualities, concerning the shape of the robot's feet and base in order to overcome various obstacles and maintain the static mechanical equilibrium. Improvements in the mechanical design are provided to improve reliability and enhance the scope of robot's applications. A new IoT framework for creating Human-robot interaction applications based on Node-RED "wiring" of Emotiv Brain Computer Interface (BCI) and Arduino based robot is designed, developed and tested. An educational application how to train the joint attention of children by a mind control method based on neurofeedback from beta oscillation in the right temporoparietal region is illustrated in a Node-RED flow. The neurofeedback is exposed on the walking robot.

1. Introduction

Applications of walking robots include rescue operations, work and inspection in harmful and dangerous environments, military purposes, etc. Typically, they move in an environment with obstacles whose positions and dimensions are not known in advance. Their mechanical and control systems are designed to be able to avoid or overcome obstacles. In some cases, the obstacles can dynamically change in time. This leads to more complicated designs of the walking robots compared to robots on wheels or tracks. They have more degrees of freedom and are slower. A common problem is the task of climbing and descending stairs [1-5]. Experimental robots are developed with a small number of degrees of freedom and a special shape of their feet that overcome obstacles while maintaining static stability. [7]. For these reasons, alternative solutions are investigated [5], [6] and [7]. In [7], the author presented a low-budget two-legged robot which is able to maintain static equilibrium. Other simple solutions are also investigated, such as: passive-dynamic two-legged walking [7] and different variants inspired by nature [1], [2] and [8].

Walking robots are also used for educational purposes. The results in the scientific literature show positive reactions and improvement in the attentional and positive emotional state of children with special educational needs (SEN) [9]. Nonhumanoid walking robot Big-Foot is successfully implemented to support the education of such children in two day-care centers for children with SEN in Bulgaria [10]. Some ideas have been challenged with the special educators - how the walking robot to become more intelligent and personal, in order to act as a mediator for learning and socializing because these children show deficits in early social communication skills such as Joint Attention (JA), social requesting and referencing. Joint attention is the shared focus of two individuals on an object and gaze shifting and behavioral response are the most used measures to assess the establishment of JA [11]. However, children with ASD avoid eye contact or lose focus on humans quickly. In this context, our hypothesis is that the robot has the potential to establish JA better because these children trust robots more than humans [10] and we can use this as a pure social consequence of sharing an experience. Furthermore, the JA can be made visible on a robot by exploiting a neuro physiological approach instead of gaze tracking as an indicator for a shared focus of two individuals on a same object. By integrating walking robot with a brain-aware device we propose an innovative concept for establishing and assessing JA in more objective way. For example, the Big-Foot will climb stairs only if the active brain-patterns correlated to attention system of the human brain are observed and evaluated.

*Anna Lekova, +359887435648, alekova.iser@gmail.com

The proposed nontraditional brain-inspired Robotics intervention should run anytime and anywhere with a remote supervision of the special educator over the WiFi in order to aid children not only in schools but in family environment. Therefore, the proposed framework should comply to the concept for human-robot personal communication by augmenting intelligence to robot, however not digitally by pushing a button, clicking, dragging or speaking, but biologically and continuously through emotions, mental intentions or even via chemically released by neurons rewards [12].

All this imposes technical challenges concerning robot design, proper neuroscience computing and ubiquitousness. Movement of walking robots is accomplished in two ways: motion with maintaining static stability or use of dynamic gaits. When the number of legs is small, less than 4, maintaining static stability is not an easy task as it is necessary to change the center of gravity, which requires additional degrees of freedom. Overcoming obstacles with few supports is even harder. This article discusses a robot design that has only three supports. In this case the problem of maintaining static stability is overcome by increasing the area of these supports and using suitable shapes and materials.

The technical challenges concerning ubiquitous computing are how to merge people, processes, devices and technologies with sensors and actuators. We exploited the idea behind the Internet of Things (IoT) and the innovative" Cloud Computing" infrastructure [13]. Thus, all sensing, computation, and memory can be integrated into a single standalone Socially-assistive Robotics system. Node-RED [14] is an open source development tool built by IBM, which allows to wire up IoT as nodes in flows. Node-RED is built on Node.js and can run anywhere if the applications are capable of hosting node.js, such as small single board computers like the Raspberry Pi, personal laptops or in cloud environments, such as the IBM Cloud. The Node-RED connectivity allows nodes to collect and exchange data ubiquitously and its flow-based programming is an ideal solution to wire up the biological brain intelligence to robots anytime and anywhere. Based on the idea behind IoT, that uniquely addressable "things" communicate with each other and transfer data over the existing network protocols, we propose how the information channel between the human brain and external devices to be applied for IoT brain-to-robot control. By analogy to Visual servoing [15] and Tactile servoing [16], we defined a term "Brain servoing" (or brain-based robot control) that uses EEG feedback information extracted from the brain EEG sensor to control the motion of a robot. The control tasks intend to translate a specific brain activity interaction patterns in robot commands ubiquitously. The control instructions are transferred from continuously decoded JA performance metric into robot commands and are sent to the robot actuators via a Node-RED set-up Emotiv Brain-Computer Interface (BCI) to Arduino.

In this study we illustrate a non-traditional control method where the brain electrical activity is captured by EMOTIV brain-listening headset [17] and specific spatial and temporal brain frequencies correlated to JA are translated into commands to control the walking robot Big-Foot. Because the robot Big-Foot has a simple and innovative design, children with specific needs find it attractive, and it does not create feelings of anxiety and discomfort when interacting with it. This helps the robot function as a mediator between the children and the therapist. The 3D model of the robot is cheap and easy to be controlled by children. It is extremely maneuverable and can climb stairs. These features allow it to be used in educational games for children. To the best of our knowledge we first propose an IoT framework for creating Human Robot Interaction (HRI) applications based on Node-RED "wiring" of Emotiv BCI [18] and Arduino based robot.

2. Mechanical design and improvement of the robot

Fig. 1 presents the design of the robot and Table 1 lists its main components. Two engines are mounted in the body 5. The rotor of motor 2 is connected to and rotates the circular base 1. This allows the robot to change its orientation and turn when the feet 4 are raised above the ground. The motor 3, via a connecting shaft, drives the arms 7. At the end of the arms 7, the feet 4 are mounted. The feet maintain a constant orientation with respect to the base 1 and the body 5 by means of two gear mechanisms 6 with a gear ratio of $i=1$. This design allows the robot to rotate more than 360 degrees around axis $R1$. There is also no limitation of the rotation around axis $R2$. The two rotations are reversible, making the robot extremely maneuverable.

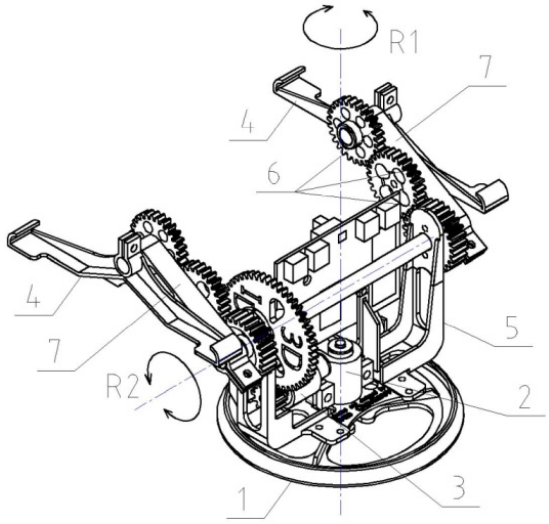


Figure 1: 3D model of the robot Big-Foot

Table 1. Part list of the main components

Position	Part description
1	Circular base
2	DC motor for rotating the base
3	DC motor for rotation of the arms
4	Feet
5	Body of the robot
6	Gear transmission mechanisms
7	Rotating arms

When the robot moves on a flat terrain it passes through two main phases. During the first phase, the feet 4 are stationary on the ground. The arms 7 move, and the body 5, together with the base 1, is moved at a distance of one step S (Fig. 2a). All points on the robot's body move along trajectories that represent arcs of circles, for example, the trajectories of points B and P on Fig. 2. The radius r_{AB} is determined by the length AB of the arms 7. During this phase, the arms 7 are rotated at an angle α

Similarly, to the first phase, the trajectories of the points of the feet are also moving along arcs of a circle with radius r_{AB}. In this case, the angle of rotation of the arms 7 is $(2\pi - \alpha)$ (Fig. 2b). During this phase, the robot does not move forward but can rotate around axis *R1* (Fig. 1). The second phase ends when the feet touch the ground. The movements during the two phases are cyclically repetitive.

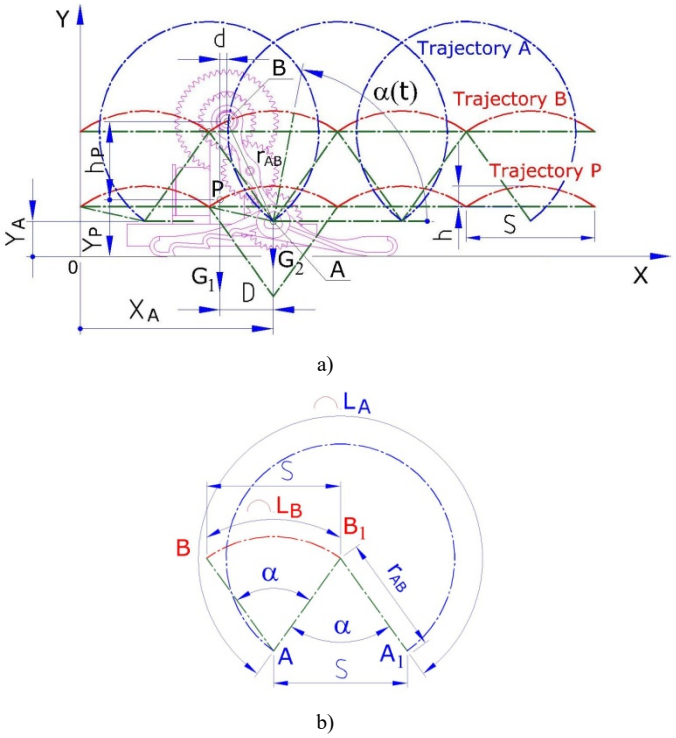
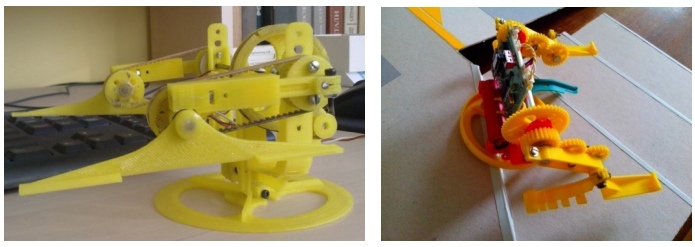


Figure 2: Trajectories of different points of the robot during the two phases of motion - a) and b) step S and angles of rotation of arm AB.

During the second phase, the robot's feet rise to a relatively high height, which helps to overcome high obstacles. The behavior of the robot has been verified through computer simulation and experiments with 3D printed prototypes.

More detailed descriptions of the kinematics and behavior of the robot in the various phases of its movement are described in [8], [19] and [20].

In the original idea described in patent [19], maintaining of parallelism between the feet and the base is achieved by a belt mechanism. The 3D printed prototype (Fig. 3a) showed that the belt drive is not suitable due to the need for considerable tensioning. The drawbacks of this model, described in more details in [20], have led to the need of improvement of the mechanical design.



a) b)

Figure 3: a) First 3D printed model, b) model with improved design climbing stairs.

Improvements have been made to the mechanical structure for a more reliable torque transmission. A new type of joint between the shaft and the feet is used (Fig. 4, position 8), which significantly reduces the stress concentration. This joint is realized by a smooth transition from a polygon to a circle, and the 3D printing technology allows its physical implementation and application in our model [20], [29].

Changing the shape of the feet and the round base makes it possible to overcome higher obstacles (Fig. 4, positions 9 and 10).

Another approach is used to improve the capabilities of the robot, namely to add additional elements. Two tails have been added (Fig. 5, position 11), which prevent the robot from rolling over when overcoming high obstacles.

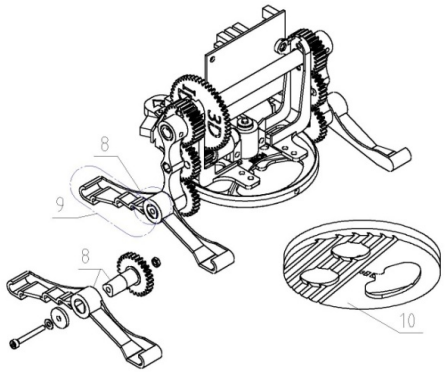


Figure 4: Improvements in the design of the Big-Foot robot

The addition a platform (Fig. 5, position 12) is used for applications of the robot for developing games for children with specific needs.

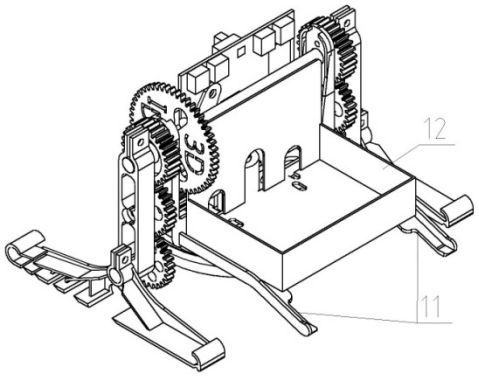


Figure 5: Walking robot with tails and a paltform.

3. Human mind control method of robot with Node-RED EmotiveBCI-to-Arduino interface

The most common way to control a robot is the joystick or mobile device. However, for some people the joystick or mobile control is difficult or impossible for many reasons and other (non-traditional) control methods have being developed based on the recent innovative sensors and technologies, such as motion sensing devices, etc. In this section, we present a non-traditional control method that rely on "Brain servoing" which uses feedback information extracted from a brain sensor (EEG feedback) to control the walking robot.

A traditional bi-directional DC control of two motors (operating at a voltage of 2.7-10.8V) is used. The current is about 1.2A per motor and the load on motor is up to 2A for few seconds. The H-bridge DC motor control is used in order to rotate motors in both directions. Four transistors in the circuit are controlled in pairs and play a role of switches to control the motor to rotate in both directions. Fig. 6 presents the connecting scheme and Fig. 7 the code uploaded on the ARDUINO MICRO. The end-user interfaces via dialog box for laptop or from mobile device are presented in Fig. 8.

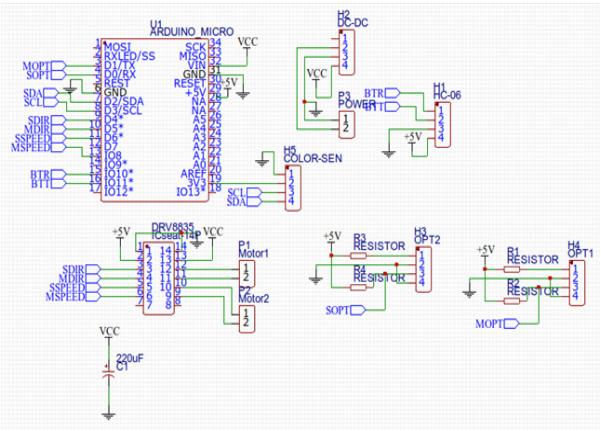


Figure 6: Connecting scheme

```
bigFoot2 | Arduino 1.8.8 (Windows Store 1.8.19.0)
File  Edit  Sketch  Tools  Help
bigFoot2 §

#define MOTOR_1_L 3
#define MOTOR_1_R 4
#define MOTOR_2_L 10
#define MOTOR_2_R 11
char  temp;
void motor_drive(byte stat_m1_r, byte stat_m1_l, byte stat_m2_f,
  // SET 1 or 0 to pins -> motors
  // L and R
  digitalWrite(MOTOR_1_R, stat_m1_l);
  digitalWrite(MOTOR_1_L, stat_m1_r);
  // F and B
  digitalWrite(MOTOR_2_R, stat_m2_f);
  digitalWrite(MOTOR_2_L, stat_m2_b);
void setup() {
 Serial.begin(9600);
  pinMode(MOTOR_1_L , OUTPUT);  pinMode(MOTOR_1_R , OUTPUT);
  pinMode(MOTOR_2_L , OUTPUT);  pinMode(MOTOR_2_R , OUTPUT);
}
void loop() {
if(Serial.available() > 0) {
        temp = Serial.read();
switch (temp) {
  case 'F':
    motor_drive(0, 0, 1, 0);
```

Figure 7: The code for ARDUINO MICRO

The nontraditional control method intends to translate in robot commands a specific brain activity interaction patterns within a framework of IoT. Fig. 9 illustrates in general this framework for Brain-robot interaction and its deployment using a Node-RED tool. An application of EmotivBCI-to-Arduino Node-RED flow is explained in Section V. The commercial EEG devices, Emotiv EPOC or Emotiv Insight [17] could be used. The placement of electrodes are shown in Fig.10. In this study EEG data are acquired and recorded with the 14-channel neuroheadset EPOC+ and the EEG signals were sampled at the rate of 256 Hz. The data are wirelessly transmitted to a host computer through Bluetooth and further processed with EmotivBCI Node-RED Toolbox [18].

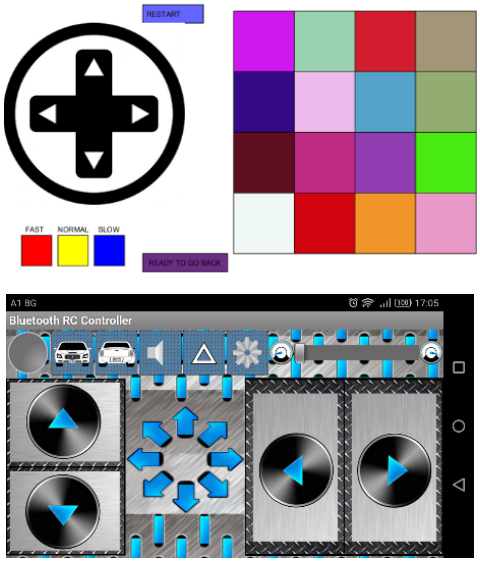


Figure 8: Traditional control methods of walking robot BigFoot



Figure 9: IoT framework for Brain-robot interaction - Node-RED EmotivBCI-to-Arduino Interface.

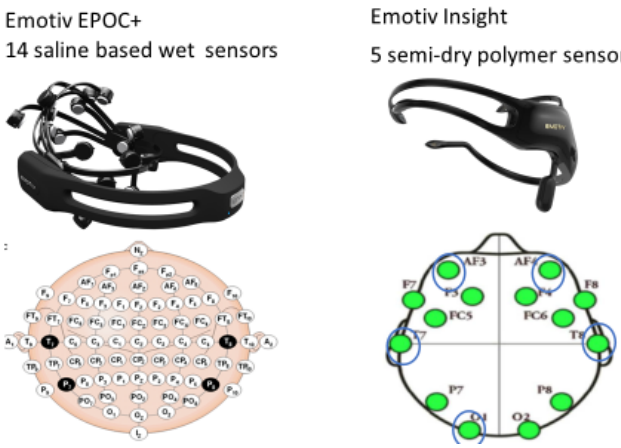


Figure 10: EMOTIV EEG headsets for EEG-based BCI - listen, record and transmit in real time the electrical activity of the brain

Node-RED uses a visual programming approach for 'wiring together' of code blocks and make up 'flows' to carry out tasks. It connects nodes as a combination of input nodes, processing nodes and output nodes in a browser-based flow editor using a wide range of nodes in the palette. The EmotivBCI Node-RED Toolbox is a custom library of input nodes for Node-RED which allow interfacing the EMOTIV technology with other Node-RED nodes and thus to create a wide variety of BCI integrations. Installation, node descriptions and use are presented in [18]. EmotivBCI Node-RED gets data from Emotiv Cortex (Cortex is built on JSON) and WebSockets for creating BCI applications and integrates data streams from the human headset with third party software or hardware. In a browser-based flow in a Web page, a user is

allowing to control the BigFoot rotation and direction of both motors using one of the following brain activity correlated patterns: in the frequency domain, correlated with human cognitive, facial and emotional states or directly by preliminary trained mental commands. EMOTIV offers the opportunity for the user to create and execute a number of Mental Commands [21], such as: push, pull, lift, drop, left, right, rotate left, rotate right, rotate forwards, rotate backwards, rotate clockwise, rotate anticlockwise, and disappear. The detected facial expressions are blink, left wink, right wink, raised eyebrows (surprise), furrowed brows (frown), smile and clenched teeth. The Emotiv technology currently offers five performance metrics detections: Engagement/Boredom, Frustration, Meditation, Instantaneous Excitement, and Long-Term Excitement, that are based on universal in nature brainwave characteristics and don't require an explicit training of the user [22].

All these correlated frequency patterns or detections are interpreted in the Node-RED function blocks and map to commands in order to control the walking robot by sending dynamically information to the serial port where the Arduino is connected. There are several ways to interact with an Arduino using Node-RED. As the Arduino appears as a Serial device, the Serial in/out nodes can be used to communicate with it after adjusting the serial port speed (baud rate) to be the same at both ends. We wired the EmotivBCI nodes to serial port and Arduino Software (IDE). We program the Arduino with the IDE, and then send and receive input over the serial port to interact with BigFoot. Two browser-based flows in a Web page, as well as Node-RED Graphical User Interface (GUI) are designed to allow the user to control the BigFoot motors dynamically by its own brain activity patterns in the frequency domain corresponding to the joint attention. Every 1 or 2 s a string for one of the following commands - 'S' for stop, 'F' for forward, 'B' for backward, 'L' for left rotation and 'R' for right rotation is sent. Except the neurofeedback exposed on the BigFoot, the GUI shows the rotation and direction as a feedback in the web page.

4. Application of the IoT framework for brain- robot control method

The walking robot is used in the development of educational play-like activities of children with SEN. Trough playing with the robot, the children improve their special orientation abilities, they learn easily and with fun by controlling the robot [11]. The interaction of children with the walking robot engages them to communicate with each other and develop their joint attention (JA). By using the non-traditional mind control methods, a neurofeedback rehabilitation is possible and will be effective for training the attention or emotion self-regulation of the brain function. We place the child's neurofeedback in the play-with-robot interventions and expose it on the walking robot.

Joint attention precedes the development of children mentalization skills [23]. Mentalization is the ability to understand the mental state, of oneself or others, that underlies evident behavior. Brain activity patterns in time and frequency domains correlated with JA are discussed in many papers [23], [24] and [25]. According to studies [25] and [26] the oscillatory brain activity in the alpha and beta ranges in the right temporoparietal region correlates with the anticipation and prediction of another person's responses and preferences. Authors in [23] have tested whether neuronal activity preceding JA correlates with mentalization in typically developing (TD) children and whether this activity is impaired in children with autistic spectrum disorder (ASD) who evidence deficits in JA and mentalization skills. TD children shown beta rhythm (15-25 Hz) in the temporoparietal region preceding the JA behavior, while ASD children did not show an increase in beta activity. In the study [23] statically significant difference in increasing the beta band power in the right parietal group of channels is found and again the data analysis suggested that the right temporoparietal region and the middle/superior frontal gyrus are the main brain regions contributing to the beta power differences between the two groups. Based on these neuroimaging findings we designed a play where a child can move the BigFoot to climb the stairs through a neurofeedback from the oscillatory brain activity in the alpha and beta ranges in the right temporoparietal region. Thus, we try to teach and train children joint attention by learning to modulate their rhythm power in the beta frequency band [15-25 Hz] in order to move the walking robot. The power spectrum of right temporoparietal lobe intensity is obtained by EmotivBCI tool and the changes in JA are calculated based on functional brain imaging in terms of event-related desynchronization (ERD) or event-related synchronization (ERS) [26].

According to Pfurtscheller [26] sensory and cognitive processing results in changes of the ongoing EEG in form of ERD/ERS that are highly frequency-band specific. For example, oscillations with 10 Hz comprise more synchronized neurons than oscillations with 40 Hz. The ERD is interpreted as a correlate of an activated cortical area with increased excitability and the ERS in the alpha and lower beta bands can be interpreted as a correlate of a deactivated cortical area [26]. Furthermore, the frequency of brain oscillations depends on the percentage of a population of neurons synchronized. With an increasing number of synchronized neurons, the average frequency becomes slower and if only 10% are synchronized, the amplitude is 10-fold the activity of the 90% of not-synchronized neurons.

In terms of information theory, a desynchronized system represents a state of maximal readiness and a maximum of information capacity [27]. The neuroscience explanation is the ERD, i.e. in the underlying neural network small areas of neurons involved in a particular neural computation (neural ensemble), work in a relative independent or desynchronized manner.

We use (8) to obtain the percentual decease (ERD) or increase (ERS) in the band power during a test (activation) interval compared with a baseline (reference) interval. The ERD/ERS index foe alpha and beta bands are the respond to different levels of JA:

$$ERD/ERS\% = (A-R) / R*100 \quad (8)$$

where A is the power within the frequency band of interest in the activity period and R is the preceding baseline or reference period. Positive numbers are obtained for ERS% and negative - for ERD% that reflect synchronization and a state of band power decrease. We tested our research hypothesis in pilot experiments with five students (right handed male and female in average age about 18 years old) by measuring whether the amount of alpha ERD increase and the amount of beta ERS increase with higher level of

JA and mentalization skills. The mean threshold of 5% for ERS in the beta rhythm was defined experimentally. We plan to perform the experiment explained in [23] with children with ASD in order to compare the establishment and assessment of JA based on the intervention and observations of the gaze shifting and behavioral response from practitioners versus establishment and assessment of JA by the brain-inspired robot control and a neurofeedback response.

The control commands are transferred from continuously decoded JA performance metric into robot commands and are sent to the robot actuators via Node-RED EmotivBCI-to-Arduino Interface. For this application, only the right temporal T8 and right parietal P8 electrodes are used for right-handed participants and their beta and alpha power are analyzed during the neutral or joint attention states. The power spectrum density for beta band is received continually from the EMOTIV Node-RED EEG streaming data. The command 'S' is sending to COM4 by default. In a function node, the ERS and ERD are calculated for each electrode and when alpha ERD significantly decrease and beta ERS increase (thresholds in %, specific for the participant), the corresponding command 'F' or 'B' is sent to COM4. The results showed that robot could be successfully navigated by the positive values of 5% for ERS in the beta rhythm, which is a correlate with joint attention on walking robot in order to complete a navigation task to climb stairs by human intention. In the future we will assign the percentual increase into fuzzy sets in order to define levels of difficulty for climbing stairs.

The Node-red flow for brain-robot interaction based on EmotivBCI toolbox is shown in Fig.11. After streaming the row data from the Emotiv node in the first Node-RED flow, the power of frequency bands of interest are stored in global variables. They are accessed in the second flow (Fig.12) to set-up the baseline settings in the frequency band of interest for 1 or 2 min. How the robot commands are mapped and sent to serial port can be seen in the first Node-RED flow.

We intend to test the proposed framework with a neurofeedback from the frontal theta for training and remembering the orientation in space and implicitly to expose it by the walking robot. The increase in theta power during successful encoding of new information is discussed from neurological point of view in [28] with the relationship with hippocampal theta induced in the cortex.

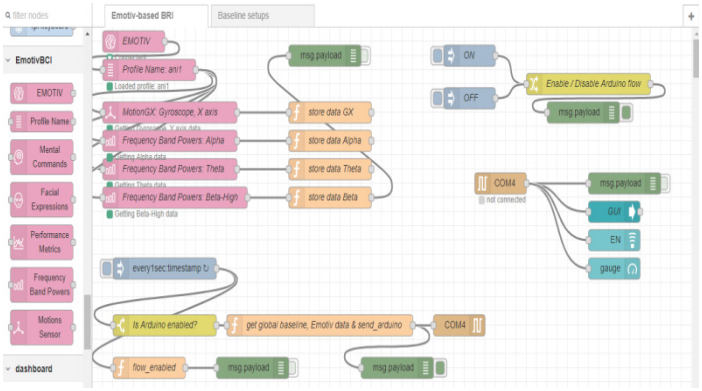


Figure 11: Node-red flow for EmotivBCI-Arduino interaction

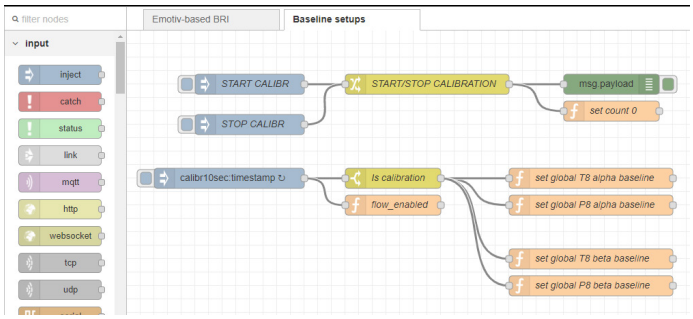


Figure 12: Node-red flow for baseline setting in the frequency band of interest

5. Conclusion

Although the proposed walking robot has a minimum number of mechanical components, it is extremely maneuverable. It has only two motors that makes it simple to control and suitable for educational or rehabilitation purposes. The robot is not expensive and easy to manufacture with a 3D printer. Several control methods are proposed, one of them is "Brain servoing" that uses EEG feedback and brain activity patterns for mind-based robot control. The proposed IoT framework for creating Human Robot Interaction (HRI) applications based on Node-RED "wiring" of Emotiv BCI and Arduino based robot has been applied for neurofeedback training. The results showed that robot could be successfully navigated by the level of human attention to complete a navigation task for climbing stairs.

In the future, the robot's design and control system will be enhanced by collecting and recording information from more sensors. Since, the proposed framework for mind-robot control is enough general, it will be easily applied for other areas of neurons involved in a particular neural computation, as well as other brain sensors and different humanoid or nonhumanoid robots in the IoT.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research is supported by the National Scientific Research Fund, Project N DH17/10. We thank to all high school students from the "National Professional High School for Computer Technologies and Systems", Pravets, Bulgaria, who participated in the experiments and implementation of the Arduino hardware.

References

[1] Ben-Tzvi P., Ito S., Goldenberg A.A. (2009), A mobile robot with autonomous climbing and descending of stairs, Robotica 27(2), pp. 171-188, https://pdfs.semanticscholar.org/0550/5591df73849ecba97e8754d518eda1bead41.pdf

[2] Poramate Manoonpong, Sakyasingha Dasgupta, Dennis Goldschmidt, Florentin Wörgötter, Reservoir-based online adaptive forward models with neural control for complex locomotion in a hexapod robot, 2014, International Joint Conference on Neural Networks (IJCNN), Pages: 3295 – 3302.

[3] Liu Juan Xiu, Chen Qing Wei, Dynamics and control study of a stair-climbing walking aid robot, 2010 International Conference on Mechanic Automation and Control Engineering, Pages: 6190 – 6194.

[4] Chih-Hsing Liu, Meng-Hsien Lin, Ying-Chia Huang, Tzu-Yang Pai, Chiu-Min Wang, The development of a multi-legged robot using eight-bar linkages as leg mechanisms with switchable modes for walking and stair climbing,

2017, 3rd International Conference on Control, Automation and Robotics (ICCAR), Pages: 103 – 108.
[5] B. Jeon, Y. Jeong, K. Kwak, S. Yeo, D. Ha, S. Kim (2009), Bio-Mimetic Articulated Mobile Robot overcoming stairs by using a slinky moving mechanism, Proceedings of ICAD2009, The Fifth International Conference on Axiomatic Design, Campus de Caparica, March 25-27, 2009, pp.173-179.
[6] Fei Sun, He Hua Ju, Ping Yuan Cui, A new 12 DOF biped robot's mechanical design and kinematic analysis, Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, 2011, Pages: 2396 – 2400.
[7] N. E. Nava Rodriguez* G. Carbone† M. Ceccarelli Design Evolution of low-cost humanoid robot CALUMA, 12th IFToMM World Congress, Besançon (France), June18-21, 2007 University Cassino, Italy, https://pdfs.semanticscholar.org/bde8/0efbb0328483ff611f82ad035981ab3a133e.pdf
[8] Chavdarov I. , Walking robot realized through 3d printing, Comptes rendus de l'Acad´emie bulgare des Sciences, Tome 69, No 8, 2016, pp. 1069-1076, http://www.proceedings.bas.bg/cgi-bin/mitko/0DOC_abs.pl?2016_8_13
[9] Kruse T., Pandey A., Alami R., Kirsch A.. Human-Aware Robot Navigation: A Survey. Robotics and Autonomous Systems, Elsevier, 2013, 61 (12), pp.1726-1743. hal-01684295
[10] Weisberg H., Jones E. Individualizing Intervention to Teach Joint Attention, Requesting, and Social Referencing to Children with Autism. Behavior Analysis in Practice (2019) 12:105–123
[11] Dimitrova, M., Lekova, A., Kostova, S., Roumenin, C., Cherneva, M., Krastev, A., Chavdarov, I. (2016) A multi-domain approach to design of CPS in special education: Issues of Evaluation and Adaptation. Proceedings of the 5th Workshop of the MPM4CPS COST Action, November 24-25, 2016, Malaga, Spain, pp.196-205.
[12] Lekova A., Pavlov V., Chavdarov I., Krastev A., Datchkinov P., Stoyanov I.. Augmented Intelligence For Teaching Robots By Imitation, International Scientific Journal "Industry 4.0" Web Issn 2534-997x; Print Issn 2543-8582, Year Ii, Issue 5, P.P. 201-204 (2017), pp. 201-204
[13] Q. Zhang, L. Cheng, and R. Boutaba, ``Cloud computing: State-of-theartand research challenges," J. Internet Services Appl., vol. 1, no. 1,pp. 718, 2010.
[14] IBM NodeRED Flow-based programming for the Internet of Things. URL: https://nodered.org/
[15] Agin, G.J., "Real Time Control of a Robot with a Mobile Camera". Technical Note 179, SRI International, Feb. 1979
[16] Q. Li, C. Sch¨urmann, R. Haschke, and H. Ritter, "A control framework for tactile servoing," in Robotics: Science and Systems, 2013
[17] Martinez-Leon, J., Cano-Izquierdo, J. & Ibarrola, J. Are low cost Brain Computer Interface headsets ready for motor imagery applications? Expert Syst Appl 49, 136 (2016).
[18] EmotivBCI Node-RED Toolbox.https://emotiv.gitbook.io/emotivbci-node-red-toolbox/node-descriptions-and-use
[19] I. Chavdarov, T. Tanev, V. Pavlov Patent application 111362. „Walking Robot", Published summary - Billetin No. 6, 30.06.2014., page 11, in Bulgarian, http://www.bpo.bg/images/stories/buletini/binder-2014-06.pdf
[20] I. Chavdarov, B. Naydenov, S. Kostova, A. Krastev, A. Lekova, "Development and Applications of a 3D Printed Walking Robot - Big-Foot", 2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), https://ieeexplore.ieee.org/document/8555843/references#references
[21] Create and execute a number of Mental Commands in EMOTIV. URL: https://emotiv.zendesk.com/hc/en-us/articles/201216335-Training-Mental-Commands
[22] Understanding the Performance Metrics Detection Suite. URL: https://emotiv.zendesk.com/hc/en-us/articles/201444095-Understanding-the-Performance-Metrics-Detection-Suite
[23] Soto-Icaza P., Vargas L., Aboitiz F., Billeke P, Beta oscillations precede joint attention and correlate with mentalization in typical development and autism, Cortex, Volume 113, 2019, Pages 210-228
[24] Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013). Oscillatory brain activity correlates with risk perception and predicts social decisions. Cerebral Cortex (New York N.Y. 1991), 23(12), pp. 2872-2883
[25] Park, J., Kim, H., Sohn, J.-W., Choi, J., & Kim, S.-P. EEG beta oscillations in the temporoparietal area related to the accuracy in estimating others' preference. Frontiers in Human Neuroscience, 12(February). 2018
[26] Gert Pfurtscheller, F. H. Lopes da Silva. Functional Brain Imaging. Hans Huber Publishers, 1988, 264 pages.
[27] Thatcher, R. W., McAlaster, R., Lester, M. L., Horst, R. L., & Cantor, D. S. (1983). Hemispheric EEG asymmetries related to cognitive functioning in children. In A. Perecuman, Cognitive Processing in the Right Hemisphere (pp. 125–145). New York: Academic Press.
[28] Klimesch W, Doppelmayr M, Russegger H, Pachinger T. Theta band power in the human scalp EEG and the encoding of new information. NeuroReport 1996;7:1235-1240.
[29] I. Chavdarov, Patent application № 112346, "3D printed joint between a shaft and a link", Published summary - Bulletin 01.2 31.01.2018., page 13-14, in Bulgarian http://www.bpo.bg/images/stories/buletini/binder-2018-01-31.pdf

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 221-229 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Optical Braille Recognition Software Prototype for the Sinhala Language

Shanmuganathan Vasanthapriyan*, Malith De Silva

Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, 70140, Sri Lanka

ARTICLEINFO

Article history:
Received: 31 May, 2019
Accepted: 24 July, 2019
Online :30 July, 2019

Keywords:
Braille
Visually impaired
Sinhala braille
Optical braille recognition

ABSTRACT

Braille system is purposely made for visually impaired people, to support their literal communication in order to share their knowledge. Louis Braille introduced the braille system consists of series dots which are embossed to read by touching. Early days Braille papers are made manually, but at current days braille documents are made using machines. Due to lack of perceiving on braille symbols and characters, it was highly needed fact to develop Braille system to different languages. In the Sri Lankan context, we found that the mostly inconvenience are happening inside of Sri Lankan education system. Such as in Special Education centers, Colleges, and universities. Written Braille scripts are evolution by a limited number of people who are specialized in the Sinhala braille system. Also, the process of marking braille documents are not effective and efficient. The focus of this research is to address the issue of literal communication gaps between society and the blind people in Sri Lanka. Average quality single-sided composed Braille dot characters are identified with maximum accuracy by using several novel methodologies. Obtained results denote the proposed methodologies are with the highest accuracy and system performance are more efficient as promised. The research presents executable software prototype, includes proposed methods which align with optical braille recognition in order to transpose the recognize braille characters. Introduce of new binaries cell transcription method of Braille character from a Braille document and decoding them into Sinhala characters. The proposed cost-effective system can display decoded braille characters by normalizing to Sinhala text which is in a human-understandable form.

1. Introduction

1.1. Background

Society consists of different people who have various capabilities. Among them, there are some people who are less capable of doing some daily activities due to various reasons. Visual illness is one of the causes. Blindness is a visual illness, occur due to various physical or neurological factors. It can be occurred at birth, due to an accident or due to other disabilities or diseases. Most of the time these conditions make a serious effect on their lives and change the whole lifestyle. According to WHO (World Health Organization), around 314 million people worldwide are living with serious vision impairments. Among them, 37 million people are suffering from blind and 124 million people are having low vision. Moreover, 90% of those blind people are living in low economic countries.

*Shanmuganathan Vasanthapriyan, & email: priyan@appsc.sab.ac.lk

When we consider Sri Lanka, we are having a population of 21 million. Among the population, around 150,000 people are having eye blindness and around four hundred thousand (400,000) people are having a low vision. According to the report of "*National Survey of Blindness, Visual Impairment, Ocular Morbidity and Disability in Sri Lanka*" [1], the prevalence of blindness among those aged 40+ years is 1.7% and 15.4% for moderate visual impairment.

As we all know, we all are gifted with inborn talents. It's the same for visually impaired or low sighted people. Most of the time they are searching for their capabilities. In addition rest of the people are expecting them to play an integral role in the society by using their inherited talents [2]. But in most of the cases, the blindness inhibits those talents and search other's help. Blind people need to put an extra effort than the rest in order to live normal life [3]. As they all are part of humans, we have a responsibility to make their lives easier. We have to facilitate them to socialize with the world. The key for it is communication. In

addition, we have to help them to reach universal knowledge. We should share their wiseness within the world and help them to come after in their lives.

To keep flow their lives with an ordinal pattern of the society, they need to have a clear knowledge of with other ordinal lifestyles. When they come to get to know these facts, they face the major problem in communication, knowledge sharing, and knowledge gathering. Instead of achieving everyday needs, there ought to be a few techniques to make emerged their abilities and keep up them on with their gifts by giving them information and let them impart their insight into society.

It's critical to keep data in the written format since it assumes obviously job with regards to knowledge storing and sharing. For that composed learning from easygoing note to training, recorded data to encoded figure content we need images to trade that learning and data. As ordinary individuals, we have our very own written mechanism as indicated by our languages, just as outwardly visually impaired individuals likewise need some sort of images and method for the game plan of images so as to express their emotions, contemplations, information and so forth [4]. To come up with these challenges the world has come up with several mechanisms like braille system to make it happen [5].

This issue is especially clear in the schooling framework and University frameworks, where these days' blind students are educated in standard classes. A significant number of these blind students perform the evaluation, tests, and schoolwork composing utilizing the Braille medium. Notwithstanding, most educators of these understudies are not Braille proficient [6]. One strategy by and by used to beat this trouble is that the blind students work is first sent to trained Braille transcribes, where the Braille is translated to literary text and afterward sent back to the educator before it is marked. This makes pointless postponements and cost for the blind students, educator, and government. Issues additionally exist in the working environment where any data composed by a Braille client that will be deciphered by other Braille uneducated people, should be first interpreted by the Braille client themselves [7].

Also, blind people feel helpless when it comes to situations like filling application forms. Even we don't realize, application forms play a major role in human's life from birth to death. Instances such as have a birth certificate, Death certificate, enroll to a school, to enter an university, to get an occupation, to go abroad, to have visa likewise most of the important events are linked with application forms and filling an application means we are providing our personal and confidential details to another party in order to get an opportunity or a service [8].

Even they know Sinhala, English, Tamil or other languages, they don't know how to read or write in normal characters. Also, normal people don't how to read or write in Braille characters. So there is a huge gap between normal and blind people and society make feel those people as really disable and less confident.

When it's come to knowledge gathering and knowledge sharing, literally communication is a major medium in knowledge dissemination. In order to do communication, visually impaired people have many systems to accomplish this intention. The Braille system is the reliable and most famous system for this purpose. Braille system is a communication system which is literal. It enables blind and partially sighted people to read and write through using touch stimuli. They use this braille system for calculations, menus, signs, elevator, and books [3].

Louis Braille, a French teacher invented Braille in 1825. He formulated this concept by using the military secret codes. These secret military codes are known to be night writing. This technique is used by military soldiers at dark times to communicate with each other. These secret codes have twelve dot cell six dot height and two dots wide [9]. The only problem is Louis Braille was faced on when making the Braille system that visually impaired people cannot feel all the dots at one touch. Also, the finger doesn't have that much sensitivity for recognizing dots at on time. He made it to practical by making this cell 2 dots wide and 3 dots in height. In order to make it use as character set, made up of different combinations of raised dots (tiny palpable bumps) as above mentioned. In order to represent different characters or sequences of characters, the 3-by-2 (3 rows and 2 columns) arrangement was used [10, 11].

Sinhala is known as the official native language of Sinhalese. According to the Sri Lankan context, there have been new to the braille system by comparing to other languages. The Sinhala language is often considered as two alphabets or an alphabet within an alphabet. It is because of the presence of two sets of letters. The core set, known as the "*Suddha Sinhala*" (pure Sinhalese) or "*Eḷu Hoḍiya*" (Eḷu alphabet) can represent all native phonemes [12]. In order to deliver Sanskrit and Pali words, the Misra Sinhala (mixed Sinhala), an extended set was introduced by ancient scholars. Current Sinhala alphabet uses 60 letters in the alphabet. There are 18 vowels and 42 constants contain in today's standard Sinhala alphabet. However, there are 57 characters are used in present [12].

1.2. Sinhala language braille writing style

Sinhala braille and English braille are two different alphabets. They can't be combined. We can write only single alphabets letters only at once. Numbers are recognized from a pre identification character. The numbers are identified by the braille character "⠼". If "⠼" appears before any of the braille character the following character series until space appears is considered as numeric characters

Sinhala braille system has a different way of writing comparing to the normal Sinhala writing system. As previously mentioned with the combination of 2 by 3 arrangement of the braille dots we have 64 of different combinations use when braille writing. Braille diacritics are use made up with combining constant letters along with the vowels.

In Sri Lanka, most of the blind people are not going through any kind of education mechanism. There are fewer education centers which having knowledge of Braille system in Sri Lanka. People who have knowledge of the braille system only can read the braille documents and understand what visually impaired people have written, what they try to express. Most of the normal people don't have any knowledge on how to read braille documents and as a society, we fail to understand what they express in the literal way of their feelings and knowledge. There are some people write their own braille books and exposed to the world. In Sri Lanka, most of the visually impaired people stay at home, and the parents don't have knowledge of how to read or write Braille. Even if visually impaired people have to write braille, people who are at home cannot understand the braille documents. There is a clear gap between literal communication between sighted people and visually impaired people when it's come to share their knowledge.

This identified gap is addressed through this research using the software prototype.

There is a research conducted by Perera et. al, [13] that identify Sinhala Braille characters in single-sided Braille document and translates to the Sinhala language. This system was also capable of identifying Grade1 English Braille characters, numbers, capital letters and some words in Grade 2 English Braille system.

The main objective is to make a platform to understand braille fonts and convert it to Sinhala fonts in order to formalize communication between visually impaired people and the rest of other normal people. If there is a system which can translate written braille to Sinhala, It will be an excellent communication tool for sighted people (who do not know Braille) with blind writing. And it will reduce the time and money which needed to translate such documents; also, it will secure blind people's privacy in situations like filling application forms. In order to overcome above-mentioned problems research have to follow various researches and have found there are fewer studies have done come over to address these problems. By using latest Sinhala braille alphabet [14], documents are converted with high accurate and efficiency.

2. Methodology

The overall methodological framework is shown in the following figure 1.

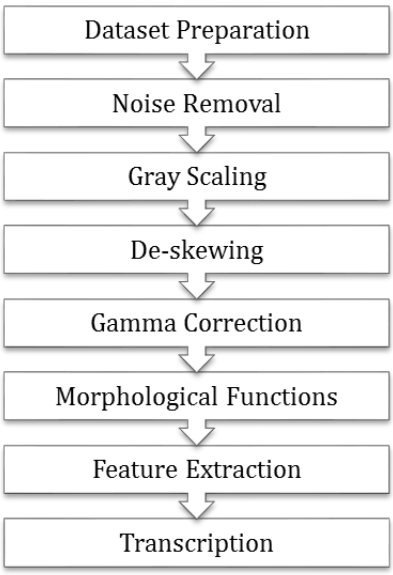


Figure 1: Methodological framework

2.1. Dataset preparation

As the first step in the research, the braille documents were collected. Both handwritten documents and typed documents were collected. The braille slate was used to write the handwritten document and braille typewriter was used for the typed documents. The size of the handwritten document was 9x12 inches and type documents were 8.5x11.8 inches and 9.5x12.5 inches.

After Braille documents have been collected, scanning was performed. In Optical Braille Recognition (OBR) scanning braille documents enable us to perform digitization which produces a digital image of the scanned Braille document. A flatbed scanner has been used for the digitization process. This is because it has been found as a cheap alternative to scan Braille images [8]. Figure 2 and Figure 3 shows the difference in light source distribution when grayscale.

Horizontal resolution and vertical resolution of 300 dpi and of 200dpi have been used for the scanning process. Most of the researches have been conducted in this area used 100 to 200 dpi resolution when scanning their documents so it is recommended to get quality images and the images are stored in JPEG format. Using this configuration Braille documents that are collected from different sources are scanned and prepared for subsequent processes.

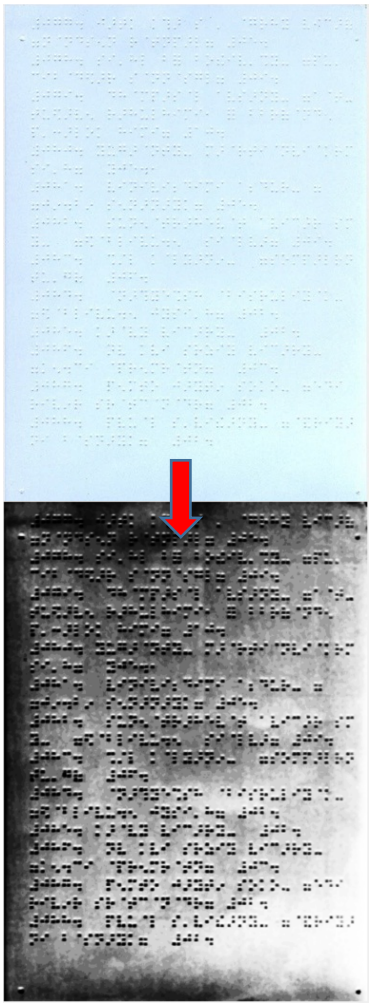


Figure 2: Braille Accusation Using Camera and After Grayscale

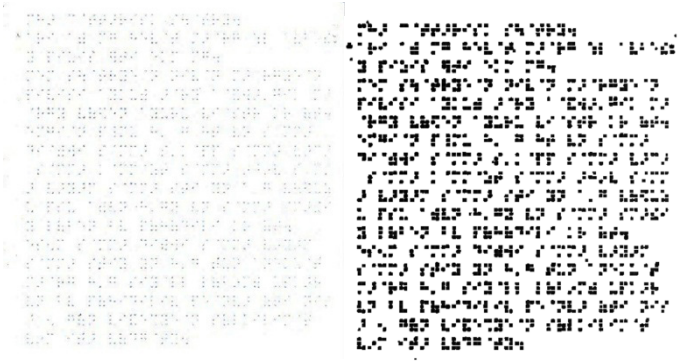


Figure 3: Braille Accusation Using Scanner and After Grayscale

Braille document property description

- Braille sheets: 12
- Resolution: 200 dpi, 300 dpi
- Digital format: Colour
- Image sizes: 9 X 12 inches, 8.5x11.8 inches, 9.5x12.5 inches
- Braille type: Single-sided
- Image format: JPEG

2.2. Pre-processing

Noise removal

Before starts to do any compilation on the scanned digital image, it's recommended to do pre-process part on the image. Before that, it's necessary to compare some technical aspects before coming to a conclusion which needs to select. There are several techniques on noise removal such as mean filtering, Median filtering, and Gaussian blur are some of it. In here research got selected the Gaussian blur for noise removal and later here it discussed why select Gaussian algorithm [15].

If the window size is small the effect of filters will reduce because it is not capable remove noise much effectively. If the window size is got increased it will remove the noise, but as well it will blur the braille bots and also the edges of the documents. Which lead to reduce the detail of the image where do not contain noise also.

There can be several ways of getting the braille document noisy. When considering the handwritten documents, according to the way of writer use the paper may lead to noise. Sometimes the environment may lead to noises. Another way of getting noise is paper may itself having some patches or print issues. However here below Figure 4 give a clear idea on how handwritten braille document have exposed for noise. Type documents are can be exposed to noise when they read by the visually impaired people because of the dust and other unnecessary things having on the fingers at the time of reading braille image [16]. Finally, in the end, there may be noise when getting through the scanner.

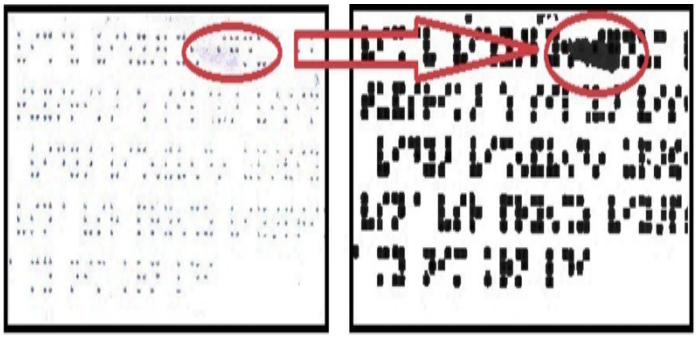


Figure 4: Noise on Handwritten Image after Pre-Processing

But in here its need to focus on both side at this stage. When using powerful noise removal algorithm to removing the noise it may affect to respective needed details on the documents which are not affected by the noise at high scale. With the above facts no need to use powerful noise removing algorithms because they may get the effect to somewhat on context [4].

Mean filter

Mean filtering is done using read pixel by pixel of the image and assign the mean value of values for its pre-defined kernel's structure center pixel. When the kernel is large, the center pixel may have a high effect through other pixel values. Even a pixel which does not contain noise also gets affected. Also, it should not affect the braille cell segmentation and features extraction. But this will lead to unnecessary complexity on future steps.

Median filter

This mechanism is a little bit ahead than the mean filter because the median filter is replacing a middle pixel with the median value of the kernel's center pixel. Which leads to damage the edges of the characters and edges of the fields [17].

Gaussian blur

Gaussian blur was used to remove the noise. Instead of utilizing other low pass filtering algorithms, Gaussian blur channel does not lessen braille dot details. However, it makes decrease noise in an incredible way. Figure 5 denotes how the Gaussian blur work on noise removing. As indicated by the Gaussian blur work it weighs more on closest pixels of the utilized kernel canter. Figure 6 shows how weight is distributed around the center pixel. Equation 1 shows Gaussian blur for one dimension.

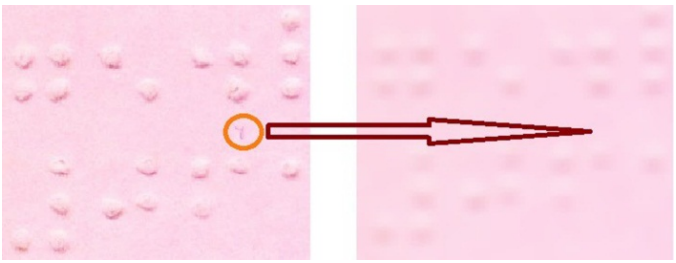


Figure 5: Gaussian Blur Apply On Noised Document

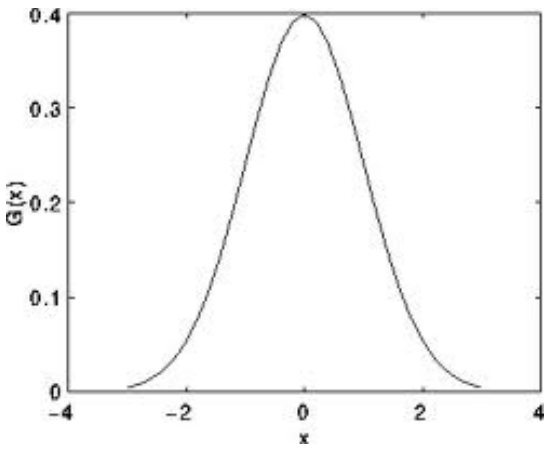


Figure 6: Pixel Weighting According To Kernel Center

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \quad (1)$$

A filter using EmguCV was applied to the input in order to perform the smoothing function [18]. Equation 2 shows how EmguCV has done for the above mathematical operation (equation 1). Equation output pixel's value g (i, j) is determined as a weighted sum of input pixel values.

$$g(i,j) = \sum_{k,l} f(i+k, j+l)h(k,l) \quad (2)$$

The coefficients of the filter are the kernel which is represented by h (k, l). It helps to visualize a channel as a window of coefficients sliding over the image. Weighing of the pixel was done through f (i+k, j+l). Even though it is not the fasted filter it is good to filter for working with images. Gaussian filtering was done by convolving each point in the input array with a Gaussian kernel and then summing them all to produce the output array [18]. Table 1 describes the above mention details as a summary.

Table 1. Reasons Summary of Gray Scale, Mean Filter, and Gaussian Blur

Median Filter	Mean Filter	Gaussian Blur
Low Computational Cost	High Computational Cost	High Computational Cost
Preserve Image Details	Very Low Capabilities to Preserve Image Details	High Capabilities to Preserve Image Details
Low Memory Consumption	High Memory Consumption	Low Memory Consumption
Simple Implementation	Simple Implementation	Simple Implementation
Analytical Process Becomes Simpler	Analytical Process Becomes Relatively Complex	Analytical Process Becomes Relatively simple

2.3. Grayscale

The expectation of using gray scaling technology is not only for reducing noise but also several other results are expected. Since the resolution and quality of the image will higher according to the dpi of the input image. The time which takes for the process relatively large when quality is high, which is negatively affected for systems performance. But when do gray scaling it is an extremely small amount of time takes for the entire process.

On the other hand to conduct analyzing processes such as edge detection, segmentation, and feature extraction and classification techniques effectively it is necessary to reduce the intensity of the image since inputs to the system are color images. Since gray scaling contains only 0-255 values on pixels it is useful to use gray scaling such that it is easy to conduct analytical processes [19]. As the second step, convert the color image into a grayscale image. Figure 7 shows the result when the input image goes through gray scaling.

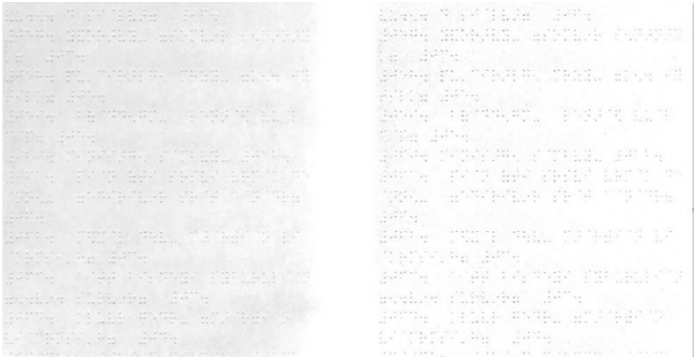


Figure 7: Gray Scale of Input Image

2.4. De-skewing

Rotation is used to de-skew braille documents in place of horizontal and vertical projection profiles.. In some cases the document may not have skewing errors but in general when braille documents are composed there is a possibility for skewing errors to occur. Instead of squandering increasingly computational capacity to de-skew documents, the work utilizes moderately fixed positions to de-skew the documents using rotation angle. Figure 8 demonstrates De-skewed image Utilizing Rotation.

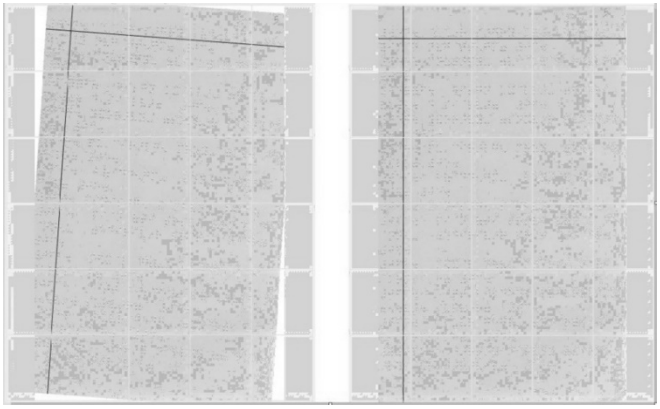


Figure 8: De-Skewed Image Using Rotation

2.5. Gamma correction

Gamma correction was done on the gray image to enhance subtleties of the braille spots. Gamma correction can be utilized in non-direct connections and in encoding or interpreting the luminance in images. Equation 3 and 4 demonstrate the calculation.

$$V_{out} = V_{in}^{\gamma} \tag{3}$$

$$\gamma = \frac{d \log (V_{out})}{d \log (V_{in})} \tag{4}$$

As this connection is nonlinear, the impact won't be equal for each pixel and will rely upon their original value. Whenever $\gamma<1$, the first original dark regions will be brighter and the histogram will be moved to one side and vice versa with $\gamma>1$. Here work utilizes the gamma= 4.0 and Figure 9 depicts the eventual outcomes on a grayscale image.



Figure 9: Gamma Correct on Gray Scale Image

2.6. Morphological erosion

Morphologic capacity disintegration is utilized to sharpen the edges of the braille spots. Unlike other methods, their morphologic capacities disregard the flaws which happen to the surface and state of the picture. Furthermore, administrators such as segregation and

joining of individual pixels and different pixels help locate the decorated braille dot. Disintegration impact as per the organizing component was utilized. In contrast to paired disintegration, grayscale disintegration improves the dot subtleties [9]. Indicating a picture by *f(x)* and the grayscale organizing component by *b(x)*, where B is the space that *b(x)* is characterized, the grayscale disintegration of *f* by *b* is given as in the equation (5). Image after Morphological Erosion is shown in Figure 10.

$$(f\theta b)(x) = \inf_{y \in B}[f(x+y) - b(y)] \quad (5)$$

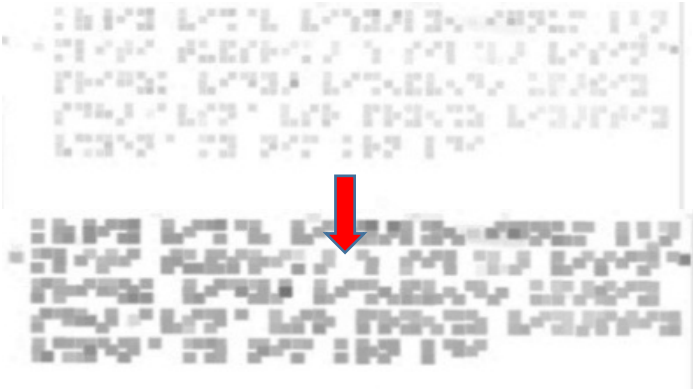


Figure 10: Image after Morphological Erosion

2.7. Feature extraction

Feature extraction is an illustrative system of the Braille image. The principle function of this procedure is to extract the Braille dots from the binaries image. The image capturing, pre-handling and division stages cause the image to be appropriate for various element extraction calculations. Some component extraction calculations are only concerned with the forms of the image while a few calculations ascertain each pixel of the image. Conversely, the underlying image might be noise influenced or obscured by different reasons..[7].

Simultaneously, the nature and the yield of the image pre-processing and division steps are determined by the picture feature and the extraction strategy. Feature extraction phase can be considered as the most important part of braille character translation because the accuracy of the translated character heavily depends on the extracted features of the cells.

Thresholding

Before the segment, the braille cell into six equal compartments, need perfect detection of braille dots. For this work have experimented on various thresholding techniques. First attempt was using straightforward thresholding strategies like Binary, To Zero thresholding and Binary inverse thresholding etc. The most intricate task was to recognize limit esteem in light of the fact that the reports have various hues and when checking, pixel luminance may change. To overcome the issue, in this research versatile thresholding Otsu strategy was utilized. Since the attempt was to make binomial pictures along the way of preparing part and it is the greatest contribution to Otsu thresholding, Otsu thresholding calculation located the bimodal qualities and attempted to discover the boundary that limits the weighted intra-class change. Bimodal picture is a picture whose histogram has two pinnacles. From the calculation, it takes the center of those histogram crests and computes the intra-class fluctuation.. Figure 11 shows the result of Otsu thresholding.

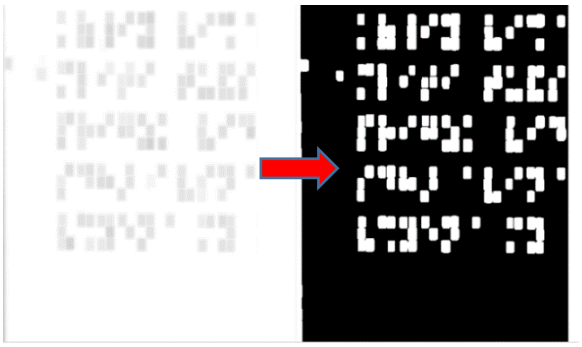


Figure 11: Image after Otsu thresholding

It actually finds a value of t which lies in between two peaks such that variances to both classes are minimum. After making thresholding, cell segmentation done using the standard braille dimensions.

The vertical as well as horizontal (however not corner to corner) separation between two focuses of neighboring dots in a specific cell is 2.5 mm. The separation between focus to focal point of relating spots in vertically contiguous cells is 5.0 mm, in the interim it is 3.75 mm for the horizontal neighboring cells.

2.8. Transcription

As presented in Figure 12 braille cells are sectioned and checked for the normal white pixels over 60%. On the off chance that the normal of white pixels is over 60%, it's perceived as substantial braille dot and marked as 1. If it's not, it is considered as there is no braille spot, which means the rate ascribed is 0.

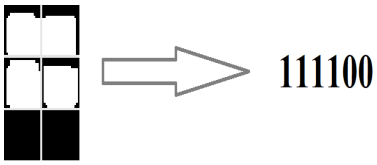


Figure 12: Cell transcription to binary values

Here below algorithm written for the binaries cell.

```
Inputs: erodeimg1, erodeimg2,
projection_x, projection_y; img,
resolutiondpi_x
For counter = projection_x to
projection and counter less than
height of erodeimg1 then
For counter = projection_x to
projection x + resolutiondpi_x and
less than width of erodeimg1 then
For counter = 0 and less than height
of img1 then
For counter =0 and less than width
of img1 then
do calculation:
Check row 1, column 1;
Check row 1, column 2;
Check row 2, column 2;
Check row 3, column 1;
Check row 3, column 2;
End for
End for
End for
End for
```

When binaries string array is extracted from the above way, the mapping to Sinhala Unicode is done using SQL Server. All the particular related string array contain it's Unicode in the database. The database connects to the system via Entity Framework in C#.

As a summary, for an example, Sinhala quote "ෟ ෟ ෟ

ඹ	⠽ (braille pattern dots-126)	10	10	100
ඩ	⠳ (braille pattern dots-346)	11	9	81.81818
ච	⠨ (braille pattern dots-14)	14	14	100
ඡ	⠄ (braille pattern dots-16)	10	9	90
ජ	⠩ (braille pattern dots-245)	18	16	88.88889
ට	⠹ (braille pattern dots-23456)	20	20	100
ඨ	⠫ (braille pattern dots-2456)	10	9	90
ඩ	⠻ (braille pattern dots-1246)	13	13	100
ඪ	⠿ (braille pattern dots-123456)	10	10	100
ත	⠮ (braille pattern dots-2345)	55	53	96.36364
ථ	⠷ (braille pattern dots-1456)	10	10	100
ද	⠸ (braille pattern dots-145)	30	30	100
ධ	⠬ (braille pattern dots-2346)	17	16	94.11765
න	⠝ (braille pattern dots-1345)	72	70	97.22222
ප	⠏ (braille pattern dots-1234)	28	28	100
ඵ	⠞ (braille pattern dots-156)	11	10	90.90909
බ	⠂ (braille pattern dots-12)	22	21	95.45455
භ	⠈ (braille pattern dots-45)	12	11	91.66667
ම	⠍ (braille pattern dots-134)	73	71	97.26027
ය	⠽ (braille pattern dots-13456)	65	59	90.76923
ර	⠗ (braille pattern dots-1235)	80	79	98.75
ල	⠇ (braille pattern dots-123)	48	46	95.83333
ළ	⠌ (braille pattern dots-456)	12	12	100
ව	⠧ (braille pattern dots-1236)	70	70	100
ශ	⠯ (braille pattern dots-12346)	11	7	63.63637
ෂ	⠶ (braille pattern dots-146)	13	11	84.61538
ස	⠎ (braille pattern dots-234)	80	77	96.25
හ	⠓ (braille pattern dots-125)	30	30	100
ෆ	⠤ (braille pattern dots-124)	11	9	81.81818
ණ	⠫ (braille pattern dots-1356)	17	15	88.23529
ඟ	⠒⠶ (braille pattern dots-23)(braille pattern dots-1245)	11	9	81.81818
ඬ	⠒⠻(braille pattern dots-23)(braille pattern dots-1246)	11	10	90.90909
ඳ	⠒⠸(braille pattern dots-23)(braille pattern dots-145)	12	10	83.33333
ඹ	⠒⠂ (braille pattern dots-23)(braille pattern dots-12)	11	10	90.90909
ඥ	⠿ (braille pattern dots-12345)	10	9	90
ක්	⠁ (braille pattern dots-4)	118	113	95.76271
කං	⠲ (braille pattern dots-3)	40	34	85

Table 3: Punctuation Wise Identification Results and Accuracy Rate

Punc-tuati-on	Symbol	# of sam-ples	# of samples correctly identified	Perce-ntage (%)
.	⠲ (braille pattern dots-256)	23	21	91.30435
,	⠂ (braille pattern dots-2)	11	10	90.90909
;	⠆ (braille pattern dots-23)	12	12	100
:	⠒ (braille pattern dots-25)	10	10	100
!	⠖ (braille pattern dots-235)	12	11	91.66667
[]	⠶ (braille pattern dots-2356)	20	18	90
“	⠦ (braille pattern dots-236)	10	9	90
”	⠴ (braille pattern dots-356)	10	9	90
/	⠤ (braille pattern dots-36)	21	21	100
Space	(no braille dots)	76	73	96.05263

Table 4: Identified Numerals Results with Accuracy Rate

#	Symbol	# of sam-ples	# of samples correctly identified	Perce-ntage (%)
1	⠼⠁ (braille pattern dots-3456) (braille pattern dots-1)	25	25	100
2	⠼⠃ (braille pattern dots-3456) (braille pattern dots-12)	17	16	94.11765
3	⠼⠉ (braille pattern dots-3456) (braille pattern dots-14)	16	16	100
4	⠼⠙ (braille pattern dots-3456) (braille pattern dots-145)	16	15	93.75
5	⠼⠑ (braille pattern dots-3456) (braille pattern dots-15)	17	15	88.23529
6	⠼⠋ (braille pattern dots-3456) (braille pattern dots-124)	20	18	90
7	⠼⠛ (braille pattern dots-3456) (braille pattern dots-1245)	21	21	100
8	⠼⠓ (braille pattern dots-3456) (braille pattern dots-125)	22	22	100
9	⠼⠊ (braille pattern dots-3456) (braille pattern dots-24)	21	21	100
0	⠼(braille pattern dots-3456) (braille pattern dots-356)	70	68	97.14286

3.1. Threats to validity

With several approaches have done related to the work there is many limitations have found. Image acquisitions using different types of equipment would have an effect on results receiving. While the acquired image is not with the minimum dpi will affect results. With the lightning disturbances without equally spread on documents make the threat to results which observed. Cropping document with the desired way of the human pattern will lead to making half braille characters and direct effect on results we obtain.

4. Conclusion and Future Works

The Braille framework is a tactile technique broadly utilized by visually impaired individuals to read and write or peruse and compose. Braille documents contain lines of characters, where each character has six spots masterminded in three lines and two segments (three rows and two columns) and each dot can either be raised or be level as indicated by the corresponding character. Braille is reasonably understandable by visually impaired individuals; notwithstanding, sighted individuals need not have the option to comprehend these codes. Braille recognition system can connect the correspondence communication gap between visually impaired and sighted individuals. A ton of exertion has been made worldwide by specialists to connect this gap. In this investigation,

an endeavor has been made on the element extraction and characterization modules of Sinhala Braille acknowledgment as indicated by the Sri Lankan context.

This proposed programming model is fit for extracting Braille characters from a Braille document pursued by decoding them into Sinhala characters and after that standardization of the decoded Sinhala characters into readable Sinhala content. The transformation from Braille documents into Sinhala content is another zone where much research has not been completed.

An algorithm to distinguish dots in a picture of embossed Braille material obtained by an optical scanner was proposed. Despite the fact that the Braille dots have a similar shading as the foundation, they cast delicate shadows when checked with a standard flatbed scanner. These shadows are utilized to find the dots on the page.

In general, the methodology indicates achievability as a practical, cost-effective, quick and simple strategy to identify dots in Braille records. It doesn't require costly or complicated equipment. It utilizes a flatbed scanner which can be shared with different applications. Robustness to adapt to low-quality scans and imperfect reports are worked in at various levels. The outcomes got were promising during trials performed on single-sided embossed records, with over 95% accuracy.

The recognition rate for character recognition has the opportunity to get better. One approach to improve it is to actualize a Sinhala spell check algorithm. Any misrecognized image can be recognized since the word and sentence that the character has a place which would never again make sense.

It is important that every one of the examples in the investigations are sensibly all well-formed. A conceivable future upgrade in this undertaking is utilizing various allegations with numerous gadgets like cell phones. This may incorporate adding a few algorithms to progressively change different edges for various pieces of the framework, to represent fluctuation, for example, the shade of the paper. On account of slanted pictures, a few components to deal with the revolution will be automated. Indeed, even there are severe principles about the size of the spots utilized just as the dispersing between them, size of the braille specks in the wake of filtering is vary from record to report as indicated by its shading. If the feature extraction algorithm automatically adopted according to its color of the document there will be a much better extraction of features for classification.

Author's Contributions

Shanmuganathan Vasanthapriyan: Designing of experiment (75%), Laboratory and experiment (75%), Interpretation and data analysis (75%), Writing up (75%).

Malith De Silva: Designing of experiment (25%), Laboratory and experiment (25%), Interpretation and data analysis (25%), Writing up (25%).

References

[1] Ministry_of_Health. National Survey of Blindness, Visual Impairment, Ocular Morbidity and Disability in Sri Lanka. 2015; Available from: https://www.iapb.org/wp-content/uploads/National-Survey-of-Blindness-A-Report-2014-2015.pdf.

[2] Srinath, S. and C.R. Kumar, A novel method for recognizing Kannada Braille: Consonant-Vowels. International Journal of Emerging Technology and Advanced Engineering, 2013. 3(1): p. 596-600.

[3] Al-Salman, A.S., et al. An efficient braille cells recognition. in 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM). 2010. IEEE.

[4] Hentzschel, T. and P. Blenkhorn, An optical reading system for embossed Braille characters using a twin shadows approach. Journal of Microcomputer Applications, 1995. 18(4): p. 341-354.

[5] Mennens, J., et al., Optical recognition of Braille writing using standard equipment. IEEE Transactions on Rehabilitation Engineering, 1994. 2(4): p. 207-212.

[6] Al-Salman, A., et al., An arabic optical braille recognition system. 2007.

[7] Wajid, M., M.W. Abdullah, and O. Farooq. Imprinted Braille-character pattern recognition using image processing techniques. in 2011 International Conference on Image Information Processing. 2011. IEEE.

[8] Mousa, A., et al., Smart braille system recognizer. International Journal of Computer Science Issues (IJCSI), 2013. 10(6): p. 52.

[9] Zhang, H., J. Li, and J. Yin. A Research on Paper-Mediated Braille Automatic Extraction Method. in 2010 International Conference on Intelligent Computation Technology and Automation. 2010. IEEE.

[10] Antonacopoulos, A. and D. Bridson. A robust braille recognition system. in International Workshop on Document Analysis Systems. 2004. Springer.

[11] Li, J. and X. Yan. Optical braille character recognition with support-vector machine classifier. in 2010 International Conference on Computer Application and System Modeling (ICCASM 2010). 2010. IEEE.

[12] Educational_Publications_Department, සිංහල භාෂාව හා සාහිත්‍යය. 2018. 1 - 213.

[13] Perera, T. and W. Wanniarachchi, Optical Braille Translator for Sinhala Braille System: Paper Communication Tool Between Vision Impaired and Sighted Persons. The International Journal of Multimedia & Its Applications (IJMA) Vol, 2018. 10.

[14] Sinhalese Braille. 2016; Available from: https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/Sinhalese_Braille.html.

[15] Dubus, J., et al. Image Processing techniques to perform an autonomous System to translate relief Braille back into ink called LectoBraille. in IEEE 10th International Conference in Medicine and Biology Society, New Orleans. 1988.

[16] Falcon, N., et al. Image processing techniques for braille writing recognition. in International Conference on Computer Aided Systems Theory. 2005. Springer.

[17] Hu, Y. and H. Ji. Research on image median filtering algorithm and its FPGA implementation. in 2009 WRI Global Congress on Intelligent Systems. 2009. IEEE.

[18] Moshnyaga, V.G. and K. Hashimoto. An efficient implementation of 1-D median filter. in 2009 52nd IEEE International Midwest Symposium on Circuits and Systems. 2009. IEEE.

[19] Padmavathi, S., S.S. Reddy, and D. Meenakshy, Conversion of braille to text in english, hindi and tamil languages. arXiv preprint arXiv:1307.2997, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 221-227 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Virtualization in Cloud Environment: Bandwidth Management

Isaac Odun-Ayo*, Blessing Idoko, Temidayo Abayomi-Zannu

Department of Computer and Information Sciences, Covenant University, Ota, 112233, Nigeria

ARTICLE INFO

Article history:
Received: 22 January, 2019
Accepted: 20 March, 2019
Online: 27 March, 2019

Keywords:
Cloud computing
Virtualization
Bandwidth Management
Applications
Hypervisor
Virtual Machines

ABSTRACT

Cloud computing recently emerged as an interesting model that enables computing and other related internet activities to take place anywhere, anytime. Cloud service providers centralize all servers, networks, and applications to allow their users' access at any time and from any location. Cloud computing uses already existing resources like server, CPU and storage memory but runs on a new technology known as virtualization. The core idea of virtualization is to create several virtual versions of one single computing device or resource. This enables many user operating systems to work on such a single underlying piece of device. Network bandwidth is one of the critical resources in a cloud environment. Bandwidth management involves the use of techniques, technologies, tools, and policies to help avoid network congestion and ensure optimal use of the subscribed bandwidth resources while also being a bedrock of any subscription-based access network. Bandwidth management is being utilized by organizations to allow them to efficiently utilize their subscribed bandwidth resources. Bandwidth management deals with the measurement and control of packets or traffic on a network link in order to refrain from overburdening or overloading the link which can lead to poor performance and network congestion. In this paper, the highest development with respect to virtualization in cloud computing is presented. This study review papers available on cloud computing and relevant published literature in multiple areas like conferences, journals etc. This paper examined present mechanisms that enable cloud service providers to distribute bandwidth more effectively. This paper is therefore a study of virtualization in cloud computing, and the identification of bandwidth management mechanisms in the cloud environment. This will benefit forthcoming cloud providers and even cloud users..

1. Introduction

This paper expands upon the findings of the research carried out in the 2017 international conference on next-generation computing and information systems [1]. It expands on cloud bandwidth management, its mechanisms, and challenges. Cloud computing means "almost anything can be accessed" [2]. This allows for ubiquitous, convenient, on-demand network access to customizable computing resources e.g. storage, networks, servers, services, applications etc. that service providers could quickly render with minimal intervention. Cloud computing is powerfully changing the manner in which organizations and enterprises perform IT-related activities. Cloud computing features like elasticity, scalability, multi-tenancy, resource pooling, lower initial investment, easy management, faster deployment, location independent, device independent, reliability, and security make it attractive to business owners and IT users [3],[4]. Cloud computing model permits outsourcing of computational resources in a way that is more effective. The huge computing power and storage size of cloud resources permit day-to-day internet users to perform their tasks on pay-as-you-go terms. Cloud data centers will keep expanding, as their client base becomes larger.

In general, the Cloud massive network is built using an approach known as virtualization which permits the installation of different applications and operating systems unto physical hardware. It creates a different layer for the operating system (OS) to seat on, thereby separating the OS from the physical hardware and the core OS. Virtualization is done on the physical hardware by installing a hypervisor. A Hypervisor is a common technique used for the implementation of Virtualization. It permits for different operating systems to be installed on the same hardware. Though the operating systems coexist on the same physical hardware, they behave as each OS have its dedicated resources. Some of the types of Hypervisors include VMware ESX, KVM, Xen, and Hyper-V. Another emerging approach for implementing

*Isaac Odun-Ayo, Covenant University, +2348028829456
isaac.odunayo@covenantuniversity.edu.ng

virtualization is containerization commonly known as Container-based virtualization or operating system virtualization [5].

Cloud data centers normally have mechanisms for planning how their resources like computing space, memory, and bandwidth are released to the users. Bandwidth is required to distribute resources across different cloud networkS. It is very important to maintain cloud bandwidth to meet up with the increasing request from cloud resources across data centers. On a cloud network, the bandwidth size requested by a customer is usually not assured [2].

For a satisfactory performance and fair usage of the network, bandwidth management is used to achieve this with the implementation and creation of network policies which ensures that the sufficient amount of bandwidth being required is readily available for those time-sensitive and mission-critical applications. It also avoids opposition between lower priority traffic and the critical applications for the limited network. With cloud tenants making greater use of cloud networking infrastructure in a shared manner, there is an increased desire and a growing concern on how to arrange, reserve and monitor the bandwidth in the cloud computing environment [6], [7].

Generally, bandwidth management involves the use of techniques, technologies, tools, and policies put together by an organization in order to ensure optimal utilization of the available bandwidth [8]. The reason for a network bandwidth management is to ensure the accurate bandwidth size is available and assigned to the accurate users and application at the right location and time [9]. Generally, without an active plan for ensuring bandwidth management, the available bandwidth regardless of its size can never be sufficient for the ever-increasing request of the users [10]. Also, Malicious Cloud Bandwidth Consumption (MCBC) which attack is a new type of attack that aims to consume the bandwidth maliciously causing the financial burden to the cloud service host can be mitigated or avoided when utilizing bandwidth management [11].

Different operating systems (Oss) are running simultaneously on a physical server is a process known as Virtualization and is carried out utilizing a hypervisor or virtual machine manager (VMM). The physical network resources require sufficient bandwidth to assign to the virtual machines (VMs) and optimal bandwidth allocation to VMs is very critical especially in today's cloud data centers with massive networked VMs. Bandwidth management poses a significant tension especially within a cloud-computing environment that permits the idea of VMs to have specific aims like load balancing and failure tolerance. High latency and a sharp drop in network speed are expected if there is no sufficient bandwidth after VM-migration. Every initiated data transfer that must take place over the internet requires bandwidth. For example, server to VMs data transmission over the internet requires bandwidth; this is one reason why network administrators must ensure accurate bandwidth management.

The objective of this study is to analyze bandwidth management in a cloud-based computing and virtualization environment. Various area of virtualization and cloud computing will be discussed. This research will impart to the knowledge of cloud-based bandwidth allocation approaches and management. The remaining section of the research is as follows. Section 2 examines related work. Section 3 discusses virtualization in a cloud environment and bandwidth management mechanisms. Section 4 concludes the paper and suggests future work.

2. Related work

In [12] the author suggested a solution in Bandwidth Management on Cloud Computing Network, to distribute available bandwidth according to user priorities and actual demand of data transfer to-from the cloud is proposed. Three key metrics: throughput, response time and utilization for monitoring and analyzing the network performance were presented. In [13], the authors proposed an approach that will only permit the guaranteed bandwidth size when providing the requested service. It also discussed how Linux kernel module called Linux TC can be used to manage the allocation of network bandwidth. In [14], the authors proposed having a formal definition for cloud-based applications that are free and open to all. The authors built a Cloud Resources Allocation Model named CRAM4FOSS for these free and opened applications. According to the authors, the proposed solution is to accurately certify the stability of allocating Cloud resources. In [15], the authors discussed the idea of an open source virtualization. Several other virtualization ideas are addressed and afterward, analysis of two open sources IaaS systems was presented. [16] discussed the idea of skewness to strengthen the manipulation of servers. [17] proposed an approach to introduce separation of VMM concept instead of kernels separation concept since it will enable suitable separation of cloud VMs. In [18], the authors concentrated on the idea of VM migration for cloud providers and reviewed VM placements in data centers alongside its limitations and advantages were conducted. 1n [19], the authors focused on virtualization and its application in radiology. In [20], the authors highlighted how virtualization affects information security. In [21], the authors proposed a modular kind of the bandwidth manager that could rapidly adapt to unexpected changes that can be induced by the rates of dirty inactive applications or due to network congestion. The authors' goal is to optimize bandwidth consumption and latency during live VM migration. [22] proposed a combined algorithm for optimal allocation which enables network bandwidth and VMs to minimize the cost affiliated with the users. A decision is made by the algorithm to restrain network bandwidth and VMs from some specified cloud providers. [23] Argued on the importance of stability and fair bandwidth allocation. The authors proposed a method to prevent interference between the underlying hardware and VMs. In [24], the authors proposed an algorithm for dynamic bandwidth shaping while still maintaining the limitations on network resources. In [25], "Towards a Tenant Demand-aware bandwidth allocation strategy in the cloud data center" is discussed. The authors propose a strategy to handle the issue of over bandwidth subscription cloud datacenter. The main functions of the proposed system involve bandwidth prediction, bandwidth pre-allocation, and bandwidth gathering. In [26], the authors developed SpongeNet; which is a bandwidth allocation proposal that's made up of three determinants. Component one is to provide an accurate, flexible and simple way for users' requirements definition. Component two is an algorithm for merging policies to handle multiple goals. Component three is to guarantee fairness between dedicated and non-dedicated bandwidth users. In [27], the authors proposed a framework for predicting resource management to overcome the drawbacks of the reactive cloud resource management approach. In [28], the authors discussed the current trend and development in cloud computing and open source software. The authors found out that Open Source Software (OSS) like OpenStack provides the most comprehensive infrastructure in cloud computing and open source software.

3. Virtualization in Cloud Environment and Bandwidth Management

3.1. Virtual Machines and Containers Technology

A. Concept of Virtual Machines

The infrastructure layer is the host of the virtual machines (VMs) within the cloud network. It is a component within cloud computing. The core concept of virtualization is to separate OS from the underlying hardware, it divides a physical hardware machine such as a server into multiple virtual machines. This makes it possible to run all types of operating like Linux, Unix, Mac OS, and Windows on a host or physical machine, that is the coexistence of different operating systems on the same physical machine. Bandwidth size that closely matches the actual task and needs of each VM can be assigned. It allows for maximum utilization of the host machine. Hypervisors also knew as Virtual Machine Managers (VMSs) are used to create and control VMs. The virtual machine is then tied to the hypervisors and not directly to the physical hardware. There are two types of hypervisors. Type-1 hypervisors remove an extra layer between the underlying host hardware and the virtual machines because it executes directly on physical hardware without a base OS and acts as the operating system on the hardware. Here, physical hardware without a base OS can be known as a bare-metal. Type-2 hypervisors are hosted hypervisors since the installation is done on top of the existing OS on the physical hardware. These hypervisors execute as applications and the host machine won't need to be set up separately for setting up the virtual machines. Figure 2 depicts the Type-1 hypervisor and Figure 3 depict the Type-2 hypervisor [29].

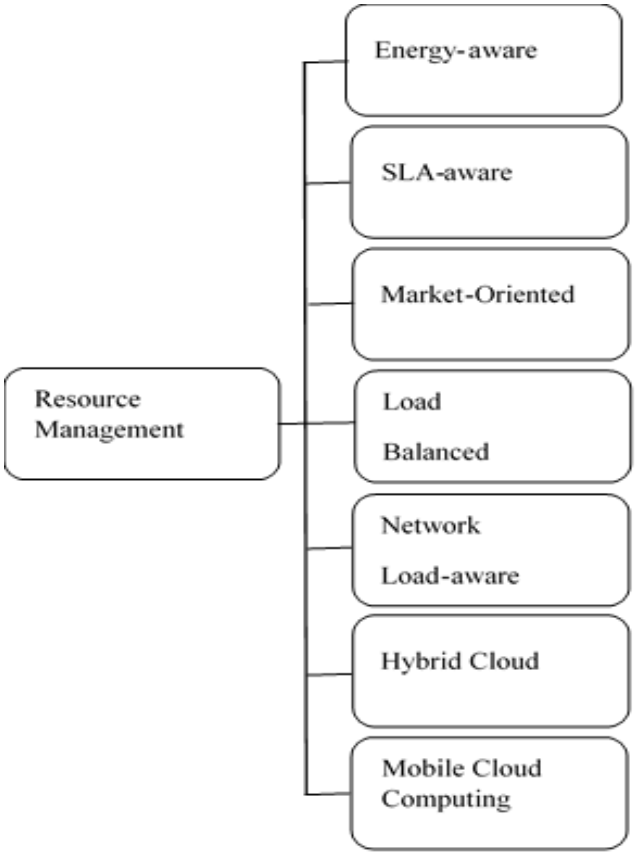


Figure 1: Classification of Resource Management Techniques [32]

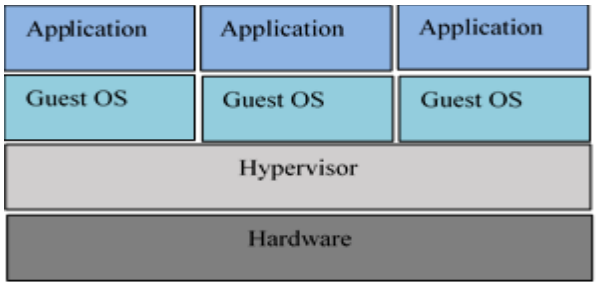


Figure 2: Type-1 Hypervisor [29]

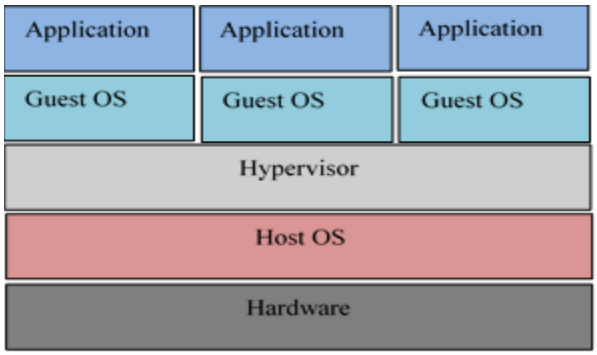


Figure 3: Type-2 Hypervisor [29]

B. Concept of Containers

Containerization is known as a virtualization technique which occurs within the OS level instead of the hardware level. Containers allow for easy migration between platforms and hosts compared to virtual machines and share a single OS kernel within an isolated environment [30]. This creates isolation confines within the application area instead of the server area which makes sure that if an error occurs within the single container, it won't affect the whole server or VM but only that individual container. Container technology is continuously evolving and guarantees an easy to deploy, streamlined and secure means of enacting certain infrastructure requirements while offering offer an alternative to VMs and mitigates the compatibility issues between applications which exists within the same OS [31].

Cloud containers are trending within the IT world and are modeled to virtualize just one application. Linux is the only server OS that supports cloud containers presently and Hyper-V containers are to be integrated to Microsoft Azure after being introduced by Microsoft [1].

C. Characteristics, Application, Benefits of Virtualization

In [1] it is highlighted the following characteristics of virtualization and its application areas as follows:

- Partitioning: This is a key feature in virtualization, used to logically divide a physical hardware resource for the installation of different/several operating systems.
- Isolation: Virtual machines coexist on the same physical hardware but each virtual machine exists separately from the core physical hardware and other virtual machines. With this feature, if there is a breakdown/downtime on one VM, the other active VMs will not be affected. Also, there is no data-sharing between VMs.

- Encapsulation: The encapsulated process maybe a business service and VMs can be stored and labeled as a file for identifying its service type. This averts the issue of hindrance amongst the applications. Storage, memory, networks, application, operating systems, and hardware are some of the application areas of virtualization. Here, to separate a software implies placing that software in a separate virtual machine space or in a container so that it does not coexist with other operating systems.

Virtualization of hosts or physical machines provides the following benefits [1]:

- Functional Execution Isolation: The hypervisor is assigned the responsibility of maintaining protection between the VMs and applications deployed on multiple VMs. Privileges may be given to users in their VM without violating the host integrity or isolation.
- Enhance Reliability: Hypervisors provides greater reliability of hosted virtualized applications due to their live migration capabilities enabling them to be independent and reliable.
- Customized Environment: Virtualization permits for the creation of customized space for a dedicated user, upon request for a customized and dedicated environment can be created and provided to handle a specific need.
- Testing and Debugging Parallel applications: Testing parallel applications can leverage virtualized environments, as a fully distributed system could be emulated inside a single physical host.
- Easier Management: Custom-made run-time environment can be migrated, started-up, shut down in a considerably variable way which depends upon the requirements of the person that is responsible for the essential hardware.
- Ability to coexist with legacy applications.
- VMs help to preserve binary compatibility in the run-time environment for legacy applications.

3.2. Resource Management Techniques

Cloud Resource management involves the procurement and release of cloud resources like physical hardware, storage space, virtual memory, and network bandwidth. Cloud resources can be physical or virtual components of limited availability. Resource management in virtualization environment and cloud computing is very critical. [32] used SLA-awareness, load balancing, energy efficiency etc. in the classification of resource management techniques, also depicted in figure 1.

- Energy-aware RM techniques: This involves minimizing the energy consumption by merging workload on a few numbers of physical servers. Virtual machines (VMs) on physical hardware with less load are migrated to other hardware that can accommodate more workload. This less loaded and idle hardware is then switched off to reduce the power expenses incurred by service providers and decreases Carbon dioxide (CO2) emission. [33] Presented a centralized control system for the allocation and management of all the concerned cloud resources. Resource Management (RM) choices are done hourly in order to reduce the subsequent effects. The choices can be based on the shutdown of the server, VM migration between multiple servers and powering up of the server.
- Load-balanced RM techniques: This is an important feature of the system in any computing environment. It is made use of in order to explain the idea of distributing the workload or traffic request between many resources and can maximize the performance of available bandwidth. Operations are moved amongst the physical machines after the load balancing algorithm have been applied and the system efficiently balances the load amongst the cloud resources while the usage by each server is examined. Once there is an overload of server resources, it migrates some of these workloads to an underutilized server but in a situation whereby both are underutilized, one server gets all the workload transferred to itself while the other server will be left unutilized or on standby. If hypothetically speaking that the workload migration failed, the server itself can be scaled down or up accordingly.
- SLA-aware RM techniques: Cloud-based Service Providers (CSPs) must strongly discourage the issue of violations, maintaining regular check during service provisioning to users. A service level agreement (SLA) is signed between the CSPs and customers which includes information such as the price, penalty clause that is enforced in an event of agreement violation and required a level of service(s). Authors in [34] proposed algorithms for capacity allocation in the hope of guaranteed SLA and handle fluctuating workloads which interfaces with the resource controllers that are geographically distributed while redirecting the network traffic load whenever congestion is observed. An application can also execute on different workloads and VMs equally allocated on the VMs if necessary. A workload analyst is utilized in order to predict future workload necessities throughout the workloads fluctuation and the capacity is altered on the premise of resultant predictions. Additionally, SLA violations can be mitigated by making sure that the response time is low when inter-VM communications are ongoing but if there is a higher response time, the VM is carried/moved to another physical machine.
- Market-oriented RM techniques: This involves solutions that are beneficial to cloud service providers and market-oriented. [35] Presented an RM technique where a SaaS service provider employs an IaaS service provider to serve its clients. Clients pay to the SaaS service provider for the services received and depending on the satisfaction level shown by the clients, a one-of-a-kind optimal function is utilized to calculate the amount that the user has to pay. According to [36], each cloud server has a dynamic voltage/frequency scaling (DVFS) module and it is assumed that a server cannot be switched on or off, also there is a common cost of VM migration

- Network load aware RM: This technique reduces high network traffic that can degrade the overall network performance. It allows service providers to choose a data center that could fulfill the interests of the clients. A user sends a request to the service provider and resources are assigned using an adaptive resource allocation algorithm. These algorithms achieve this by selecting a data center either based on the distance between the user or based on the distance between the data center. After the data center is chosen, the work is allocated to one of the servers thanks to the VM.
- RM techniques for the hybrid/federated cloud: In this technique, private cloud users are offered the benefit of utilizing the resources in a public cloud environment whenever there are insufficient resources to handle the needs of it, in-house users. To address the issue of decision-making on when to use public cloud resource, the authors in [37] presented a solution that is rule-based for hybrid clouds. Users' requests are divided into two which are the critical data/tasks request that is granted higher precedence and secondary data/tasks request that are granted lower priority. The critical data/tasks are hosted on the private clouds for security while the task with low priority can utilize the resources of both private and public cloud. When the resources of the private cloud are exhausted, the public cloud resources would be utilized.
- RM techniques for mobile clouds: Here, an energy efficient solution is presented to solve the problem of energy which is a major challenge in mobile systems. The authors in [38] proposed a technique to reduce the total power consumption. All the mobile devices primary job is to move the workload to one of the available servers in order to reduce the overall energy consumption.

3.3. Cloud Bandwidth Allocation Mechanisms

According to [2], bandwidth mechanisms used by Cloud providers in allocating bandwidth resource and ensuring load balancing include:

- Fair-Share Bandwidth allocation: In this approach, the specific time interval is set for the flow of the defined data size between the host machine and virtual machines. The Throughput and Network weight are used to ensure fairness during the data flows. When the calculated fair share values reach the random acceptable values, the packets are transferred otherwise they are dropped.
- Transmission Control Protocol (TCP): Here, packets are divided into segments and assigned a sequence number each to specify the order of the data. Transmission takes place at a specific time interval. An acknowledgment is received for every successful transmission otherwise, retransmission takes place. This approach is connection oriented and for end-to-end congestion controls.
- Bandwidth Capping: in this approach, a specific bandwidth limit is set for the VM when transmission of data is happening simultaneously.
- Secondnet: This approach guarantees the assignment of network bandwidth amongst every VM pair and the communication patterns between multiple pair of the VM can vary due to the time as well as the data received/transferred amongst them.
- Netshare: Utilized in big data centers and function as a central system for allocating bandwidth in a virtual cloud network. It lessens the tasks in data centers. Each Cloud customer is connected to various cloud network resources. Division of the network into slices is done and each of the slices is assigned bandwidth via the link connecting it to the cloud network.
- Approximate Fairness- Quantized Congestion Notification (AF-QCN): This is a congestion control approach for data centers, delivered via a switch. The network bandwidth is shared among the transmitter's connection to the VM.

3.4. Bandwidth Management Benefits

In [39] the author highlighted the following benefits using the Zscaler Bandwidth Control solution:

- Prioritization of business applications: It limits the possible impact of bandwidth-intensive applications like streaming media, file sharing, and social media business applications. Bandwidth constraints and policies are checked and reconfigured to meet user requirement
- Deliver a better user experience: features like segmentation of network traffic and retransmission provide a faster Internet connection and avoid loss of packet after a successful data transmission.
- Reduce costs and simplify IT: Eliminate the need to install and manage additional hardware in-house and prevent bottlenecks due to the enforced cloud policies.

In [26] the author highlighted the following benefits:

- Work-conservation
- Guarantees bandwidth
- Practical enforcement
- Fairness between tenants
- Applicability to different contexts
- Easy deployment
- Precise and flexible requirement expression
- Efficient bandwidth saving

In [25] the author presented the following benefits:

- Provides dynamic network management
- Improves network efficiency
- Easy integration.

3.5. Cloud Bandwidth Management Challenges

According to [40], as VMs scale bigger, utilizing more memory, higher bandwidth, the need to maintain optimal performance and health management tools become difficult for the existing management software. These tools require multiple

information about the different features of each VM, the underlying host storage infrastructure, machine, and networking. Supporting these tools and coordinating the activities of a large number of simultaneous users with management privileges require a secure connection in real-time with a high level of consistency and must be backward compatible. Scaling the bandwidth of the ever-increasing hundreds of VMs means improving the existing code for optimal performance. Since users are getting used to virtualization and migrating every one of their in-house server workloads/applications to virtual environments, the need to scale bandwidth capacity will continuously get bigger for the predictable future. The management software needed to scale bandwidth to reach the level of the users' needs is highly challenging. [9] Presented the following challenges in the utilization of bandwidth and can become a problem for the existing management tools:

- The absence of administrative support
- Uncontrolled users' downloads
- Virus attacks
- Higher demand compared to available bandwidth
- Unreliable Internet service provider
- Technical breakdowns
- Different user bandwidth requirements
- Too many network users

4. Conclusion

Cloud computing enables many organizations store, compute and retrieve/access their application from any location over the internet. This has shifts management core concern of maintaining in-house hardware resources and reduced the cost incurred. The concept of VMs-migration in cloud computing requires that the right bandwidth size must be made available on the host or physical machine. For example, when a VM is unable to serve its users request due to a technical failure on the host machine, the VM is moved to another physical machine to provide the needed services. VMs-migrated enables better management and optimization of the data centers and host machines. This paper focused on Cloud bandwidth management. An analysis of the concept of VM-migration and bandwidth management approaches was done. In conclusion, with the increasing number of cloud users and data centers, further research work for enhancing the bandwidth allocation approaches and management is required. This will further enhance the resource-sharing feature of the VMs.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation, and Discovery (CUCRID).

References

[1] I. Odun-Ayo, O. Ajayi, and C. Okereke, "Virtualization in Cloud Computing: Developments and Trends," in 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), 2017, pp. 24–28. https://doi.org/10.1109/ICNGCIS.2017.10

[2] K. Chandrakanth and S. Gayathri, "Examining Bandwidth Provisioning on Cloud Computing and Resource Sharing Via Web," Int. J. Eng. Sci. Adv. Technol. [IJESAT], vol. 2, no. 5, pp. 1372–1376, 2012.

[3] M. Peter, and G. Timothy, "The NIST Definition of Cloud Computing," NIST Special Publication pp. 800-145, 2011. https://doi.org/10.6028/NIST.SP.800-145

[4] M. Ahmed, A. Sina, R. Chowdhury, M. Ahmed, and M. H. Rafee, "An Advanced Survey on Cloud Computing and State-of-the-art Research Issues," vol. 9, no. 1, pp. 201–207, 2012.

[5] Y. Al-dhuraibi, F. Paraiso, N. Djarallah, P. Merle, and C. Computing, "Elasticity in Cloud Computing : State of the Art and To cite this version : Elasticity in Cloud Computing : State of the Art and Research Challenges," 2018.

[6] C. T. Yang, J. C. Liu, R. Ranjan, W. C. Shih, and C. H. Lin, "On construction of heuristic QoS bandwidth management in clouds," Concurr. Comput. Pract. Exp., vol. 25, no. 18, pp. 2540–2560, Dec. 2013. https://doi.org/10.1002/cpe.3090

[7] A. Amaro, W. Peng, and S. McClellan, "A survey of cloud bandwidth control mechanisms," in 2016 International Conference on Information and Communication Technology Convergence (ICTC), 2016, pp. 493–498. https://doi.org/10.1109/ICTC.2016.7763520

[8] G. Hong, J. Martin, and J. Westall, "Adaptive bandwidth binning for bandwidth management," Comput. Networks, vol. 150, pp. 150–169, 2019. https://doi.org/10.1016/j.comnet.2018.12.019

[9] C. Technology, "Bandwidth management in universities in Zimbabwe : Towards a responsible user base through effective policy implementation Lockias Chitanana Midlands State University, Zimbabwe," vol. 8, no. 2, pp. 62–76, 2012.

[10] Y. Liu, V. Xhagjika, V. Vlassov, and A. Al Shishtawy, "BwMan: Bandwidth Manager for Elastic Services in the Cloud," in 2014 IEEE International Symposium on Parallel and Distributed Processing with Applications, 2014, pp. 217–224. https://doi.org/10.1109/ISPA.2014.37

[11] M. Chidananda, A. Manjunatha, A. Jaiswal, and B. Madhu, "Detecting Malicious Cloud Bandwidth Consumption using Machine Learning," Int. J. Eng. Technol., vol. 8, no. 5, pp. 2199–2205, Oct. 2016. https://doi.org/10.21817/ijet/2016/v8i5/160805210

[12] E. Randa, I. Mohammed, A. Babiker, and A. N. Mustafa, "Bandwidth Management on Cloud Computing Network," vol. 17, no. 2, pp. 18–21, 2015. https://doi.org/10.9790/0661-17221821

[13] G. F. Anastasi, M. Coppola, P. Dazzi, and M. Distefano, "QoS Guarantees for Network Bandwidth in Private Clouds," Procedia - Procedia Comput. Sci., vol. 97, pp. 4–13, 2016. https://doi.org/10.1016/j.procs.2016.08.275

[14] S. Jlassi, A. Mammar, I. Abbassi, and M. Graiet, "Towards correct cloud resource allocation in FOSS applications," Futur. Gener. Comput. Syst., vol. 91, pp. 392–406, 2019. https://doi.org/10.1016/j.future.2018.08.030

[15] A. Kovari and P. Dukan, "KVM & OpenVZ virtualization based IaaS Open Source Cloud Virtualization Platforms : OpenNode, Proxmox VE," pp. 335–339, 2012.

[16] Z. Xiao, W. Song, and Q. Chen, "Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment, IEEE Transaction On Parallel And Distributed Systems," pp. 1–11, 2013.

[17] J. McDermott, B. Montrose, M. Li, J. Kirby, and M. Kang, "The Xenon Separation VMM: Secure Virtualization Infrastructure for Military Clouds", Naval Research Laboratory Center for High Assurance Computer Systems, Washington. https://doi.org/0.1109/MILCOM.2012.6415673

[18] M. Masdari, S. Shahab, and V. Ahmadi, "An overview of virtual machine placement schemes in cloud computing", Journal of Network and Computer Applications, vol. 66, pp. 106–127, 2016. https://doi.org/0.1016/j.jnca.2016.01.011

[19] R.P. Patel, "Cloud computing and virtualization technology in radiology", Clinical Radiology, vol. 67, pp. 1095 -1100, 2012.

[20] L. Shing-Han, C. David, C. Shih-Chih, P.S. Chen, L. Wen-Hui and C. Cho, "Effects of virtualization on information security", Computer Standards & Interfaces, vol. 42, pp. 8, 2015. https://doi.org/10.1016/j.csi.2015.03.001

[21] E. Baccarelli, D. Amendola, and N. Cordeschi, "Minimum-energy bandwidth management for QoS live migration of virtual machines," Comput. Networks, vol. 93, pp. 1–22, 2015. https://doi.org/10.1016/j.comnet.2015.10.006

[22] J. Chase, R. Kaewpuang, W. Yonggang, and D. Niyato, "Joint Virtual Machine and Bandwidth Allocation in Software Defined Network (SDN) and Cloud Computing Environments," pp. 2969–2974, 2014.

[23] H. Tan, L. Huang, Z. He, Y. Lu, and X. He, "Journal of Network and Computer Applications DMVL : An I / O bandwidth dynamic allocation method for virtual networks," J. Netw. Comput. Appl., vol. 39, pp. 104–116, 2014. https://doi.org/10.1016/j.jnca.2013.05.010

[24] X. Chen and S. Mcclellan, "Improving Bandwidth Allocation in Cloud Computing," no. August, 2016.

[25] J. Cao, Z. Ma, J. Xie, X. Zhu, F. Dong, and B. Liu, "Towards tenant demand-aware bandwidth allocation strategy in cloud datacenter," Futur. Gener. Comput. Syst., 2017. https://doi.org/10.1016/j.future.2017.06.005

[26] J. P. D. Comput, H. Yu, J. Yang, H. Wang, and H. Zhang, "Towards predictable performance via two-layer bandwidth allocation," J. Parallel Distrib. Comput., vol. 126, pp. 34–47, 2019. https://doi.org/10.1016/j.jpdc.2018.11.013

[27] M. Balaji, C. Aswani, and G. S. V. R. K. Rao, "Predictive Cloud resource management framework for enterprise workloads," J. King Saud Univ. - Comput. Inf. Sci., vol. 30, no. 3, pp. 404–415, 2018. https://doi.org/10.1016/j.jksuci.2016.10.005

[28] I. Odun-ayo, A. Falade, and V. Samuel, "Cloud Computing and Open Source Software : Issues and Developments " C," vol. I, pp. 140-145, 2018.

[29] S. C. Mondesire, A. Angelopoulou, S. Sirigampola, and B. Goldiez, "Simulation Modelling Practice and Theory Combining virtualization and containerization to support interactive games and simulations on the cloud," Simul. Model. Pract. Theory, no. August, pp. 1–12, 2018.

[30] V. Chang et al., "Journal of Network and Computer Applications," vol. 77, no. May 2016, pp. 87–105, 2017. https://doi.org/10.1016/j.jnca.2016.10.008

[31] U. Agreement, R. Issued, B. Internet, T. Advisory, G. Technical, and W. Group, "Port Blocking," no. August 2013.

[32] S. Mustafa, B. Nazir, A. Hayat, R. Khan, and S. A. Madani, "Resource management in cloud computing : Taxonomy, prospects, and challenges q", Comput. Electr. Eng., vol. 47, pp. 186–203, 2015. https://doi.org/10.1016/j.compeleceng.2015.07.021

[33] D. Ardagna, B. Panicucci, M. Trubian, and L. Zhang, "Energy-aware autonomic resource allocation in multitier virtualized environments," IEEE Trans Serv Comput; vol. 5, pp. 2–19, 2012.

[34] D. Ardagna, S. Casolari, M. Colajanni, and B. Panicucci, "Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems." J Parallel Distrib Comput; vol 72, pp. 796–808, 2012.

[35] L. Chunlin, and L. Layuan, "Multi-Layer RM in cloud computing. J Network Syst Manage", 2013. http://dx.doi.org/10.1007/s10922-012-9261-1

[36] S. Ali, S. Jing, S. Kun, "Profit-aware DVFS enabled RM of IaaS cloud. Int J Comput Sci Issues (IJCSI)", vol. 10 pp. 237–47, 2013.

[37] R.K. Grewal, and P.K. Pateriya, "A rule-based approach for effective resource provisioning in hybrid cloud environment. Adv Intell Syst Comput", vol. 203, pp. 41–57, 2013.

[38] Y. Ge, Y. Zhang, Q. Qiu, and Y. Lu, "A game theoretic resource allocation for overall energy minimization in mobile cloud computing system, In: Proceedings of the 2012 ACM/IEEE international symposium on low power electronics and design ISLPED", vol. 12,. pp. 279–84, 2012.

[39] D. Sheet, "Zscaler Bandwidth Control."

[40] X. Xu, Q. Zhang, S. Maneas, S. Sotiriadis, and C. Gavan, "Simulation Modelling Practice and Theory VMSAGE : A virtual machine scheduling algorithm based on the gravitational effect for green Cloud computing," Simul. Model. Pract. Theory, no. June, pp. 1–17, 2018. https://doi.org/10.1016/j.simpat.2018.10.006

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 224-233 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Fuzzy Simulation of Historical Associative Thesaurus

Olena Basalkevych[*,1], Olexandr Basalkevych[2]

[1]*Danylo Halytsky Lviv National Medical University, Department of Medical Informatics, 79005, Ukraine*

[2]*Product Engineering and Development Globallogic Inc., 79016, Ukraine*

ARTICLE INFO

Article history:
Received: 03 June, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
fuzzy associative field
stimulated fuzzy associative field
fuzzy associative word structure
fuzzy associative map
fuzzy historical associative thesaurus
associative antipodes
associative experiment
Osgood's semantic differential
fuzzy associative differential
fuzzy associative stratification
Older Scots.

ABSTRACT

Research is dedicated to fuzzy reconstructions in the diachronic associative linguistics, and encompasses several simulation practices. The principles of an associative memory model and fuzzy computing with words are adopted in the working methodology. The evolved mathematical model reanimates the older patterns of Scottish associative map by means of fuzzy logic. The compiled fuzzy associative fields are connected by epidigmatic relations represented by fuzzy associative word structures in a fuzzy associative map constituting the grounds of the proposed thesaurus. The simulation of the historical associative thesaurus enables extended associative scrutiny: tracing associative antipodes, reconstructing an associative experiment for any map associate, finding words' similarities and distances, as well as associative differentiating and stratifying the Older Scottish lexis. The special interest represents the fuzzy associative differential, an electronic tool for estimating the lexis through its stable syntagmatic associates. All the simulation processes are fuzzy initiated and quantified. The proposed practice in its final configuration has no precedents, representing the novelty of theoretical interpretation and empirical application. The proposed terminology is mainly introduced by the researchers.

1. Introduction

Ethnoconsciousness of older speakers is the matter of research in the represented work. Associative maps of the modern ethnoses are mainly grounded on an associative experiment, which allows ascertaining the lingual associations through collecting the reactions of respondents to stimuli, the practice that is absolutely impossible for literal realization at the older language-states. Logically, to reconstruct the map and subsequently to simulate an associative experiment in the diachronic psycholinguistics seems real under the study of lexical relations present in the synchronic associative map. Such relations could be artificially reconstructed, quantified and canonically combined to reanimate the mental lexicon and all the related issues for the periods more or less remote in time.

In our research, the capability of fragmental reconstructing and activating the mental lexicon of Older Scots is realized with fuzzy modelling. The first turning point of the research is compilation of the diachronic synonymic sets of Older Scottish adjectives. The source of compilation is the Dictionary of Older Scottish Tongue (DOST) [1], an explanatory dictionary, providing the English description to Older Scottish lexis. The incentive is a synonymic set of English adjectives, any of which becomes an explaining semantic feature for the extricated Older Scottish synonyms. All those synonyms are gathered into a set, and ordered chronologically. Such sets are characterized by the connotative, stylistic and ethnic variability. The second decisive moment is the concluding of associativity nature of the compiled sets. The fuzzification of the sets is enabled by Zipf's regularity allowing to treat the diachronic measure, calculated for any set synonym, as that of associativity. In our research the compiled fuzzy sets are called the fuzzy associative fields (FAFs). The revealed fuzzy associative word structures (FAWSs) connect FAFs into the network, called an associative map or associative lexicon of Older Scots. Another associative relationship – antonymy – is brought into the structure, joining any FAF with its antonymic one. Thus a two-wing associative map is evolved at the stage. The fuzzy marking of the associative map allows us to

[*]Olena Basalkevych, 12 Zelena str., Lviv, 79005, +380955365348, olena.basalkevych@gmail.com

run the evolved network, realizing multitask simulation of the mental Older Scottish lexicon. Thus the first type of the run enables finding the associative chains through positive and negative wings of the map, called associative antipodes.

The following iconic associative relationship – syntagmatic one – attaches the available in the dictionary collocates to any adjective-knot of the network. Then the associative experiment in its nominal format is ready to be reproduced, bringing to any adjective of the map the words from its fuzzy associative structure; all existing till the moment associates within the major field; time antonym with its dominant and, finally, an available collocate. All the entries are collected in one fuzzy set, called a stimulated fuzzy associative field (SFAF). The simplest fuzzy metrics of similarity and distance could be found only within the same FAF, since they are not context- and turn-sensitive. The junction of the two antagonist-fields is regarded as a fuzzy differential, which enables the differentiation of the available Older Scottish lexis throughout the network. It gathers for any word the fuzzy subset of all its syntagmatic collocates available within the associative fragment. The new recommended method introduces into a traditional Osgood's semantic differential the expected denotative meanings. The electronic tool, created for the purpose, we call a fuzzy associative differential. The fuzzy stratification of Older Scottish lexis is another issue of the same approach. The electronic resource realizing all the facilities for the activated part of the associative map is called a fuzzy historical associative thesaurus (FHATh). All the techniques are grounded on the classical principles, notified in the following chapter, but represent an independent system, compiled in the series of previous author's works [2-4], mentioned throughout the paper. The absolute novelty of the practice lies in its possibility to imitate the older mental lexicon patterns on the basis of the proposed row of reconstructions, which was not applied in the diachronic linguistics before. The created electronic resource completely corresponds to the principle of computing with words: the user is pressing a word in the map and software returns the estimating phrase (supported with illustrating numbers) in response [5].

The classical sources whose concepts make the background of the developed mathematical model are analyzed in the following Chapter 2; and the comparison with the resonant works is realized afterwards (Chapter 3). The stages of development of the mathematical model are described in Chapter 4. The realization of the model resulted in different research techniques is explained in Chapter 5. The software packet and its principles of computing with words are described at the very end of the chapter. Finally, the received results are discussed (Chapter 6), and general conclusions are drawn (Chapter 7).

2. Report on Scientific Sources

2.1. Associative thesaurus of J.Kiss

One of the cotemporary and highly professional patterns of an associative map, reflecting the associative structure of words, was developed in 1972 under the aegis of British Medical Research Council by J.R. Kiss in his "An associative thesaurus of English", which is defined as a word association network represented by an oriented linear graph with labelled arcs, where the nodes are words and labels on the arcs give estimates of stimulus-response probabilities. As to its organization: "Two different approaches to the semantic structuring of the thesaurus data can now be distinguished. One is aimed at the discovery of semantic components for a word (covering all its word senses). The other is aimed at the discovery of groups of words which are relevant to each other because the concepts they label go together (although may not be similar to each other)" [6]. The concept becomes the first foundation of the proposed work. Our oriented graph of a network configuration called a fuzzy associative network (FAN) is organized following both processing approaches, the latter is well represented by FAFs gathering Scottish words whose main semantic components "go together" in synonymic sets of English adjectives, and the former completely corresponds to FAWSs.

2.2. Age-Frequency Regularity of G. Zipf

The second foundation is the regularity of G.K. Zipf stating the relationship between frequency and age of a word: "Hence, in sum, we have found not only a direct relationship between relative age and relative frequency, but also an inverse relationship between relative age and relative size in words of the same frequency" [7]. His conclusion on the English lexis time-age correlation states that the rank of the most frequent modern English words contains mostly the words originated from Old English [7]. In our research, Zipf's statement allows us to treat a diachronic measure as a frequency measure and therefore as an associativity one.

2.3. Fuzzy Logic

The state of affairs actualizes the third foundation of the work that is fuzzy logic, which shares the mathematical apparatus ready to fix the partial state of truth of a statement in our mind or speech. The honoured founder of the theory Lotfi Zadeh was working with the human language when fuzzy logic emerged, the logic implying the numerousness of partial truths on the way from nonsense to absolute. Lotfi Zadeh argued that probability lacks sufficient "expressiveness" to deal with uncertainty in the natural language. According to [8], in his "Computing with Words" Zadeh writes: "Humans have many remarkable capabilities. Among them there are two that stand out in importance. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information and partiality of truth. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations. In large measure, Computing with Words is inspired by these remarkable capabilities."

A linguistic variable refers to the set of values more or less representing the emerged truth. The degrees of such a representation, or the measures of truth partiality, are caught by *membership functions*, normally, real numbers lying in the interval from 0 to 1 inclusively. The set is called a *fuzzy set*.

The formal definition: Let E be a countable or uncountable set and x be an element from E. Then *a fuzzy subset A of the set E* is characterized as a set of the ordered pairs $\{(x, \mu_A(x))\}, \forall x \in E$; where $\mu_A(x)$ is a characteristic fuzzy function (or a membership function) that takes its values in totally ordered set M or within [0;1], and indicates the level, or degree, of the element x belonging to the subset A. A linguistic variable alters in names, taking them from the set of verbal (word) characteristics,

representing all the sorts of it. Normally, such a set is called *a term set* [9, 10].

The fuzzification of our structure starts from the associativity measure that corresponds to a membership function, reflecting the degree of associativity of field members with a dominant. In short terms, Zipf's regularity enables fuzzy simulation of the diachronic associative thesaurus.

The fuzzy metrics of distance and similarity borrowed from the same theory [9] are available for all the elements of the represented associative map fragment.

Granularity in fuzzy logic implies dealing with data points within a granule as a whole rather than individually [11]. In our work it is used in consolidation phase while differentiating Older Scottish words. Available word-characteristics with their fuzzy estimates are grouped into six concept sets or granules with deriving the general probability and fuzzy measures for any granule. Fuzzy stratification is another way to address to the point data through the stratum which contains them [12]. The approach seems to be very profitable in the pay of the historical lexicography, naturally laying its verbal wealth layer over layer.

Thus the mathematical model of the represented work is extendedly grounded on fuzzy logic.

2.4. Osgood's Semantic Differential

The constructive element of FHATh is a junction of two antonymic FAFs which is constructed and recommended by our research to be used as a fuzzy associative differential (FAD). A FAD could be treated as a fuzzy diachronic modification of Ch. Osgood's semantic differential. FAD represents a lot of direct associates to different qualities, while the scales of the original Osgood's differential contain the intensity markers of these qualities. Preferably, literary meanings, since all of them are taken from the dictionary, are used in the scales of FAD, bearing classical denotative meanings with different connotative shades. Therefore, Osgood's conflict of 'connotative emotions' versus 'denotative precision' and the symbolic character of his differential scale-intervals [13] are moderated in our development. Meanwhile, the impartiality of Osgood's intensity markers whose applicability span could not be overestimated lacks in FAD. The most popular scales of Osgood's differential are Good-Bad, Strong-Weak and Active-Passive, representing so-called cross-cultural universals of Evaluation, Potency and Activity of the considered words. Our research adopts the universals for its measuring purposes.

2.5. Associative Network Memory Model

The principle of the network performance is comparable with "spreading activation" of an associative network memory model: when the concepts of interest are activated, "excitation" penetrates through the links to the associated concepts. More frequent in the past associations are stronger and easier to be excited. If the emitted excitation is potent enough to activate some of those associations and surpass the threshold, "activation is spread" and new concepts are brought to mind. The associative memory structure is highly metaphorical but still the only manner of simulating the physiology of the brain [14]. Normally, the network is evoked episodically in response to the stimulus, visualizing the associative map fragments which combine logogens as images of words, containing the information on the word valency, including stable associations. In the present work, the map is also activated episodically. Information on an associativity level of the map words as well as the smart orientation ability is hidden in the fuzziness of the network patterns.

3. Comparison Analysis

3.1. Fuzzy Hindi WordNet

In line with the proposed research, the fuzzy interpretation of Hindi WordNet [15] is realized with a fuzzy graph, and represents an ideal comparison pattern with similar underlain principles and inner graph morphology but a different fuzzification method. Our original work [2], which appeared earlier than the regarded comparison source, concentrated more on the paradigmatic and epidigmatic associativity of the proposed map. The most salient and fundamental difference of Fuzzy Hindi WordNet lies in treating the fuzzy graph vertices, synonymic sets, so-called synsets of nouns, verbs, adjectives, and adverbs as the sets of open-class words where the strength of every node is always 1 [15]. In our study the associative fields of adjectives are represented by fuzzy sets, employing the inherited vagueness of linguistic variable values that are normally represented by adjectives as degree-measurers [16]. Evidently, two works differ much by the calculus of membership functions: the classical expert opinions used in Fuzzy Hindi WordNet [15] seem to be hardly reachable for older languages, therefore we reasonably adapt a diachronic measure for the purpose. The weighed in the mentioned manner semantic and lexical connections in [15] are organized among synonymic sets and words, respectively, through association, hypernymy, hyponymy, meronymy, holonymy, antonymy, entailment, troponymy, gradation, and causative ties, whilst in [2] the epidigmatic relations are combining the different associative fields through the entries of the same words. As to the antonymic relation, in our case it immediately connects diachronic antonyms and the whole antonymic fields, containing them. The junction is realised through the antonymy of their dominants, following the proved associativity of antonyms [17]. The goals of fuzzy upgrade also diverge, since we represent a fuzzy reconstruction device whilst [15] is more about fuzzy metrics allowing to extricate the context word values in the modern Hindi discourse and text. As to the research sway, our work is rather an illustrated recommendation to a new type of historical thesaurus while Fuzzy Hindi WordNet represents an accomplished modern lexicography development.

3.2. Fuzzy Neural Networks

In the work, the morphology of the mentioned network is sophisticated in stipulation but simple in performance, and probably with much higher structural accuracy could be realized by means of a fuzzy neural network. It is equipped with fuzzy-valued connection weights and morphological neurons good at logical operations of conjunction and disjunction or dilation and erosion in terms of [18] among the others. All the more argument is that a neural network seems to be the most appropriate for simulating the associative reasoning [19]. Nevertheless, at the current stage of the research the proposed in the article fuzzy

model suffices completely the purposes of the set lexicological research.

4. Mathematical Model

The exposed interdisciplinary correlation of facts and dogmas flows out into the model serving the diachronic associative linguistics with fuzzy logic. The principles were originally conceived in the work [2], and developed in the following works, including [3, 4]. There could be discerned several stages in the evolution of the model.

4.1. Compilation of the Diachronic Synonymic Sets of Older Scottish Adjectives

The technique starts from the English synonymic adjectival sets taken from a dictionary of synonyms of Oxford Learner's Thesaurus (Oxford Thesaurus 2008) [20] or WordNet (WordNet 2006) [21]. They become the search stimuli in the DOST: the resulting Scottish adjectives contain them as defining semantic features. The reactions to all the English set adjectives are compiled into the set of Older Scottish synonyms. Their frame registers the first citations they were mentioned in, or so-called diachronic text prototypes [22], the year of the citations, or so-called advent time, their source, encyclopedic meanings and etymological notes. Synonyms are ordered in a chronological order, and any of them is attributed with a diachronic measure, calculated due to the formula:

$$W_i^{(y)} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \left(\frac{1}{n+1} - 1 \right) + 1, \qquad (1)$$

where ($y_{\max} - y_{\min}$) is a diachronic range of a set, ($y_i - y_{\min}$) is a distance to the appearance of the concrete word in a set, and n – the number of words in a set [23]. The organization and type of the sets imitate the Historical Thesaurus of the Oxford English Dictionary [24]. Primarily, the sets are called diachronic synonymic sets, and there are 100 of them in our database. They result in very heterogeneous collections of words affiliated by the same or comparable semantic features but reflecting the notable wideness of connotative, stylistic and ethnic coloration, which evidently leads them out of ordinary synonymy.

4.2. Associativity Nature of the Constructed Diachronic Synonymic Sets

The Dictionary of Older Scottish Tongue (DOST) is a resource most readily associated with Older Scots knowledge in people's mind, according to the dictionary definition given by Budanitsky [13]. The historically forged layers of lexis are stored there, keeping the primary syntagmatic associations reflected in numerous collocations and sentences, which contribute to the secondary paradigmatic associations [25]. Since all the words are dated, and represented with diachronic text prototypes, the historical dictionary like the DOST is expected to be the people's memory enfolded. The organization of memory is hypothetically comparable with an associative map as is mentioned in Chapter 2.5.

Therefore, the set goal of the research is to unfold fragmentally the associative map of Older Scots from the DOST. The paradigmatic collections within the received diachronic synonymic sets reflect the associative relationship, since *the process of their compilation is regarded as a discrete associative experiment with historical reaction* (the traditional associative experiment counting numerous reactions to a stimulus is called a discrete associative experiment with prolonged reaction [14]). The English synonyms, the DOST search items, play the role of stimuli. Thus at the stage, we rename the diachronic synonymic sets into synonymic associative fields.

4.3. Synchronization of the Associative Structures

The projection of the diachronic synonymic sets into the synchronic plane is realized by the mentioned in Chapter 2.2. Zipf's *age-frequency* correlating regularity, which states that the oldest words of the lexicon are the most frequent ones. On the other hand, the first words given during an associative experiment are the characteristics most frequent in mind, according to [14, 26]. Both facts directly influence the nature of the diachronic measure (1): it is converted into a frequency measure, and then in an associativity one. The hypothetical syllogism leads to the treatment of diachronic coefficients (1) as values of associativity of the field members with their dominant.

4.4. Fuzzification

Following the tenets of Chapter 2.3., E is a huge countable set of all the Scottish words (or adjectives only). From E, we could discern the fuzzy subsets A_j of words rendering specific semantic features or dominants, represented by English adjectives, which give the names to A_j. Membership functions or the degrees of reflecting these semantic features are completely expressed by associativity measures (1). Consequently, a Dominant is a linguistic variable and takes the values from its term set: *Dominant={Good, Bad, Strong, Weak, Active, Passive, Flexible, Inflexible, Brave, Cowardly, Intelligent, Stupid, Large, Small, Lucky, Unlucky,…}*. Altogether it contains 100 name-values (j=1:100). In our research, fuzzy subsets A_j are called *fuzzy associative fields* [2-4] or FAFs hence. For example, FAF *Active*= (douchty/1, fere/1, deliver/1, strenthy/1, licht/1, forthirwarde/0.99, unswere/0.87, agile/0.87, fery/0.87, active/0.78, wirkand/0.7, operative/0.66, lifly/0.63, tait/0.63, ȝape/0.63, restles/0.63, frak/0.63, trig/0.63, laborious/0.59, nocturnall/0.59, slepry/0.53, liver/0.52, throuche/0.49, gangand/0.46, actual/0.43, feat/0.42, ferdie/0.36, ȝauld/0.32, suddan/0.3, spritly/0.12, vigent/0.03).

From the totality of FAFs, we can extricate an associative word structure [2-4], epidigmatic relationship, stating the level of word associativity with different Dominants. We call it a *fuzzy associative word structure* [2-4] or FAWS hence. The example of FAWSs: Douchty=(Active/1, Vigorous/1), Agile=(Active/0.87, Vigorous/0.87), Unswere=(Active/0.87, Vigorous/0.86, Helpful/0.83).

Commenting the given instances, the variability of belonging for words Douchty, Agile and Unswere is not large, we observe the same Dominants with high associativity degrees in all the three cases, which states the intension nature of the mentioned semantic features.

4.5. Fuzzy Association Map

The unfolding of the associative map hidden in the historic dictionary starts from the connection of the evolved fuzzy

associative relations into the structure corresponding to the requirements of an associative thesaurus of English by J.Kiss [6]. The structuring approaches are well described in Chapter 2.1. and are completely sufficed by the combination of the developed FAFs and FAWSs into a map. One more strong associative relation proved by an associative experiment is an antonymic one [17, 27]. In our visualization, technically saying, we connect three positive FAFs Good, Strong, and Active with their antonymic ones Bad, Weak, and Passive, creating two wings of the *fuzzy associative map* or FAM [3, 4]. Any map wing contains a lot of between-field connections, combining the same words located in different fields. Some of the words and connections are distinctly represented in Figure1. The between-field connections themselves introduce into the map the epidigmatic relationship and are given by FAWS.

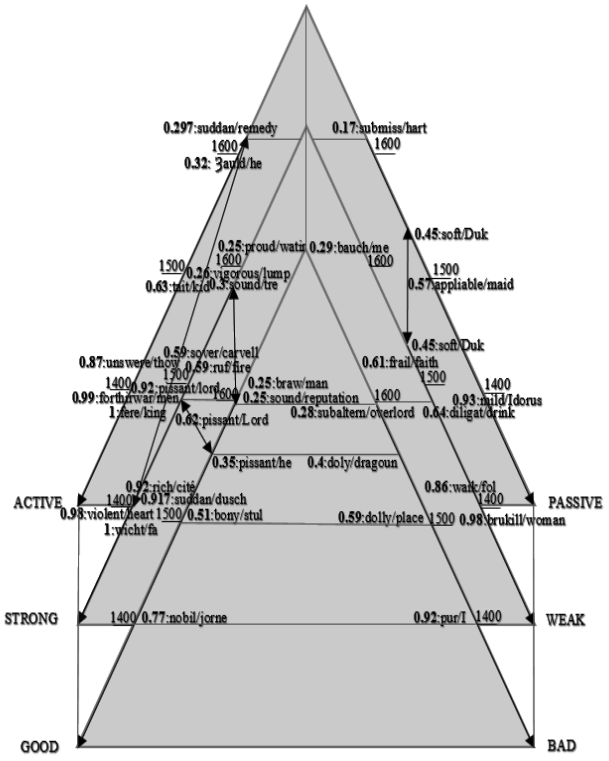


Figure 1: Fragment of FHATh.

The given fragment of a fuzzy associative map of Older Scottish adjectives covers the scales of the cross-cultural universal involved in the associative techniques of lexis Potency (Good-Bad), Activity (Active-Passive) and Evaluation (Good-Bad) ascertaining. Far not all the associates are visualized there: technically complete scales are represented by software forms.

5. Simulating Techniques

The compiled map constitutes the ground for FHATh. In a final condition it is lined up with syntagmatic associations: any adjective of the set is paired with noun or pronoun taken from the diachronic text prototypes, represented by the DOST. The diachronic text prototype is the oldest quotation registered in the literature, containing the considered word. Following Zipf's regularity, we proved that the oldest word is the most frequent one and therefore contains the greater associativity strength. The same concerns the gained by the word syntagmatic relations, documented in the diachronic text prototype, which are the most prototypical and thus the most associatively stable. (Find the collocations in Figure 1.) Their fuzzy strength due to propagation acquires the fuzziness of the adjective.

The associativity of synonymic, antonymic and syntagmatic relations is supported by Deese in his «The structure of associations in Language and Thought», described in [17]. So the complete associative map is justified and thus ready to run. The activation of the associative map corresponds to the "spreading activation", caused by the "fired" stimulus, and is well-directed, since the map is represented by an oriented graph of a network configuration. The fuzziness plays its selective function in the process of network navigation. An associative map activated in the manner is called by our research a *fuzzy associative network* or FAN, and the process of its target activation is hence named the simulation of a target.

5.1. Associative Antipodes

The first target is an associative environment. Simulation of the associative environment ends up finding two associative chains through positive and negative map wings, called associative antipodes. To find the associative environment to the chosen on the map word, algorithm traces this word FAWS, detecting the word-place with the maximum membership function, after falls to the dominant of the found word as one with the maximum associative measure or membership function, and repeats the same scenario for this dominant till no more connections could be revealed, finding the n-association chain to a word-stimulus at the 2n-th step. The fuzziness of a chain acquires the minimum membership function of the passed places, according to conjunction in fuzzy logic:

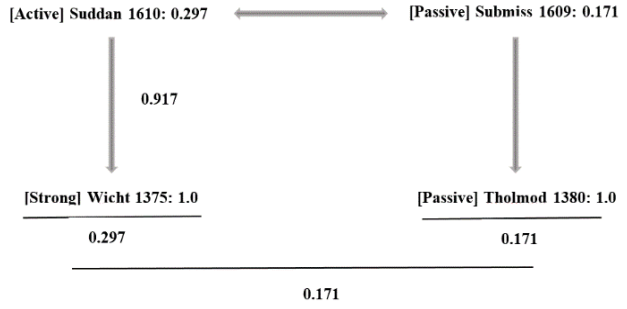
$$\mu(\mathrm{x}_1 \cap x_2) = \min[\mu(x_1), \mu(x_2)]. \quad (2)$$

Then the same figurant is connected with its time-antonym, and the run is repeated within the opposite map wing till at the 2m-th step it reaches the m-association chain with the minimum membership function. Both chains constitute the n/m associative environment for the chosen word, or associative antipodes. The activated in the manner map we call *2 run fuzzy associative network* (2RFAN) [4]. In Figure 1, two pairs of antipodes are traced according to the algorithm. They are graphically derived in Figure 2.

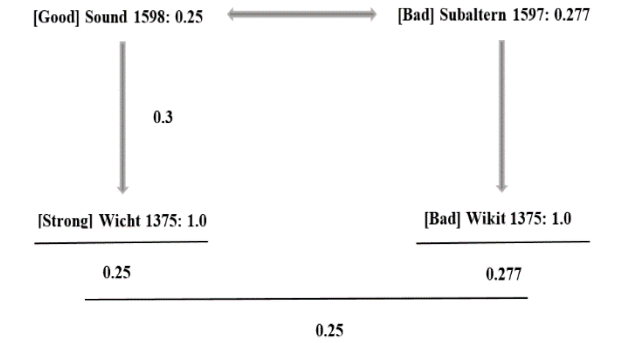
The program realization of any thesaurus function, including the tracing of an associative environment, meets the principle of computing with words: a user is pressing a word in the associative map (Figure 3) and receives the evaluating phrase about the word's relations. The phrase is constructed on the basis of fuzzy values involved in estimating the associativity level or degree of belief of the received results (Figure 4).

The console (Figure 4) states that the positive associative environment for the word Suddan from the field Active is supposed to be Wicht from the field Strong with Low degree of belief (0.3). Its negative associative environment is supposed to be Submiss from the field Passive and the dominant Tholmond

from the same field, this time with Very low degree of belief (0.17). In its turn, the positive environment for Sound from Good is also Wicht from Strong but with lower degree of belief (0.25), and the negative antipode here is Subaltern from Bad and Wikit from Bad with Low degree of belief (0.28). The data support the graphics in Figure 2.



(a)



(b)

Figure 2: Associative antipodes: (a) [Suddan/Wicht]-[Submiss/Tholmod]; (b) [Sound/Wicht]-[Subaltern/Wikit]

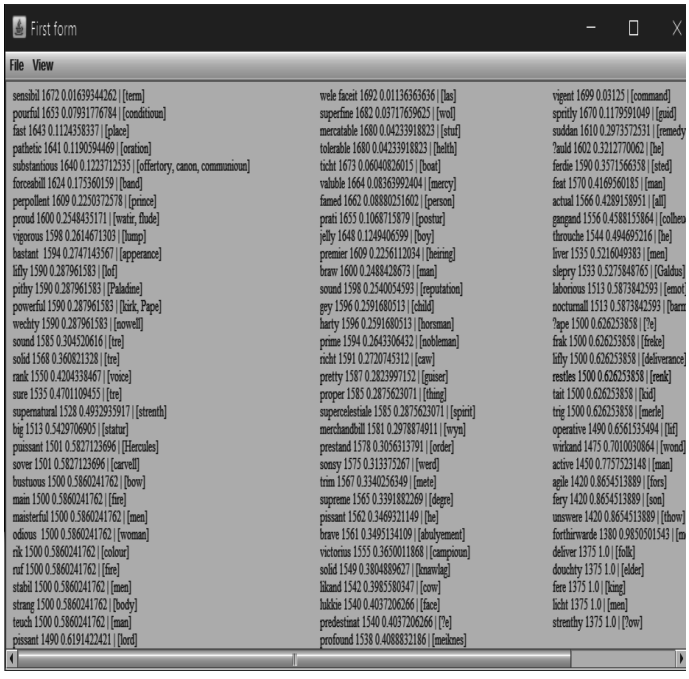


Figure 3: Map for selecting the words and realizing the simulation techniques.

In both cases we observe one semantic shift through positive map wing and no one but falling to the dominant in the negative wing. The same links could be singled out in Figure 1.

The found associative ramifications reveal the deepness of the associative map. On the limited fragment it shows quite a good coherence tendency.

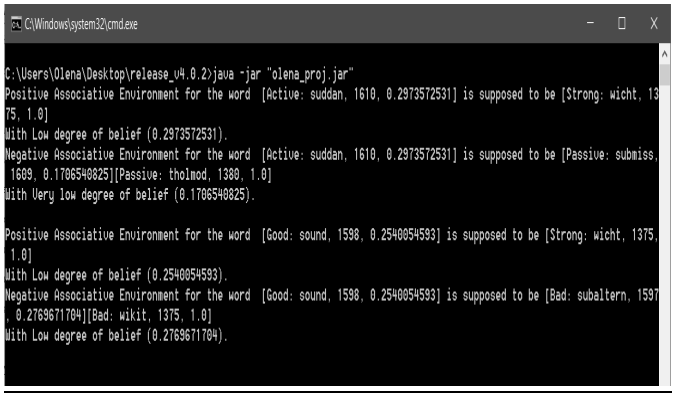


Figure 4: Console with results of tracing an associative environment for the word Suddan from the field Active, and Sound from the field Good.

5.2. Associative Experiment

According to the revealed regularities of the real associative experiment [14, 17, 25, 26], its schematic reproducing includes associative structure of the word (epidigmatic components), and from the field where its associativity is maximal: older synonyms of the word (paradigmatic components), collocates (syntagmatic components), time antonym and the dominant of a time antonym (paradigmatic components). All the entries are grouped into one fuzzy set or a *stimulated fuzzy associative field* or SFAF. Herewith, the dominant of the most associative field is entering the set with its proper fuzzy value 1 for subjective reasons of immersing in its mental plan. The membership function of a collocation is suggested by the adjective fuzziness. In such a case, an experiment could be treated as independent. The context-dependent experiment extracts the infield associations from the FAF of word localization. For instance, on pressing Pissant at the scale Good, within the framework of the independent associative experiment the algorithm traces all dominants Pissant belongs to [wicht/0.62, quhite/0.35]. The strongest association is observed in the FAF Strong, so an algorithm flexibly changes the field and extracts all the older synonyms, collocate Lord and time antonym Diligat with its dominant Tendir from FAF Weak, adds to the whole "community" the remaining dominant Quhite from the "calling" FAF Good, and orders the received fuzzy set. The program release of an associative experiment simulation for the stimulus Pissant is in Figure 5:

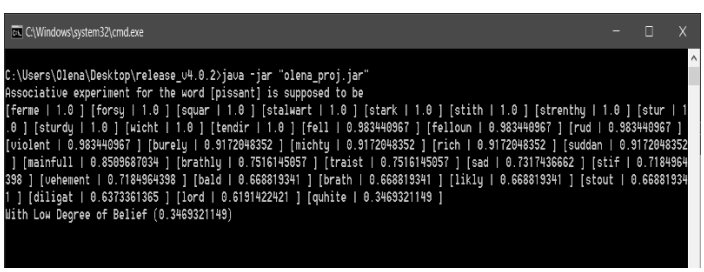


Figure 5: Console with the results of simulating an associative experiment for the stimulus Pissant.

The stability of the given SFAF or the associative reaction to the stimulus is read from the minimum membership function of its entries, normally that of the last element of the ordered SFAF. In our case, it is 0.35: with Low degree of belief the Older Scot would react to the stimulus Pissant in the proposed manner. If one needs more trustful result, the set should be cut at the required fuzziness level, expelling the newest and thus the rarest words.

5.3. Distance and Similarity

So far the infield trivial fuzzy metrics have been employed by FHATh to quantify the similarity and distance between associates. Thus the distance between associate x_i and associate x_j is

$$D = |\mu(x_i) - \mu(x_j)|, \quad (3)$$

where $\mu(x_i)$ and $\mu(x_j)$ are their membership functions or associativity measures, respectively, and their similarity is

$$S=1-D. \quad (4)$$

The classical fuzzy metrics' formulas are employed, and since they are not semantic-shift sensitive, we apply them only within a field. The between-field associativity metrics seem to be more complicated, and already regarded associative antipodes bring the maximum information on the direction and value of existing associativity among different fields. The program release for Distance and Similarity calculus for word-associates Valuble and Victorius from FAF Good is given in Figure 6:

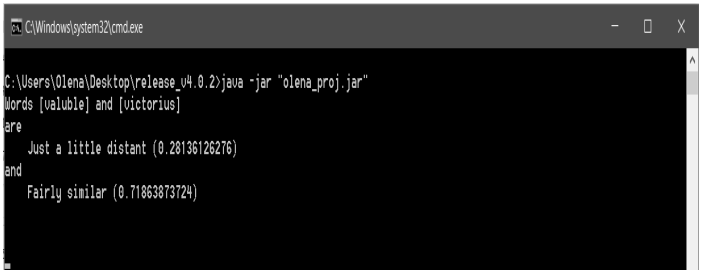


Figure 6: Release of fuzzy distance and similarity calculation for associates Valuble and Victorius from the FAF Good.

The Distance between Valuble and Victorius is rather small 0.28, that is why the software concludes that they are just a little distant. Therefore, the Similarity is quite high 0.72, with the conclusion that they are fairly similar.

5.4. Fuzzy Associative Differential

The function of the proposed by authors FAD is to provide as many associations to the considered quality as possible to describe it naturally and softly. The junction of antonymic FAFs serves the best for the purpose [3, 4]. All possible subtleties are gathered there and measured with associativity degrees. Three universals of Evaluation, Potency and Activity are depicted in Figure1: the space allows us to provide just sectional semantic values, the completeness of them is visible in the program forms (one of them is given in Figure 3). The scales are ready to serve any lexical request, and the tiniest shades of the quality could be communicated by them to the regarded word. In the case of the older language patterns, the service reflects the stable language collocations, provided by the diachronic text prototypes, and therefore treated as the most strengthened in time. So we gather all the possible characteristics for the collocate, regarded in the map, throughout all the six differential's scales and consolidate them into six fuzzy sets, or six granules in terms of fuzzy logic, mentioning the proper membership function $\mu_{Granule}$ and probability $P_{Granule}$ for any granule:

$$Word=\{Potency:[(Strong, \mu_{Strong}): P_{Strong}] \ OR/AND \ [(Weak, \mu_{Weak}): P_{Weak}]; \ Evaluation:[(Good, \mu_{Good}): P_{Good}] \ OR/AND \ [(Bad, \mu_{Bad}): P_{Bad}]; \ Activity:[(Active, \mu_{Active}): P_{Active}] \ OR/AND \ [(Passive, \mu_{Passive}): P_{Passive}]\} \quad (5)$$

Where, two antonymic granules, positive and negative, correspond to any criterion. Altogether, there are three universal criteria: Potency, Evaluation, and Activity.

For any of them:

$$P_{Granule}+P_{AntGranule}=1 \quad (6)$$

Positive and negative antonymic granules constitute a sample space: the probability of being strong is opposite to the probability of being weak.

The probability of any granule is equal to the sum of probabilities of its term-units:

$$P_{Granule}= = P_{Granule}(x_1)+ P_{Granule}(x_2)+...+P_{Granule}(x_k) \quad (7)$$

$$P_{AntGranule}= =P_{AntGranule}(x_1)+ P_{AntGranule}(x_2)+ ...+P_{AntGranule}(x_l) \quad (8)$$

Where, the number of term-units in both antonymic granules makes n: $n=k+l$, (9)

with k as the number of term-units in the positive granule, and l as that in negative one.

Then for any term-unit x_i either from positive or negative granule, the probability is:

$$P_{Granule}(x_i)=P_{AntGranule}(x_i)= 1/n \quad (10)$$

The membership function of the positive granule, following the disjunction in fuzzy logic, constitutes:

$$\mu_{Granule} = =max(\mu_{Granule}(x_1), \mu_{Granule}(x_2), ... \mu_{Granule}(x_k)) \quad (11)$$

Where, $\mu_{Granule}(x_1), \mu_{Granule}(x_2), ... \mu_{Granule}(x_k)$ are the membership functions of linguistic term-units of the positive granule.

The same for the negative granule:

$$\mu_{AntGranule}= = max(\mu_{AntGranule}(x_1), \mu_{AntGranule}(x_2), ... \mu_{AntGranule}(x_l)) \quad (12)$$

Where, $\mu_{AntGranule}(x_1), \mu_{AntGranule}(x_2), ... \mu_{AntGranule}(x_l)$ are the membership functions of linguistic term-units of the negative granule.

Thus we can create six differentiating granules for any collocate of the map, narrowing its evaluation to traditional

Osgood's coordinates, but implicitly contributing to them both denotative and connotative estimates of the criteria, to balance the situation when: "Osgood's semantic differential was an attempt to represent words as entities in an n-dimensional space, where measuring the distance between them could naturally follow from our knowledge of Euclidean geometry. Unfortunately, after extensive experimentation, Osgood found his system to rely on "connotative emotions" attached to a word rather than its "denotative meaning", and discontinued further research." [13] Our evaluation absorbs the extended psycholinguistic analysis, since according to J. Deese, associations reveal the latent mental lexicon, the hidden word's value [14, 17].

The example of fuzzy granule-differentiation for some Scottish lexis (He, Kid, and Thow) is notified in Figure 7.

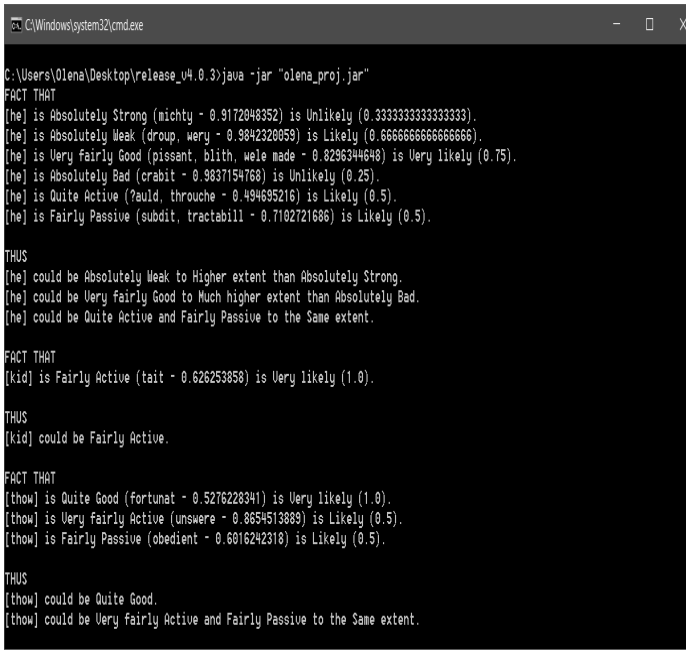


Figure 7: Console with the Older Scottish mental representations for He, Kid, and Thow.

The software reveals how Strong, Weak, Active, Passive, Good, and Bad He, Kid and Thow (in English 'thaw') are on the given map fragment:

Fact that

He is Absolutely Strong (michty - 0.92) is Unlikely (0.33)

He is Absolutely Weak (droup, wery - 0.98) is Likely (0.66)

He is Very Fairly Good (pissant, blith, wele made - 0.83) is Very Likely (0.75)

He is Absolutely Bad (crabit - 0.98) is Unlikely (0.25)

He is Quite Active (ʒauld, throuche 0.49) is Likely (0.5)

He is Fairly Passive (subdit, tractabill - 0.71) is Likely (0.5)

Thus, the software summarizes:

He could be Absolutely Weak to Higher Extent than Absolutely Strong

He could be Very Fairly Good to Much Higher Extent than Absolutely Bad

He could be Quite Active and Fairly Passive to the Same Extent

Thus the richest piece of the mental portrait on early and medieval *Him* has been represented by our fuzzy associative differential.

Kid and Thow are not so amply characterized on the fragment but still well-determined. According to the differential, "The fact, that *kid* is fairly active is very likely". As to *thaw*, the differential says that "Thow could be quite good" and "Thow could be very fairly active and fairly passive to the same extent."

The proposed by us electronic tool, called a fuzzy associative differential, seems to be very beneficial in the (re-)constructing of mental representations, and is highly recommended to the attention of cognitivists.

5.5. Fuzzy Lexis Stratification

On the other hand, the vertical cuts of the associative map in Figure 1 represent stratification figures of Older Scottish lexis [3, 4]. According to the late concept of Lotfi Zadeh, the special way of fuzzy classification was proposed as the way of organizing dictionary, encyclopedias and notebooks. It reveals the special relevance for the historical time-oriented thesaurus like ours.

The fuzzy stratification could open possibilities of ascertaining the lexis stratum gravity poles, time antonymy and significance status of the strata, all enabled by the manipulation with border membership functions [3, 4].

Any stratum in Figure 1 counts 100 years, and there are four of them within our figures for time-structuring of historical vocabulary of Potency, Evaluation, and Activity themes. The proposed by us fuzzy variable:

$$Gravity = \\ = (PositiveConcept/\ \mu(x_i);\ NegativeConcept/\ \mu(x_j)) \quad (13)$$

is introduced in any stratum, where $\mu(x_i)$ and $\mu(x_j)$ are the membership functions of the border (between-strata) antonymic terms x_i and x_j, respectively. The adopted error (the difference in advent time of the border words) makes 50 years.

The stratification for Potency was represented by authors in [3, 4].

Strong-Weak:
1400, (Burely/0.92) – 1420, (Smal/0.86);
1500, (Rik/0.59) – 1500, (Frail/0.61);
1600, (Proud/0.25) – 1600, (Bauch/0.29).
1300-1400 stratum: Gravity = (strong/0.92; weak/0.86);
1400-1500 stratum: Gravity = (strong/0.59; weak/0.61);
1500-1600 stratum: Gravity = (strong/0.26; weak/0.29).

In the same manner we characterize the sectional time antonyms, gravity or concept predominance within the remaining two structures:

Good-Bad:
1400, (Nobil/0.77) – 1400, (Pur/0.92);
1500, (Bony/0.51) – 1500, (Dolly/0.59);
1600, (Braw/0.25) – 1597, (Subaltern/0.28).

1300-1400 stratum: Gravity = (good/0.77; bad/0.92);
1400-1500 stratum: Gravity = (good/0.51; bad/0.59);
1500-1600 stratum: Gravity = (good/0.25; bad/0.28).

Active-Passive:

1420, (Unswere/0.87) – 1400, (Mild/0.93);
1500, (Lifly/0.63) – 1500, (Appliable/0.57);
1602, (Ʒauld/0.32) – 1609, (Submiss/0.17).
1300-1400 stratum: Gravity = (active/0.87; passive/0.93);
1400-1500 stratum: Gravity = (active/0.63; passive/0.57);
1500-1600 stratum: Gravity = (active/0.32; passive/0.17).

The balancing of values or the small predominance of negative concepts (the negative concepts being more alarming are traditionally more developed in equal development conditions) is observed in all the cases, except the latest stratum, which reveals the considerable gravity to the Active flank, pointing out to the notable development of the concept Active throughout the 16-th century. Linguists could speculate on historical reasons, for example, influence of Renaissance ideas in the case. According to Sapir, the language changes remain behind the cultural ones evoking them [28]. Therefore, approximately in the 14-15 centuries, at the time of High Renaissance, the concept Active became comparably more demanded.

5.6. Software Description

The architecture is the same as in [4] and based on MVP design template (*Model-View-Presenter*), delivering the visual reflection and event-procession behaviour into different classes, namely *View* and *Presenter*. In the case of 2RFAN, two XML files of the model DOM are chosen, for positive and negative runs of the 2RFAN.

The software performs altogether 5 main actions. They can be executed by the following button combinations realized in the given map (Figure 3):

- Left Click: Associative antipodes tracing.
- Left Click + Alt: Fuzzy Associative Word Structure evolving.
- Left Click + Shift: Word Differentiation through six scales of the FAD with final evaluative granularity.
- Left Click + Ctrl: Associative Experiment simulation for a chosen stimulus.
- Right click on two words of the same row: Distance and Similarity calculation for any two associates of the same field.

The provided interface is very sensitive and maximally friendly, making its evaluation in words. The correspondence between verbal estimators and membership functions (MFs) is given in Table 1.

The intervals are adopted from [8]. The proposed by us terms differ depending on words they modify: the first column modifies the degree of belief (modifiers for nouns); the second – all the six criteria as well as words Distant and Similar (modifiers for adjectives). In the proposed practice, the conclusion depends on the closeness of MFs to the centric characteristics, in favour of a more intensive modifier if MFs are equidistant from the two centers.

Table1: Correspondence between numeric and lexical evaluation.

Modifying terms for degrees of belief	Modifying terms for good, bad, active, passive, strong, weak; distant and similar	Range	Center
No	Not at All	0.17	0
Very Low	Just a Very Little	0.00 – 0.33	0.17
Low	Just a Little	0.17 – 0.50	0.33
Medium	Quite	0.33 – 0.67	0.50
High	Fairly	0.50 – 0.83	0.67
Very High	Very Fairly	0.67 – 1.00	0.83
Perfect	Absolutely	0.83	1

The terms for granule-probabilities depend on the probability distribution across two antonymic scales. There are four general terms and they are ranged by us in Table 2.

Table 2: Correspondence between granule-probabilities and terms.

Terms for Granule-Probabilities	Range
Very Unlikely	0.00 – 0.24
Unlikely	0.25– 0.49
Likely	0.5 – 0.74
Very Likely	0.75 – 1

The typical granule-estimating Zadeh's phrases [11] are adopted in our fuzzy associative differential (Chapter 5.4 and Figure 7).

The comparative terms for final conclusions on lexis differentiation depend on the number of shifts between the given above four probability ranges for the compared granules, and are given in Table 3.

Table 3: Comparative estimators for final conclusions on lexis differentiation.

Number of Shifts	Comparative Terms
0	the Same Extent
1	Higher Extent
2	Much Higher Extent
3	the Highest Extent

Thus the lexis and syntax of the phrases in Figures 4-7 are clarified by the aforesaid. The principle of computing with words seems to be utterly justified and profitable in linguistic conclusions like these.

6. Discussion

The electronic resource of the type recommends the method of associative modeling for the older languages, and thus extracts the linguistic knowledge of the remote times. The unfolded fragment of the associative map becomes the canvas for speculations on the Older Scottish consciousness. Many-sided conclusions on lexis development could be drawn from the realized on the map simulations. Associative antipodes' extraction reveals the reachability, coherence and deepness of the

map and, on the other hand, between-field associative communication. The chains count the semantic shifts of associations, notifying the reachability potency by the number of between-field transitions. The stability of the found positive and negative associative environment is characterized by the determined fuzziness values. The simulated associative experiment provides the most complete associative answer to the stimulus, meanwhile being sensitive to the forced turnovers in favour of the higher associativity relations. The constructed associative map makes it possible to derive an associative experiment for any map element, equipping it with paradigmatic (synonymic and antonymic), syntagmatic and epidigmatic associations. The fuzzy metrics' identification gives precise figures on similarity and distance between associates inside the field. The lexis differentiation practice provides the complete assessment analysis of nouns and pronouns of the map. The software groups the found characteristics for the regarded lexis into the fuzzy sets that correspond to the six scales Good, Bad, Active, Passive, Strong and Weak, constituting cross-cultural universals of Osgood's differential. The fuzzy sets are supported with joined probability-values and make the fuzzy granules, introducing the partial meanings of both connotative and denotative nature to the criteria of the scales, and breaking the limits of Osgood's differential. The special significance in this respect is referred to our fuzzy associative differential, designed for the purpose. The fuzzy stratification structures help us to draw the historical and lexical conclusions concerning the strata development. The fuzzy facilities are numerous and make the efficient research device, called an electronic FHATh, organized after the principle of computing with words, providing verbalized characteristics to the requested relations.

7. Conclusions

The fragment of a fuzzy model of associative memory provides complete and well-quantified associative relationship, episodically reflecting the world lingual picture of the older speakers, their communicational habits, lingual preferences, inbuilt attitudes and dispositions. The deepness of associative correlation and coherence of the area constitute the subject for the research conclusion. The prospect is observed in multiplication of the simulated network patterns ready to serve different stereotypes. The final combination of them in the larger network would enrich artificial intelligence with cognitive, behavioristic and evolutional linguistic data and mechanisms of their mining, which should broaden its historical prognostication abilities.

References

[1] The Dictionary of the Scots Language, Edinburgh: Scottish Language Dictionaries, 2007. https://www.dsl.ac.uk

[2] O. Basalkevych, "Fuzzy Modelling of Associative maps" IV International Scientific Practical Conference "Linguocognitive and Sociocultural Aspects of Communication", Scientific records of "Ostroh Academy" National University, Ostroh, Ukraine, 56, May 2015.

[3] O. Basalkevych, O. Basalkevych, "Fuzzy Reconstructions in Linguistics" in Proceedings of IEEE Second International Conference on Data Stream Mining and Processing, Lviv, Ukraine, 21-25 August 2018. DOI: 10.1109/DSMP.2018.8478451

[4] O. Basalkevych, "Fuzzy Diachronic Differential" in Proceedings of the IV Ukrainian-German conference "Informatics. Culture. Technics.", Odesa, 106-108, 2016.

[5] J.M.Mendel, "Computing with words: Zadeh, Turning, Popper and Occam" IEEE Computational Intelligence Magazine, 2 (4), 10-17, 2017. https://doi.org/10.1109/MCI.2007.90668978

[6] G.R. Kiss, Ch. Armstrong, R. Milroy, J. Piper, "An associative thesaurus of English and its computer analysis" In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), The Computer and Literary Studies, University Press: Edinburgh, 153-165, 1973.

[7] G. Zipf, "Prehistoric 'cultural strata' in evolution of German: the case of Gothic" Modern Language Notes, 62 (8), 522-530, 1947.

[8] L. Martinez, R.M. Rodriguez, F. Herrera, The 2-tuple Linguistic Model. Computing with Words in Decision Making, Springer International Publishing, Switzerland, 2015.

[9] М. Сявавко, Інформаційна Система «Нечіткий експерт», Л.: Видавництво ЛНУім.Франка, 2007.

[10] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I" Elsevier Information Sciences, 8 (3), 1975. https://doi.org/10.1016/0020-0255(75)90036-5

[11] L. Zadeh, "Fuzzy sets and information granularity" Advances in Fuzzy Systems-Applications and Theory Vol 6. Fuzzy Sets, Fuzzy Logic and Fuzzy Systems. Selected Papers by Lotfi A. Zadeh, G.J. Klir and B. Juan (eds.) 433-448. Singapore: World Scientific, 1996.

[12] L. Zadeh, Stratification, Target Set Reachability and Incremental Enlargement Principle. [Electronic Video Resource] The Design of Robotics and Embedded systems, Analysis and Modeling Seminar, 250 Saturdja Dai Hall, UC Berkeley, 2016. https://www.youtube.com/watch?v=Ok0vJ0uykCA&list=PL8092C965854FA5BB&index=30

[13] O. Budanitsky, "Lexical semantic relatedness and its application in natural language processing" Technical report CSRG-390. Computer system research group, 5, 1999.

[14] О.О. Селіванова, Сучасна Лінгвістика. Напрями та Проблеми, Полтава: Довкілля-К, 2008.

[15] A. Jain, D.K. Lobiyal, "Fuzzy Hindi WordNet and Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures" ACM Trans. Asian & Low-Resource Lang. Inf. Process, 15(2): 8:1-8:31, Dec. 2015. https://dx.doi.org/10.1145/2790079

[16] W. Banks, "Linguistic variables: clear thinking with fuzzy logic" Byte Craft Limited, Waterloo (Ontario, Canada), 2013.

[17] M.L. Murphy, Semantic Relations and the Lexicon, Cambridge University Press, 2003

[18] P. Sussner, M. Valle, "Fuzzy associative memories and their relationship to mathematical morphology" In the project of Fuzzy and lattice-based associative memory models, 2012.

[19] U. Priss, "A classification of associative and formal concepts", Napier University Press, 2002.

[20] Oxford Learner's Thesaurus. A Dictionary of Synonyms. Oxford University Press, 2008.

[21] WordNet. A lexical database for the English language. Princeton University, 2006. http:// wordnet.princeton.edu

[22] J. Schafer, Documentation in the O.E.D.: Shakespeare and Nashe as Test Cases, Oxford: Clarendon Press, 1980.

[23] M. Bilynskyi, "Synonymous string as a diachronic reconstruction: the OED Middle English Evidence for verbs and deverbal coinages", Inozemna Philologia, Lviv, 121, 2009.

[24] Historical Thesaurus of the Oxford English Dictionary, Oxford English Dictionary, Oxford University Press, 2019. — www.oed.com

[25] О.В. Бісікало, Концептуальні Основи Моделювання Образного Мислення Людини, Вінниця, 2018.

[26] D. Nelson, C. McEvoy, S. Dennis, "What is and what does free association measure? Free Association" 3-46, 2002.

[27] A. Mochida, "Antonymy and synonymy in adjectives in L2 mental lexicon: Associative or conceptual relations?" The University of Queensland, 2003. https:// www.geocities.jp/akiramochida33/researchpro.html

[28] E. Sapir, "Language and environment" American Anthropologist. New Series, 14 (2), 226-242, 1912.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 3, 227-234 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Multi-Band Hand of God Antenna with Functionality Variation with Use of Slots and Feed Position

Mervat Akram Madi*, Karim Youssef Kabalan

Electrical and Computer Engineering Department, American University of Beirut, 1107 2020, Lebanon

ARTICLE INFO

Article history:
Received:02 May, 2019
Accepted:25 May, 2019
Online: 18 June, 2019

Keywords:
Multiband antenna
Slotted antenna
Patch Antenna

ABSTRACT

This article investigates different versions of a new multi-band micro-strip antenna design. NASA inspired the antenna design from "Hand of God" galaxy discovery. A multi-band operation between 0.15 and 8 [GHz] is obtained and captures up to 10 frequency bands. The antenna has a small area of 4 x 5 cm^2 and is printed on 1.6 mm epoxy-FR4 substrate. Adding slots and changing the feed position, led to different designs and increase in multi-band operation. Moreover, varactors are mounted on antenna surface, which resulted in frequencies as low as 50 MHz and as high as 8 GHz when. S11 simulations results were successfully compared with measured ones. Applications include different mobile generations, TV and different IoT systems.

1. Introduction

Multi-band antennas capture various segments of the frequency spectrum simultaneously. Yet, a challenging aspect of maintaing the required gain appears [1-6]. Of course, an antenna should have high efficiency to be able to capture many frequencies. Moreover, new technology requires smaller front end devices, and hence antenna size follows. Moreoever, selected frequencies are targetted, to cope with licenced frequency bands. Hence, research is still developing in this area.

Multiple frequency operation has become recently desired to engage IoT applications in same device. It allows performing technologies that require and or scan wide range of frequencies. For example, it is desirable to work with two mobile generations simultaneously [7]. Adapting to wearable technologies, [8], presented a relatively large antenna operating for different applications with up to six frequency bands. Alternatively, different bands related to same generation can be combined in same antenna [9]. Moreover, switches were mounted to achieve multiple resonances, while changing the switching scenario changes operation band being the L or S band [10,11]. Metamaterials were also exploited in [12] for multi-band operation.

The need to minimize space occupied by modern devices such as phones or sensors renders multi-band antennas more attractive since they replace many antennas [10], while still consistently frequencies in MHz range are mainly emitted by large antennas that are quarter wavelength, i.e. more than 12 cm in length and width [4,11,13,14]. This creates a challenge if high frequency ratio is targeted within same antenna. supporting different active parts of wireless spectrum. Low

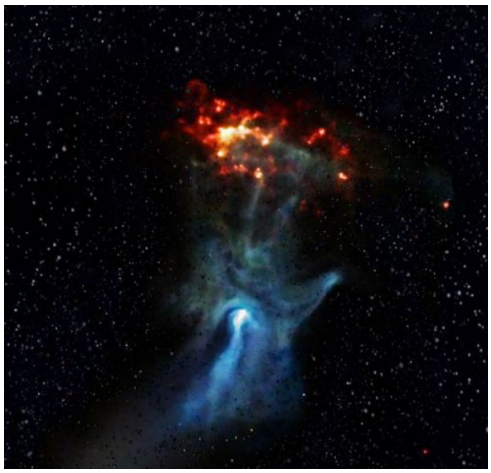


Figure 1: Galaxy named 'Hand of God' by NASA

This paper suggests a novel design that covers multiple low and high communications frequency bands, serves IoT applications, and maintains a compact size. The design is inspired by a photo of a galaxy captured recently by NASA, entitled 'Hand of God' [15]. Hence, this antenna duplicates a cosmic shape using

* Mervat Akram Madi, mm102@aub.edu.lb

electromagnetic radiation modelling, in an articulate way. Ultimately, the antenna presented here, resonates at frequencies near 0.1 [GHz] up to 8 [GHz], knowing that different techniques are used to enhance these aspects.

First, as power input position is varied, operating frequencies shifted. Hence, position of feed is an important parameter that is thoroughly investigated in the paper. Second, slots were easily incorporated in the design and enhanced increase of frequency operation in addition to bandwidth ratio. Third, varactors are added across slots to study their effect on frequency reconfigurability. As a result, frequency is tuned in the MHz range and resonance at 100 MHz is achieved, in spite of the very small physical length of the antenna.

The design of multiple extensions given by a hand shape is elaborated in this paper, since naturally each frequency is linked to a radiating element. Yet, the operation bands obtained outnumber the dipole, finger like extensions.

The antenna has compact volume of 4 x 5 x 0.16 cm^3, a new design that succeeds in combining more than six bands and captures ten bands in one specific scenario, with high frequency ratio of 53 starting at 0.15 [GHz], if no diodes are used. Frequencies above 8 [GHz] were not included in this research. 3G, 4G and IoT frequency applications are still main concern in antenna design. Authors present the 'Hand of God' for the first time, novelty is achieved in design, size and range of application [16]. In following sections, the antenna design significance is shown FEM modeling in HFSS. Third section shows the simulations results for feed through one chosen position. Section Four details all interesting scenarios in feeding positions and displays a comparison in S11 (return loss) results between measurements and simulations for the various fabricated prototypes. In section five, the effect of mounting varactors while DC voltage control is applied is shown. Sections six and seven show the current density and the corresponding radiation patterns for the resonant frequencies relative to the first discussed prototype.

2. Antenna Design Methodology

Due the importance of monopole, L and U shaped antennas in microstrip antenna design, the hand design holds importance in relation of design and shape. As the first draft of the patch design had shown multiple band characteristic, parametric sweeps that modify the shape for better resonance were done in FEM software tool HFSS. In fact, there is a conjecture that physical shapes in nature cope well with a certain frequency, and these shapes manifest certain cosmic frequencies, which is literally a direct link between shape and frequency. In fact, both ends of physics and metaphysics meet at the antenna design platform.

Moreover, the effective relation between having multiple extensions from a basic feed to a multiple frequency operation has immediately inspired me to recreate the hand shape on the patch after seeing the NASA photographs of a far galaxy. Yet, the use of FEM calculation tool for antenna design has facilitated turning the antenna conceptual design into an operating one. Here, experience in antenna design and persistence in using the HFSS software to explore all parametric influences on their own or in relation to other parameters has led to good results and even unexpectedly 9 resonant bands when slots were used.

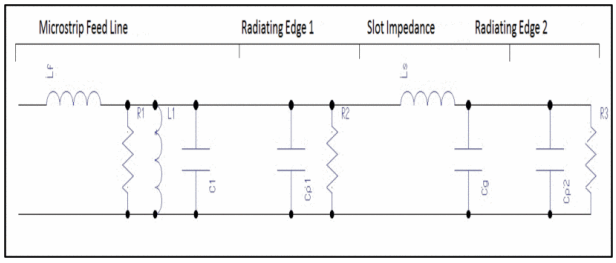


Figure 2: Equivalent circuit of slot loaded antenna

The 'Hand of God' [15], is source of inspiration of this design due to the photo captured by NASA telescope. The finger shapes and hand is a renovated form of dipole antenna array, with main 'palm' space acting as a relatively large capacitance that stabilizes antenna radiation operation at large frequencies in GHz range. This is no the first time a cosmic/natural shape is successfully exploited in electromagnetics, examples include the spiral shape (inspired from a galaxy [17]) and many others showing the beauty of science in relation to nature.

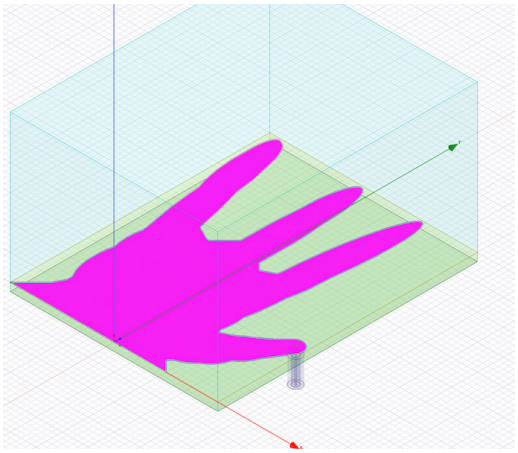


Figure 3: Simulated feed at 1st fingertip

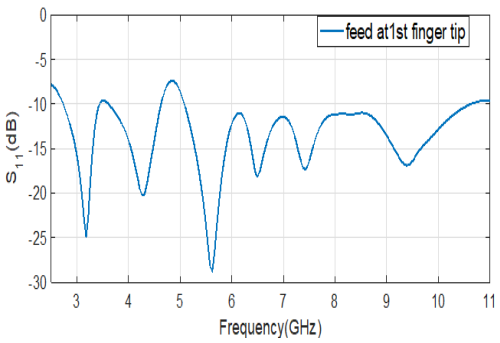


Figure 4: S11 simulated results of feed at 1st fingertip in Figure 3.

The antenna with four fingers combines inductive and capacitive effect of each extension. Yet, it forms a new way of

combining four dipole antenna extensions in a compact design. These were responsible for multiple frequency operation. In particular, each element has different length and shape, and spaces between elements play crucial role in minimizing destructive interference, since targeted frequencies have wavelengths comparable to antenna size. After performing many simulations on common area in the patch and length of the extensions, final design was achieved as shown in Figure 3. Initial setting of the model was even varied in total area and substrate material, but a double-sided FR4 substrate was most successful. Finally, small area of 4 x 5 cm ^{2} accommodated the multiple 'finger like' extensions to achieve the targeted multi-band operation.



Figure 5: First feed scenario prototype at 4th fingertip.

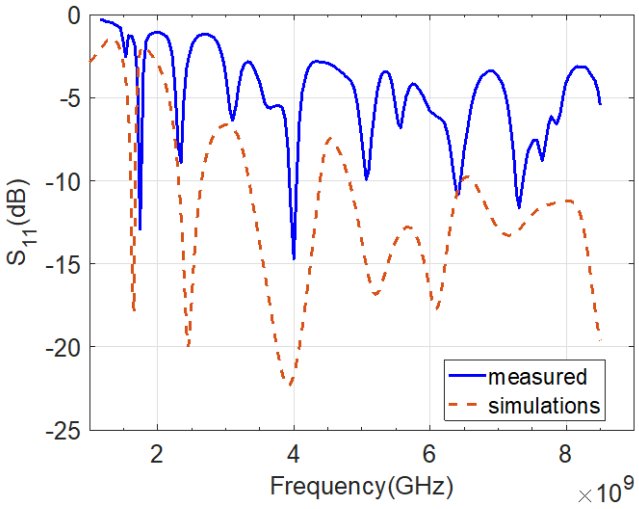


Figure 6: Comparison of S11 obtained in measurements and simulations results for 4th prototype (Figure 5).

Slots are used to increase bandwidth. Loading the radiating patch with slots or stub elements can extend bandwidth of the patch antenna. Additionally, varying the dimension of cutting slots on different locations and adjusting the shape of the slots like rectangular, circular, octagonal, fractal, and modified/partial ground plane, can help to obtain the desired bandwidth at estimated resonant frequency [18]. Slots are also used in size miniaturization of antenna [19] and increasing number of resonant frequencies [20].

Slot impedance is considered as series inductance and parallel capacitance, which adds to the total inductance and capacitance as shown in Figure 2, [21]. In fact, slots helped increase the number of resonant frequency bands over [0.1-8] GHz to nine with tunability over frequency range [0.2-0.8] GHz, and total area = 4 x 5 cm ^{2}. In a similar approach, [22] has used 3 slots to achieve three resonant bands tunable at 3 frequencies over [0.6-2.7] GHz, and an area = 5 x 10 cm^2. Slots with modes controlling method were used in [23]. Starting with a monopole at 0.25 λ then adding an open slot at 0.25 λ then a tuning pad, resulted in 2 bands [0.8-1.02] GHz and [1.69-2.73] GHz, knowing that the area = 10 x 6 cm^2. Multiple slits on a rectangular patch were also used in [20], to achieve 4 bands over [5.8-7.87] GHz and antenna area of 2.6 x 7.75 cm^2. Other methods can be listed here with corresponding characteristics, for example using multiport and theory of characteristic modes (TCM) to isolate frequency bands, resulting in 3 bands over [0.8-2.2] GHz [24]. In comparison, [25] mounted a 3D loop folded antenna (volume 60×5×5 mm^3) on 100×60×0.8 mm^3 FR4 substrate with full ground plane, for a mobile handset, to design a 3 bands tunable antenna over [0.8-2.5] GHz. Consequently, the present method is better improved over other methods due to carving the slots on the antenna patch along the extended finger shapes, which enhanced the resonant characteristic modes by doubling the mode effect being a result of splitting into two arms in a U shape.

3. Simulations basic design and effect of feed position

The antenna is a microstrip patch, with full ground, an area of 4 x 5 cm ^{2} and 1.6 mm thick epoxy FR4 substrate. The dimensions of finger shapes (see Figure 5) are 25.8 mm x 6 mm for 4th finger, 19.6 x 6 mm for 3rd finger , 21.5 mm x 6 mm for 2nd finger , 14 mm x 9.5 mm for 1st finger. The antenna feed is a coaxial cable as shown in Figure 3, where axes x, y, and z are in red, green, and blue respectively for later reference. In fabrication, an SMA connects the cable to the antenna at the variable positions to be discussed, where 50-Ω matching is considered.



Figure 7: Antenna prototype with the feed at the 2nd fingertip.

The design is modeled in FEM software, HFSS. In order to preserve the universal shape, and improve results and easily accommodate the shape in small area for better performance, simulations were the main guide to refine the design, where there

is no imperial formula to apply in this case. The underlying design

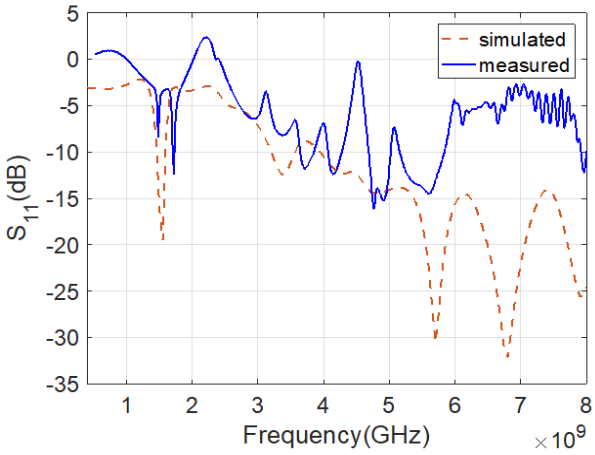


Figure 8: Comparison of S11 obtained in measurements and simulations results for 2nd prototype (Figure 7).

is an overlap of geometrical shapes like polygons, ellipses, circles, and lines in coherence with basics of drawing depicted from human anatomy. The area being restricted to 4x5 cm^2 has less extension of the wrist, which has shown no effect on multi-band aspect.

Feeding has been tested over 10's of potential positions. Yet, resonance was easily achieved when the coaxial feed is placed the fingers tips. For example, an ultra-wide band (121% bandwidth) between 2.7 and 11 [GHz] is obtained when the feed is at the tip of the first finger as shown in Figure 4. The results at the remaining fingertips are shown in comparison with measurements.



Figure 9: Antenna prototype with slots with the feed at the 1st fingertip.

4. Comparison between measurements and simulations results

A. First prototype, no slots and the power input at the fourth fingertip (Figure 5) :

The first prototype of feeding position at 4th fingertip is shown in Figure 5. It is fabricated as first test of the antenna simulations results. The reflection coefficient S11 is measure using a network analyzer that spams up to 8.5 [GHz].

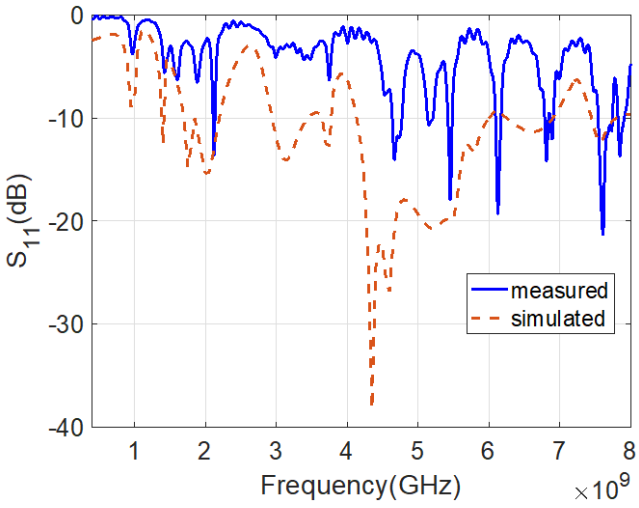


Figure 10: Comparison of S11 obtained in measurements and simulations results for prototype in Figure 9.

S11 results recorded in measurements coincide very well with simulated ones as shown in Figure 6. When comparing the resonance frequencies for the simulated and fabricated cases, small drift was obtained; i.e., (1,66, 1.74), (2.43, 2.34), (3.95, 4), (5.2, 5.06), (6.15, 6.42), and (7.2, 7.4) [GHz]. 0.27 [GHz] is the maximum difference between measured and simulated results due to many reasons, which include:

1) Modeling at a center frequency, while multiple frequencies are targeted and being discovered. As a result, slight differences



Figure 11: Antenna prototype with slots with the feed at the 2nd fingertip.

appear due targeting different frequencies while setting far field in the simulations at fixed wavelength.
2) Fabrication errors, resistive mismatch in the SMA connector, soldering and losses in measurement wires.
3) Since losses increase with frequency, discrepancy increases at

higher frequency to maximum of 0.27 [GHz].

B. Second prototype, no slots with feed at the second fingertip (Figure 7) :

For feeding position at second fingertip, comparing the S11 results as displayed in Figure 8 shows discrepancy at 6.85 GHz being a resonant frequency that did not appear in measurements unlike simulations. However, almost good matching is obtained for measured frequencies at 1.85, 3.85, 4.2, [4.8-5], 5.7 and 7.98 [GHz]. Largest differene between measurements and simulations in listed values is at 3.47 GHz obtained in simulations versus 3.78 GHz in simulations. Such discrepancies results are descirbed in paragarph A of section 4.

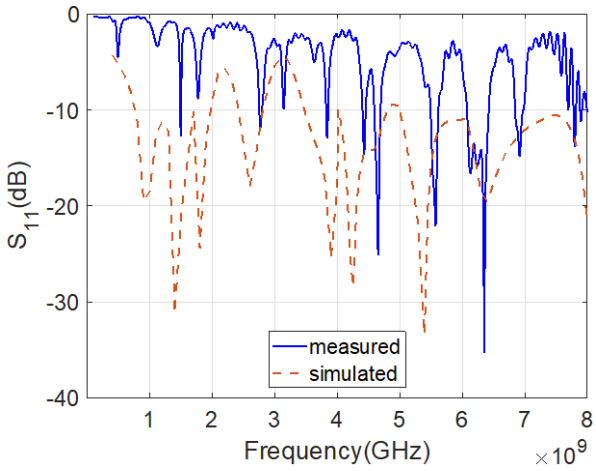


Figure 12: Comparison of S11 obtained in measurements and simulations results for prototype in Figure 11.

C. Third prototype, with slots and feed at the first fingertip (Figure 9) :

Adding slots along all fingers has created multi band operation for feed at all fingertips due to elongating the current path. Here the feed is at the first fingertip as shown simulated results show more like ultra wide band operation below 4 GHz. Common nature of multi-band operation is seen in both results with peaks almost coinciding or falling within matching region i.e for S11 belwo -10 dB. in Figure 9. Comaprision between measured and simulated results is shown in Figure 10. Measured results resonance is obtained at 2.15, 4.5, 5, 5.5, 6.05, 6.8 and 7.7 [GHz], while simulated results show more ultra wide band operation above 4 GHz in addition to several multi-band rsonances belwo 4 GHz. Multiple bands are seen in both results,



Figure 13: Antenna prototype with slots with the feed at the 3rd fingertip.

with peaks almost coinciding or falling within matching region i.e for S11 below -10 dB.

D. Fourth prototype, with slots and feed at the second fingertip (Figure 11) :

Adding slots with the coaxial feed position at the second fingertip is shown in Figure 11. The S11 measurements show good matching with simulation results (Figure 12) at the following resonant frequencies : 1.4, 1.83, 2.72, 3.9, 4.2, 4.78, 5.38, 6.32,6.97,8 [GHz] being the simulated obtained values with maximum difference of 0.25 [GHz]. Resonance at 1 [GHz] did not appear strong enough in measurements, hence it is not considered.

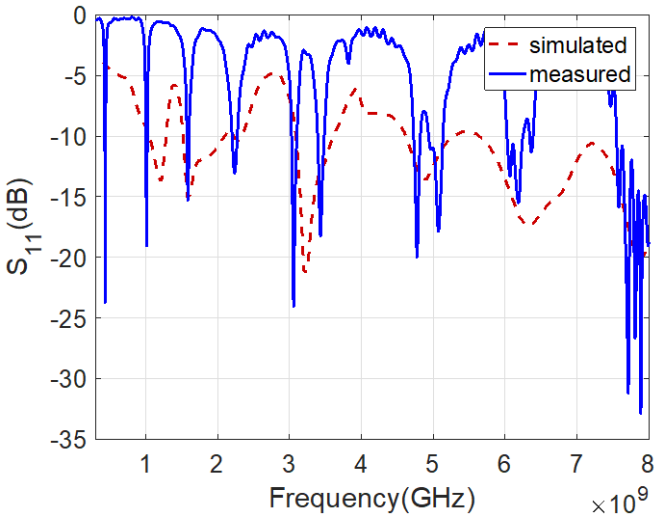


Figure 14: Comparison of S11 obtained in measurements and simulations results for prototype Figure 13.

E. Fourth prototype, with slots and feed at the third fingertip (Figure 13) :

Adding slots with the coaxial feed position at the third fingertip is shown in Figure 13. The S11 measurements show good matching with simulation results in Figure 14, at the

following resonant frequencies : 0.15, 1, 1.65, 2.25, 3, 3.4, 4.8, 5.08, 6.2, and [7.6-8] [GHz]. Resonance in measurements at 0.4 [GHz] could not be obtained in simulated results and hence it is not considered.

5. Dc biasing of mounted varactors

A simple experimental approach of adding varactors across slots was done as an arbitrary trial to check resonances and reconfigurabilty with DC biasing [26]. Simulations were not done to approve results as this experiment falls in further studies to explore the full frequency reconfigurability upon adding varactors. Moreover, experimental tools did not allow DC biasing to the full range of voltage, which is up to 60 V due to summing up the voltage across varactors, which take up to6 V each. As shown in Figure 15, varactors were added at the slots carved through second and third fingers. In order to DC bias the 13 varactors. Inductors were added across slots in 1st and 4th fingers as shown in Figure 15. Hence DC current was drawn through the inductors, which act as AC block at the same time [26]. S11 resonance at low frequencies is remarkable considering the size of the antenna. Frequency reconfigurability over [2.15-2.25] and [0.2-0.8] GHz is obtained upon DC biasing the antennas between [0-10] V as shown in Figure 16. Future studies are to be made by further increasing the DC voltage biasing range. Applications to such frequencies fall in DVB, TV broadcast, LTE, and many other telecommunications standards, which allow -6 dB S11 at maximum return loss [14]. Additional new frequency resonances are obtained at1, 1.6 and 2.15 [GHz] due to adding varactors as shown in Figure 16. The multi-band feature of the antenna at wide range of frequencies enables additional applications such as IOT [6] and energy harvesting [4,27].



Figure 15: Antenna prototype with slots with the feed between 1st and 2nd fingertip and using varactors along 2nd and 3rd fingers slots.

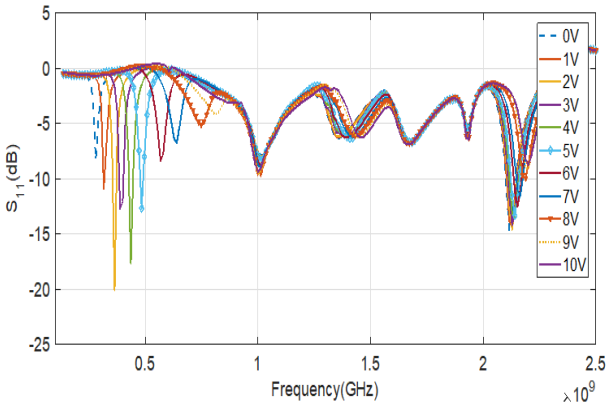


Figure 16: Comparison of S11 obtained in measurements and simulations of prototype in Figure 15 with frequency tunability upon varying the DC bias voltage.

6. Surface current density

Current flow at the antenna surface changes with frequency, consistently with the electric length required for frequency resonance [28]. Surface current density distribution is plotted from simulations for four frequencies as a demonstration. The first antenna prototype in Figure 6 is chosen as an example. Here the feed is at the 4th fingertip. Figure 17 shows the four current graphs corresponding to 1.66, 3.95, 5.2 and 7.2 [GHz] with same color scale with maximum at 6 [A.m^{-1}] and minimum at 0.012 [A.m^{-1}]. However, at 1.66 [GHz] the current density achieves highest value of 60 [A.m^{-1}]. Normally, lowest frequency corresponds to largest electric length and current density values. The current is first

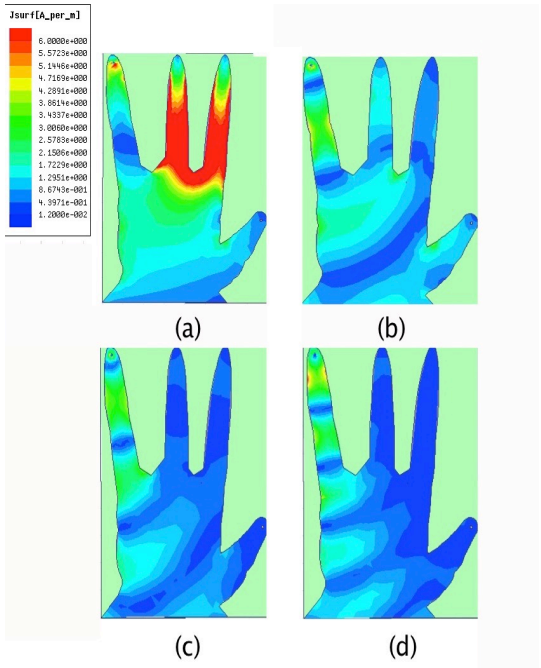


Figure 17: Simulated results for the current distribution for the 4 resonant cases corresponding to 1.66 GHz (Figure a), 3.95 GHz (Figure b), 5.2 GHz (Figure c), and 7.2 GHz (Figure d).

strong especially over the mid part of the hand 1.66 [GHz] in graph(a) of Figure 17; then the current surface area diminishes as

frequency increases as shown in graphs (b,c,d) of Figure 17. Of course, longest current path from input feed to furthest points in the patch assures the lowest frequency resonance. And obviously, the current is drawn to smallest area around the fourth fingertip where the feed is placed, at the highest frequency of 7.2 [GHz] as shown in graph (d) in Figure 17.

7. Radiation pattern

The current flow distribution in Figure 17 radiates the different patterns shown in Figure 18, [28]. The radiation pattern at the 2.43 GHz frequency is shown in Figure 18 (a). The common observed feature at low resonant frequencies is the symmetric current distribution with respect to the right diagonal resulting in a common radiation pattern at 1.66 and 2.43 [GHz]. Here, the axes are same as in Figure 5. Figure (b) has many peaks due to non-symmetric current density at 6.15 [GHz]. Figures 18 (a, b) show the main trend of radiation pattern at low (1.66 and 3.95 [GHz]) and high frequencies (above 5.2 [GHz]) respectively, as a

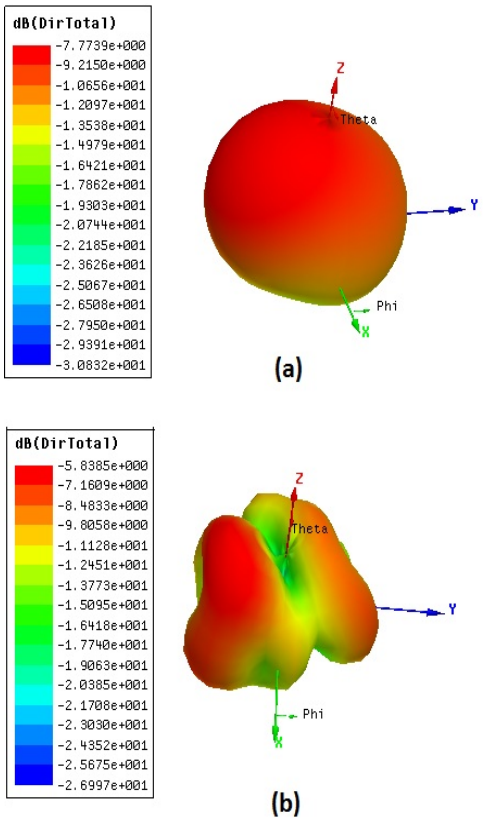


Figure 18: Radiation pattern at 2.43 and 6.15 [GHz] in (a) and (b) respectively.

consequence of change of current density symmetric distribution on the surface. A more description plot of the radiation pattern at 2.43 [GHz] is shown in Figure 19.

8. Conclusion

The 'Hand of God' motif has been successfully translated in antenna design modeling and served in achieving a multi-band operation. Slots are incorporated in the antenna showing an enhanced multi-band operation. In fact, 10 resonant frequencies were obtained over the interval [0.15-8] GHz for one prototype having slots and with the feed at third fingertip. Interesting increase in electrical length of the antenna results when varactors are added across slots. Consequently, resonance at low frequencies of 100 MHz is obtained. Moreover, frequency reconfigurability is observed at frequencies below 1 [GHz]. Finally, the shape, size, and wide range of antenna operation frequencies render it as multi-functional and suitable for different modern applications.

References

[1] Balanis, C. A. 'Antenna Theory'. (Wiley and Sons, INC., 1997).
[2] Ghosal, A., Majumdar, A., Das S. K., & Das A. (2018). 'A multiband microstrip antenna for communication system'. 2018 Emerging Trends in Electronic Devices and Computational Techniques (EDCT), Kolkata, 1-4.
[3] Xu, K. D., Li, D., Liu, Y., & Liu, Q. H. (2018). 'Printed Quasi-Yagi Antennas Using Antennas Using Double Dipoles and Stub-Loaded Technique for Multi- Band and Broadband Applications'. IEEE Access, 6, 31695-31702.
[4] Chen, Y., Liu, Q., Tang, X., Mo, Z., Li, C., & Li, F. (2018). 'Compact triple band circularly polarized directional antenna for UHF/ISM RFID mobile readers with GNSS band'. 2018 International Workshop on Antenna Technology (iWAT), Nanjing, 1-3.
[5] Li, Y., Sim, C., Luo Y., & Yang G. (2018). 'Multiband 10-Antenna Array for Sub-6 GHz MIMO Applications in 5-G Smartphones'. IEEE Access, 6, 28041-28053.
[6] Mansour, A. M. et al. (2016). 'Compact reconfigurable multi-size pixel antenna for cognitive radio networks and IoT environments'. 2016 Loughborough Antennas & Propagation Conference (LAPC), Loughborough, 1-5.

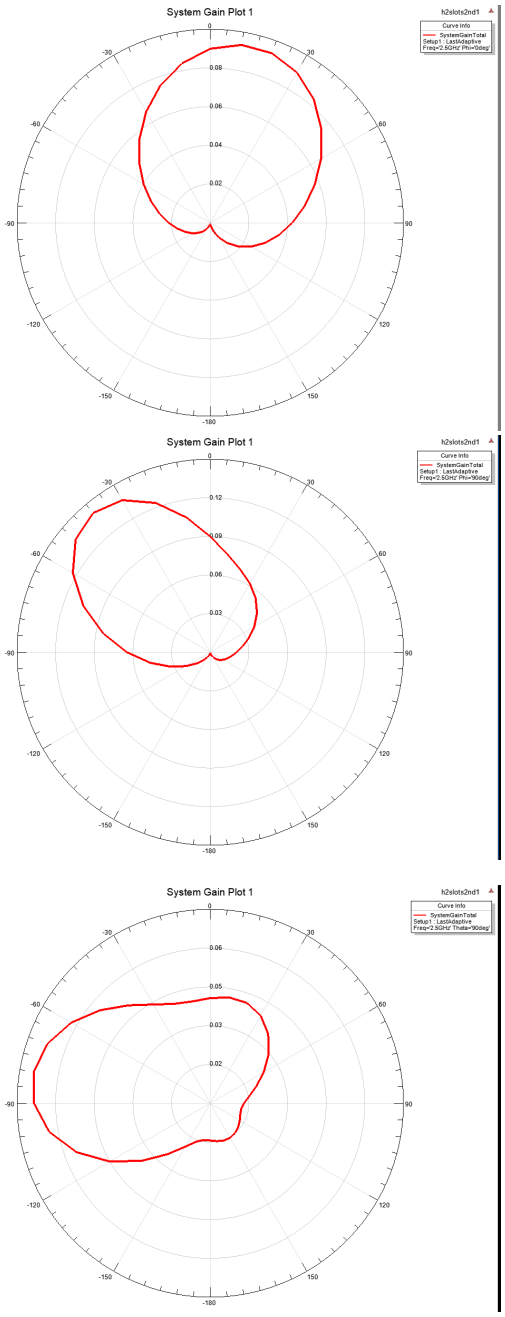


Figure 19: 2D radiation Pattern in XOZ, YOZ, and XOY planes from top to bottom respectively.

[7] Hui, D. (2017). 'A new type of multi-band mobile phone antenna for 3G&4G network', 659-63.

[8] Wang, J. J. H., and Adley, J. C. (2017). '30-2000 MHz multi-band body wearable antenna (MBWA)', 1771-2.

[9] Ozbakis, B., Okuyucu, S., Seemen, M. & Yegin, K., (2017). 'Multi-band frequency tunable LTE antenna for mobile phone applications', 939-43.

[10] Shahzad, M. U., Ahmed, W., Bukhari, S. M. R., Hassan F., & Shahid, H. (2017). 'Multiple band reconfigurable dual H-shaped patch antenna'. 2017 International Symposium on Recent Advances in Electrical Engineering (RAEE), Islamabad, 1-4.

[11] patents.google.com/patent/US8629811B2/en, 2014.

[12] Labidi, M., Salhi, R., and Choubani, F., (2017). 'A design of metamaterial multi-band bowtie antenna based on omega-shaped resonator', Applied Physics A, 123(5), 1-6.

[13] Zhang, Q., & Gao, Y. (2017). 'Compact low-profile uwb antenna with characteristic mode analysis for uhf tv white space devices'. IET Microwaves, Antennas Propagation, 11(11).

[14] Al-Husseini, M., Kabalan, K. Y., El-Hajj, A., Christodoulou, C. G., 'Reconfigurable microstrip antennas for cognitive radio', in Kishk, A., 'Advancement in Microstrip Antennas with Recent Applications'. (InTech, 2013).

[15] en.wikipedia.org/wiki/Hand_of_God(art), 2018.

[16] Mahesh, A., Shushrutha, K. S. & Shukla, R. K. (2017). 'Design of multi-band antennas for wireless communication applications'. 2017 IEEE Applied Electromagnetics Conference (AEMC), Aurangabad. 1-2.

[17] https://i.pinimg.com/564x/91/84/14/91841435effbcaddf955790f09ed4402.jpg

[18] K. Sharma, B. V. R. Reddy and A. Mittal, "Slot Loaded Microstrip Patch Antenna for WLAN and WiMax Applications," 2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, 2015, pp. 597-599.

[19] S. E. Muthumani, Vallikannu R and H. R. Patnam, "Compact slot loaded Koch fractal microstrip patch antenna," 2013 IEEE Applied Electromagnetics Conference (AEMC), Bhubaneswar, 2013, pp. 1-2.

[20] S. Verma, J. A. Ansari and M. K. Verma, "A novel compact multi-band microstrip antenna with multiple narrow slits," Microwave and Optical Technology Letters, vol. 55, (6), pp. 1196-1198, 2013.

[21] S. Ghosal and S. R. B. Chaudhuri, "Analysis of a rectangular slot on a microstrip patch antenna with an equivalent circuit model," 2013 IEEE Applied Electromagnetics Conference (AEMC), Bhubaneswar, 2013, pp. 1-2.

[22] Q. Bai et al, "An Independently Tunable Tri-Band Antenna Design for Concurrent Multiband Single Chain Radio Receivers," IEEE Transactions on Antennas and Propagation, vol. 65, (12), pp. 6290-6297, 2017.

[23] C. Deng et al, "A Novel Low-Profile Hepta-Band Handset Antenna Using Modes Controlling Method," IEEE Transactions on Antennas and Propagation, vol. 63, (2), pp. 799-804, 2015.

[24] K. K. Kishor and S. V. Hum, "Multiport Multiband Chassis-Mode Antenna Design Using Characteristic Modes," IEEE Antennas and Wireless Propagation Letters, vol. 16, pp. 609-612, 2017.

[25] Y. Li et al, "Compact Heptaband Reconfigurable Loop Antenna for Mobile Handset," IEEE Antennas and Wireless Propagation Letters, vol. 10, pp. 1162-1165, 2011.

[26] Madi, M., Al-Husseini, M. & Kabalan, K. Y. (2018). 'Frequency Tunable Cedar-Shaped Antenna for WIFI and WIMAX'. Progress In Electromagnetics Research Letters. 72, 135–143.

[27] Alsharoa, A., Neihart, N. M., Kim, S. W., & Kamal A. E., (2018). 'Multi-Band RF Energy and Spectrum Harvesting in Cognitive Radio'. Networks. 2018 IEEE International Conference Communications (ICC), Kansas City, MO. 1-6.

[28] Madi, M. & Kabalan, K. Y. (2018). 'Multi-Band Hand of God Antenna for Wireless Communications'. ICT conf., St Malo, France, July, IEEE proc.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 228-236 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Cloud Service Level Agreements and Resource Management

Isaac Odun-Ayo*, Blessing Udemezue, Abiodun Kilanko

Department of Computer and Information Sciences, Covenant University, 112233, Nigeria

ARTICLE INFO

Article history:
Received: 22 January, 2019
Accepted: 24 March, 2019
Online: 27 March, 2019

Keywords:
Cloud Computing
Cloud Providers
Resource Management
Service Level Agreements

ABSTRACT

Cloud computing is a technical "as-a-service" usage model utilizing virtualization. Virtual machines are the core of cloud computing that runs as independent machine grouped into different networks within the hypervisor. Practically done by deployment of clouds computing hosts enterprise servers in virtual machines on an array of high-end servers. Managing cloud resources has to do with controlling and limiting access to the pool of available resources. This brings about the conception of an agreement between the resource services cloud providers and the customers otherwise known as Service Level Agreement, in other to restrict access to provisioned resources. Resource provisioning is a flexible on-demand pay-as-you-go package that is negotiated; and signed based on SLAs between customers and cloud providers. SLAs enable cloud providers to evade costly SLA consequences payable when there are violations, optimizing the performance of customers' applications and professionally manage resources to reduce cost. These documents are the agreements called Service Level Agreements (SLAs). SLAs outline the expectations which are: terms, conditions, and services of the clients from their service providers regarding availability, redundancy, uptime, cost, and penalties in cases of violations. These ensure clients' confidence in the services offered. Managing resource is an on-going major issue in cloud computing. Considering the limitation of resources, it poses a challenge for cloud service providers to make provision for all resources as needed. This paper seeks a solution to problems that relates to the present trends and developments of what cloud Service Level Agreements are and it ensures so by reviewing current literature. Thus, this research is a study of cloud Service Level Agreements, cloud resource management and their challenges. This paper has made provision to act as guidance for future research and it is anticipated to be beneficiary to potential cloud end users and cloud service providers.

1. Introduction

This paper is an extension of work originally presented in 2018 International Conference on Next Generation Computing and Information Systems [1]. This expands on resource management on Cloud computing and its challenges. Cloud computing makes provision for services by cloud service providers to simplify computing, storage and applications development by an end user using the Internet. Cloud computing is a computing method that enables ubiquitous service which is accessible to a mutual group of computerised resources. . Cloud computing is not limited to networks, services, storage and applications which can be swiftly made provision for and made available with minimal supervision involved [2].

In a cloud environment, resource management is challenging because of the scalability of present data centers, the heterogeneity in types of resources, the interconnection amongst these resources, the inconsistency and irregularity of the load, and the variety of aim of varied end users [3]. Cloud computing and management of resource provisioning are actualized at:

a. Infrastructure-as-a-Service (IaaS) - resources not limited to physical machines, virtual machines, and storage devices are offered as a service to customers.

b. Platform-as-a-Service (PaaS) - offers pre-installed software to customers for application development and testing.

c. Software-as-a-Service (SaaS) - offers a service instance with customizable interfaces to many customers in a cost-efficient manner [2].

Cloud resources management and cloud services are made available based on established Service Level Agreement (SLA) indicated and contracted between Cloud service providers and their end users describing the relations of what is agreed, which

*Isaac Odun-Ayo, Covenant University, +2348028829456
isaac.odunayo@covenantuniversity.edu.ng

also does not exclude non-functional requests, such as quality of service (QoS). In [4], the authors discussed that cloud end users prefer to choose a cloud service appropriate for them from the major service providers that can offer services with adequate quality of service guaranteed.

Lock manager is an independent algorithm that runs on cloud for one manager per server. Algorithm is a set of steps used to solve computing problems such as consistency in maintaining distributed files in cloud computing. There are many client requests made to different servers, requesting for granting simultaneous access to the same file. These requests run independently of each other and pool resources with one another to maintain uniformity of the files in the cloud. This algorithm contributes solution to the major concerns for availability and scalability of file resources [5]. Making the provisioning of resource management flexible and reliable and SLAs are very important to Cloud providers and their clients. Flexible and dependable provision of resource management and SLA agreements in SaaS, PaaS and IaaS are important to the Cloud providers and their clients. This enables cloud providers to evade exorbitant SLA penalties billed for infringements, to professionally manage resources, and lessen charge while elevating the performance of clients' applications. To assure SLA of customer applications, cloud providers need refined resource and monitoring of SLA approaches to accomplish sufficient data for the management procedure [6].

Cloud computing makes use of the concept of virtualization and multi-tenancy to provide service for their end users. Enterprises can migrate some of their applications and utilize services based on appropriate agreements. Cloud computing is made up of four types of deployment: the public, private, community and hybrid clouds. Organizations own private cloud. The amenities can either be in-house or off and the management is by a third party. Private clouds are often perceived to be more secure. Service providers that offer their services with a cost to cloud users majorly own public clouds. Public cloud make use of big data centres most times across many geographical locations. They are perceived to be less secure. Community cloud are run by some firms that have shared mutual interest. The management could be a community or a third party. The combination of either private, community or public cloud is known to be Hybrid cloud. Hybrid computing is cloud utilizes on the benefits of the different cloud types.

Service Level Agreement (SLA) is an official agreement between service providers and clients to assure consumers the quality service expectancy can be realised [6], Service Level Agreement lifecycle involves six phases: ascertain the cloud service providers, outline the Service Level Agreement, institute an agreement, observe likely SLA violators, end SLA and implement consequences [6]. Service Level Agreements are used to officially define the services that offered, the Quality of Service anticipated from cloud provider, duties of both parties involve and possible fines [7]. According to [8], Service Level Agreements ought to comprise of five simple matters. They are: set of services that are made available by the cloud service providers; an explicit and imminent evidence terms of services offered; set of Quality of Service metrics for assessing service delivery points; a way of monitoring these metrics; and means of make a decision during disagreements that may arise as a result of failure in meeting the Service Level Agreement terms. It is imperative for cloud end users to delight in the resources and service assured by the cloud service provider, as the cloud service provider should gain optimally from the resources being provided [9].

The aim of this paper discusses cloud service computing, SLA and resource management. The paper discusses SLA in depth, looks at issues that relate to cloud computing and resource management from the industry perspective. The remaining paper is as follows. Section 2 studies related work. Section 3 converses Service Level Agreement in terms of resource management. Section 4 concludes the paper.

2. Related Works

Managing resources for cloud computing environments is an active field in research. We now present relevant work in Service Level Agreement, resource monitoring and management, framework, method, and model.

In [9], the authors presented dynamic data-centred virtual reality method that prevents Service Level Agreement defilements in combined cloud settings. This research offered an architecture that enables the release of end users resources without problems or violation of the Service Level Agreement. This achieved was through the use of several cloud service providers consumed by end users. SLA in Cloud Computing was offered in [10]; whereby the authors scrutinized SLAs between end users and cloud service providers. A Web Service Level Agreement structure was projected, applied, and confirmed. In [8], a structure for dialoguing SLA of Cloud-based Services was offered. The framework decides the best appropriate cloud service provider for cloud end users. These writers as well emphasised the significant characters of cloud dealers for optimal resource consumption by the cloud end user. The role of supremacy and other Service Level Agreement concerns in cloud environment were made known in [11]. The work untaken, includes main characteristics to be well thought-out when drafting up cloud service legal contracts in cloud computing. The problem of cloud domination was also evaluated with an understanding to determine right ways of handling information with least consequences. In [12], the writers proposed a method that concurrently provided for Quality of Service observing and resource optimal usage in cloud data centres. These authors offered a methodology which gathered user workloads into groups and used these groups to avert Service Level Agreement violation and at the time make better resource consumption, thus profiting both the cloud consumers and cloud providers.

In [2][13], the authors presented a resource monitoring framework which is push-pull technique for Cloud computing environments. This method is established in addition to dominant push and pull methodology monitoring method in Grids to Cloud computing with the aim to create a capable monitoring framework to be able to share resources in Clouds. In [14], the authors defined cloud federalism as being able to bring together services from different cloud vendors to provide a solution; an example is Cloud Burst. According to [15], on resource management, performing live migration is a recent feature concept within Cloud systems for active fault tolerance by flawless movement of Virtual Machines from waning hardware to unwavering hardware with the consumer not being aware of any change in a virtualized environment. Resource consolidation and management can be performed though the use of virtualization technologies. With the use of hypervisors contained by a cluster environment, it enables the consolidation of some standalone physical machines into a virtualized environment, therefore needing fewer physical resources when compared to a system without virtualization capabilities. Nonetheless, this

improvement is still inadequate. Big Cloud deployments have need of physical machines in thousands and megawatts of power.

Data-driven models can be used to manage resource cloud complex systems. The author in [16] discussed his dissertation and argued the advancements in machine learning for the management and optimization of today's resource systems through arising insights from the performance and utilization data these systems create. In other to bring to reality the vision of model-based resource management, there is need for the key challenges data-driven models raise to be dealt with such as uncertainty in predictions, cost of training, generalizability from benchmark datasets to real-world systems datasets, and interpretability of the models. Challenges of data-driven models includes:

a. Robust modelling procedures being able to work with noisy data, and to develop methods for model uncertainty estimates [17].

b. Data-driven model has to incorporate domain knowledge into the modelling process in order to increase the model accuracy and trust in the modelling results [17].

c. Transforming or migrating data from a relational schema to schema-less is difficult to accomplish due to data modeling challenges [18].

In [2], the authors designed a run-time monitoring framework name LoM2HiS to perform monitoring fixated towards increasing the scalability and practice in distributed and parallel environments. The run-time system observer target is to observe the services established on the negotiated and contracted Service Level Agreements by mapping rules; Cloud Service Provider's (CSP) clients request the provision of a contracted service; and the run-time monitor loads the service Service Level Agreement from the decided Service Level Agreement depository. Infrastructure resources are based on the service provisioning that represents hosts and network resources in a data center that host Cloud computing services. The design of this run-time monitor framework is aimed at being highly scalable.

There is a practice of dynamically allocating and de-allocating resources on physical machines in the public cloud environments [16]. As all kind businesses move to cloud, this leads to increase in varied workloads running in the Cloud; and each has diverse performance requirements and financial trade-offs [19]. According to [16], there are varied collection of Virtual Machine (VM) cloud providers, and the main problem of accurately and economically selecting the best VM for a specified workload resources is addressed. Considering the speed of growth in Cloud computing, some VM management platform solutions including OpenStack [13] has come into play.

Apache VCL in [20] is an open-source system for flexible resource providing and reserving of computation resources for different applications in a data center by the use of simplified web interface. In [16], the author discussed about Manjrasoft Aneka a service platform to build and deploy distributed applications on Clouds that makes provision for many applications for the transparency, exploiting distributed resources; and Service Level Agreement focused on allocating resource that distinguishes requested services based on their available resources and hence provides for consumers' needs.

3. Cloud Computing SLA and Resource Management

3.1. Service Level Agreement Concepts

1) National Institute of Standards and Technology (NIST)

The NIST describes cloud computing synopsis and recommendations in its publication [21], viewed from a viewpoint of cloud service provider. Based on this report, a usual profit-making cloud service level agreement has to consist of:

a. These agreements which should clearly include an agreement addressed to the consumer regarding constant availability of service, resolution of dispute steps, preservation of data, and legal protection of consumers' private information.
b. Cloud service provider has limitation of service offered due to effects of natural disasters that are beyond the control of cloud providers', outages in service, and updates. Cloud service providers are nonetheless obligated to provide realistic warning to their end users.
c. The anticipations from cloud consumers, which consist of receiving of terms and conditions including a fee for used services.

This NIST's point of view to Service Level Agreement is quite inflexible and baised in support of the Cloud providers. For a case in point, it does not take into contemplation choice for consumers to converse on the amendment to service level agreements with service cloud providers if the default SLA terms do not address all of the consumers' needs. The report also appeals consumers to know that Service Level Agreements may change at the providers' choice with practical reasons ahead of time notification. In order to be ready to transfer workloads to alternate cloud service providers in case of changes that might be undesirable. This though is not an easy activity, because of vendor lock-in and lack of standardization that could enable interoperability between cloud providers is still an on-going issue in cloud computing.

3.2. Service Level Agreement Levels

Some of the important levels related with cloud Service Level Agreements discussed:

a. Facility Level: The cloud service provider delivers on Service Level Agreement that covers the data centre services needed to conserve the client generated data and/or applications. These consist of things including electrical power, onsite generator, and refrigerating device. Service Level Agreements do assure high availability, fault tolerance, and data replication services.
b. Platform Level: There is need for physical servers, virtualization infrastructure, and network associated hardware possessed by the cloud provider and consumed by the cloud clients. Service Level Agreement at this level would contain information regarding physical security for the rejection of illegal accessibility to the building, facility, and computing resource. Checking background and analysing the character of staff should be carried out before employment by cloud service providers. [22] Every end user has control over his or her applications and some resources, while the CSP has complete control over the platform.
c. Operating System Level: A cloud service provider typically would convey certain volume of services managed to their clients. This additional service authorizes

the cloud provider to assure that the Operating System is properly continuous in operation so that it is reliably available. Service Level Agreements would enclose information regarding updates on security, system patches, confidentiality, or encryption, user authorization, and audit trails log files.

d. Application Level: This level provides safety against problems that associated with application level data. At this point, the cloud provider ensures the readiness, stability, and performance of their cloud user software which they are hosting. This is sometimes difficult to assure specifically in Infrastructure-as-a-Service and Platforma-as-a-Service whereby the end-user is only responsible for the application they put on the cloud.

3.3. Cloud Provider Service Level Agreement

The usual Service Level Agreement of cloud service providers include:

a. Service Assurance: This metric states the service level which a provider is obligated to over a contracted time frame.
b. Service Assurance Time Period: This explains the period of a service assurance ought to take place. The duration can be billed per month, or as contracted upon by mutual parties.
c. Service Assurance Granularity: This outlines the measure of resources for cloud providers to specify services assurance. For instance, the granularity can be as per service, per data centre, per instance, or per transaction basis. In relation to time of the assured service it can be fixed if granularity of service assurance is fine–grained. Service guarantee granularity can also be set up as accumulative of the considered resources such as contacts.
d. Service Guarantee: Oversights are occurrences which are not included in the service guarantee metric calculations. The oversight usually will contain mismanagement of the system by a consumer or an interruption connected with the programmed care.
e. Service Recognition: Is the account attributed to the customer or geared in the direction of forthcoming expenses if the service guarantee does not get realised. The total can be an inclusive or regulated acknowledgement of the consumer reimbursement for an overestimated service.
f. The Service Violation Measurement and Reporting: This explains how and who processes and informs of the violation of service assured respectively. The users typically do this but cloud service provider could also perform this role. In certain instances, a regulating service-monitoring third party could be assigned this.

3.4. Service Level Agreement Benefits and Challenges

2) **Service Level Agreement Benefits**

It is important to have comprehensive contract between the client and cloud provider to guarantee trust and confidence on both sides. The following are a number of the benefits of having a cloud Service Level Agreement [11]:

a. SLA supports solid understanding of cloud service and responsibilities of all stakeholders involved.
b. SLA assists the consumer to realise their understanding.
c. SLA emboldens clarity, accountability, and trustworthiness.
d. Make available information on team performance, capabilities, and staffing judgment.
e. SLA sees to the establishment of reassuring and communal functioning.

3) **Service Level Agreement Challenges**

All single agreement has certain challenges that requires both parties to deal with. These challenges include:

a. Service Level Agreements are difficult to achieve in cloud computing for the reason that some infrastructure and circumstances such as network and force majeure are beyond the control of both the cloud service provider and consumer, therefore challenging to draft a Service Level Agreement.
b. In circumstances where multi-tenancy is in use, Service Level Agreements relating to service separation and high obtainability may be challenging to for the cloud service provider for assurance.
c. Regarding cloud Software-as-a-Service model, Service Level Agreements are challenging to achieve due to the fact that it is almost impossible for the cloud service provider to assess all potential user software/application with diverse system configuration ahead of time.
d. It is challenging to come to an agreement on a cloud Service Level Agreement that is covering security. Distributed Denial of Service (DDoS), keystroke timing and side-channel attacks have been identified as some of the greatest common attacks in cloud environment [15] and continue to be a continuous challenge. As a result of the dynamic and always changing nature of these attacks, a cloud service provider can only at best give a general security-based Service Level Agreement and may not be capable of guaranteeing it.
e. In distributed and multi-cloud application deployment, Service Level Agreements are challenging to agree on as the several providers apply various standards. These standards are often proprietary and unable to interoperate therefore single Service Level Agreement for all parties concerned may be difficult to realise.

3.5. Resource Management and Challenges

4) **Resource Management Models**

Cloud provides a working tool towards achieving high efficiency in resource management and low cost service provisioning [2]. Sophisticated monitoring techniques are needed to cope with resources and the enforcement of SLAs by Cloud providers. Presently, most Cloud providers offer multi-tenant software as a service where provisioning of an application instance is made available to many customers. This has presented problems to monitor especially on the problem of how to ensure Service Level Agreements for diverse customers. The implementation of any application may affect the presentation of the others when Cloud providers make use of sharing resource for provisioning customer applications as a result of different economic reasons. Monitoring a Cloud environment can move to large number of nodes in a data center or distributed within geographical locations which is not a minor undertaking considering that the Cloud provider makes sure the monitoring processes does not make weak the performance of the applications being provisioned. In [23] [24] the authors discussed about ant colony algorithm, a job

programmed algorithm of cloud computing. Considering the cloud computing architecture and the that of ant colony algorithm analysis, managing cloud computing resources is intended to have varied rights management for its users, be it admins. It involves inspecting and the application status of host VM, monitor the task application and task performance scheduling, and the submission of computed tasks.

Sharing resources in cloud computing technology is established on virtualization of physical machines in order to attain most importantly security and performance separation. Virtual machines do share resources of physical machines when they are operational. These virtual machines provide individual customer services, and are shared among customers. It is quite challenging to effectively handle shared resources in order to decrease cost and achieve high utilization. According to [25], Autonomic resource management makes sure that Cloud providers serves huge number of demands without breaking terms of the SLA. With dynamism it manages the resources though the use of VM migration and consolidation. The upcoming computing paradigm is projected to be *utility computing;* that is, making computing services available when users require them, that is making computing services to become commodity utilities, just as internet, electric power, fuel, and telephone.

In [16], the authors pointed out that workload resource requirements are impervious, and during various points at production, it is tough finding out the resources critical for utmost performance. It has been difficult for hosted computing services like AWS Lambda for the load handled as black box functionality. It is understood that the performance detected may be due to primary environmental implementation, the allocation policies, issues performance isolation, or non-linear dependency of performance on resources accessibility. Monitoring resources expended while running task might aid in ascertaining resources for the utilization for that run, but will not indicate performance impacts in controlled set ups, or changed hardware or software.

The author Singh in [24], presented resource allocation model that evaluates users' request centered on resource requirements and resources that are available in order to get least run time. Information on resource represent if resources are available and configured. The resource allocation proposed algorithm assigns resource to negotiate the requirements for minimal run time. Client needs varies and heterogeneous therefore it is considerable to allocate using optimal solution. Availability of resource decides the allocation decision and performance, which brings about pool of various resources. Workload linked with the computing power available of the central processing unit is considered as below demand and present consumption of resources. In uncertain cloud environment, manage resources efficiently; allocation issue executed with swarm optimization tactic made better, the likelihood to look for best appropriate resource for task with least run time possible. Swarm practice is built on nature motivated and artificial intelligence system which involves of self-sufficient agent with combined activities with the use of decentralized system[24] [26].

These authors in [27] presented Ant Colony Optimization (ACO) based energy aware solution to address the problem of allocating resource. This proposed Energy Aware Resource Allocation EARA methodology aims to optimize the allocating of resources so as to make energy efficiency better in the cloud infrastructure while meeting up to the QoS requirements of the end users. The allocation of resources to jobs is based on their QoS requirements. EARA uses ACO at two levels which are: the level where ACO allocates virtual machines resources to jobs, and the level where ACO allocates physical machines resources to Virtual Machines. EARA is handled from the cloud service provider's and end users' point of view. This reduces: the overall energy consumption for the benefit of cloud service provider, and the overall execution cost and entire time of execution for the clients' satisfaction. Benefits of ant colony algorithm includes [28]:

a. Making use of decentralized approach to provide good solution to its single point of failure
b. Collecting information faster
c. Reduces load on networks.

Challenges of ant colony algorithm includes [28]:

a. Dispatching a lot of ants which result to network overloads.
b. Some operating parameters are not taken into consideration which might result to poor performance.
c. Points of initiating ants and the number of ants are not established.

In [29], the authors Zhang and Li proposed a user utility oriented queuing model to handle task scheduling in cloud environment. These authors modelled task scheduling as an M/M/1 queuing system, classified the utility into time utility and cost utility, and built a linear programming model to maximize total utility for time utility and cost utility. They also proposed a utility oriented algorithm to make the most of the total utility. They described the randomness of tasks with the $M/M/1$ model of queuing theory. This utility model involves one server, some schedulers, and some computing resources. When users' tasks are submitted, the server analyzes and schedules the tasks to different schedulers, and adds them to local task queue of the corresponding scheduler. This model can reschedule remaining tasks dynamically to get the maximised utility. Finally, every scheduler schedules its local tasks to obtainable computing resources. The main aim of utility computing is to make provision for the computing resources, which includes: computational power, storage capacity and services to its users while charging them for their usage [30]. In [31], the authors stated that utility computing makes provision for services and computing resources to customers, such as storage, applications, and computing power. This utility computing model services is the foundation of the shift to on demand computing, software as a service and cloud computing models. Benefits of utility computing:

a. The end user have no need to buy all the hardware, software and licenses needed in order to conduct business with cloud service provides.
b. Companies have the option to subscribe to a single service and use the same suite of software throughout the whole client organization.
c. It provides compatibility of all the computers in large companies.
d. It offers unlimited storage capacity

Disadvantages of utility computing:

a. Reliability issue. This means the service could be halted from the utility computing organization for any reason such as a financial trouble or equipment problems.
b. There are many servers across various administrative domains

Some utility computing driving factors [30]:

a. Virtualization enables the abstraction of computing resources so that a physical machine can function as a set of multiple logical virtual machines.
b. Resource pricing in other to develop a pricing standard that can gainfully support economy based utility computing paradigm.
c. Standardization and commoditization are important use of utility services. Standardized information technology services facilitate transparent use of services and make services ideal candidate for utility computing. The commoditization of hardware and software applications such as office suites is increasing with the acceptance of the open standards.
d. Resource allocation and scheduling handle user requests to satisfy user quality of service and maximize profit. A scheduler determines the resources based on what is required of a job which is a user request, performs the resource allocation, and then maps the resources to that job.
e. Service Level Agreement defines service constraints required by the end users. Cloud service providers know how users value their service requests, hence it provides feedback mechanisms to encourage and discourage a service request.

In [32], the author Kim proposed a resource management model that CloudStack supports. CloudStack is an effective resource model that enables scalability, multi-tenant, and it is a cloud computing software that manages the creation and deployment of cloud infrastructure. The system manager is unable to monitor VMs computing resources, but increases visibility into resource allocation and can look up the status of Virtual Machines such as if it is on or off at the console of cloud management. Comprehensive information regarding the status of the Virtual Machine chosen by the system manager can be monitored; examples are the central processing unit use rate, disk, and network activities. The main aim of this management model for resource is to evaluate unprocessed data, then report the statistical information about the Virtual Machine usage volume for the computing resources contained by the resources constraint. What this resource management model specifically does is that:

a. It enables system manager to look into the statistical analysis and the status from the log analysis framework from communication among VMs.
b. It aids system manager in generation of rules for the present constraint of obtainable computing resources gotten from the log analysis framework. An instance is when users make request and create Virtual Machines using a quad core central processing unit, the system manager has the choice of scaling down from a quad core to a dual core for the sum of central processing units constructed on the statistical report from the log analysis framework.
c. Based on firewall policy, it provides the chance to enable the system manager to script enabling the cloud management console to automatically, measure the virtual machines upward or downward.
d. Enables the system manager to produce policies during emergencies for service stability of the virtual machines.

5) **Resource Management Challenges**

Traditional resource management methods are not enough for cloud computing because they are on virtualization technology with distributed nature [33]. Considering the growth in Cloud computing, it has challenges for managing resource because of heterogeneity in hardware capabilities, per request service model, pay-as-you-use model and assurance to satisfy quality of service. Managing resources dynamically will reduce the resource controversy, scarcity of resources, fragmentation of resource, and overflow / underflow problems. To circumvent these difficulties resources should be adjusted with dynamism to manage resources efficiently in the cloud architecture [34].

There are challenges in the capacity to monitor resource and to manage SLAs in cloud computing, they include [2]:

a. Distributed Application Execution – NetLogger is a distributed system that monitors in order to observe and collect information of computer networks. In [2], the authors discussed ChaosMon, an application that monitors and displays performance information for parallel and distributed systems. This monitoring system is aimed to support application developers to enable them specify performance metrics and to also observe them visually in order to identify, evaluate, and comprehend performance logjams. This system is a distributed monitoring that has a centralized control.
b. Enforcement of the Agreed SLA – Enforcing SLAs is challenging and it involves complex processes whereby both parties might have to keep going back and forth until a mutual agreement is achieved. Implementation of different technologies poses challenges to evaluate Service Level Agreement documents, multifaceted deployment procedures, and scalability concerns. Ensuring uninterrupted monitoring and dynamic resource provision can be employed to put into effect agreed SLA.
c. Energy Management - Managing the consumption of energy is becoming challenging. Statistics and report show unchecked energy consumption might be a key problem in few years' time. According to [14], Stanford University made known that there is an annual increase of 56% in the consumption of electricity in data centers between 2005 and 2010.
d. Multi-Tenancy Application – This provision is becoming standard amongst Cloud providers to cost efficiently provide their sole instant application to numerous consumers. The technology approach is still challenging in its realization. In [2], the authors discussed proposed SLAs that is multi-tenant oriented in monitoring performance, in detection, and to schedule architecture with their clients. The writers aimed towards separating and guaranteeing optimized performance of each client in the Cloud space. This is complex and challenging to manage such resource because the cloud provider is making provision for every tenant with the same

application instance whereby the Service Level Agreements for the applications are contacted on per tenant basis. In situations like this, the workload of some tenants' applications will affect the unusual resource consumptions by other tenants considering the Cloud hardware, and all tenants share software resources. To solve this issue, a dynamic SLA mechanism architecture is designed, which monitors application grades, discerns what is not normal, and dynamically schedules shared resources to make sure of tenants' SLAs and optimize system performance.

e. High Availability - This is the server's availability for use for its intended purpose. High availability main objective is to ensure a system continues to function regardless of the failure of any components. This can be achieved through redundancy and elimination of single point of failure [14].

f. The Scalability of the Monitoring Mechanism - Key factors of cloud computing includes scalability of resource and application provisioning on-demand in a pay-as-you-go method. Cloud environments measure thousands of computing devices in large-scale environments; hence, the need for scalable monitoring tools. In [35], the authors proposed some methods to handle the challenges in setting up virtualized datacenter management systems that are scalable. The plan is to strategise a high performance and strong management tool that will be able to scale from some hosts to a large-scale Cloud datacenter. These authors indicated how they came about a scalability management tool by noting these challenges:
    a. Performance / Fairness;
    b. Security;
    c. Robustness;
    d. Availability; and
    e. Backward compatibility.

3.6. Trending Practices in Cloud Resource Management

Daily growing attributes and advancements in cloud computing are promising the future and luring organizations and stakeholders towards the use of cloud computing. This fast growth of cloud computing is in need of stricter security and large quantity of resources like memory, process and storage to meet the Service Level Agreement standards. The trending practices include:

a. Live Migration – The use of live migration of virtual machines is an indispensable attribute of virtualization, enabling a seamlessly and transparent migrating of virtual machines from a source host to another destination host, even though the VM is still running without interrupting VMs [36]. Cloud live migration is on the increase in cloud computing with the aim to gain the cloud benefits from non-cloud situated applications. Live migration is a recent concept within cloud computing that is utilized during practical fault tolerance for instance by faultlessly moving Virtual Machines away from faulty hardware to stable hardware without the consumer noting any adjustment in a virtualized space. This is so because in the use of live migration, the end user is unaware there is only a delay of 60 mini seconds to 300 mini seconds [15]. Virtual Machine live migration can be performed via static migration or Dynamic migration. In order to fulfill Service Level Agreement, static Virtual Machine migration is used as required resources gets carefully chosen as per requirement in SLA, but in dynamic Virtual Machine provisioning, the resources change dynamically to handle unforeseen workload changes [37]. Major benefit for live VM migration includes [36]:
    a. Load balancing, the moving of over-loaded servers to light-loaded servers in order to release congested hosts.
    c. Proactive fault tolerance and online maintenance

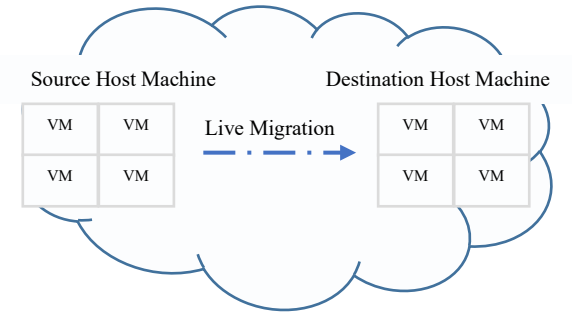


Figure 1: Live Migration [27]

b. Power Saving in Green Computing - the authors in [38] discussed a proposed resource-utilization- aware energy saving server consolidation algorithm for the provision of improved utilization of resources as it reduces the total number of virtual machines during live migrations. Experimental results demonstrated its ability to reduce the consumption of energy and violations in Service Level Agreement in cloud data center. In [39], the authors describes green data center as a depository of a competent management of the system with a reduced amount of power consumed environment. The power consumption by an average data center can be so enormous that it can serve as a power source for approximately 25,000 private homes. In [40], Jena presented an optimization algorithm that arrange job with the aim of achieving an optimized energy usage and overall computation time using Clonal Selection Algorithm (CSA). Dynamic Voltage Frequency Scaling (DVFS) technique designed to reduce the usage of cloud resources by managing the power voltage as well as the regularity of the processor without ruining the performance. In [41], the authors presented an optimization framework of two modules that run concurrently: the Datacenter Energy Controller that reduces the use of datacenter energy without degrading the Quality-of-Service (QoS) and; the Green Energy Controller that brings about renewable sources. This Green Energy Controller computes the anticipated energy budget for the datacenter by the use of the lithium-ion battery as an added energy reserve or the grid of both banks in which the Hybrid Electric Systems can drain. For cloud providers to make their services become green computing, they have to spend in renewable energy sources to generate power from renewable sources of energy, like wind, solar, or hydroelectricity [42]. Merging

resources can advance utilization and make provision for increase in space, power and cooling capacity within the same facility environment. Energy preserving approaches includes [43]:

a. Nano Data Center is a distributed computing platform that makes provision for computing and storage services and adopts a managed peer-to-peer model to form a distributed data center infrastructure.
b. Reduction of Central Processing Unit Power Dissipation adapts free cooling so that power dissipation reduces.
c. Server Consolidation is the migration of server roles from various underutilized physical servers to VMs for the reduction of hardware and energy consumption.
d. Energy Saving Strategy of System Software is a dynamic energy usage, so the operating system can dynamically manipulate the system unit for obtaining the minimal power consumption without degrading already assigned the task or any job in performance.
e. Live Migration of Virtual Machine is about the movement of a running Virtual Machine from one host to another.

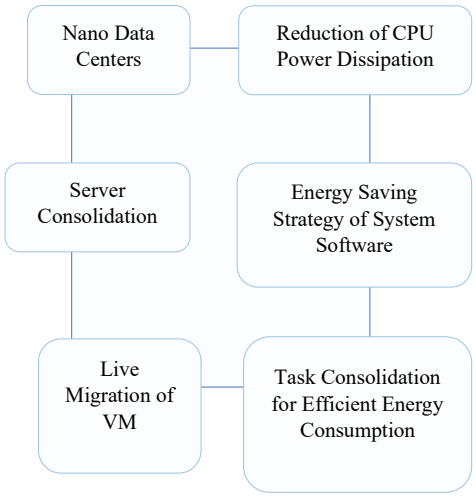


Figure 2: Energy Preserving Approaches [34]

3.7. Resource Grid Architecture Management

In [34], the authors proposed a Resource Grid Architecture for Multi Cloud Environment (RGA) for allocation and management of the resources with dynamism in virtual manner. Resource Management in Single Cloud Architecture is an ongoing area in research and RGA for Multi Cloud Environment is a current practice that has high scalability to manage resource at single cloud environments. This architecture presents a resource layer with logical resource grid and the usages of the VMs to project the resources in contrast to physical systems of a cloud. Experimentations have showed that their Resource Grid Architecture can efficiently manage the resources against numerous clouds and support the green computing. Logically, RGA is grouped into four layers to manage resource efficiently:

a. Cloud Layer - This layer investigates the assigned resources to cloud application in order to keep log of record for resource usage and directs the information to cloud Resource Prediction Unit for hot spot detection and cold spot detection. Local Resource Manager is a component that verifies the resource allocation and utilization only at cloud level and is a proxy of Global Resource Manager.
b. Network Layer - This layer supports high-speed commutation among cloud and is the backbone for the experimented architecture for the reason that intra cloud communication is very much in demand in the implementation for resource sharing and process sharing.
c. Virtual Machine Layer - This layer is a separate group of virtual machine for sole cloud architecture to accomplish the operational feasibility. Each group of virtual machine is a set of virtual machines for physical system representation and are constantly accessible to resource grid layer monitoring and resource mapping.
d. Resource Grid Layer – This layer is an assembly of management modules of this architecture such as Global Resource Manager; these modules are interwoven for the monitoring, analyzing, allocating and tracking of resource information of multiple clouds at a centralized virtual location.

4. Conclusion

Cloud computing makes provision for scalable, on demand, elastic, multi-tenant and virtualization of services to customers on the Internet through cloud service providers. The types of service are the SaaS which provides applications with customizable interfaces, PaaS provides platform for application development and testing, and IaaS which makes provision of storages and computing infrastructure to users and extends to physical machines, virtual machines. These services provided are based on the service level agreements between the cloud service providers and the client. The service level agreement stipulates the terms of the services makes provision for mutually beneficial transaction between the parties. This paper presented a survey of recent developmental trends and issues in resource management in cloud computing. Cloud computing is aimed at providing the best resource to its users. Many researchers are focused on request allocation and resource management for Cloud Computing. All-encompassing use of learning empowers systems to be dynamic in handling workloads and execution environments. This paper documented techniques for creating scheduling and resource allocation choices for cloud computing and the datacenters. This paper concluded with a review of resource management on Cloud computing and cloud SLAs.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation, and Discovery (CUCRID).

References

[1] I. Odun-ayo, A. Olasupo, and N. Omoregbe, "Cloud Service Level Agreements – Issues and Development," Int. Conf. Next Gener. Comput. Inf. Syst., pp. 1–6, 2017.

[2] V. C. Emeakaroha and R. N. Calheiros, "Achieving Flexible SLA and Resource Management in Clouds," Achiev. Fed. self-manageable cloud infrastructures theory Pract., pp. 266–287, 2012.

[3] N. M. Gonzalez, T. C. M. de B. Carvalho, and C. C. Miers, "Cloud resource management: towards efficient execution of large-scale scientific applications and workflows on complex infrastructures," J. Cloud Comput. Adv. Syst. Appl., vol. 6, no. 1, 2017.

[4] I. Odun-ayo, O. Ajayi, and A. Falade, "Cloud Computing and Quality of Service : Issues and Developments," Proc. Int. MultiConference Eng. Comput. Sci., vol. I, pp. 179–184, 2018.

[5] A. Koci and B. Cico, "ADLMCC-Asymmetric Distributed Lock Management In Cloud," Int. J. Inf. Technol. Secur., vol. 10, 2018.

[6] L. Wu, S. K. Garg, and R. Buyya, "Journal of Computer and System Sciences SLA-based admission control for a Software-as-a-Service provider in Cloud computing environments," J. Comput. Syst. Sci., vol. 78, no. 5, pp. 1280–1299, 2012.

[7] K. Buck, D. Hanf, and D. Harper, "Cloud SLA Considerations for the Government Consumer," Syst. Eng. Cloud Comput. Ser., pp. 1–65, 2015.

[8] R. El-awadi and M. Abu-rizka, "A Framework for Negotiating Service Level Agreement of Cloud-based Services," Int. Conf. Commun. Manag. Inf. Technol., no. 65, pp. 940–949, 2015.

[9] F. Faniyi, R. Bahsoon, and G. Theodoropoulos, "A dynamic data-driven simulation approach for preventing service level agreement violations in cloud federation," Int. Conf. Comput. Sci. ICCS, pp. 1167–1176, 2012.

[10] S. B. Dash, H. Saini, T. C. Panda, A. Mishra, and A. B. N. Access, "Service Level Agreement Assurance in Cloud Computing : A Trust Issue," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 2899–2906, 2014.

[11] M. Cochran and P. D. Witman, "Governance and Service Level Agreement Issues in A Cloud Computing Environment," J. Inf. Technol. Manag., vol. XXII, no. 2, 2011.

[12] Olasupo O. Ayayi, F. A. Oladeji, and C. O. Uwadia, "Multi-Class Load Balancing Scheme for QoS and Energy Conservation in Cloud Computing," West African J. Ind. Acad. Res., vol. 17, pp. 28–36, 2016.

[13] Y. Jiang, H. Sun, J. Ding, and Y. Liu, "A Data Transmission Method for Resource Monitoring under Cloud Computing Environment," Int. J. Grid Distrib. Comput., vol. 8, no. 2, pp. 15–24, 2015.

[14] O. Ajayi and F. . Oladeji, "An overview of resource management challenges in cloud computing," UNILAG Annu. Res. Conf. Fair, vol. 2–3, pp. 554–560, 2015.

[15] A. J. Younge, G. Von Laszewski, L. Wang, S. Lopez-alarcon, and W. Carithers, "Efficient Resource Management for Cloud Computing Environments," Int. Conf. Green Comput. IEEE , 2010.

[16] N. Yadwadkar, "Machine Learning for Automatic Resource Management in the Datacenter and the Cloud," Ph.D. Diss. Dept. Electr. Eng. Comput. Sci. Electr. Eng. Comput. Sci. Univ. California, Berkeley, U.S. Accessed Jan. 6, 2019. Available http//digitalassets.lib.berkeley.edu/etd/ucb/text/Yadwadka, 2018.

[17] A. Ostfeld, "Data-Driven Modelling : Some Past Experiences and New Approaches approaches," J. Hydroinformatics, vol. 10, no. 1, 2008.

[18] A. Schram and K. M. Anderson, "MySQL to NoSQL Data Modeling Challenges in Supporting Scalability," Proc. 3rd Annu. Conf. Syst. Program. Appl. Softw. Humanit., pp. 191–202, 2012.

[19] C. Reiss, G. R. Ganger, R. H. Katz, and M. A. Kozuch, "Heterogeneity and Dynamicity of Clouds at Scale : Google Trace Analysis," Proc. Third ACM Symp. Cloud Comput., 2012.

[20] G. Blanchard, G. Lee, and C. Scott, "Generalizing from Several Related Classification Tasks to a New Unlabeled Sample," Proc. 24th Int. Conf. Neural Inf. Process. Syst., pp. 2178–2186, 2011.

[21] L. Badger, R. Patt-corner, J. Voas, L. Badger, R. Patt-corner, and J. Voas, "Cloud Computing Synopsis and Recommendations," Natl. Inst. Stand. Technol. Spec. Publ. 800-146, 2012.

[22] I. Odun-ayo, N. Omoregbe, and B. Udemezue, "Cloud and Mobile Computing – Issues and Developments," Proc. World Congr. Eng. Comput. Sci., vol. I, 2018.

[23] F. G. Fang, "Study on Resource Management Based on Cloud Computing," MATEC Web Conf., vol. 63, 2016.

[24] H. Singh, "Efficient Resource Management Technique for Performance Improvement in Cloud Computing," Indian J. Comput. Sci. Eng., vol. 8, no. 1, pp. 33–39, 2017.

[25] R. Buyya, S. K. Garg, and R. N. Calheiros, "SLA-Oriented Resource Provisioning for Cloud Computing : Challenges, Architecture, and Solutions," Int. Conf. Cloud Serv. Comput., 2011.

[26] E. Pacini, C. Mateos, and C. García Garino, "Distributed job scheduling based on Swarm Intelligence: A survey," Comput. Electr. Eng., vol. 40, no. 1, pp. 252–269, 2014.

[27] K. Ashok, K. Rajesh, and S. Anju, "Energy aware resource allocation for clouds using two level ant colony optimization," Comput. Informatics, vol. 37, pp. 76–108, 2018.

[28] B. Ghutke, "Pros and Cons of Load Balancing Algorithms for Cloud Computing," Int. Conf. Inf. Syst. Comput. Networks, 2014.

[29] Z. Zhang and Y. Li, "User utility oriented queuing model for resource allocation in cloud environment," J. Electr. Comput. Eng., vol. 2015, pp. 1–8, 2015.

[30] I. Chana and T. Kaur, "Delivering IT as A Utility- A Systematic Review," Int. J. Found. Comput. Sci. Technol. (IJFCST)., vol. 3, no. 3, 2013.

[31] S. Mawia, I. Omer, and A. B. A. Mustafa, "Comparative study between Cluster, Grid, Utility, Cloud and Autonomic computing," IOSR J. Electr. Electron. Eng., vol. 9, no. 6, pp. 61–67, 2014.

[32] A. Kim, J. Lee, and M. Kim, "Resource management model based on cloud computing environment," Int. J. Distrib. Sens. Networks, vol. 12, no. 11, 2016.

[33] S. Parikh, N. Patel, and H. B. Prajapati, "Resource Management in Cloud Computing : Classification and Taxonomy," Comput. Res. Repos., 2017.

[34] C. Aruna and R. Prasad, "Resource Grid Architecture for Multi Cloud Resource Management in Cloud Computing," Emerg. ICT Bridg. Futur. - Proc. 49th Annu. Conv. Comput. Soc. India, vol. 1, pp. 631–640, 2015.

[35] S. Vijayaraghavan and K. Govil, "Challenges in Building Scalable Virtualized Datacenter Management," ACM SIGOPS Oper. Syst. Rev., vol. 44, no. 4, pp. 95–102, 2012.

[36] M. Noshy, A. Ibrahim, and H. A. Ali, "Optimization of live virtual machine migration in cloud computing : A survey and future directions," J. Netw. Comput. Appl., vol. 110, no. March, pp. 1–11, 2018.

[37] S. B. Rathod and K. R. Vuyyuru, "Secure Live VM Migration in Cloud Computing : A Survey," Int. J. Comput. Appl., vol. 103, no. April 2016, 2014.

[38] B. Saha, "Green Computing : Current Research Trends," Int. J. Comput. Sci. Eng., vol. 6, no. June, pp. 3–6, 2018.

[39] S. Sharma and G. Sharma, "A Review on Secure and Energy Efficient Approaches for Green Computing," Int. J. Comput. Appl., vol. 136, no. May, 2016.

[40] R. K. Jena, "Energy Efficient Task Scheduling in Cloud Environment," 4th Int. Conf. Power Energy Syst. Eng., vol. 141, pp. 222–227, 2017.

[41] A. Pahlevan, R. Maurizio, G. D. V. Pablo, B. Davide, and A. David, "Joint computing and electric systems optimization for green datacenters Handbook of Hardware / Software Codesign Soonhoi Ha and Jürgen Teich," Jt. Comput. Electr. Syst. Optim. Green Datacenters. Ha S., Teich J. Handb. Hardware/Software Codesign. Springer, Dordr., no. November, 2016.

[42] M. K. Aery, "Green Computing : A Study on Future Computing and Energy Saving Technology," An Int. J. Eng. Sci., vol. 25, no. November 2017, 2018.

[43] A. D. Borah, D. Muchahary, and S. K. Singh, "Power Saving Strategies in Green Cloud Computing Systems," Int. J. Grid Distrib. Comput., vol. 8, no. February, pp. 299–306, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 234-242 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Transformation of Solar Energy to Electricity with Photovoltaic Systems, Reduction of Electrical Consumption and Installation of LED Lamps in the ITSAO

Jeremias Bravo Tapia*, Juan José Rivera Toledo, Francisco Ramos Guzmán, Yoxkin Estévez Martínez

Instituto Tecnológico Superior de Acatlán de Osorio, División de la Carrera de Ingeniería Electrónica, Carretera Acatlán-San Juan Ixcaquixtla Km. 5.5 Unidad Tecnológica C. P. 74949, Acatlán de Osorio, Puebla, México.

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 26 August, 2019
Online: 08 October, 2019

Keywords:
Electricity consumption
LED lamps
Photovoltaic system
Solar planter

ABSTRACT

In this work of technological development and research is presented the work done in the ITSAO to contribute to the use of clean energy such as photovoltaics and the implementation of activities to reduce electricity consumption by the technological community, the monitoring of electricity consumption is made on a monthly basis and an annual report is prepared in which the consumption monitoring and the reduction percentage thereof are shown by tables and graphs. The students of electronic Engineering carry out activities of optoelectronics and sustainable development where they intervene with the design and installation of photovoltaic systems that are part of their academic formation. It has a photovoltaic system that produces 2.16 kW-h / d and another will be installed to have photovoltaic energy in the tables of the ITSAO planters; these actions are part of the operational control related to the energy of the Integral Management System (SGI).

1. Introduction.

This paper is an extension of work originally presented in 2018 XXXI International Summer Meeting on Power and Industrial Applications (RVP-AI) [1] off the IEEE Sección México; in the Instituto Tecnológico Superior de Acatlán de Osorio, Puebla, México; work continues with the development of projects related to the saving of electrical energy with renewable energy applications and the replacement of fluorescent tube lamps by LED's lamps; this is how the Academy of Electronic Engineering has intervened in the installation and replacement of LED's tube lamps.

In [2] it presents a work aimed at reducing the consumption of electrical energy delivered by the conventional electricity network, through the use of photovoltaic solar energy. It is proposed the design of an electronic transfer system that facilitates the use of DC electric power from the solar panel, thus eliminating the use of the voltage inverter, commonly used in this type of applications and which increases the cost of the use of solar energy.

Alternative energies have always been present in the daily life of humanity, although their existence has often been unknown because of the need to put the economic over the environmental benefit. In addition to being free, they are clean and protective of the environment [2]. It is necessary to implement an adequate ecological culture in the present generations so that it contributes to the care of the global environment and in this way the new generations are already developed in an environmental system from which it will no longer be necessary to learn to adapt to the environment.

Currently as an example of the application of energy efficiency is in a house at least there is a device that requires the use of electrical energy for its operation, so in some cases there is a high demand and even waste of energy, when there is a considerable amount of these consumer elements, it affects the areas of the rural sector and certain cases of a greater proportion of electricity consumption, because although they have subsidized rates, they suffer increases in costs, thus paying the demand for generated consumption by other sectors, without worrying about the high value in the payment of the energy service. [2].

The study of renewable energies is currently presented as a very useful tool, due to the scarcity of conventional energy sources that have been exploited for several years, without thinking that this would lead humanity to a point where those resources would be exhausted [2]. This allows the development of research for the generation of electricity from clean energy sources in small, medium and large-scale production.

* Jeremias Bravo Tapia., Email: jrmsbt@hotmail.com

Energy efficiency is a primary consideration when designing off-grid renewable energy systems including portable micro-grids. This study focuses on characterizing the potential benefits associated with using energy-efficient exterior area lighting commonly associated with remote installations. Light-Emitting Diode (LED) luminaires are becoming more commercially available, and this study compares two LED products designed for exterior lighting to traditional metal halide lamps. The energy efficiency benefit of the LED luminaires is offset by their lower illuminance. However, a comparison of lighting standards for specific purposes such as security lighting indicates that LEDs may be appropriate for applications where a metal halide system would provide significantly more illumination than required at a much higher energy cost [3].

Measures to improve energy efficiency in small and medium enterprises do not apply due to financial reasons, lack of information and limited internal skills. Suggested information measures, audits, capacity building. As potentially effective measures to support the implementation of energy efficiency measures. Companies in Austria, Bulgaria, Cyprus, Italy, Romania, Slovakia and Spain have identified and implemented a wide range of energy efficiency measures, which in part consist of no-cost options, but are also partly innovative technical solutions [4]. This must be implemented throughout the world starting from our home to be able to impact the environment around us.

Energy efficiency has a vital part to play in all future energy scenarios. Improved energy efficiency limits energy demand growth to one-third by 2040 even though the global economy grows by 150%. Energy policy has traditionally been almost entirely concerned with energy supply and most often focused on the electricity system. The close link between electrification and modernization has given the electricity supply industry in all countries a strong political voice. Energy efficiency, energy conservation or “energy savings” has tended only to be given a high priority in times of crisis, when energy prices are particularly high or when economies face electricity shortages. This was most notable during the “oil crises” of the 1970s. Until recently, energy efficiency policy has always had a lower priority in energy policy development compared to energy supply [5].

All the energy used is ultimately to meet human needs. Natural gas is not burned to heat the room in which one lives by simply heating, but it is done to feel comfortable in that room. Energy is used to produce the materials that are needed to make consumer goods, for example, steel is produced to make cars, paper for printing books and fertilizers to grow more food so as not to be hungry. Energy is not a necessity in itself, it is one of the goods used to meet human needs that can be realized through the use of energy called energy services. [6].

Energy consumption in homes and buildings is large and, in the future, it tends to be unsustainable for our planet, both due to the depletion of natural resources and the irreversible damage it causes to the ecosystem. To help the sustainable development of Humanity it is essential to create an energy education that allows, without waste, continue using fossil fuels for the development of new energy technologies more efficient and in harmony with the environment. The option of renewable energies offers an alternative to achieve a social development that tends to the ecological and that in the future an economic benefit is obtained as well. The energy obtained from photovoltaic panels, specifically the Sun Irradiance, is proposed as an aid to minimize energy consumption in both houses and buildings. Currently, photovoltaic energy is still not profitable, but it seems to be the most viable option until now to solve the great problem of generating electricity without using fossil fuels [7].

The photovoltaic parks have taken great importance as renewable energy generation systems in individual buildings integrating from the design of the building involving the architecture and installers of photovoltaic systems; thus, the photovoltaic industry is a technology that is already included in the architectural design process [8], [9] & [10].

In ITSAO, a photovoltaic system will also be implemented in the gardens of the esplanade so that the technological community has the possibility of connecting their electronic equipment such as laptops and mobile devices outdoors on the rest tables.

2. Determination of solar radiation.

For the development of the photovoltaic part there is location data of the ITSAO in Google Maps [11]. Data provided by NASA corresponding to the solar radiation received by the solar panels, the hours of solar radiation were used to calculate the design of the battery bank and the temperature to see the performance of the solar panels [12].

The coordinates of the location of the ITSAO at a latitude of 18.226883 and a longitude of -98.040143 are shown in Figure 1 a). Figure 1 b) shows the location coordinates on the NASA platform in Prediction of Worldwide Energy Resource Prepare.

Next, Figure 2 shows the data that must be identified for the calculation and design of the photovoltaic system as the minimum radiation for solar panels inclined towards Ecuador, the days without sun in a month to determine the size of the battery bank and the maximum and minimum temperature at two meters high.

Figure 3 a) shows the behavior of the monthly days without solar radiation to add the days of autonomy in the design of the size of the battery bank, in Figure 3 b) the maximum temperature is shown and in Figure 3 c) the minimum temperature is shown at two meters high. this to determine the efficiency behavior of solar panels that depend on the temperature.

One of the very important parameters for the sizing of photovoltaic systems is the solar radiation that is received in the place of the installation of the solar panels, for this data is taken provided by NASA that as mentioned above are reliable data to work sizing, performance, days of autonomy and storage capacity.

Figure 4 shows parameters the solar radiation values in the ITSAO; days without sun, maximum temperature, minimum temperature; in blue are the radiation data for different inclinations are monthly data and an annual average.

Figure 5 shows the optimal peak solar radiation hours (HSP) to calculate the size of the sizing of the photovoltaic system on a monthly and annual basis; The annual average is 5.87 HSP with this the daily production that a solar panel will have is calculated according to its power provided by the manufacturer.

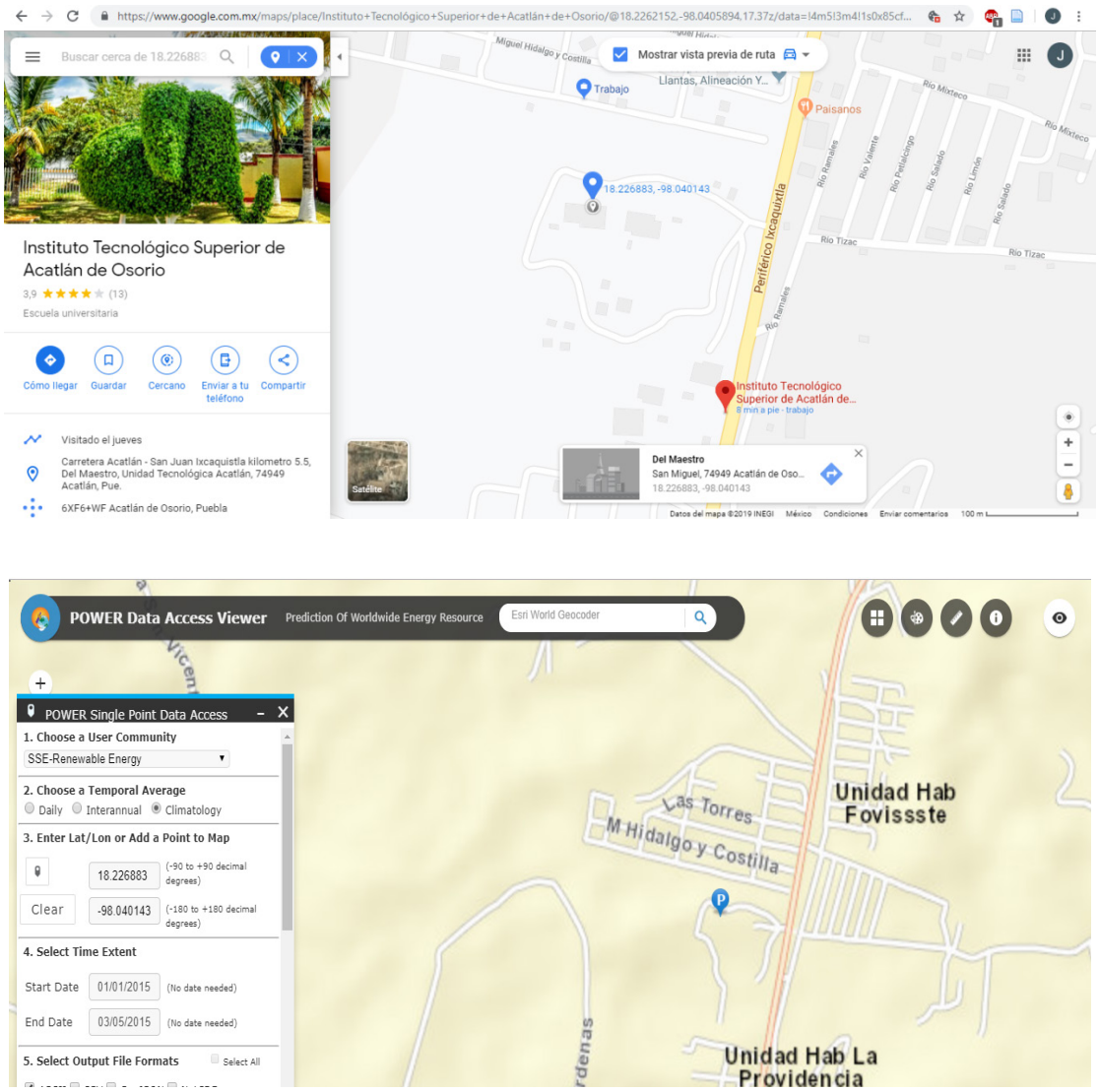
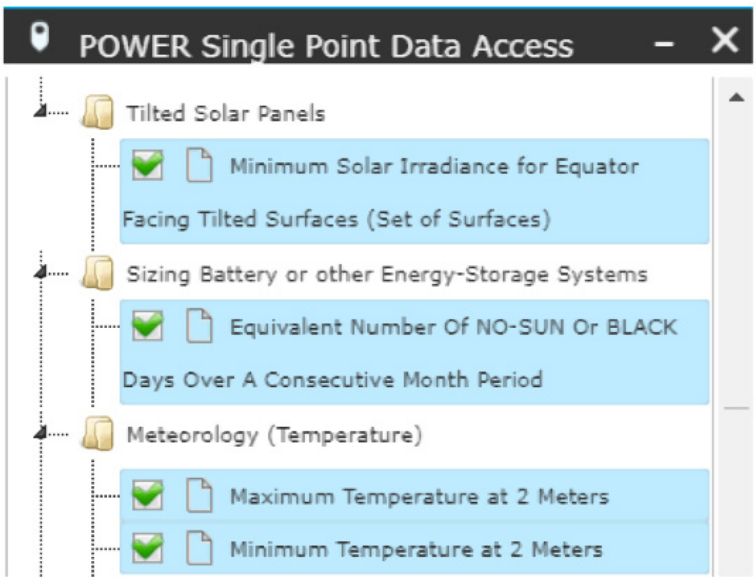
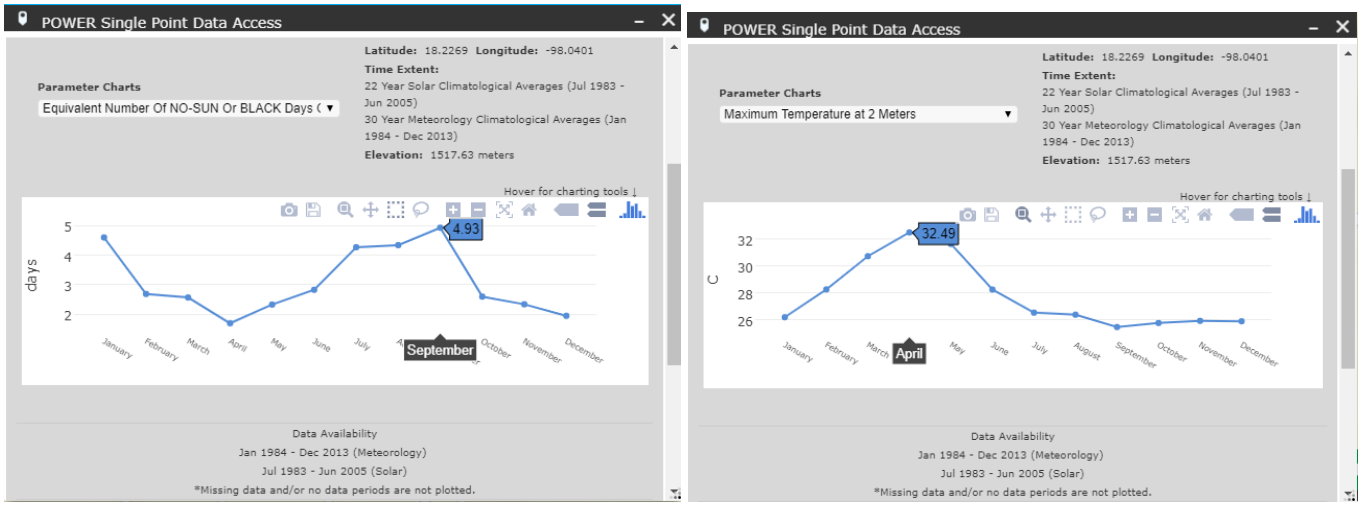


Figure 1: Location coordinates of ITSAO. a) Google maps [11]. b) NASA platform [12].



[12].



b)

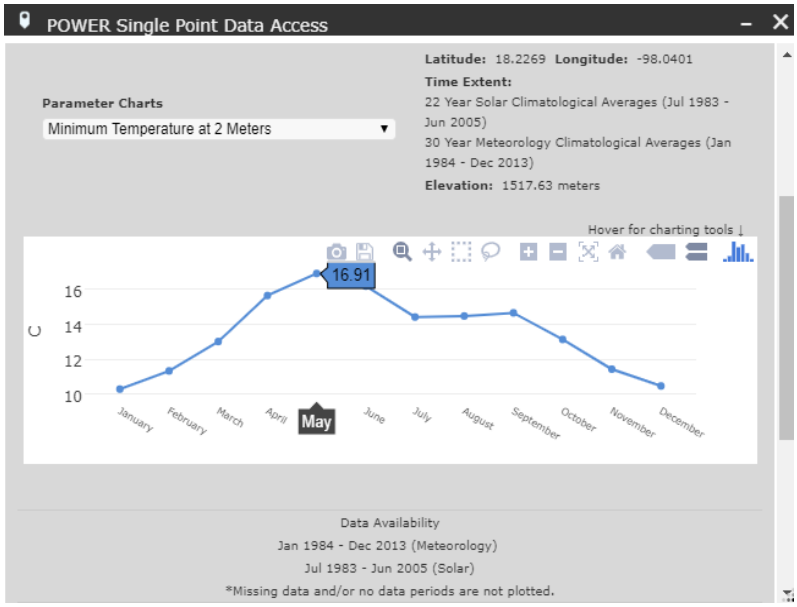


Figure 3: Parameter graphs for calculating photovoltaic systems, a) days without solar radiation, b) maximum temperature and c) minimum temperature [12].

https://power.larc.nasa.gov/downloads/POWER_SinglePoint_Climatology_018d23N_98d40W_15095925.txt

```
-BEGIN HEADER-
NASA/POWER SRB/FLASHFlux/MERRA2/ 0.5 x 0.5 Degree Climatologies
22-year Additional Solar Parameter Monthly & Annual Climatologies (July 1983 - June 2005), 30-year Meteorological and Solar Monthly & Annual Climatologies (January 1984 - December 2013)
Location: Latitude  18.2269   Longitude -98.0401
Elevation from MERRA-2: Average for 1/2x1/2 degree lat/lon region = 1517.63 meters   Site = na
Climate zone: na (reference Briggs et al: http://www.energycodes.gov)
Value for missing model data cannot be computed or out of model availability range: -999
Parameter(s):
EQVLNT_NO_SUN_BLACKDAYS_MONTH    SRB/FLASHFlux 1/2x1/2 Equivalent Number Of NO-SUN Or BLACK Days Over A Consecutive Month Period (days)
T2M_MAX                          MERRA2 1/2x1/2 Maximum Temperature at 2 Meters (C)
T2M_MIN                          MERRA2 1/2x1/2 Minimum Temperature at 2 Meters (C)
SI_EF_MIN_TILTED_SURFACE         SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance for Equator Facing Tilted Surfaces (Set of Surfaces) (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL                SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL_ANG            SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal Angle (Degrees)
SI_EF_MIN_TILTED_ANG_ORT         SRB/FLASHFlux 1/2x1/2 MinimumSolar Irradiance Tilted Surface Orientation (N/S Orientation)
PARAMETER                 JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC     ANN
-END HEADER-
T2M_MAX                 26.18   28.25   30.72   32.49   31.63   28.24   26.52   26.37   25.45   25.75   25.91   25.88   27.78
T2M_MIN                 10.30   11.34   13.02   15.65   16.91   16.19   14.42   14.48   14.66   13.14   11.44   10.48   13.50
EQVLNT_NO_SUN_BLACKDAYS_MONTH 4.60  2.69    2.57    1.70    2.33    2.83    4.27    4.34    4.93    2.60    2.34    1.95    -999
SI_EF_MIN_TILTED_SURFACE_0  4.37    5.40    6.20    6.50    6.12    5.63    5.66    5.36    4.70    5.08    4.91    4.57    5.38
SI_EF_MIN_TILTED_SURFACE_18 5.14    6.06    6.52    6.39    6.03    5.67    5.66    5.17    4.76    5.52    5.77    5.56    5.69
SI_EF_MIN_TILTED_SURFACE_3  4.53    5.54    6.29    6.52    6.14    5.67    5.69    5.35    4.73    5.18    5.08    4.76    5.46
SI_EF_MIN_TILTED_SURFACE_33 5.48    6.25    6.39    5.92    5.62    5.39    5.35    4.74    4.55    5.56    6.12    6.04    5.62
SI_EF_MIN_TILTED_SURFACE_90 4.01    3.90    3.05    1.90    1.97    2.36    2.20    1.60    2.07    3.17    4.31    4.68    2.94
SI_EF_MIN_OPTIMAL       5.53    6.25    6.52    6.52    6.14    5.70    5.71    5.36    4.78    5.58    6.16    6.17    5.87
SI_EF_MIN_OPTIMAL_ANG   42.00   33.00   20.00   5.00    -6.00   -11.00  -9.00   0.00    12.00   27.00   40.00   46.00   16.00
SI_EF_MIN_TILTED_ANG_ORT    S       S       S       S       N       N       N       S       S       S       S       S       S
```

[12].

https://power.larc.nasa.gov/downloads/POWER_SinglePoint_Climatology_018d23N_98d40W_1e021619.txt

```
-BEGIN HEADER-
NASA/POWER SRB/FLASHFlux/MERRA2/ 0.5 x 0.5 Degree Climatologies
22-year Additional Solar Parameter Monthly & Annual Climatologies (July 1983 - June 2005), 30-year Meteorological and Solar Monthly & Annual Climatologies (January 1984 - December 2013)
Location: Latitude  18.2269   Longitude -98.0401
Elevation from MERRA-2: Average for 1/2x1/2 degree lat/lon region = 1517.63 meters   Site = na
Climate zone: na (reference Briggs et al: http://www.energycodes.gov)
Value for missing model data cannot be computed or out of model availability range: -999
Parameter(s):
EQVLNT_NO_SUN_BLACKDAYS_MONTH    SRB/FLASHFlux 1/2x1/2 Equivalent Number Of NO-SUN Or BLACK Days Over A Consecutive Month Period (days)
T2M_MAX                          MERRA2 1/2x1/2 Maximum Temperature at 2 Meters (C)
T2M_MIN                          MERRA2 1/2x1/2 Minimum Temperature at 2 Meters (C)
SI_EF_MIN_TILTED_SURFACE         SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance for Equator Facing Tilted Surfaces (Set of Surfaces) (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL                SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL_ANG            SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal Angle (Degrees)
SI_EF_MIN_TILTED_ANG_ORT         SRB/FLASHFlux 1/2x1/2 MinimumSolar Irradiance Tilted Surface Orientation (N/S Orientation)
PARAMETER                 JAN      FEB      MAR      APR      MAY      JUN      JUL      AUG      SEP      OCT      NOV      DEC      ANN
-END HEADER-
T2M_MAX                   26.18    28.25    30.72    32.49    31.63    28.24    26.52    26.37    25.45    25.75    25.91    25.88    27.78
T2M_MIN                   10.30    11.34    13.02    15.65    16.91    16.19    14.42    14.48    14.66    13.14    11.44    10.48    13.50
EQVLNT_NO_SUN_BLACKDAYS_MONTH 4.60  2.69     2.57     1.70     2.33     2.83     4.27     4.34     4.93     2.60     2.34     1.95     -999
SI_EF_MIN_TILTED_SURFACE_0    4.37  5.40     6.20     6.50     6.12     5.63     5.66     5.36     4.70     5.08     4.91     4.57     5.38
SI_EF_MIN_TILTED_SURFACE_18   5.14  6.06     6.52     6.39     6.03     5.67     5.66     5.17     4.76     5.52     5.77     5.56     5.69
SI_EF_MIN_TILTED_SURFACE_3    4.53  5.54     6.29     6.52     6.14     5.67     5.69     5.35     4.73     5.18     5.08     4.76     5.46
SI_EF_MIN_TILTED_SURFACE_33   5.48  6.25     6.39     5.92     5.62     5.39     5.35     4.74     4.55     5.56     6.12     6.04     5.62
SI_EF_MIN_TILTED_SURFACE_90   4.01  3.90     3.05     1.90     1.97     2.36     2.20     1.60     2.07     3.17     4.31     4.68     2.94
SI_EF_MIN_OPTIMAL             5.53  6.25     6.52     6.52     6.14     5.70     5.71     5.36     4.78     5.58     6.16     6.17     5.87
SI_EF_MIN_OPTIMAL_ANG         42.00 33.00    20.00    5.00     -6.00    -11.00   -9.00    0.00     12.00    27.00    40.00    46.00    16.00
SI_EF_MIN_TILTED_ANG_ORT      S     S        S        S        N        N        N        S        S        S        S        S        S
```

[12].

It is possible to determine that there is an annual average of 5.87 hours of minimum solar radiation equivalent to 5.87 kWh/m^2/day. For the photovoltaic calculations applying the rounding, 6 hours of solar radiation are considered.

3. A photovoltaic system installed with LED lamps.

It has a photovoltaic system installed with three 120 W solar panels giving 360 W-h between the three with a daily output of 2.16 kWh/d taking into account 6 hrs of solar radiation for this region. Figure 6 shows the module installed with the three solar panels.



Figure 6: Photovoltaic module installed in the ITSAO.



Figure 7: Battery bank.



Figure 8: Charge controller, batteries and inverter.

The LED lamp has a consumption of 30 W with a reflector installed at an approximate height of 10 m; By having 2.16 kWh/d of production, up to 7 lamps can be connected to work all night, considering a continuous operation of 10 hrs per day with the consumption of 300 W per night of each lamp. Figure 9 shows the installed lamp.



Figure 9: 30 W lamp with reflector installed.

The operational control of energy has the purpose of establishing guidelines and activities for the control of electrical energy in the ITSAO facilities applicable to all activities, processes, products and services that consume electricity.

To reduce electricity consumption to date, we have worked with the acquisition of LED lamp tubes to replace the fluorescent tubes that consume the most electrical energy; in all areas of the ITSAO the lamps have been replaced with the participation of the technological community. Figure 10 shows the LED lamps corresponding to the interior lighting of the ITSAO buildings.

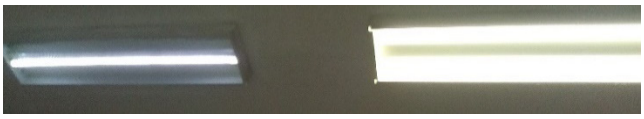


Figure 10: Lamps of LED tubes in operation.

Monthly monitoring of electricity consumption is made by comparing quantitatively the monthly consumption of the current year with the previous one so that the energy savings obtained through an annual report to the representative of the general management can be identified and analyzed annually.

Table 1 shows the monitoring of electricity consumption, in this case, corresponds to the annual report of consumption monitoring in 2016, Table 2 shows the monitoring of electricity consumption for 2017, Table 3 shows the monitoring of electricity consumption for the year 2018 and Table 4 shows the monitoring of electricity consumption for the year 2019.

Table 1: Monitoring of electricity consumption in 2016.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2016													
PERIOD	12/29/2014 TO 01/27/2015	01/27/2015 TO 02/26/2015	02/26/2015 TO 03/27/2015	03/27/2015 TO 04/28/2015	04/28/2015 TO 05/28/2015	05/28/2015 TO 06/26/2015	06/26/2015 TO 06/27/2015	06/27/2015 TO 08/27/2015	08/27/2015 TO 09/25/2015	09/25/2015 TO 10/26/2015	10/26/2015 TO 11/26/2015	11/26/2015 TO 12/28/2015	ANNUAL SAVINGS
	12/29/2015 TO 01/27/2016	01/27/2016 TO 02/26/2016	02/26/2016 TO 03/27/2016	03/27/2016 TO 04/27/2016	04/27/2016 TO 05/26/2016	05/26/2016 TO 06/27/2016	06/27/2016 TO 07/25/2016	07/25/2016 TO 08/25/2016	08/25/2016 TO 09/26/2016	09/26/2016 TO 10/25/2016	10/25/2016 TO 11/25/2016	11/25/2016 TO 12/26/2016	
MONTHLY CONSUMPTION kW-h	8960	12180	14000	12040	16800	5180	21280	14840	17640	14000	16100	10640	163660
	8260	11060	9800	11620	18340	13720	12320	15540	16240	19640	15400	10360	162300
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	7.81%	9.20%	30.00%	3.49%	-9.17%	-164.9%	42.11%	-4.72%	7.94%	40.29%	4.35%	2.63%	0.83%

Table 2: Monitoring of electricity consumption in 2017.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2017													
PERIOD	12/29/2015 TO 01/27/2016	01/27/2016 TO 02/26/2016	02/26/2016 TO 03/27/2016	03/27/2016 TO 04/27/2016	04/27/2016 TO 05/26/2016	05/26/2016 TO 06/27/2016	06/27/2016 TO 07/25/2016	07/25/2016 TO 08/25/2016	08/25/2016 TO 09/26/2016	09/26/2016 TO 10/25/2016	10/25/2016 TO 11/25/2016	11/25/2016 TO 12/26/2016	ANNUAL SAVINGS
	12/29/2016 TO 01/23/2017	01/23/2017 TO 02/23/2017	02/23/2017 TO 03/27/2017	03/27/2017 TO 04/26/2017	04/26/2017 TO 05/25/2017	05/25/2017 TO 06/26/2017	06/26/2017 TO 07/26/2017	07/26/2017 TO 08/24/2017	08/24/2017 TO 08/25/2017	08/25/2017 TO 10/25/2017	10/25/2017 TO 11/24/2017	11/24/2017 TO 12/26/2017	
MONTHLY CONSUMPTION kW-h	8260	11060	9800	11620	18340	13720	12320	15540	16240	19640	15400	10360	162300
	7420	14280	17920	15120	22540	16520	16520	15120	15400	13440	12880	9520	176680
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	10.17%	-29.11%	-82.86%	-30.12%	-22.90%	-20.41%	-34.09%	2.70%	5.17%	31.57%	16.36%	8.11%	-8.86%

Table 3: Monitoring of electricity consumption in 2018.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2018													
PERIOD	12/29/2016 TO 01/23/2017	01/23/2017 TO 02/23/2017	02/23/2017 TO 03/27/2017	03/27/2017 TO 04/26/2017	04/26/2017 TO 05/25/2017	05/25/2017 TO 06/26/2017	06/26/2017 TO 07/26/2017	07/26/2017 TO 08/24/2017	08/24/2017 TO 08/25/2017	08/25/2017 TO 10/25/2017	10/25/2017 TO 11/24/2017	11/24/2017 TO 12/26/2017	ANNUAL SAVINGS

	12/26/2017 TO 01/25/2018	01/25/2018 TO 02/23/2018	03/23/2018 TO 03/26/2018	03/26/2018 TO04/24/2018	04/24/2018 TO 05/25/2018	05/25/2018 TO 06/25/2018	06/25/2018 TO 07/26/2018	07/26/2018 TO 08/24/2018	08/24/2018 TO 09/25/2018	09/25/2018 TO 10/25/2018	10/25/2018 TO 11/26/2018	11/26/2018 TO 12/26/2018	
MONTHLY CONSUMPTION kW-h	7420	14280	17920	15120	22540	16520	16520	15120	15400	13440	12880	9520	176680
	6720	10360	16940	13160	17080	18340	16520	15400	20720	19460	14000	10640	179340
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	9.43%	27.45%	5.47%	12.96%	24.22%	-11.02%	0.00%	-1.85%	-34.55%	-44.79%	-8.70%	-11.76%	-1.91%

Table 4: Monitoring of electricity consumption in 2019.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2019													
PERIOD	12/26/2017 TO 01/25/2018	01/25/2018 TO 02/23/2018	03/23/2018 TO 03/26/2018	03/26/2018 TO04/24/2018	04/24/2018 TO 05/25/2018	05/25/2018 TO 06/25/2018	06/25/2018 TO 07/26/2018	07/26/2018 TO 08/24/2018	08/24/2018 TO 09/25/2018	09/25/2018 TO 10/25/2018	10/25/2018 TO 11/26/2018	11/26/2018 TO 12/26/2018	ANNUAL SAVINGS
	12/26/2018 TO 01/26/2019	01/26/2019 TO 02/22/2019	02/22/2019 TO 03/25/2019	03/25/2019 TO 04/24/2019	04/24/2019 TO 05/27/2019	05/27/2019 TO 06/24/2019	06/24/2019 TO 06/25/2019	06/25/2019 TO 08/24/2019	08/24/2019 TO 08/25/2019	08/25/2019 TO 10/25/2019	10/25/2019 TO 11/26/2019	11/26/2019 TO 12/26/2019	
MONTHLY CONSUMPTION kW-h	6720	10360	16940	13160	17080	18340	16520	15400	20720	19460	14000	10640	179340
	7140	9940	20300	15960	22120	17920	20720						114100
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	-6.25%	4.05%	-19.83%	-21.28%	-29.51%	2.29%	-25.42%	100.00%	100.00%	100.00%	100.00%	100.00%	36.38%

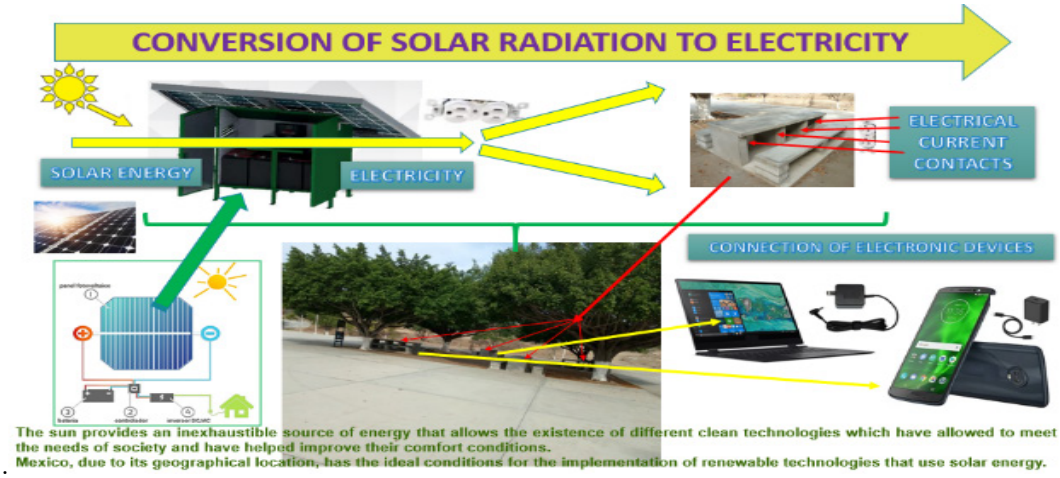


Figure 11: Infographic of the photovoltaic system for the ITSAO planters.

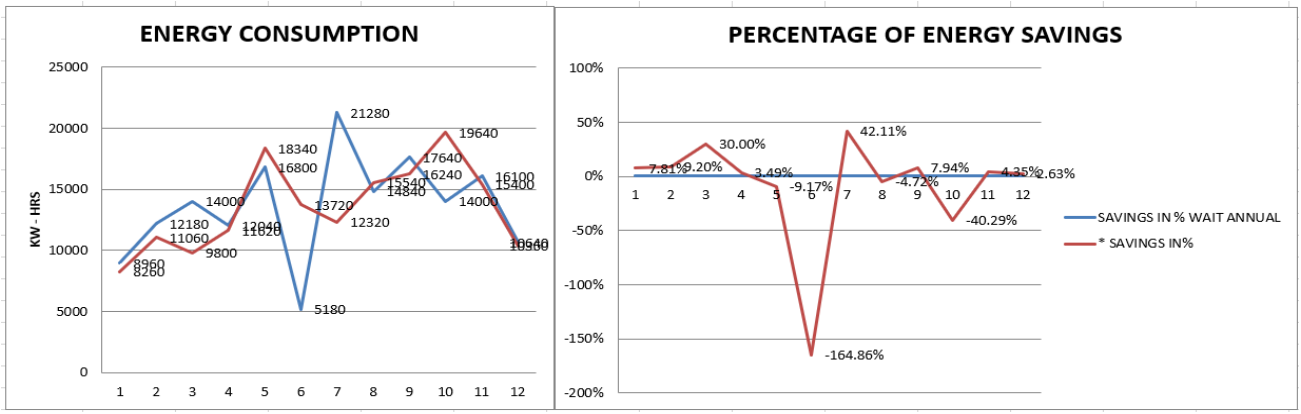


Figure 12: Graph of consumption in 2016 compared to 2015 and Graph of the reduction percentual for electricity consumption in 2016 compared to 2015.

One of the activities to be implemented is a photovoltaic system with the purpose of supplying electricity to the tables located in the planters of the esplanade; the conversion of solar energy to electricity will be applied so that the technology community can use it to power their electronic devices such as laptops and mobile devices while they rest or gather in the flower beds. Figure 11 shows an infographic that describes the operation and location of the photovoltaic elements and electrical power outlets

5. Results of the electric follow-up corresponding to the reports of 2016, 2017, 2018 and the current follow-up in 2019.

Figure 12 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2016 compared to 2015.

Figure 13 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2017 compared to 2016. In this case, electricity consumption increased due to the installation of air conditioning equipment in salons that did not have this electrical equipment.

Figure 14 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2018 compared to 2017. In this case, electricity consumption increased due to the use of air conditioning equipment in salons that did not have this electrical equipment.

Figure 15 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2019 compared to 2018. In this case, to date only monitoring is carried out until July the follow-up is in process.

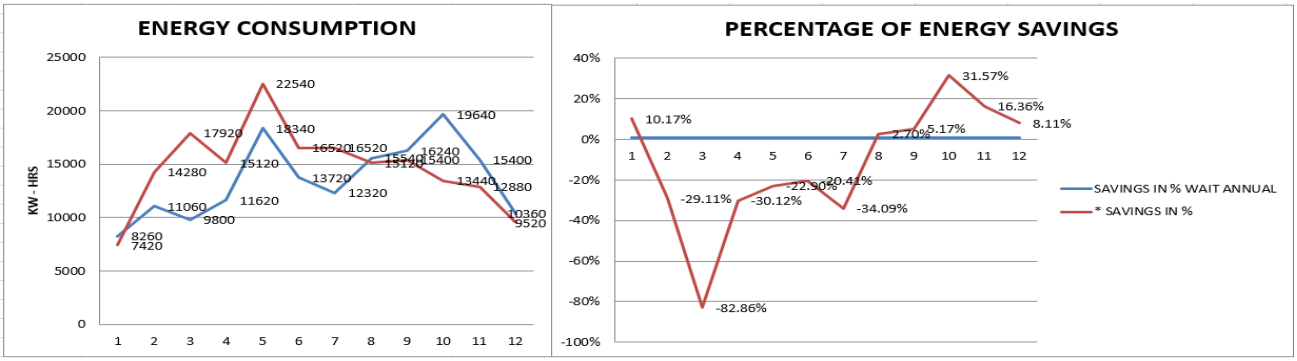


Figure 13: Graph of consumption in 2017 compared to 2016 and Graph of the reduction percentual for electricity consumption in 2017 compared to 2016.

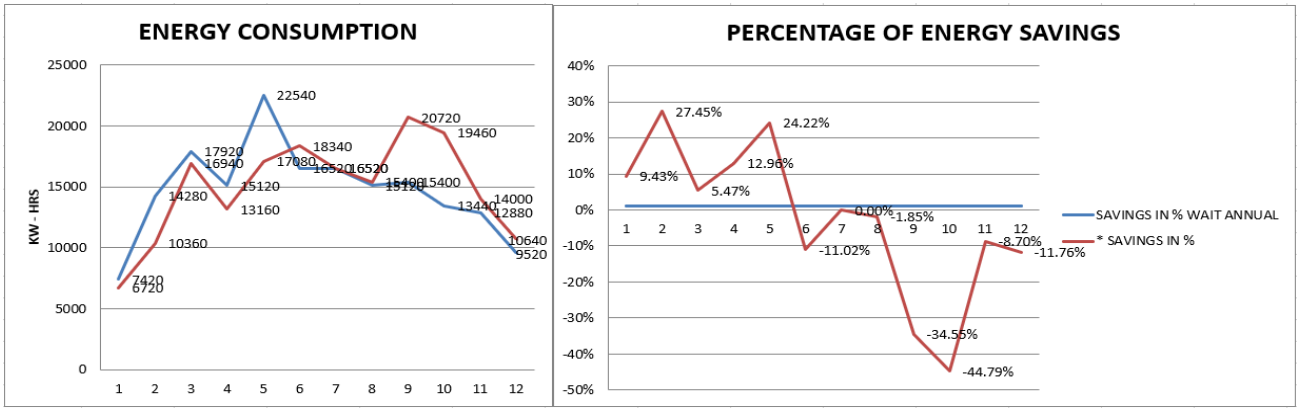


Figure 14: Graph of consumption in 2018 compared to 2017 and Graph of the reduction percentual for electricity consumption in 2018 compared to 2017.

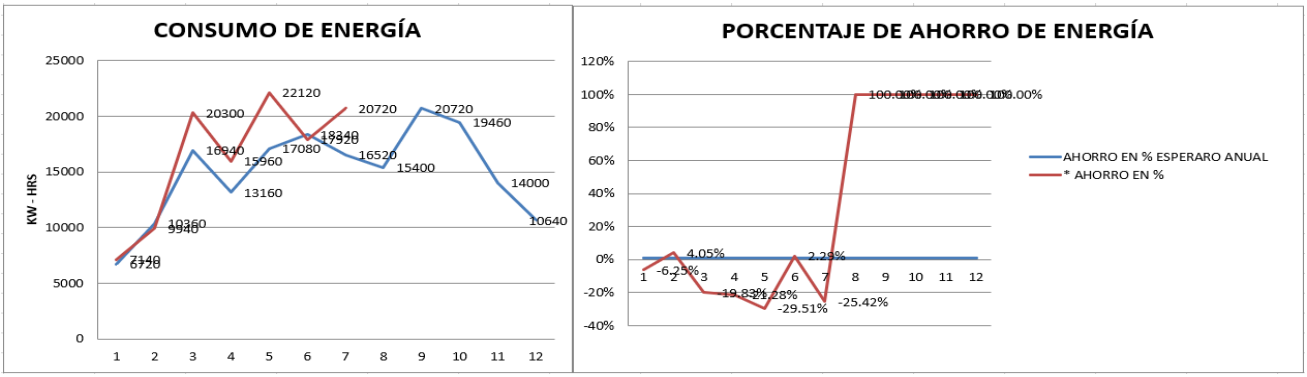


Figure 15: Graph of consumption in 2019 compared to 2018 and Graph of the reduction percentual for electricity consumption in 2019 compared to 2018.

The activities implemented are not only part of the Integrated Management System of the ITSAO, but they are also part of the academic and professional training of the students and the intervention of the Electronic Engineering Academy. They are also part of the research area corresponding to the research line "Renewable Energy and Energy Efficiency".

6. Conclusions.

With the execution of the activities carried out, the emphasis is mainly on the use of clean energies such as solar as an alternative means to obtain electricity for the external lighting of the ITSAO facilities, with the implementation of LED lamps it is possible to reduce the electricity consumption with sustainable approach; the entire technological community will be given the opportunity to use photovoltaic energy for their outdoor electronic devices. The students manage to carry out sustainable development practices, research foundations, research workshops and optoelectronics to strengthen their academic training. The activities implemented to reflect a good result by observing the graphs in the reduction of electricity consumption.

Conflict of Interest.

The authors declare no conflict of interest.

Acknowledgement.

The research was carried out with the support of the Electronic Engineering area and the ITSAO technological community within the framework of the implementation of the Environmental Management System and the research area.

References.

[1] B. T. Jeremias, E. M. Yoxkin, J. R. T. Juan, y R. G. Francisco, "Photovoltaic Energy, LED Technology and Energy Efficiency in the ITSAO", en *2018 XXXI INTERNATIONAL SUMMER MEETING ON POWER AND INDUSTRIAL APPLICATIONS (RVP-AI)*, 2018, pp. 19-22.

[2] J. C. Cruz-Ardila, J. C. Cardona-Gómez, y D. M. Hernández-Porras, "Aplicación electrónica para el ahorro de energía eléctrica utilizando una energía alternativa", *Entramado*, vol. 9, n.º 2, pp. 234-248, jul. 2013.

[3] A. N. Sperber, A. C. Elmore, M. L. Crow, y J. D. Cawlfield, "Performance evaluation of energy efficient lighting associated with renewable energy applications", *Renew. Energy*, vol. 44, pp. 423-430, ago. 2012.

[4] J. Fresner, F. Morea, C. Krenn, J. Aranda Uson, y F. Tomasi, "Energy efficiency in small and medium enterprises: Lessons learned from 280 energy audits across Europe", *J. Clean. Prod.*, vol. 142, pp. 1650-1660, ene. 2017.

[5] Steven Fawkes, Kit Oung, y David Thorpe, *Best Practices and Case Studies for Industrial Energy Efficiency Improvement. An Introduction for Policy Makers*. DTU Library, 2016.

[6] J. de Beer, *Potential for Industrial Energy-Efficiency Improvement in the Long Term*. Springer Netherlands, 2000.

[7] Soraya Navarro Rayas, José Antonio González Moreno, y César M. López Andrade, "Implementación de un sistema fotovoltaico para la alimentación de un edificio de usos múltiples". 2016.

[8] S. Medio, "Photovoltaic Design Integration at Battery Park City, New York", *Buildings*, vol. 3, n.º 2, pp. 341-356, abr. 2013.

[9] M. Wichliński y J. Staszkiewicz, "Photovoltaic installation for a roofed car park in Egypt as an example of effective use of solar energy", *E3S Web Conf.*, vol. 49, p. 00127, 2018.

[10] C.-G. Haba, "Monitoring photovoltaic parks for damage prevention and optimal operation", en *2017 International Conference on Electromechanical and Power Systems (SIELMEN)*, Iasi, 2017, pp. 321-326.

[11] "Google Maps", *Google Maps*. [Online]. Available in: https://www.google.com.mx/maps/place/Instituto+Tecnol%C3%B3gico+Superior+de+Acatl%C3%A1n+de+Osorio/@18.2248389,-98.0409622,17z/data=!3m1!4b1!4m5!3m4!1s0x85cf4182fb37cc4f:0xf1a9691f53f1acf!8m2!3d18.2248389!4d-98.0387735. [Accessed: Mar 03, 2019].

[12] "POWER Data Access Viewer". [Online]. Available in: https://power.larc.nasa.gov/data-access-viewer/. [Accessed: Mar 03, 2019].

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 237-243 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Application of The Half-Sweep Egsor Iteration for Two-Point Boundary Value Problems of Fractional Order

Rostang Rahman[1,*], Nur Afza Mat Ali[1], Jumat Sulaiman[1], Fatihah Anas Muhiddin[2]

[1]*Mathematics with Economics Programme, Faculty of Science and Natural Resources, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

[2]*Faculty of Computing and Mathematical Sciences, Universiti Teknologi MARA Sabah, 88997 Kota Kinabalu, Sabah, Malaysia*

ARTICLE INFO

Article history:
Received: 03 January, 2019
Accepted: 07 March, 2019
Online: 29 March, 2019

Keywords:
4HSEGSOR
Caputo's fractional operator
Finite difference method
Two-point boundary value problem

ABSTRACT

The point of this study is to explore and elucidating the performance of the four-point Half-Sweep EGSOR (4HSEGSOR) iterative method to solve fractional two-point boundary value problems by using Caputo's fractional operator and family of finite differences (FD) schemes. To apply the iterative methods, linear system needs to be constructed via the discretization process with fractional order to get the approximation equation of the linear fractional two point boundary value problem by using the Caputo's derivative operator. Then the generated linear system has been solved using the proposed 4HSEGSOR iterative method. In the addition, the formulation and application of the 4HSEGSOR method to solve the problems are also presented. Three numerical examples and comparison are used to illustrate with tested FSSOR and HSSOR methods. The numerical results reveals the effectiveness of 4HSEGSOR method compared with tested iterative methods.

1. Introduction

This paper inspired from numerous researchers in science and engineering where there have been discuss the steady-state problems which refer to Fractional Boundary Value problems (FBVPs) since this problem become more attractive in the recent years in many applications such as mathematics, engineering, economy, and other fields [1,2]. Since the rapid growth of computer technology, the numerical techniques are used to solve the large size problem. Following that, there are many researcher have been proposed numerical methods to solve the FBVPs such as Fix et al [3] applied the Least Squares Finite-element methods, Li et al [4] applied the Reproducing Kernel method, Odibat et al [5] applied the Modified Homotopy Perturbation method, and Diethelm et al [6] applied the Extrapolation method. For instance, Sunarto et al [7] started the study on the finite different method to solve the unsteady-state problems using application of the SOR method. In addition, Sunarto et al [8] extended the study with the application of the full-sweep AOR iteration concept to the same problem. Furthermore, this study extend the study to steady-state problems and focus to two-point Fractional BVPs. The problem will be solved numerically and represented as follows [9]:

$$d(x)\frac{\partial^{\beta}v(x)}{\partial x^{\beta}}+a(x)\frac{\partial^{2}v(x)}{\partial x^{2}}+b(x)\frac{\partial v(x)}{\partial x}+c(x)v(x)=F(x),\ x\in[\gamma,\mu] \qquad (1)$$

with the Dirichlet boundary conditions,

$$u(\gamma)=\gamma_0, \qquad u(\mu)=\mu_1.$$

where $a(x)$, $b(x)$, $c(x)$ and $d(x)$ are known functions or constants respectively. Then the derivative term of fractional order is $\frac{\partial^{\beta}v(x)}{\partial x^{\beta}}$ and the parameter β represent fractional order with the range in this study consider as $1\le\beta\le 2$. To get the approximate solution of two-point FBVPs must be discretized to form an approximation equations. Based on the family of FD schemes and Caputo's fractional derivative operator, the approximation equations use to construct a linear system at each point.

For solving the linear systems which have large and sparse, there are several concept iterative methods from previous researchers have been discussed [10,11]. Other than that, Rahman et al. [12] studied the numerical solution of two-point FBVPs based on the iterative method where SOR method has been used

*Corresponding Author: Rostang Rahman, Mathematics with Economics Programme, Faculty of Science and Natural Resources, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia, Contact No:+60 14 654 1569, rostang054@gmail.com

and AGE method in studied Rahman et al. [13]. From the studied, this paper focus to expand the iterative method to the half-sweep scheme. The idea come from Abdullah (1991) who initiated of Half-sweep iteration which is ones of the common utilized use iterative techniques to solve any linear systems. Following that, in this paper proposed 4HSEGSOR iterative method as linear solver to increase the iteration convergence rate so as to solve the two-point FBVPs on the Caputo's FD approximation equation. To demonstrate the capability of the 4HSEGSOR method, the implementation of the Half-Sweep SOR (HSSOR) iterative method also considered as comparison method in this study and Full-Sweep SOR (FSSOR) iterative methods react as a control method.

After that, the approximation equation will be construct for the problem (1) based on the HS second-order Caputo's fractional operator via fractional derivative theory. Therefore, let discuss some definitions that can be applied to construct the approximation equation towards problem (1) by using the Caputo's fractional derivative operator.

Definition 1 [14]. The definition of fractional integral operator for Riemann-Liouville, ${}^{*}J^{\beta}$ of order $-\beta$ given by:

$$ {}^{*}J^{\beta}F(x)=\frac{1}{\Gamma(\beta)}\int_{0}^{x}(x-t)^{\beta-1}F(t)dt,\quad \beta>0,\quad x>0 \tag{2}$$

Definition 2 [14]. The definition of Caputo's fractional derivative operator, D^{β} of order $-\beta$ given by:

$$ D^{\beta}F(x)=\frac{1}{\Gamma(m-\beta)}\int_{0}^{x}\frac{F^{(m)}(t)}{(x-t)^{b-m+1}}dt,\quad \beta>0 \tag{3}$$

Where parameter β refer to $m-1<\beta\le m,\ m\in N,$ and $x>0.$

Definition 3 [15]. The function of Gamma denoted by $\Gamma(\beta)$ is a generalization of factorial function for complex argument with positive real part it is defined as:

$$ \Gamma(\beta)=\int_{0}^{\infty}x^{t-1}e^{-x}dx \tag{4}$$

To solve the problem (1) as mentioned in early section, we get numerical approximations by using fd scheme via the Caputo's derivative definition with Dirichlet boundary conditions and fractional derivative operator. In addition to that, the existence and uniqueness of the solution for the problem (1) can be referred and discussed by Diethelm in 2002 in the study Analysis of Fractional Differential Equations [16].

Theorem 4 (existence) [16]. Assume that $D:=\left[0,x^{*}\right]\times\left[x_0^{(0)}-\eta,\ x_0^{(0)}+\gamma\right]$ with some $x^{*}>0$ and let the function $F:D\to\mathbb{R}$ be continuous. Furthermore, define $x:=\min\left\{x^{*},\left(h\Gamma(\beta+1)/\|f\|_{\infty}\right)^{1/\beta}\right\}.$ Then there exists a function $v:[0,x]\to\mathbb{R}$ solving the initial value problem (1).

Theorem 5 (uniqueness) [16]. Assume that $D:=\left[0,x^{*}\right]\times\left[x_0^{(0)}-h,\ x_0^{(0)}+h\right]$ with the $x^{*}>0$. Besides that, let the function $F:D\to\mathbb{R}$ be bounded on D and fulfill a Lipschitz condition the second variable.

Before solving the problem (1), the solution domain of the problem has been confined to the finite domain $a\gamma\le x\le\mu$, with $1\le\beta\le 2$ whereas the parameter β refers to the fractional order derivative. In order to solve the problem (1), let consider the Caputo's fractional derivative of order β as:

$$ \frac{d^{\beta}v(x_i)}{dx^{\beta}}=\frac{1}{\Gamma(2-\beta)}\int_{0}^{x_n}\frac{dv^{2}(s)}{dx^{2}}\left(x_n-s\right)^{1-\beta}ds \tag{5}$$

2. Half-sweep Approximation Equation

Before discretizing problem (1), let the solution domain of the problem be partitioned consistently to facilitate in discretizing process. In order to discretize, we consider some positive integers N which is the number of subintervals $[\gamma,\mu]$ and then the length of grid size are defined as.

$$ h=\Delta x=\frac{\gamma-\eta}{N} \tag{6}$$

and

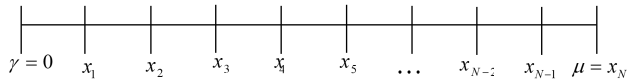
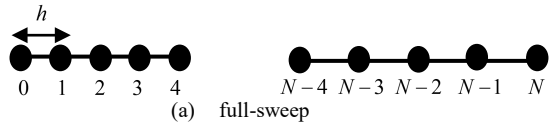
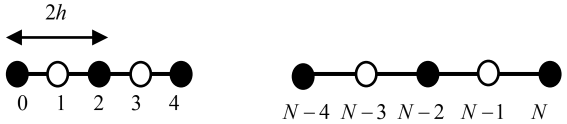


Figure 1: The distribution of interior node points over the finite grid network.

From the Figure 1, we develop the consistently grid network of the solution domain where the polar grid of the solution domain will be shown as $x_i=\gamma+ih,\ i=0,1,2,\ldots,N$ and the values of function $v(x)$ at point x_i are given as $v(x_i)=v_i$. As mentioned in Section 1, the half-sweep concept is imposed to improve the convergence rate. Here the different values of h from the current point to next point between full-sweep and half-sweep is shown in Figures 2 which is Figure 2(a) shows the implementation of the full-sweep iteration which all interior nodes are considered one by one point in which its distance for each point is h. However, for half-sweep iteration with its distance $2h$ can be seen in Figure 2(b).



(a) full-sweep



(b) half-sweep

Figure 2: The distribution of uniform grid network

After that, to obtain the approximation equation of problem (1), now the equation (5) will be consider to derive based on the

formulation of Caputo's fractional derivative operator which can be gotten by a straightforward quadrature formula as follows.

$$\begin{aligned}\frac{d^{\beta}v(x_i)}{dx^{\beta}} &= \frac{1}{\Gamma(2-\beta)}\int_0^{x_n}\frac{dv^2(s)}{dx^2}\left(x_n-s\right)^{1-\beta}ds \\ &= \frac{1}{\Gamma(2-\beta)}\sum_{j=0}^{i-1}\int_{jh}^{(j+1)h}\left(\frac{v_{i-j+1}-2v_{i-j}+v_{i-j-1}}{h^2}\right)\left(x_nh-s\right)^{1-\beta}\partial s \\ &= \frac{1}{\Gamma(2-\alpha)}\sum_{j=0}^{i-1}\left(\frac{v_{i-j+1}-2v_{i-j}+v_{i-j-1}}{h^2}\right)x\int_{jh}^{(j+1)h}\left(x_nh-s\right)^{1-\beta}ds \\ &= \frac{h^{-\beta}}{\Gamma(3-\beta)}\sum_{j=0}^{i-1}\left(v_{i-j+1}-2v_{i-j}+v_{i-j-1}\right)\left((j+1)^{2-\beta}-j^{2-\beta}\right)\end{aligned} \tag{6}$$

Then the discrete approximation of equation (6) can be given as

$$\frac{d^{\beta}v(x_i)}{dx^{\beta}} = \sigma_{\beta,h}\sum_{j=0}^{i-1}g_j^{\beta}\left(v_{i-j+1}-2v_{i-j}+v_{i-j-1}\right)+O(h) \tag{7a}$$

Where can be define by

$$\sigma_{\beta,h} = \frac{h^{-\beta}}{\Gamma(3-\beta)} \tag{7b}$$

and

$$g_j^{\beta} = \left(i+1\right)^{1-\beta}-i^{1-\beta}. \tag{7c}$$

However, this study don't use the operator (7) which is shown as the full-sweep Caputo's fractional derivative since this study focus to the half-sweep concept. Refer to the study of Sonarto et al [17] over the problem, we have propose the half-sweep concept to the Caputo's fractional derivative operator based on the operator (7) as follows:

$$\frac{\partial^{\beta}v(x)}{\partial x^{\beta}} \cong \sigma_{\beta,k}\sum_{j=0}^{N}g_j^{\beta}(v_{i-j+2}-2v_{i-j}+v_{i-j-2}) \tag{8a}$$

where we have the following expressions

$$\sigma_{\beta,h} = \frac{2h^{-\beta}}{\Gamma(3-\beta)} \tag{8b}$$

and

$$g_j^{\beta} = \left(\frac{i}{2}+1\right)^{2-\beta}-\frac{i}{2}^{2-\beta}. \tag{8c}$$

As mentioned in Section 1, using the second-order and first-order half-sweep central difference discretization scheme and half-sweep Caputo's fractional derivative operator (8), the second-order Half-Sweep Caputo's fd approximation equation for problem (1) given as:

$$\begin{aligned}d_i\sigma_{\beta,h}\sum_{j=0}^{j-2}g_{\beta}^{j}(v_{i-j+2}-2v_{i-j}+v_{i-j-2})+a_i(\frac{v_{i+2}-2v_i+v_{i-2}}{4h^2}) \\ +b_i(\frac{v_{i+2}-v_{i-2}}{4h})+c_i(v_i)=F_i\end{aligned} \tag{9}$$

Now, the approximation (9) is known as the half-sweep FD approximation equation it means that the value for each point considers at $i=2,\ 4,\ 6,\ \ldots,\ N-2$ refer to the figure 2(b) which show the distribution of grid network. By simplifying (9), the approximation equation can be written as:

$$a_i^{*}v_{i-2}+b_i^{*}v_i+c_i^{*}v_{i+2}-R_i=F_i, \qquad i=2,4,6 \tag{10}$$

Let us define

$$\left.\begin{aligned}&\lambda_i=\mathrm{d}_i\sigma_{\beta,h}, \quad a_i^{*}=\alpha_i-\mu_i-\lambda_i, \quad b_i^{*}=c_i-2\alpha_i+2\lambda_i, \\ &c_i^{*}=\alpha_i+\mu_i-\lambda_i, \quad R_i=\lambda_i\sum_{j=1}^{j-2}g_j^{\beta}(u_{i-j+2}-2u_{i-j}+u_{i-j-2})\end{aligned}\right\} \tag{11}$$

Again, besides the values of $i=2,4$, and 6 , the simplify of approximation equation (9) should be appeared for $i=8,10,12,\ldots,N-2$. To simplify the approximation equation, the second-order Half-Sweep Caputo's FD approximation equation can be expressed as follows:

$$R_i^{*}+p_iv_{i-6}+q_iv_{i-4}+r_iv_{i-2}+s_iv_i+z_iv_{i+2}=F_i, \quad i=8,10,12,\ldots,N-2 \tag{12}$$

where

$$\left.\begin{aligned}&q=\lambda_ig_2^{\beta}-2\lambda_ig_4^{\beta}, \quad p=\lambda_ig_4^{\beta}, \quad r=a_i^{*}-2\lambda_ig_2^{\beta}+\lambda_ig_4^{\beta}, \\ &s=b_i^{*}+\lambda_ig_2^{\beta}, \quad R_i^{*}=\lambda_i\sum_{j=6}^{j-2}g_j^{\beta}(u_{i-j+2}-2u_{i-j}+u_{i-j-2}), \quad z=c_i^{*}.\end{aligned}\right\} \tag{13}$$

According to equations (10) and (12), the linear system can be constructed in matrix form to facilitate in solving system of second-order half-sweep Caputo's approximation equations. As a result, the system of linear equations in the form can be given as:

$$A\underline{v}=\underline{f} \tag{14}$$

where can be define,

$$A=\begin{bmatrix}s_2 & z_2 & 0 & 0 & 0 & 0 & 0\\ r_4 & s_4 & z_4 & 0 & 0 & 0 & 0\\ q_6 & r_6 & s_6 & z_6 & 0 & 0 & 0\\ p_8 & q_8 & r_8 & s_8 & z_8 & 0 & 0\\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0\\ 0 & 0 & p_{N-4} & q_{N-4} & r_{N-4} & s_{N-4} & z_{N-4}\\ 0 & 0 & 0 & p_{N-2} & q_{N-2} & r_{N-2} & s_{N-2}\end{bmatrix}_{\left(\frac{N}{2}-1\right)\times\left(\frac{N}{2}-1\right)} \tag{15}$$

and

$$\underline{v}=\begin{bmatrix}v_2 & v_4 & v_6 & \cdots & v_{N-4} & v_{N-2}\end{bmatrix}^T \tag{16}$$

and

$$\underline{F}=\begin{bmatrix}F_2+p_2v_2\\ F_4+q_4v_4\\ F_6+r_6v_6\\ F_8+R_i\\ \vdots\\ F_{N-4}+R_{N-4}\\ F_{N-2}+R_{N-2}-z_{N-2}v_N\end{bmatrix} \tag{17}$$

3. Iterative Method

From previous discussion in Section 2 and refer to the linear system equation (14), the formulation of 4HSEGSOR iterative method will be discussed in this section to test the examples than record the numerical results of linear system. Based on the characteristics of the coefficient matrix A, it's showed the matrix have parse and large-scale. Following to the characteristics, the iterative methods are suitable option to solve the linear system [10]. Besides to improve the consequence rate of iterative method

for solving the linear system numerically, there are iterative methods have been proposed from previous researchers such as Jacobi, GS, and SOR methods. However, this study considers the application of 4HSEGSOR method as linear solver as mentioned in early this section.

Particularly, the 4HSEGSOR method is essentially the extension to half-sweep concept for the Four-point Explicit Group SOR (4EGSOR) iterative method which is combination between EG method developed by Evans [18] and SOR method developed by Young [10]. From previous study, Saudi and Sulaiman [19] have been discuss the explicit group to apply into the robot path planning and elucidated the EGSOR iterative method using Nine-Point Laplacian [20] in order to improve the consequence rate of iterative method in the study.

The aim of EGSOR method is to reduce the computational time of the convergence rate by uses small constant size groups strategy based on mesh point and weighted parameter "ω" with its range value given as $\omega \in [1,2)$. However, the main propose of the half-sweep iteration concept in this study is to reduce the computational complexities during iteration process. Following to that, the linear system was divided into several completed small group of four points ($p=4$) start with $i=8,16,24,\ldots,N-8$. However, for the first three points $i=2,4,$ and 6 can be treated as ungroup case [21]. Since this paper deals with application 4HSEGSOR iterative method for solving the linear system (14), it can be stated as:

$$\begin{bmatrix} v_i \\ v_{i+2} \\ v_{i+4} \\ v_{i+6} \end{bmatrix} = (1-\omega)\begin{bmatrix} v_i \\ v_{i+2} \\ v_{i+4} \\ v_{i+6} \end{bmatrix} + \omega \begin{bmatrix} s_i & z_i & 0 & 0 \\ r_{i+2} & s_{i+2} & z_{i+2} & 0 \\ q_{i+4} & r_{i+4} & s_{i+4} & z_{i+4} \\ p_{i+6} & q_{i+6} & r_{i+6} & s_{i+6} \end{bmatrix}^{-1} \begin{bmatrix} S_2 \\ S_4 \\ S_6 \\ S_8 \end{bmatrix}, \qquad (18)$$

for $i=8,16,24,\ldots,N-8$. To show the illustration, implementation of the 4HSEGSOR iterative method can be shown in Figure 3. Again the first three points, $i=2,4,$ and 6 can be stated as ungroup case in which these point can be used to construct the 3-point HSEGSOR method.

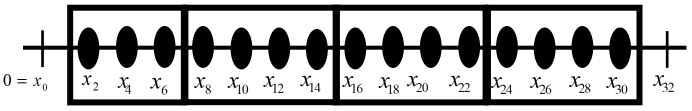


Figure 3: Implementation of the half-sweep 4EGSOR iterative method at solution domain m=32.

Based on Figure 3, Algorithm 1 is the implementation of 4HSEGSOR method that be summarized where the optimum values for the parameters, "ω" depends on the minimum number of iterations from several executions. The tolerance error set as $\varepsilon = 10^{-10}$.

Algorithm 1: 4HSEGSOR iteration

i. Set the value of parameters, $\underline{v}^{(0)} \leftarrow 0, \varepsilon \leftarrow 10^{-10}$
ii. Calculate the coefficient matrix, A.
iii. Calculate the vector, $\underline{F}$.
iv. Calculate the value of $\underline{v}^{(k+1)}$
    a) For $i=2$ calculate an ungroup cases.
    b) For $i=8,\ 16,\ 24,\ \ldots,\ N-8,$ calculate again equation (10).
v. Perform the convergence test, $\left|\underline{v}^{(k+1)} - \underline{v}^{(k)}\right| \le \varepsilon = 10^{-10}$. If yes, move to step (vi). Otherwise repeat the step (iv).
vi. Display approximate solution.

4. Numerical Experiment

This paper demonstrates three examples of numerical experiments for the comparison of two-point FBVPs where the exact solution for each example already given at $\beta = 1.50$. Besides that, this paper considers others two different values to analyse the performance of Caputo's Fractional operator in which $\beta = 1.25$ and $\beta = 1.75$ refer to Section 1 the range parameter "β" is $1 \le \beta \le 2$. Other than that, three different iterative methods will be implemented such as FSSOR, HSSOR and 4HSEGSOR methods. Beyond that point, considered such as number of iterations (K), computational time measured in second (s) and maximum absolute error (error) as measurement parameters in the numerical experiments need to be recorded while the convergence test considered the tolerance error which is fixed as $\varepsilon = 10^{-10}$. The following are three examples of numerical experiments for problem (1).

Example 1

By considering the fractional two point BVP below [22]

$$\frac{\partial^{\beta} v(x)}{\partial x^{\beta}} + \frac{\partial^2 v(x)}{\partial x^2} + v(x) = x^2 + x + \frac{4}{\sqrt{\pi}}\sqrt{x} + 3, \qquad 0 \le x \le 1 \qquad (19)$$

With the subject boundary condition as $v(0)=1,$ and $v(1)=3$, and the exact solution for the problem given by $v(x) = x^2 + x + 1$ when the value of $\beta = 1.50$.

Example 2

By considering the fractional two-point BVP below [23]

$$\frac{\partial^2 v(x)}{\partial x^2} + \frac{\partial^{\beta} v(x)}{\partial x^{\beta}} + v(x) = \frac{15}{4}x^{0.5} + \frac{15}{8}\sqrt{\pi}x + x^{2.5} + 1, \qquad 0 \le x \le 1 \qquad (20)$$

With the subject boundary condition as $v(0)=1,$ and $v(1)=2$, and the exact solution for the problem given by $v(x) = x^{2.5} + 1$ when the value of $\beta = 1.50$.

Example 3

By considering the fractional two-point BVP below [23]

$$a\frac{\partial^{\beta} v(x)}{\partial x^{\beta}} + b\frac{\partial^2 v(x)}{\partial x^2} + cv(x) = 4\sqrt{\frac{x}{\pi}} + x^2 + 2, \qquad 0 \le x \le 1 \qquad (21)$$

With the subject boundary condition as $v(0)=0,$ and $v(1)=1$, and the exact solution for the problem given by $v(x) = x^2$ when the value of $\beta = 1.50$.

The implementation for the numerical experiments consider the C programing language as a tools. Following to that, all the

Table 1. Numerical results on iterative methods for problem 1.

M	Method	$\beta = 1.25$			$\beta = 1.50$			$\beta = 1.75$		
		K	Time	Max Error	K	Time	Max Error	K	Time	Max Error
128	FSSOR	751	0.10	2.1230e-02	769	0.13	1.3915e-05	1130	0.12	1.7118e-02
	HSSOR	385	0.02	2.1242e-02	320	0.01	1.3887e-05	385	0.01	1.7188e-02
	HS4EGSOR	132	0.01	2.1242e-02	114	0.01	1.3889e-05	125	0.01	1.7188e-02
256	FSSOR	1483	0.78	2.1228e-02	2051	0.77	1.3924e-05	3547	1.25	1.7081e-02
	HSSOR	765	0.08	2.1230e-02	769	0.09	1.3915e-05	1130	0.10	1.7118e-02
	HS4EGSOR	264	0.02	2.1230e-02	239	0.03	1.3917e-05	285	0.04	1.7118e-02
512	FSSOR	2975	5.62	2.1229e-02	5603	10.68	1.3912e-05	11600	15.90	1.7063e-02
	HSSOR	1522	0.83	2.1228e-02	2130	1.14	1.3922e-05	3525	1.88	1.7082e-02
	HS4EGSOR	512	0.20	2.1228e-02	551	0.19	1.3929e-05	691	0.25	1.7082e-02
1024	FSSOR	6020	33.26	2.1230e-02	15229	58.89	1.3862e-05	11600	199.24	1.7054e-02
	HSSOR	2975	6.25	2.1229e-02	5603	11.70	1.3912e-05	5700	24.29	1.7063e-02
	HS4EGSOR	1005	1.34	2.1229e-02	1309	1.77	1.3933e-05	1883	2.49	1.7063e-02
2048	FSSOR	11785	439.15	2.1230e-02	41071	1281.11	1.3708e-05	123730	1717.04	1.7048e-2
	HSSOR	6020	49.89	2.1230e-02	15229	125.68	1.3862e-05	38179	313.63	1.7054e-02
	HS4EGSOR	1973	10.25	2.1230e-02	3189	16.57	1.3928e-05	5465	28.20	1.7054e-02

Table 2. Numerical results on iterative methods for problem 2.

M	Method	$\beta = 1.25$			$\beta = 1.50$			$\beta = 1.75$		
		K	Time	Max Error	K	Time	Max Error	K	Time	Max Error
128	FSSOR	759	0.02	3.2817e-05	769	0.04	8.8991e-04	1110	0.03	3.2013e-02
	HSSOR	382	0.01	3.1865e-02	320	0.01	1.7051e-03	385	0.01	3.2604e-02
	HS4EGSOR	129	0.01	3.1865e-02	112	0.01	1.7051e-03	125	0.01	3.2604e-02
256	FSSOR	1493	0.16	3.3317e-02	2039	0.24	4.6070e-04	3384	0.40	3.1690e-02
	HSSOR	741	0.03	3.2822e-02	769	0.05	8.8708e-04	1110	0.05	3.2010e-02
	HS4EGSOR	259	0.02	3.2822e-02	239	0.02	8.8708e-04	281	0.03	3.2010e-02
512	FSSOR	2826	1.25	3.3570e-02	5478	2.39	2.4094e-04	11330	4.93	3.1521e-02
	HSSOR	1493	0.29	3.3317e-02	2039	0.37	4.6019e-04	3384	0.60	3.1690e-02
	HS4EGSOR	492	0.09	3.3317e-02	541	0.10	4.6019e-04	693	0.15	3.1690e-02
1024	FSSOR	5813	9.88	3.3699e-02	14856	25.44	1.2919e-04	37258	63.29	3.1433e-02
	HSSOR	2826	1.91	3.3570e-02	5478	3.61	2.4085e-04	11330	7.51	3.1521e-02
	HS4EGSOR	1004	0.70	3.3570e-02	1283	0.90	2.4087e-04	1845	1.27	3.1521e-02
2048	FSSOR	11427	77.53	3.3764e-02	40209	271.72	7.2583e-05	121355	820.21	3.1387e-02
	HSSOR	5806	15.14	3.3699e-02	14856	38.43	1.2917e-04	37258	96.54	3.1433e-02
	HS4EGSOR	1975	5.30	3.3699e-02	3121	8.34	1.2924e-04	5349	14.32	3.1433e-02

results of numerical experiments have been recorded based on the iterative methods which are FSSOR, HSSOR and 4HSEGSOR methods in Tables 1, 2, and 3 with three different values of β as mentioned in early section. For the grid sizes have used at five different values where the values of m = 128, 256, 512, 1024, and 2048.

5. Discussions of result

Through numerical experiments results from Tables 1, 2, and 3 by imposing the comparison between HSSOR and 4HSEGSOR iterative methods with FSSOR method react as control method as discuss in Section 1 with three different values of parameter $\beta = 1.25$, 1.50, and 1.75 , it is conspicuously that number of iterations at Table 1 for HSSOR and 4HSEGSOR iterative method have declined. The result showed the number of iterations when the parameter set to $\beta = 1.50$, at size 128 declined from 769 for FSOR to 320 for HSSOR. Again, the number of iterations declined from 320 to 114 for 4HSEGSOR. It's the same situation for another size the number of iteration has declined. To show more clearly, the percentage of number iterations declined approximately by 82.20% –83.31%, 85.18% – 92.24%, and 83.77% – 95.58% which corresponds to the HSSOR iterative method. For execution time, the result showed 4HSEGSOR method reduce the computational complexity. The execution time when the

Table 3. numerical results on iterative methods for problem 3.

M	Method	$\beta = 1.25$			$\beta = 1.50$			$\beta = 1.75$		
		K	Time	Max Error	K	Time	Max Error	K	Time	Max Error
128	FSSOR	712	0.04	2.1230e-02	768	0.07	1.3915e-05	1031	0.06	1.7118e-02
	HSSOR	385	0.01	2.1242e-02	319	0.01	1.3889e-05	378	0.01	1.7188e-02
	HS4EGSOR	126	0.00	2.1242e-02	107	0.00	1.3888e-05	115	0.00	1.7188e-02
256	FSSOR	1397	0.32	2.1228e-02	1988	0.51	1.3926e-05	3272	0.72	1.7081e-02
	HSSOR	712	0.02	2.1230e-02	768	0.06	1.3915e-05	1031	0.04	1.7118e-02
	HS4EGSOR	243	0.02	2.1230e-02	218	0.02	1.3916e-05	247	0.02	1.7118e-02
512	FSSOR	2789	2.44	2.1229e-02	5128	4.34	1.3919e-05	10600	12.24	1.7063e-02
	HSSOR	1397	0.17	2.1228e-02	1988	0.24	1.3926e-05	3272	0.39	1.7082e-02
	HS4EGSOR	473	0.10	2.1228e-02	474	0.10	1.3931e-05	578	0.11	1.7082e-02
1024	FSSOR	5489	19.7	2.1230e-02	13970	54.84	1.3863e-05	34573	136.42	1.7054e-02
	HSSOR	2789	1.24	2.1229e-02	7230	3.18	1.3899e-05	11331	4.95	1.7063e-02
	HS4EGSOR	910	0.66	2.1230e-02	1097	0.79	1.3937e-05	1563	1.19	1.7063e-02
2048	FSSOR	10845	165.55	2.1230e-02	37307	598.82	1.3708e-05	111085	1531.67	1.7048e-02
	HSSOR	5489	9.40	2.1230e-02	13970	23.85	1.3863e-05	34573	59.15	1.7054e-02
	HS4EGSOR	1892	5.28	2.1230e-02	2643	7.28	1.3931e-05	4729	13.02	1.7054e-02

parameter set to $\beta = 1.50$, a size 128 declined from 0.13 seconds for FSOR to 0.01 seconds for HSSOR. But, the number of iterations remain same from 0.01 seconds to 0.01 for 4HSEGSOR. It's the same situation for another size the execution time has reduce. To show more clearly, the percentage of execution time reduced about 90.00% – 97.67%, 92.31% – 98.71%, and 91.67% –98.75% corresponds to the HSSOR method. From the elucidating result for Table 1, the effectiveness 4HSEGSOR can be conform better compare to HSSOR.

In fact, Table 2 show that the numerical experiments for number of iterations HSSOR and 4HSEGSOR iterative method have declined. However, the result showed the number of iterations 4HSEGSOR drastically declined compare with HSSOR. It can be show when the parameter set to $\beta = 1.50$, the number of iteration declined from 5478 for FSOR to 541 for 4HSEGSOR meanwhile for HSSOR just to 2039 from 5478. It's the same situation for another size the number of iteration has drastically declined. To show more clearly, the percentage of number iterations declined approximately by 82.73%–83.00%, 85.44–92.24% and 88.74%–95.60% respectively as compared with the HSSOR method. Also for execution time, the result showed 4HSEGSOR method reduce the computational complexity. The execution time at size 2048 declined from 77.53 seconds for FSOR to 15.14 seconds for HSSOR. Again, the execution time declined from 15.14 seconds to 5.30 seconds for 4HSEGSOR. It's the same situation for another size the execution time has reduce. To show more clearly, the percentage of execution time reduced about 50.00%–93.16%, 75.00%–96.93%, and 66.67%–98.25% respectively than the HSSOR method. From the elucidating result for Table 2, the effectiveness 4HSEGSOR can be conform more better compare to HSSOR.

Lastly, from the numerical results in Table 3 show that the number of iterations for the 4HSEGSOR iterative method have declined. The result showed the number of iterations at size 512 declined from 2789 for FSOR to 1397 for HSSOR. Again, the number of iterations declined from 1937 to 473 for 4HSEGSOR. It's the same situation for another size the number of iteration has declined. To show more clearly, the percentage of number iterations declined approximately by 82.30%–83.42%, 86.07%–92.92% and 88.85%–95.74% respectively as compared with the HSSOR method. Also, implementations of computational time for the result showed 4HSEGSOR method reduce the computational complexity. The execution time at size 512 declined from 2.44 seconds for FSOR to 0.17 seconds for HSSOR. Again, the number of iterations declined from 0.17 seconds to 0.10 for 4HSEGSOR. It's the same situation for another size the execution time has reduce. To show more clearly, the percentage of execution time reduced about 93.75% - 100.00%, 96.08%–100.00% and 97.22%–100.00% respectively than the HSSOR method.

6. Conclusion

For the conclusion, this study success the numerical experiments of the application 4HSEGSOR iteration to solve the linear system generated by the steady-state problems which is two-point FBVPs. The family FD scheme and Caputo's fractional derivative operator was applied to construct the linear system. From the linear system, the iterative methods success to applied to get the approximation solution. From the numerical results recorded in Tables 1, 2, and 3 by imposing the comparison between HSSOR and HS4EGSOR iterative methods with FSSOR method react as control method, clearly the results promising two improvements in the number of iterations (K) and execution time (s). According to the numerical results are recorded, it can be showed that the HS4EGSOR method is superior and it has required a much lesser number of iterations and computational time to solve the problems.

Acknowledgment

The authors acknowledge the research grant scheme (GUG0227-1/2018) from Universiti Malaysia Sabah, Malaysia for the completion of this article.

References

[1] C. Kou, and X. Zhang, "Solutions of the fractional differential equations with constant coefficients." IEEE. Eighth ACIS International Conference on Vol. 3, pp. 1056-1059, 2007.

[2] D. Matignon, "Stability results for fractional differential equations with applications to control processing." In Comp eng in systems appl., Vol. 2, pp. 963-968, 1996.

[3] G. J. Fix, and J. P. Roof, "Least squares finite-element solution of a fractional order two-point boundary value problem." Computers & Mathematics with Applications., 48(7-8), 1017-1033, 2004.

[4] X. Li, H. Li, and B. Wu, "A new numerical method for variable order fractional functional differential equations." Applied Mathematics Letters., 68, 80-86, 2017.

[5] Z. Odibat, and S. Momani, "Modified homotopy perturbation method: application to quadratic Riccati differential equation of fractional order." Chaos, Solitons & Fractals., 36(1), 167-174, 2008.

[6] K. Diethelm, & G. Walz, "Numerical solution of fractional order differential equations by extrapolation." *Numerical Algorithms.*, *16*(3-4), 231-253, 1997.

[7] A. Sunarto, J. Sulaiman, and A.Saudi, "Application of The Full-Sweep AOR Iteration Concept for Space-Fractional Diffusion Equation." IOP Publishing, In Journal of Physics., 710(1), 012019, 2016.

[8] A. Sunarto, J. Sulaiman, and A.Saudi, "SOR method for implicit finite difference solution of time-fractional diffusion equation." In Proceedings of the 11th Seminar on Science and Technology., pp. 132-139, 2013.

[9] S. Alkan, K. Yildirim, and A. Secer, "An efficient algorithm for solving fractional differential equations with boundary conditions." Open Physics., 14(1), 6-14, 2016.

[10] Young, D.M., 1971. Iterative solution of large Linear Systems, Academic Press, London, England, Pages:570.

[11] W. Hackbusch, Iterative Solution of Large Sparse Systems of Equations, Springer-Verlag, New York, 1995.

[12] R. Rahman, N. A. M. Ali, J. Sulaiman, and F. A. Muhiddin, "Caputo's finite difference solution of fractional two-point boundary value problems using SOR iteration." In American Institute of Physics Conference Series, 2013(2), 2018.

[13] R. Rahman, N. A. M. Ali, J. Sulaiman, and F. A. Muhiddin, "Caputo's Finite Difference Solution of Fractional Two-Point BVPs Using AGE Iteration." In Journal of Physics: Conference Series, IOP Publishing, vol. 1123, no. 1, p. 012044. 2018.

[14] I. Podlubny, "Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications" Academic press, New York, USA., Pages 340, 1999b.

[15] B. R. Sontakke, and A. S. Shaikh, "Properties of Caputo operator and its applications to linear fractional differential equations." International Journal of Engineering Research and Applications 5, 5(22-29), 2015.

[16] Diethelm, K., & Ford, N. J. Analysis of fractional differential equations. Journal of Mathematical Analysis and Applications, 265(2), 229-248, 2002.

[17] A. Sunarto, J. Sulaiman, and A. Saudi, "August. Solving the time fractional diffusion equations by the Half-Sweep SOR iterative method." IEEE. In Advanced Informatics: Concept, Theory and Application (ICAICTA), International Conference of pp. 272-277, 2014.

[18] D. J. Evans, "Group explicit iterative methods for solving large linear systems." International Journal of Computer Mathematics., 17(1), 81-108, 1985.

[19] A. Saudi, and J. Sulaiman, "Block Iterative Method for Robot Path Planning." In The 2nd Seminar on Engineering and Information Technology (SEIT2009), Kota Kinabalu, Malaysia. 2009.

[20] A. Saudi, and J. Sulaiman, "Robot path planning via EGSOR iterative method using Nine-Point Laplacian." In 2012 Third International Conference on Intelligent Systems Modelling and Simulation, pp. 61-66. IEEE, 2012.

[21] D. J. Evans, and A. R. Abdullah, "A new explicit method for the diffusion-convention equation." Computers & mathematics with applications., 11(1-3), 145-154, 1985.

[22] A. El-Ajou, O. A. Arqub, and S. Momani, "Solving fractional two-point boundary value problems using continuous analytic method." Ain Shams Engineering Journal., 4(3), 539-547, 2013.

[23] A. Pedas, and E. Tamme, "Piecewise polynomial collocation for linear boundary value problems of fractional differential equations." Journal of Computational and Applied Mathematics., 236(13), 3349-3359, 2012.

[24] H. Azizi, "Chebyshev Finite Difference Method for Fractional Boundary Value Problems." Journal of Mathematical Extension., 9, 57-71, 2015.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 244-250 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Automatic Service Orchestration for e-Health Application

Anatolii Petrenko, Bogdan Bulakh*

System Design Department, National Technical University of Ukraine, Igor Sikorsky Kyiv Polytechnic Institute, 03056, Ukraine

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 13 July, 2019
Online :30 July, 2019

Keywords:
Service orchestration
Semantic technologies
e-Health

ABSTRACT

This paper describes an architectural approach to the development of dynamic service-oriented systems for e-Health using the service orchestration mechanism and semantic technologies. The main idea is the dynamic synthesis of the complex functionality required by user or by software agent. This idea should help to build and easily extend the applied loose-coupled systems without strict dependencies on concrete web services and their invocation details. A sample scenario of such dynamic orchestration is covered and analyzed, possible ways of further improvement of this approach are given.

1. Introduction

This paper is an extension of work originally presented at the 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC) [1]. Since then the original research has been continued in the direction of development of service-oriented e-Health application. This paper covers more details of dynamic service orchestration and how it can be applied to the e-Health software design.

2. Problem Overview

Modern network applications widely use web service interaction to retrieve or update data. There may be a lot of interconnected services of different types (like SOAP or REST) interacting with each other and external third-party services, thus forming so called "service ecosystem". The interaction of these services may be implemented in different ways (see Figure 1):

- "Hard-coded" interaction: direct invocation of the service at specific URL is fixed in source code with all the invocation details. In this case even simple relocation of services to another URLs requires changing (and consequent recompiling, redeployment, retesting) of all dependent software modules (other web services, for instance).
- Services could use service discovery mechanism by means of some registry (in its simplest case by using some "name server" to translate service name to concrete URL. UDDI registry is an advanced example). This makes it easier to modify and scale the service ecosystem, but services are still tightly coupled by data formats and interaction patterns.
- Service choreography approach (e.g. "publisher-subscriber" model) could be implemented by means of some message broker. This event-driven service architecture consists of loosely coupled services that can be dynamically connected to each other through a subscription to different message types.
- Service orchestration approach: all interactions are controlled by external orchestration software. In such system services could be truly agnostic to their environment while publishing only their interface description.

It must be noted that in many real-life systems composed of hundreds of services all of these approaches may be implemented within a single service ecosystem (forming "heterogeneous" service ecosystem).

In any case it is only skilled developer who is capable to organize the inter-service communication to bring some functionality to the system. It is almost never possible for the end users to create new functionality without skills and knowledge in programming. This problem is partly solved by so called "workflow management systems" (like scientific workflow systems [2] or engineering ones [3]). Most of service-oriented workflow systems typically rely on the orchestration approach.

* Bogdan Bulakh, Email: bogdan.bulakh@gmail.com

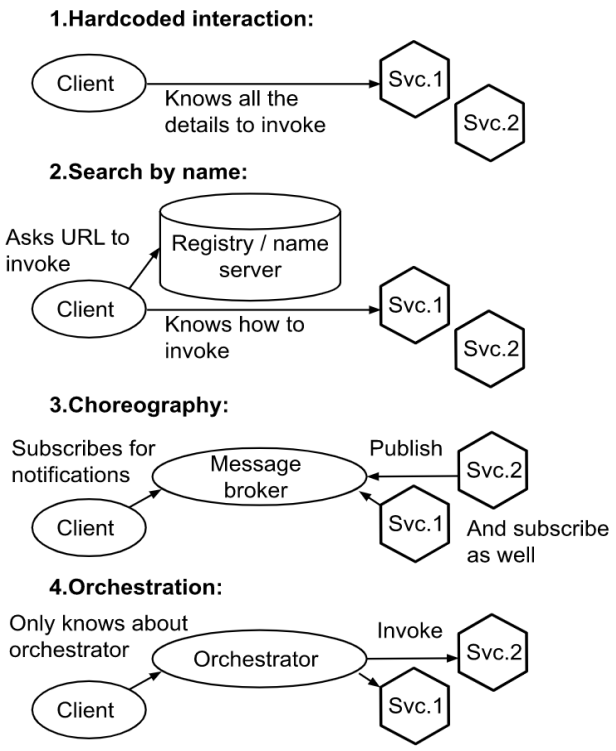


Figure 1: Some existing service interaction patterns (client could be implemented as another service as well)

2.1. Problem Definition

This paper focuses on a slightly different issue: how to make it possible for *both users and program agents* to request and execute the desired functionality *without any service-level knowledge* (existence of specific services and their interfaces' details)? In other words, dependencies between the clients and web services are forcing developers to update clients when service ecosystem changes. This costs time and extra QA efforts. We propose an approach that helps to avoid the update of the clients in case of changes in service ecosystem and thus makes it easier to maintain and improve the overall system.

Instead of making calls to specific services, or even "constructing workflows of services", software clients can issue "functionality requests" to some broker. While these requests can be fulfilled, the clients need no updates. The fulfilment of these requests could be organized using the orchestration approach, if there is no single service capable to provide the requested functionality. This is what we call here a "*dynamic semantic-based orchestration*": the automatic synthesis of a service workflows to provide the requested functionality (including some goal and output data). One of the main benefits of the dynamic orchestration is that service clients and services themselves are not coupled in any way except the functionality semantics. There is no need to keep track of web service interface changes on each client, it's only a knowledge base about services (which binds functionality semantics with service invocation details) that should be kept updated instead.

3. Possible Application Scenarios

There are a lot of application fields for dynamic semantic-based orchestration. Some of them are listed below.

Scientific and engineering workflows

There are a lot of existing 'Scientific workflow systems' and many of them support invocation of the remote services (Taverna workflows [4], Apache Airavata [5] etc.). The main problem there is a high complexity of constructing the service workflows for non-IT-domain specialists and scientists, because the workflow elements are bound to concrete services, their operations, and data formats. With the dynamic semantic-based orchestration the end users could just set the goals of the orchestration and receive the automatically built workflow. No need to deal with service-level details for the end users.

Performance monitoring dashboards

This feature can be implemented in medical (see further), scientific, engineering, industry production, financial and other domains. Given the ecosystem of services to fetch different parameters of the objects to monitor (patients' health data, the status of lab equipment or industrial equipment etc.). User needs a tool that can be used to easily construct and customize the KPIs. Possible solution: a user describes the functionality he needs in terms of some knowledge base, and new dashboard indicator will get its values from the dynamically orchestrated services providing the functionality requested.

Business analytics and data mining

Same as previous kind of scenario but it also involves the services for mining new knowledge from existing data helping to improve existing business processes. These workflows are more complex, compute intensive and provide valuable information (insights). It is a service-oriented analytics solution for business enterprises of different scale. In [6] a workflow management system for 'big data' mining is described. But its workflows are still composed manually by the end user.

Smart house and IoT

The dynamic semantic-based service orchestration can be used in controlling IoT devices, e.g. as a part of smart house. For example, anyone in the smart house environment can easily add new automation scenario, e.g. to control lighting, climate or heating or water supply in a smart way, just by describing the desired functionality. In a similar way this approach can be applied, say, to control production processes on factories or for stock management. The PROtEUS++ [7] is a promising example of a specialized workflow engine for IoT tasks, and by the way it supports dynamic service discovery using the semantic service description.

Workflows for e-Health

Workflows for patient's diagnostics and treatment (DTWf) could also be implemented with the help of automatic semantic-based service orchestration. The DTWf typically can contain invocation of basic services for diagnostics (like service to check blood pressure, controlling blood pressure data received from devices or manually entered by the patient), services of patient's health status prediction (based on the data provided by other services) and so on [8]. This application field will be mainly referred further as it is studied by authors in scope of implementation of the project of the mobile e-Health platform development.

4. Implementation

The main idea of the proposed solution is to hide as much low-level service interface specifics as possible from the client (functionality consumer), allowing the client to operate only with its goals and data. It is similar to how declarative programming languages (like SQL or SPARQL) differ from imperative ones: when working in a declarative style you need to specify what goal to reach but not how it should be reached. You actually don't worry about implementation details. Ideally the client just declares the new computational goal (e.g. "to get specific output results from specific input data" etc.) without referencing particular services. He only operates with terms and facts from the knowledge base that can help him to express his goal. The rest process is automated by means of automatic service discovery and automated orchestration.

The solution proposed is based on the following three main components (also implemented as services), as shown on Figure 2:

- *Service for execution* (SE) of functionality requests, which serves as a kind of broker and the single point of access for clients. It does not only find some concrete web services able to fulfil the client's request, but also is responsible for execution of any service operations and results provision. In other words, clients don't ask SE to find some services, they ask to do some actions, provide input data and wait for output. In order to fulfill client's request SE communicates with other two components.
- *Service registry* (SR). This service searches for available services that could provide the requested functionality (according to registered service's semantic annotations). In fact, it is an interface to the knowledge base containing facts about services and their functions. The second mission of SR is to find the possible service orchestration scenario if no single service matches the requested functionality. Then this scenario is passed to the Service orchestrator.
- *Service orchestrator* (SO). This service is responsible for execution of service workflows, created by SR. Workflows actually can consist either of a single service, or describe the orchestration scenario involving multiple services. This service complies with a SaaS model: new workflows can be deployed in a cloud infrastructure and can be interacted via the REST interface (invoked, queried about status and results, canceled, disposed etc.). It should be noted that today there are a lot of orchestration tools available. The orchestrator engine is a core part of SO and it can be implemented on top of many of existing tools, for example: BPEL engine, Taverna Workflows, Netflix Conductor (a microservices orchestration engine).

4.1. Service Registry and Knowledge Base

The SR has an extended functionality compared to UDDI registries. Instead of binding to standard fields and database schema, it is based on flexible knowledge base. Administrator-level users can add new facts or modify existing facts about services registered and extend the ontology with new classes. The SR can be queried about services in a similar way like Triple stores are queried with SPARQL queries. But the functionality of SR is not limited by simple querying the triple store. It includes the matchmaking logic helping to construct the complex service workflow according to the goals set and inputs provided. To do this the SR should contain the formalized knowledge on web service interaction details (protocols, procedures, interfaces), and the knowledge on basic data formats used and how to extract and transform the data from them (xml, JSON, CSV etc.). That's why the knowledge base should consist of domain ontology (to set goals, to describe services functionality and operations, inputs and outputs), service ontology (to allow automatic invocation, see Figure 3), and data formats ontology (to allow automatic data flow building).

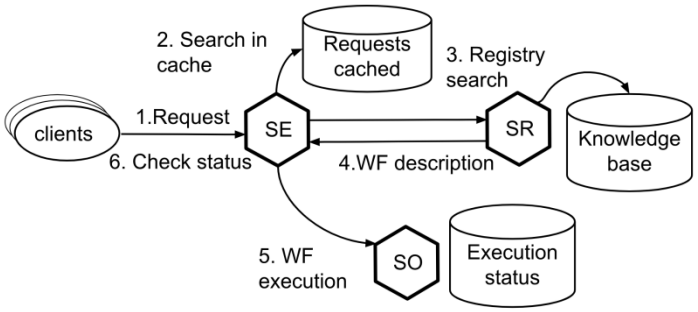


Figure 2: Main components of the proposed solution

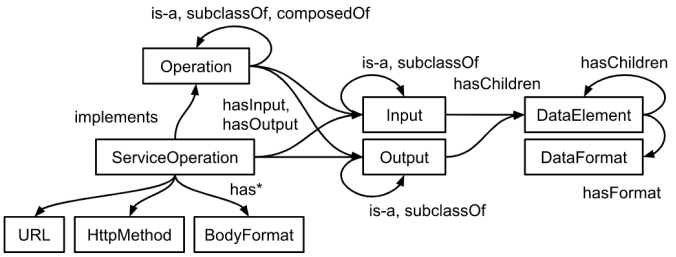


Figure 3: Service ontology (fragment)

4.2. Service Matchmaking

Existing standard solutions in service registration like UDDI are not enough for advanced semantic search. Instead it is proposed to rely on service ontologies, similar to OWL-S ontology [9], which will give more relevant results. Service matching can be done using the semantic proximity of the service ontology elements (from registry knowledge base and those from the "request for functionality" query) [10]. In order to compare these ontology elements, the following elementary proximity estimations are needed to be computed: proximity of classes, proximity of classes and instances, proximity of instances, proximity of predicates. Some service A can be called 'relevant' to some search query R only in case the proximity estimate is higher than some threshold value M_T (see further).

Another aspect is to choose what information should be included into comparison. It is proposed to use the following information categories: C (context), IOPE (inputs, outputs, preconditions, effects) and QoS.

The context C form search query can be formally defined as any information that could implicitly and explicitly impact the query generation by user (it can be profile-oriented, history-oriented, process-oriented, other context). An explicit context is

directly provided by user, while implicit one is that being gathered by any automated tool.

The service functionality can be matched by use of IOPE attributes. The *hasInput*, *hasOutput*, *hasPrecondition* and *hasEffects* query attributes (R) are matched against the corresponding ones from each service's description (A). Matching can be estimated as a one of following outcomes: "exact match", "plug-in" match when R is a subclass of A, "subsume" when R and A have common partial match, "failure" is A and R have nothing common.

QoS attributes include service price, performance, reliability, stability, scalability, security etc. The total match level estimation combines those estimations mentioned above:

$$M(R, A) = K_C M_C + K_I M_I + K_O M_O + K_P M_P + K_E M_E + \\ + K_{QoS} M_{QoS} > M_T$$

If we group IOPE and C requirements as functional requirements (FR) and QoS we refer to as non-functional requirements (NFR), and if we use mismatch-level (or level of matching error) as an indicator of matching failure (for practical reasons: to be able to use mismatch penalties instead of proximity estimations), then we'll have the following general matching error function:

$$E(R, A) = K_{FR} E_{FR} + K_{NFR} E_{NFR} < E_T$$

Both measures M and E could be used combined. First, we look for the single service with the highest M(R,A). If $M(R,A) > M_T$ then we accept it as a service providing requested functionality. If $M(R,A) < M_T$ then we need to compose a workflow of services with less possible $E(R,A) < E_T$. In order to estimate E for a workflow of services {A} we can use this general approach:

$$E(R, \{A\}) = K_{FR} E_{FR}^* + K_{NFR} E_{NFR}^*,$$

where * marks summary values for a workflow, for instance, internal inputs and outputs that are connecting services are excluded from error estimation.

During the practical implementation of a e-Health product with less than hundreds of service operations and without third-party services we found it sufficient to consider only IOE requirements (E for effects in form of functionality requested), without NFR. So, we used the following brief construct for the functionality request:

$$R = (F, \{I\}, \{O\}),$$

where F (function) – action to perform, I (input) – available inputs specification, O (output) – requested outputs specification. In a practical implementation we used JSON format, flexible enough to express complex I and O descriptions (see Figure 4). Here is a simple example of request to SE to get user name by ID:

```
{
  "operation":"get",
  "input":[
    {
      "name":"user",
      "properties":[
        {
          "name":"id",
          "value":"user001"
        }
      ]
    }
  ],
  "output":[
    {
      "name":"user",
      "properties":[
        {
          "name":"firstname"
        },
        {
          "name":"lastname"
        }
      ]
    }
  ]
}
```

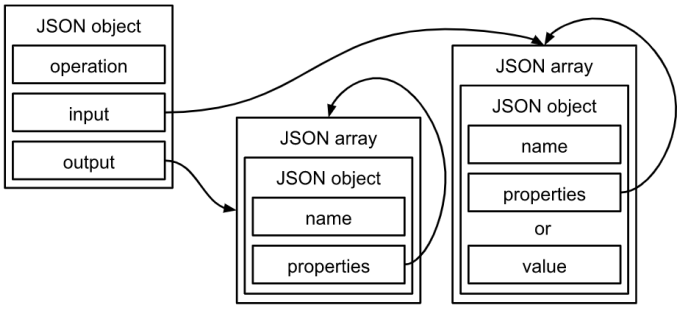


Figure 4: Inputs and outputs specification format

4.3. e-Health Application

The are many research papers studying the use of service-oriented architecture for medical software [11,12], and some of them mention service orchestration as a part of medical software, but this paper focuses on slightly different things, related to orchestration mechanism itself.

Workflow composition or, in other words, automatic semantic-based orchestration scenario synthesis, makes is easier to extend existing functionality of the system. It could be possible to get this new functionality even without development of new services: when the new functionality request could be satisfied with an orchestration scenario involving existing services. Before describing an example of such scenario, let us briefly describe the e-Health application that uses the proposed approach.

The e-Health application we develop is a system providing a virtual office for doctors and patients as an online point of their communication. It is useful in conditions when the nearest medical center is far from patient's home. In this situation the initial diagnostics can start online, and in many cases, it could be enough to help patient without real-life visit to the hospital. Also, the virtual doctor's office remembers all the patient treatment history and can provide some intelligent tool helpful for doctors and patients, like prediction of the crisis state of patients' health based on ordinary indicators like blood pressure, glucose level or complains about pain. This is an example of extended functionality that can be built by re-using the existing functions. The general

architecture blocks of this e-Heath application are presented on Figure 5.

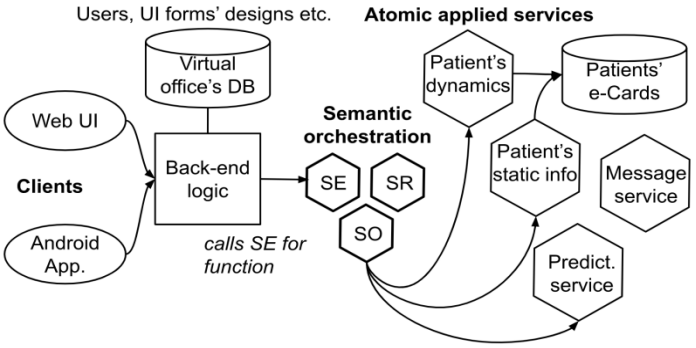


Figure 5: Virtual doctor's office using the dynamic semantic-based orchestration

Most of the office's functionality is not implemented just by back-end logic querying the database. There is a separate layer called "applied services" which is actually an API to do everything on patient's record. Conversations with doctors, diagnoses, treatment prescriptions, health indicators, reports – all of this is implemented with atomic REST services (we can call them microservices). The back-end logic's responsibility is to serve front-ends' requests and gather all necessary data for displaying to the user. Any call to operate patient's data is made as a "functionality request" with function, inputs and outputs provided to SE. This way we can truly separate back-end development and microservice API development, and only need to reflect important changes in a knowledge base of SR.

4.4. Service Workflow Composition Example

One of the examples where the dynamic orchestration takes place is the prediction of diabetes, heart problems and so on based on the last entered health indicators. This prediction is just a hint for the doctor and should only attract his attention to some possible problems (by displaying a warning), so there was no need to achieve very high accuracy of prediction. To develop such "intelligent" hint test datasets were analyzed, prepared and used for neural network training. This ended up as a separate service with one single output (binary warning flag) and a number of inputs (like weight, insulin and glucose level and much more) developed by separate team with machine learning background. In order to get the diabetes warning flag, the back-end developers just made a simple request: "given patient's ID and current date, run diabetes prediction and return warning flag". At the same time the SR's knowledge base contains information about inputs and outputs of all services, including the diabetes predictor.

Let's take a look at the process of execution of that request mentioned above. After JSON describing this request is accepted and parsed by SR, the SR itself starts to search for single service capable to satisfy the request. It founds the service operation with function briefly described like "Diabetes prediction". But this service operation has a lot of inputs, despite it has the output and function matching the request. So, no single service is found, and SR tries to build the workflow. It finds a lot of combinations of atomic services, but the best one (according to matching error penalties) is presented on Figure 6. It is executed, and the result is returned to the back-end caller (the client).

The Figures 7 and 8 shows two screens of a mobile client developed using the described diabetes prediction workflow to display a warning icon (the application's language is Ukrainian). The first screen is a menu screen for a doctor with a following action on a selected patient: indicators, chat, log, complains, diagnosis, treatment. The second screen shows some sample conversation between a doctor and his patient in a chat mode.

When a doctor selects someone of his patients, the diabetes prediction workflow is called, and the client displays or hides the warning icon (a circle with Ukrainian letter "D" inside) at the top of the screen related to that patient (seen on both screens on Figures 7 and 8). In a similar way any other predictor (e.g. for heart problems) is called via the dynamic semantic-based orchestration.

4.5. Evaluation of Results

The performance penalties for this sophisticated process of services invocation (compared with direct REST services calls) could be estimated through the workflow execution time T_{exec} :

$$T_{exec} = T_M + T_V + T_{WF} + T_S$$

where T_M is a matchmaking time, T_{WF} is a workflow execution related overhead (depends on workflow execution engine speed), T_S is a total time for the execution of the longest sequence of service invocations in a flow (web services execution time). There is also T_V which is a validation time of the composed workflow. Currently we don't use any validation technique, and fully rely on the consistency of the knowledge base during the automated workflow synthesis ($T_V = 0$). But there is a risk of the incorrect behavior of the composed workflows (like infinite loops) even if they were constructed without any inconsistency with the knowledge base, so we need formal modeling techniques. Petri nets [13] and process calculus [14] could be successfully used to model web service workflows and related tools will be further integrated to the system to improve its reliability.

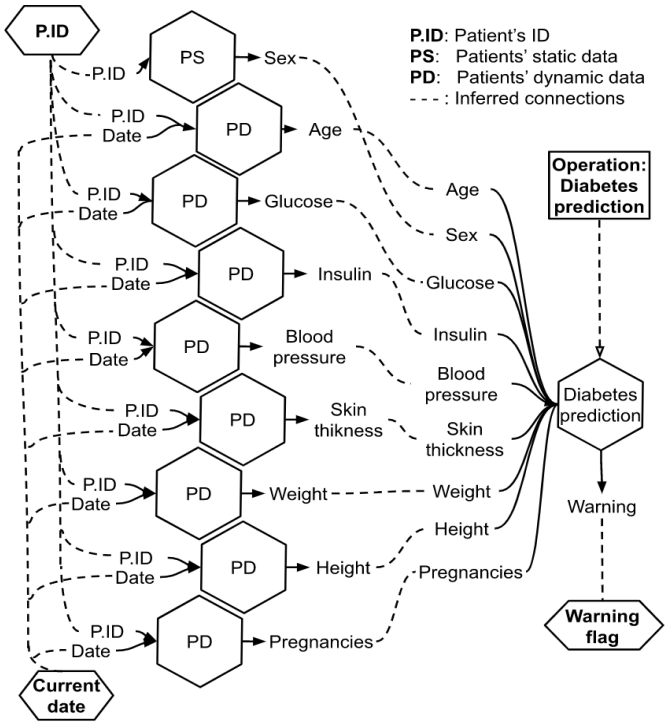


Figure 6: Diabetes prediction workflow (inputs and output are marked bold)

When comparing hard-coded scenario execution ($T_{exec}' = T_S$) with dynamic orchestration T_{exec}, the dynamic and flexible approach lasts longer by $T_{exec} - T_{exec}' = T_M + T_{WF}$, which was 3 to 5 seconds in case of small service registry (about 30 service operations) we used for diabetes prediction scenario. It should be noted that this time is seriously affected by the knowledge base facts quantity and quality (seriously impacts T_M).

In order to avoid same process of workflow synthesis each time a doctor needs a hint, the resulting workflow is cached in SR's knowledge base. The overhead is minimal and almost equal to ($T_{WF} + T_S$) because in this case $T_{exec}'' = T_C + T_{WF} + T_S$, where T_C is a cache search time and $T_C << T_M$). But in case of any changes in the knowledge base the cache should be cleared. This is a trade-off we get instead of updating the clients: the knowledge base update leads to reasoning on the updated facts and repeating the service matchmaking again from scratch. But found these knowledge base related activities less time-consuming then updating, redeploying and QA of all client applications.

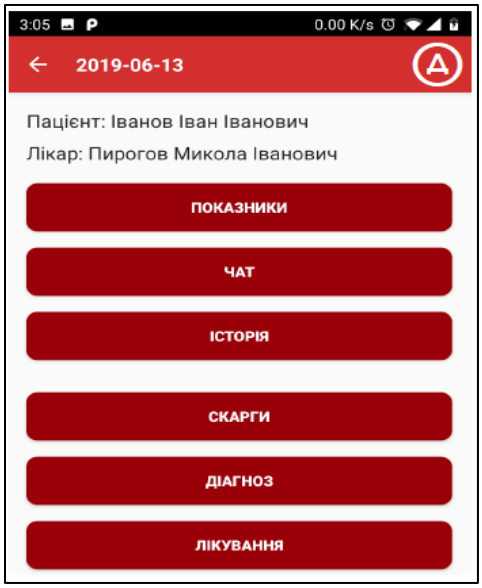


Figure 7: A 'patient menu' sample screen of the mobile e-Health application with diabetes warning icon at the top.



Figure 8: Chat screen of the mobile e-Health application, with diabetes warning icon at the top.

5. Conclusions and Future Work

This paper presented a description of a software architectural approach that aims to bring more flexibility to service-oriented systems and make service dependencies weaker. The main idea is to provide additional "abstract request layer" to the well-known service orchestration mechanism. The abstract request for functionality is being matched with the knowledge base of available web services to perform actual calculations.

The approach described has been successfully applied to the e-Health system, and it gives promising results for this application field, since e-Health applications typically have a lot of functions and data sources that could be accesses with the web service interfaces.

In our opinion this approach has the following benefits not only for e-Health domain but for other application fields as well:

- Service clients and web services are fully separated and could be developed independently by the teams of different development tools and background. The only dependency is the knowledge base (and service registry that uses it). Service ecosystem can be easily updated without changes to its clients. Only the knowledge base needs to be updated.
- It is possible to develop the UI for users to allow them easily construct the functionality requests and thus extend the functionality without help from software developers. This is one direction for our future work. It includes the development of workbench where users can test their requests and development of UI editor to control the displaying of requested results.
- In case of small service registries, the matchmaking process and workflow synthesis will not harm performance. This is true in case of avoiding usage of some existing heavyweight tools for orchestration like BPEL engines as the core for service orchestrator. The problem is that these tools like BPEL engines were basically designed for static workflows, not dynamic ones.

Of course, this vision has its drawbacks:

- The e-Health software must be very robust. But it has to be noted that the absence of hard-coded reference to concrete services there is a non-zero probability that SR will find wrong services or does not find anything thus producing wrong results returned by SE. From this point of view the only way in assuring the correct behavior is continuous testing of functions used by clients. When client gets updated with new functionality request call, this call should be added to the test plan. The improvements in QA of the solution presented are planned for our further work. At the same time formal models like Petri nets or process calculus can help to test the behavior of the dynamically generated workflows automatically.
- Another option to make sure correct services are matched to a request is to make only strict logical assertions by SR. No partial matching possible, service or workflow is either matching the request or not. In this case the clients will

need to use only strict request constructs and only correct ontology terms it their requests, thus making the new programming language to express the functionality requests. This will make it harder to add the new requests and make them work by end users and even programmers.

- In case of large registries of services, the matchmaking process could be time-consuming, since a lot of services could be "paired" according to their inputs and outputs and their functionality. So this solution will not work as expected in a system with hundreds of service operations and third-party services. And there will be problems if there are too many fine-grained microservices in the ecosystem even if it does not include external third-party services. In this case the matchmaking process should be restricted of simplified since e-Health software should operate without lags.
- Another problematic issue is the correct management of the knowledge base. Each time a new service is developed and should be registered in the system, there are a lot of work to do in the knowledge base. In fact, the knowledge base is the most important component of such system and requires very accurate updates, permanent logging of changes and intensive QA. Together with service registry it forms the fragile bottleneck of the system.
- Such complex interaction patterns could bring additional problems in security and data privacy. This is very important for e-Health applications working with personal medical data. But this aspect is out of scope of this paper.

Other possible directions for the future work are: research on the specifics of orchestrating services that control IoT devices and moving orchestrator service to the cloud infrastructure.

Acknowledgment

This paper describes partial results of the ongoing fulfillment of the project "The development of the modern service systems by the example of mobile medical system for a front-line settlements in the military conflict zone" by the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".

References

[1] A.I. Petrenko, B.V.Bulakh, "Intelligent service orchestration as a service", Proc. of 2018 IEEE First International Conference on System Analysis and Intelligent Computing (SAIC), 8-12 October, 2018, Kyiv. - pp 201-205. https://doi.org/10.1109/saic.2018.8516723

[2] B. Ludäscher, M. Weske, T. McPhillips, S. Bowers, "Scientific Workflows: Business as Usual?", Lecture Notes in Computer Science. (2009) 31-47. https://doi.org/10.1007/978-3-642-03848-8_4.

[3] V. Ladogubets, B. Bulakh, V. Chekaliuk, O. Kramar, "Employing BPEL Engines for Engineering Calculations", The Experience of Designing and Application of CAD Systems in Microelectronics: 12-th Intern. Conf. «CADSM'2013», 19-23 February 2013, Polyana-Svalyava (Zakarpattya), Ukraine: proc. – Lviv, 2013. – P. 427–430. – ISBN 978-617- 607-393- 2

[4] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers et al. "The Taverna workflow suite: Designing and executing workflows of Web Services on the desktop, web or in the cloud". Nucleic Acids Research. 41 (Web Server issue): W557–W561. doi:10.1093/nar/gkt328. PMC 3692062. PMID 23640334.

[5] M. Pierce, S. Marru, L. Gunathilake, T. A. Kanewala et al., "Apache Airavata: Design and Directions of a Science Gateway Framework," 2014 6th International Workshop on Science Gateways, Dublin, 2014, pp. 48-54. doi: 10.1109/IWSG.2014.15

[6] J. Kranjc, R. Orač, V. Podpečan, N. Lavrac, M. Robnik-Sikonja.. "ClowdFlows: Online workflows for distributed big data mining". Future Generation Computer Systems 68, 2017. pp. 38-58. doi:10.1016/j.future.2016.07.018.

[7] R. Seiger, S. Huber, P. Heisig. "PROtEUS++: A Self-managed IoT Workflow Engine with Dynamic Service Discovery". 9th Central European Workshop on Services and their Composition (ZEUS). 2017. pp 90-92.

[8] O.O. Petrenko, A.I. Petrenko. "Cyber-Physical Medical Platform for Personal Health Monitoring", Journal of Scientific Achievements (JSA), Australia, vol. 2, issue 8, 2017, pp. 24-28. ISSN: 2207-4236

[9] OWL-S: Semantic Markup for Web Services. [online] Available https://www.w3.org/Submission/OWL-S/.

[10] Петренко І.А., Петренко О.О. "Автоматизовані методи пошуку і відкриття необхідних сервісів", Вісник Університету «Україна», Серія «Інформатика, обчислювальна техніка та кібернетика», №1(17), 2015, С. 55-64

[11] S. Rodriguez Loya, K. Kawamoto, C. Chatwin, and V. Huser, "Service Oriented Architecture for Clinical Decision Support: A Systematic Review and Future Directions". Journal of Medical Systems, December 2014, 38(12): 140. https://doi.org/10.1007/s10916-014-0140-z

[12] A. Celesti, M. Fazio, F. Galán Márquez, A. Glikson, H. Mauwa, A. Bagula, F. Celesti, and M. Villari. How to Develop IoT Cloud e-Health Systems Based on FIWARE: A Lesson Learnt. Journal of Sensor and Actuator Networks. 2019, 8(1), 7; https://doi.org/10.3390/jsan8010007

[13] W. van der Aalst, K.M. van Hee. Workflow management: models, methods, and systems. MIT press, 2004. 368 p.

[14] M. Weidlich, G. Decker, M. Weske. "Efficient Analysis of BPEL 2.0 Processes using pi-Calculus". Proceedings of the IEEE Asia-Pacific Services Computing Conference (APSCC'07). – Japan, 2007. – P. 266-274.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 251-255 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Exchange Rate Modeling: Medium-Term Equilibrium Dynamics

Anton Kuzmin*

The Financial University under the Government of the Russian Federation, Department of Data Analysis, Decision-Making and Financial Technologies, Moscow, Russian Federation

ARTICLE INFO

Article history:
Received: 14 June, 2019
Accepted: 24 July, 2019
Online :30 July, 2019

Keywords:
Exchange Rate
Equilibrium Dynamics
Modeling
Export-Import Operations
Capital Flows
The Terms Of Trade

ABSTRACT

In this study, we develop the author's approach, based on the principles of modeling of the International Flows Equilibrium Exchange Rate (IFEER). As a result, a new model is built, depending on the main macroeconomic factors in the medium term. The model presents a wide system of basic factors, but many of the intermediate factors in the modeling were eliminated. The results are discussed in the context of the previously known results of economic theory.

1. Introduction

This paper is an extension of work originally presented in Eleventh International Conference "Management of large-scale system development" (MLSD 2018, IEEE Russian section) [1] .

In the scientific literature the nominal and real exchange rate models are widely represented in conditions of the dynamic general equilibrium (DSGE-models), the founders of which is the model Obstfeld-Rogoff [2] of monopolistic competition of two open economies. At the same time, the previous approaches to fundamental factors modeling were developed by Mussa [3] and Dornbush [4] in conditions of both the monetarist and Keynesian approaches.

Many authors emphasizes the effectiveness of fundamental factors models of nominal and real exchange rates (e.g., [5]).

Conditions for the existence of the equilibrium rates were allocated in 1945 by R. Nurkse [6], which paid considerable attention to the balance of payments equilibrium, the lack of restrictions on export-import operations and capital flows and internal stability of the economy.

However, from the standpoint of structural modeling, the known models of economic theory cannot provide adequate results for assessing of the equilibrium real or nominal rates level. In a significant majority the exchange rates are not expressed in terms of the equilibrium of the financial market, but in terms of the theory of purchasing power parity (PPP). Structurally, the exchange rate analysis is limited by the current balance. Herewith, it should be emphasized the role of the exchange rates in the process of regulating the external balances and the role of the indicator of competitive advantages in foreign trade flows.

To solve these problems, we propose to use the author's approach, based on the principles of the international flows equilibrium exchange rate (IFEER) modeling [1, 7].

2. Conceptual Bases of Modeling

It should be noted, that in this article we develop the IFEER-approach and the new equilibrium exchange rate model will be presented as a result.

Initially, we suppose that the world consists of two equal rights open economies and international trade flows take into account the trading competitive position of these countries. In this context, international trading bilateral competitive position is determined by the real exchange rate. It should be noted here, that approach "small open economy - the rest of the world" under IFEER-concept was used by the author in modeling of the Russian ruble exchange rate (e.g., [7]).

We consider all real market transactions at nominal exchange rates $e_i, i \in (1, L)$ on the domestic currency market that occurred in a certain period of time.

*Anton Kuzmin, Moscow, Russian Federation, a_kuzmin@rambler.ru

Then e_i, D_i, R_i in *i-th* transaction are: the bilateral nominal exchange rate, the amount in the determined foreign currency and the amount in the national currency respectively.

In this context, these variables are linked by ratios:

$$e_i D_i = R_i, \text{ and therefore, } e_i = R_i / D_i .$$

So, we define the synthetic value of the bilateral nominal exchange rate in conditions of varying degrees of free floating for the certain period as the sum of exchange rates, weighted by amounts in determined foreign currencies:

$$e \equiv \sum_{i=1}^{L} \frac{D_i}{\sum_{j=1}^{L} D_j} \times e_i , \quad (1)$$

Using the definition of the exchange rate in *i-th* transaction, we get from (1):

$$e = \sum_{i=1}^{L} \frac{D_i}{\sum_{j=1}^{L} D_j} \times \frac{R_i}{D_i} = \sum_{i=1}^{L} R_i \Big/ \sum_{j=1}^{L} D_j . \quad (2)$$

Next, on a conceptual level, we can disaggregate balance of payments flows:

$$e_t = 1 \Big/ e_t^* = (I_t + K_t^-) \Big/ (E_t + K_t^+) = \\ = (E_t^* + K_t^{*+}) \Big/ (I_t^* + K_t^{*-}), \quad (3)$$

where I_t - the demand in the national currency on imports,

E_t - the supply in the foreign currency from exports,

K_t^+ - the capital inflows in the foreign currency,

K_t^- - capital outflows in the national currency.

This formula takes into account such important components as export revenue, demand for imports, demand for foreign assets and so on.

Therefore,we are developing two-period model in the dynamics in times *t* and *t-1*.

Herewith, at time *t* value of the functions of export-import operations and capital flows will be directly determined by the international competitive position represented by the real exchange rate at the time *t-1*:

$$e_{t-1}^R = e_{t-1} P_{t-1}^* \Big/ P_{t-1} , \quad (4)$$

where P_{t-1}^* - the exernal prices index,

P_t - the internal prices index.

For the model verification the price indices can be represented by consumer prices index (CPI), GDP-deflator and so on.

In the analysis of exchange rates, many authors (e.g., M. Bahmani-Oskoee and G. Goswami [8]) determine the terms of trade, reflecting the effectiveness of export-import operations in world trade, as one of the determining factors of the dynamics.

For example, F. Caramazza [9] explored the expectations of investors of German mark against French franc dynamics – one of the main cross-rates in the world at that time. He explained the change in parity by largely fundamental macro-variables, including terms of trade and relative inflation differential. L. Bartolini [10] similarly revealed the relationship between market expectations of devaluation and CPI-indicator of trading international advantages.

In the author's model of the ruble exchange rate [11], the use of trade terms by micro-agents as the main determinant of foreign trade operations is proved and justified on the statistical series of import and export of Russia. However, they are expressed in adjusted nominal exchange rates at price levels, including actual export prices.

3. Exchange Rate Modeling: Medium-Term Equilibrium Dynamics

In the domestic foreign exchange market at time *t* export earnings as part k_E of the average real gross domestic output $(Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta}$ in foreign prices P_t^* depend on the decision of producers-exporters and the state of international trade conditions e_{t-1}^R at time *t-1*:

$$E_t = P_t^* k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta} (e_{t-1}^R)^z , \quad (5)$$

where $k_E = const$,

Q_t - real gross domestic output.

In a function (5) $k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta}$ says, that the produced exports are part of averaged real gross domestic output (e.g., real GDP).

Due to the stability Q_t, the averaging method does not give a significant contribution to the fluctuations of the functional dependence (5).

The parameter *z* is the response to change in international trading position and the index *x* is associated with indexes *z* and *y* in the functions (5) and (6): *x=(z-y).*

In a functional dependence (5) the parameter δ reflects the fact that the growth rate of export exceeds the growth of gross domestic product because of a limitation of domestic demand: $\delta \geq 1$.

In the framework of the simulation because of equal rights and symmetry of flows the export of a country is an import of counterparty.

By analogy with a functional dependence (5):

$$E_t^* = P_t^* k_I (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\lambda} (e^{*R}_{t-1})^{-y} , \quad (6)$$

where $k_I = const$.

The parameter y also is the response to change in real international trading position.

So, using the formula (4) and $e_t = 1/e_t^*$, functional dependency (6) can be rewritten

$$E_t^* = I_t = P_t\, k_I (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\lambda} (e_{t-1}^R)^y .$$

As can be seen from formula (3), capital flows and export-import operations directly affect the exchange rate dynamics.

It should be noted that many well-known models of economic theory have serious difficulties with formalizing of the impact of capital flows on the exchange rates. However, within the framework of the proposed IFEER-concept, we can directly determine these functional dependencies.

At the same time, the mechanism of formation of functional dependencies of capital flows is very similar to the mechanisms of formation of export-import operations regarding the chosen factor system. More details can be found in the author's work [1]. For example, at time t we suppose that non-residents buy part of the real averaged domestic product $k_{K^+}(Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta}$ in their prices:

$$K_t^{+} = P_t^* k_{K^+} (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1}^R)^z , \qquad (7)$$

where $k_{K^+} = const$, $\theta \geq 0$.

Capital inflows are capital outflows of its counterparty:

$$K_t^{*+} = P_t\, k_{K^-} (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho} (e^{*R}_{t-1})^{-y} , \qquad (8)$$

where $k_{K^-} = const$.

We can rewrite a functional dependence (8):

$$K_t^{*+} = K_t^{-} = P_t\, k_{K^-} (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho} (e_{t-1}^R)^y .$$

Let's substitute dependences (5) - (8) in (3):

$$e_t = \Big[P_t k_I (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\lambda} (e_{t-1}^R)^y + \\ + P_t k_{K^-} (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho} (e_{t-1}^R)^y \Big] \times \\ \Big[P_t^* k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta} (e_{t-1}^R)^z + \\ + P_t^* k_{K^+} (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1}^R)^z \Big]^{-1} =$$

$$= \Big[P_t (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho} (e_{t-1}^R)^y \times \\ \times (k_I (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\lambda-\rho} + k_{K^-}) \Big] \times \\ \times \Big[P_t^* k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1}^R)^z \times \\ \times (k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta-\theta} + k_{K^+}) \Big]^{-1}$$

Due to the significant stability of the members $(Q_{t-1}^{\delta/\delta+} Q_t^{1/\delta+1})$ and $(Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})$, compared to price indicators P_{t-1}^* and P_t, we can consider as a constant the construct:

$$\frac{(k_I (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\lambda-\rho} + k_{K^-})}{(k_E (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\delta-\theta} + k_{K^+})} = k^{x+1} = const . \qquad (9)$$

Further, it can be rewritten as:

$$e_t = k^{x+1} \frac{P_t (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho}}{P_t^* (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1}^R)^{\delta}} =$$

$$= k^{x+1} \frac{P_t (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho}}{P_t^* (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1} P_{t-1}^* / P_{t-1})^{\delta}} = .$$

$$= k^{x+1} \frac{P_t (Q_t^{*\frac{1}{x+1}} Q_{t-1}^{*\frac{x}{x+1}})^{\rho}}{P_t^* (Q_t^{\frac{1}{x+1}} Q_{t-1}^{\frac{x}{x+1}})^{\theta} (e_{t-1})^{\delta}} \left(P_{t-1} / P_{t-1}^* \right)^x$$

So after regrouping we get:

$$e_t (e_{t-1})^x = \left(k \frac{P_t}{P_t^*} Q_t^{*\rho/x+1} Q_t^{-\theta/x+1} \right) \times \\ \times \left(k \frac{P_{t-1}}{P_{t-1}^*} Q_{t-1}^{*\rho/x+1} Q_{t-1}^{-\theta/x+1} \right)^x . \qquad (10)$$

After intertemporary separation of the fundamental economic variables in formula (10):

$$e_t = k \frac{P_t}{P_t^*} Q_t^{*\rho/x+1} Q_t^{-\theta/x+1} .$$

And after the designation $\theta' = \theta/x+1$ and $\rho' = \theta/x+1$ for convenience, we get:

$$e_t(P(t),Q(t),P^*(t),Q^*(t)) = k\frac{P_t}{P_t^*}Q_t^{*\rho'}Q_t^{-\theta'}. \quad (11)$$

4. Setting the Model Parameters and Verification

The starting point of the study of the medium-term exchange rate dynamics should be characterized by stability, both internal and external economic situation. More details can be found in the author's work [1].

We suggest that setting of the parameters ρ' and θ' can use the least squares method with a normalized differences of the nominal theoretical $e_t(\theta,\rho)$ and nominal market exchange rates under constraints $\rho' \geq 0$ and $\theta' \geq 0$:

$$ND(e)_t = \frac{e_t(\theta,\rho) - e_t(\text{nominal})}{e_t(\text{nominal})}. \quad (12)$$

At the starting point of the selected period it must be put: *P(start)=1, Q(start)=1, P*(start)=1, Q*(start)=1*. Coefficient k must be equal to the nominal market exchange rate at the starting point.

Next, the sum of normalized squared deviations (12) minimizes:

$$\begin{cases} \min\limits_{\theta',\rho'} \sum\limits_t \left(ND(e)_t\right)^2 \\ \theta' \geq 0 \\ \rho' \geq 0 \end{cases}. \quad (13)$$

In order to verify the model, it is advisable to consider the results in the context of the known results of economic theory of determination of the exchange rates.

For these intentions, let us rewrite formula (11) in logarithms:

$$\log e_t = \log k + (\pi_t - \pi_t^*) + \rho'\log Q_t^* - \theta'\log Q_t = \\ = \log k + (\pi_t - \pi_t^*) + \rho' q_t^* - \theta' q_t \quad . \quad (14)$$

Using the I.Fischer equation $PQ=MV$, where M is the money supply, V is the velocity of money, we present this in logarithms:

$$\pi = m + v - q\,.$$

Assuming in the medium term the rate of the velocity of money is constant and substituting (14) symmetrically for both countries, we obtain (with $a = \log k = const$)

$$\log e_t = \log k + (m_t + v - q_t - \\ -(m_t^* + v^* - q_t^*)) + \rho' q_t^* - \theta' q_t = \quad (15) \\ = a + (m_t - m_t^*) + (\rho'+1)q_t^* - (\theta'+1)q_t = \\ = a + (m_t - m_t^*) + \rho'' q_t^* - \theta'' q_t\,.$$

Verification of dependencies of a similar type (15), having the same system of fundamental macroeconomic factors, has been successful within the framework of verification of the monetary model. Of the works on this topic it should be noted econometric studies of models of fundamental determinants of B. R. Clarida et al. [12], J. Cheung, M. Chinn, and A. Pascual [13], I. Chowdhury [14].

It is also interesting to consider predictive abilities of dependencies of a similar type. One of the most famous is the study of J. Chen and N. Mark [15]. The forecasts for the formula of type (15) at $\rho'' = \theta''$ on different time horizons gave a significantly better result than in accordance with a random walk. And it was in the medium and long term.

5. Results and Conclusion

Exogenous variables are the total products Q and Q^* and the levels of external and internal prices P^*, P in the countries of the bilateral exchange rate in this model. Other basic intermediate determinants, such as export-import operations, capital flows, the terms of trade in the modelling process have been eliminated.

It should be noted here, that author's IFEER-concept of Equilibrium Exchange Rate (EER) correlates with the widely known economical concepts: behavioural EER (BEER) [16, 17] and fundamental EER (FEER) [18].

Using national consumer price and real domestic product indices and oil prices, the main research formula (11) was applied by the author to the study of the dynamics of the Russian ruble against the US dollar in the period 2013-2015. The results of econometric verification showed a very high quality of the model.

Thus, we can state the high practicality of models of fundamental determinants in relation to different currency pairs, based on both "real" (Russian ruble, US dollar, Canadian dollar, etc.) and "past" (German mark, French franc) currencies.

Conflict of Interest

The author declares no conflict of interest.

References

[1] A. Kuzmin, "Equilibrium Exchange Rate Modeling" in Eleventh International Conference "Management of large-scale system development" (MLSD 2018), Moscow, Russian Federation, https://ieeexplore.ieee.org/document/8551843/metrics#metrics, Publisher: IEEE, 2018, Electronic ISBN: 978-1-5386-4924-4. DOI: 10.1109/MLSD.2018.8551843.

[2] M. Obstfeld, K. Rogoff, "Exchange Rate Dynamics Redux" Journal of Political Economy, **103**(3), 624-660, 1995.

[3] R. Dornbush, "Expectations and Exchange Rates Dynamics" Journal of Political Economy, Vol.84, December, 1161-1176, 1976.

[4] M. Mussa, "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy Under Regime of Controlled Floating" Scandinavian Journal of Economics, May, **78** (2), 229-248, 1976.

[5] C.M. Engel, N.C. Mark and K.D. West, Exchange Rate Models Are Not As Bad As You Think, NBER Macroeconomics Annual, Vol. 22, 2007, revised in 2011.

[6] R. Nurkse, Conditions of International Monetary Equilibrium, 1945 in The International Monetary System: Highlights from Fifty Years of Princeton's Essays in International Finance, Kenen P.B. (ed.), Boulder, Colorado: Westvie Press, 1993.

[7] A. Kuzmin, "Exchange Rate Modeling: The Case of Ruble" Review of Business and Economics Studies, **3**(3), 39-48, 2015.

[8] M. Bahmani-Oskoee, G. Goswami, "Exchange Rate Sensitivity of Japan's Bilateral Trade Flows" Japan and the world economy, **16**(1), 1–15, 2004.

[9] L. Bartolini, Devaluation and Competitiveness in a Small Open Economy: Ireland 1987-1993, IMF Working Paper WP/93/82, Washington: IMF, November, 1993.

[10] F. Caramazza, French-German Interest Rate Differentials and Time-Varying Realignment Risk, Washington: IMF Staff Papers, Vol.40, September, 1993.

[11] A. Kuzmin, "A Structural Model of Exchange Rate Dynamics" Review of Business and Economics Studies, **2**(3), 86-92, 2014.
[12] R. H. Clarida, L. Sarno, M.P. Taylor, G. Valente, The Out-of Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond, NBER Working Paper, № 8601, Cambridge, Massachusetts: National Bureau of Economic Research , 2001.
[13] Y.-W. Cheung, M. Chinn, A.G. Pascual, Empirical Exchange Rate models: Are any Fit to Survive?, IMF Working Paper WP/04/73, Washington: IMF, 2004.
[14] I. Chowdhury, "Sources of Exchange Rate fluctuation: empirical evidence from six emerging market countries" Applied Financial Economics, **14**(1), 697-705, 2004.
[15] J. Chen, N.C. Mark, "Alternative Long-Horizon Exchange-Rate Predictors" International Journal of Finance and Economics, 1, 229–250, 1996.
[16] P.B. Clark, R. MacDonald, Exchange Rates and Economic Fundamentals: A Methodological Comparison of BEERs and FEERs, IMF Working Paper 98/67, Washington: International Monetary Fund, March, 1998.
[17] P.B. Clark, R. MacDonald, Filtering the BEER a permanent and transitory decomposition, IMF Working Paper 00/144 - Washington: International Monetary Fund, 2000.
[18] J. Williamson, Estimates of FEERs, In Estimating Equilibrium Exchange Rates, Williamson J. (ed.), Washington: Institute for International Economics, 1994.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 256-264 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Event Monitoring using Distributed Pattern Recognition Approach on Integrated IoT-Blockchain Network

Anang Hudaya Muhamad Amin[*,1], Nazrul Muhaimin Ahmad[2], Subarmaniam Kannan[2]

[1]*Computer Information Sciences Division, Higher Colleges of Technology, United Arab Emirates*

[2]*Centre of Intelligent Cloud Computing (CICC), Multimedia University, Malaysia*

ARTICLEINFO

Article history:
Received: 12 June, 2019
Accepted: 17 July, 2019
Online: 08 August, 2019

Keywords:
Event Monitoring
Internet-of-Things (IoT)
Blockchain Technology

ABSTRACT

With the advancement in the field of Internet-of-Things (IoT), event monitoring applications have rapidly evolved from a simple event data acquisition towards predictive event analytics involving multi-sensory data aggregation in a distributed environment. Existing event monitoring schemes are mainly relying on inefficient centralized processing mechanism, which may lead to the common single-point of failure for the entire system. In addition, there is no proper method for verifying the event data generated by the monitoring system. In this paper, we present a distributed event monitoring scheme using a Hierarchical Graph Neuron (HGN) distributed pattern recognition algorithm. HGN is a single-cycle learning graph-based recognition scheme that is modelled for in-network deployment. In this work, event data retrieved from multi-sensory IoT devices within a distributed event monitoring network is converted into pattern. To address the event data verification problem, we integrate our proposed scheme with blockchain technology. Combining this IoT event monitoring capabilities with blockchain-based data storage and verification could leads towards a scalable event detection and monitoring model for large-scale network. The results obtained from our simulation shows that the proposed scheme offers high event detection accuracy and capable of minimizing the event storage requirements on blockchain network.

1 Introduction

Since a decade ago, Internet-of-Things (IoT) has been extensively explored as a common platform for event monitoring. In industrial manufacturing applications, IoT plays an important role in monitoring and controlling processes. The monitoring systems designed for such applications require seamless interoperability between heterogeneous distributed IoT sensor devices and processing nodes, as well as an ability to analyze different types of sensor data.

The importance of such systematic interoperability may be conceptualized through an industrial scenario, wherein the plant maintenance is automatically scheduled by coordinating and synchronizing different sub-systems including distributed engineering, control, and maintenance sub-systems. Once a maintenance event is detected, the system can directly generate suitable downstream process controls, computed according to the applications requirements as to generate a correct response within the entire system.

A conventional way of performing event detection and monitoring usually involves a set of interconnected sensor devices that are linked to a centralized processing node (sink). These sensor devices made up the IoT infrastructure, enabling data acquisition process. With this method of implementation, common issues are related to communication and processing latency. Edge computing [1] is a form of computing model that enables data acquisition and processing to be performed within the close proximity of IoT devices.

Event monitoring involving critical applications such as environmental and disaster monitoring usually requires spatio-temporal data analysis for detecting unusual events or incidents. Machine learning and neural network algorithms are commonly being applied for such use cases. Nevertheless, these approaches involve complex computational procedures to be performed on resource-abundant processing nodes. In the context of large-scale and highly distributed IoT network, these approaches may not be suitable due to the resource-constrained specification of most IoT devices. Furthermore, massive data transmission may occur between the

[*]Anang Hudaya Muhamad Amin, Higher Colleges of Technology, P.O.Box 15825, Dubai, United Arab Emirates, +9712-2064772 & aamin@hct.ac.ae

sensor nodes and the processing node, due to large volume of data to be processed continuously.

A number of approaches have been carried out to enable complex machine learning and neural network algorithms to perform in a fully-distributed manner. However, most of the existing schemes suffer from the tightly-coupled processing mechanism in detecting and classifying event/non-event. Distributed algorithms that implement a graph-based approach show a promising way towards achieving purely-distributed processing mechanism.

Graph Neuron (GN) [2] is a distributed pattern recognition algorithm that follows the principles of graph theory with (vertex, edge) representation. GN algorithm has been implemented in several different case studies, including event recognition in wireless sensor networks (WSNs). GN performs a single-cycle in-network processing in its learning mechanism. As such, this improves its scalability against other graph-based pattern recognition schemes.

This paper is an extension of work originally presented in ICIAS 2018 [3]. In this paper, we extend our proposed idea on integrating a distributed pattern recognition algorithm and blockchain technology for event monitoring in IoT network. GN-based algorithm known as Hierarchical Graph Neuron (HGN) [4] is chosen as the distributed pattern recognition algorithm to be incorporated in the proposed IoT-blockchain infrastructure, enabling event detection and monitoring to be performed in a distributed manner, close to the event data acquisition point. Integrating blockchain network and IoT infrastructure allows the event signal or data from IoT devices to be stored and verified by blockchain nodes in the form of transaction using the distributed ledger mechanism. This approach provides a way to collect and preserve significant series of events that were captured and monitored by the IoT devices. Within this infrastructure, distributed pattern recognition could be implemented by identifying patterns of multiple sensor activation on IoT devices, and thus allowing complete transactions of event data to be recorded within the blockchain network.

The outline of this paper is as follows: Section 2 provides an overview on the current approaches towards distributed event monitoring. Section 3 focuses on the Graph Neuron (GN) based approach for event detection in a distributed environment. The model for integrated IoT-Blockchain platform for event monitoring is presented in Section 4. The proposed IoT-blockchain scheme with distributed pattern recognition for event monitoring is presented in Section 5. This section also includes description of the simulation works that have been performed. Finally, Section 6 concludes the paper.

2 Distributed Event Monitoring

A typical setup for distributed event monitoring network is shown in Figure 1. In this model, each IoT device act as an observer, which observe a stream of events. These observations would then be used for analysis involving either a simple function as collecting the number of observations or other complex computations such as identifying anomalies in the input distributions. Such computations usually being carried out by the coordinator node, who is overseeing the entire observations within the network.

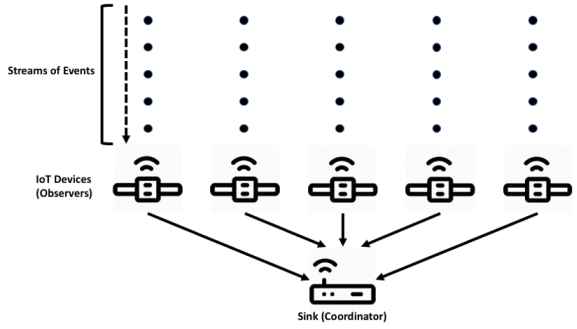


Figure 1: Distributed event monitoring setup with series of IoT devices as observers and processing node as coordinator.

There are a number of related works that demonstrate the distributed event monitoring activities carried out in different kinds of applications. These include the work by Wu et al. [5] and Ahmad Jan et al. [6].

In the survey conducted by Cormode [7], there are two approaches in deriving useful computations for the observations obtained from the sensor/IoT network. The first approach involves all the observers to simply send all the observations to a single and centralized coordinator. A drawback of such approach is such that it burdens the communication network between the coordinator and the observers. A second approach implements a periodic polling, in which at a particular time interval, the coordinator polls each observer for information on the observations and collates this information to get a snapshot of the current status since the last poll. A limitation of this approach is that it does not fit well, when considering event analysis using a non-linear computational model as in the continuous event monitoring applications.

The current centralized event monitoring systems that implement such coordinator-observer model can be seen in the works recorded by Ahmad Jan et al. [6] usually suffer from several limitations. Firstly, centralized event data processing can lead towards single-point of failure whereby failure of centralized node basically fails the entire event monitoring activity. Secondly, magnitude of data transmission to a central repository would be significantly increased as a result of massive number of IoT devices connected. This phenomena creates a connection gridlock from IoT devices to the centralized server.

Apart from the coordinator-observer model, other scheme such as presented by Basirat and Khan [8] utilizes the in-network processing capabilities that enable each IoT device to perform computation on observations within the body of the network, without relying on a centralized data processing node. With the in-network processing model, each IoT node is responsible for collecting the observations, as well as performing data aggregation. Analysis is carried out within the collaborative effort by each node, in exchanging critical information obtained from the observations.

Event monitoring with pattern recognition approach is widely applied in different application fields such as health informatics, manufacturing automation, and structural monitoring. Distributed pattern recognition algorithms such as Graph Neuron (GN) [9] and Vector Symbolic Architecture (VSA) [10] offer better schemes towards improved event

monitoring capability that runs on a distributed processing environment.

The proposed Graph Neuron (GN) based approach for event monitoring utilizes the in-network processing model that enables event identification to be conducted within the body of the network, rather than complex analysis and computations being performed at the processing node (coordinator). Discussion on how GN implements such processing model will be discussed in the next section.

3 GN-Based Approach for Event Monitoring

Graph Neuron (GN) is a single-cycle learning distributed pattern recognition algorithm that utilizes the in-network processing model [11]. The main characteristic of any distributed pattern recognition is such that it requires recognition of patterns to be implemented in a decentralized manner. Graph Neuron with similar feature, has been designed for fully-decentralized event detection scheme.

GN algorithm utilizes the graph-like pattern representation with each element of the pattern is represented with a *(value, position)* format. A variation of GN algorithm with hierarchical structure, known as Hierarchical Graph Neuron (HGN) was later developed by Nasution and Khan [4] that eliminates the crosstalk issue in GN implementation. In this paper, we present our proposed approach of utilizing HGN algorithm for distributed event monitoring on multi-sensor IoT devices.

3.1 Hierarchical Graph Neuron (HGN)

Hierarchical Graph Neuron (HGN) algorithm extends the functionality of Graph Neuron (GN) algorithm by increasing the pattern matching accuracy for highly-distorted patterns. HGN has been proven to eliminate the crosstalk issue found in GN implementation. The neuron firing in GN relies on the comparative function between stored and input elements on each neuron, which is called bias entry. These entries composed of *(value, position)* pairs that represent pattern elements. The approach is completely different from conventional neural network schemes whereby firing of neuron is based upon weight adjustment and calibration. Figure 2 shows a typical HGN network composition.

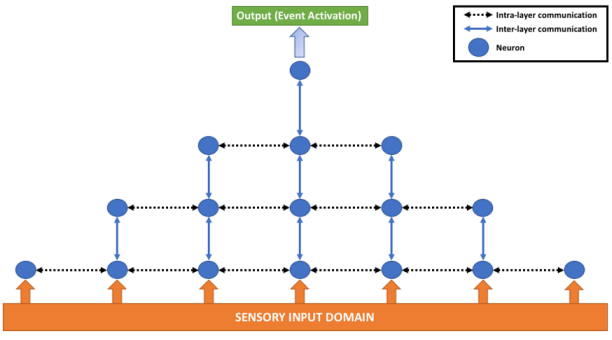


Figure 2: HGN network composition in a pyramid-like formation. Each neuron at the base layer corresponds to each input element within the sensory domain.

As shown in the figure, neurons composition in HGN network are arranged in a pyramid-like structure with base layer neurons act as input neurons to the network.

3.2 HGN Learning Mechanism

The neurons composition in each HGN network, n_{hgn} is represented by the following equation:

$$n_{hgn} = \left(\frac{p+1}{2}\right)^2 \tag{1}$$

Where p represents the number of input elements within the sensory input domain.

Following the graph-based structure, each neuron in HGN network composition forms a vertex that contains pattern element information (value or identification (ID)) while the adjacent inter-neuron communication is represented by the edge of a graph.

The HGN learning mechanism involves finding matched indices between input and stored patterns. In performing this comparison, each neuron compares its input pattern with the inputs obtained from its adjacent neurons. Adjacency information for each neuron is represented using the *(left, right)* formation. This adjacency information is known as bias entry and each neuron maintains an array of such entries. Each neurons bias array only stores the unique adjacency information derived from the input patterns. The pattern storage process involving N bias array sizes is represented in the following equation:

$$E^B = (\langle x, y \rangle; x \in N, y \in N) \tag{2}$$

Where E^B, comprises the sets of two-element ordered pair respectively, while x and y are the inputs within each entry. E^B can also be formed as one- or three-element ordered pair, depending upon the neuron location within the HGN hierarchical structure.

In retrieving matched bias entry, a linear search method is used on the adjacency information obtained for a given input pattern within each neuron through the bias array composition. The bias array entry is unique. Thus, the following equation shows a function to estimate the required number of comparisons for each input entry C_i, given n_{bEnt} number of bias entries in the array and r number of occurrences for each entry:

$$C_i = \frac{n_{bEnt} + 1}{r + 1} \tag{3}$$

In principle, HGN implements a single-cycle learning for recognizing patterns, through implementation of a graph-matching scheme on its pattern representation. Within HGN network composition, each neuron performs a forward propagation of index values obtained from the matching process as shown in Figure 3, from the base layer neurons towards the top neuron. The top layer neuron will decide on the event/non-event status based on the entire pattern recall search on the bias array composition. The procedures only involved a one-pass cycle for each input pattern, without any iterations involving value alteration as to obtain a recognition output.

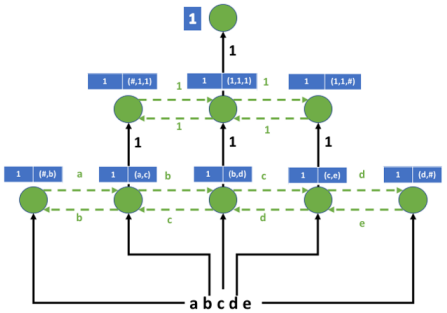


Figure 3: Neuron firing mechanism within the HGN network composition. Neuron activation is achieved through value comparison between adjacent neurons.

3.3 HGN for Event Monitoring

With distributed and lightweight features of HGN, an event detection scheme for IoT network can be carried out at the IoT device level. It could act as a front-end middleware that could be deployed within each IoT nodes in the network, forming a network of event detectors. Hence, the processing scheme minimizes the processing load at the sink and provides near real-time detection capability. Preliminary work on distributed HGN integration for distributed networks such as wireless sensor network (WSN) has been conducted by Muhamad Amin and Khan [12].

In integrating HGN within event monitoring, we have considered mapping each HGN processing cluster into each IoT node, with the assumption that each node is having more than one sensor inputs. Our proposed scheme is composed of a collection of IoT nodes and a sink. We consider a deployment of IoT network in two dimensional plane with M sensors, represented by a set $M = (m_1, m_2, ..., m_n)$, where m_i is the ith sensor. The placement for each of these sensors is uniformly located in a grid-like area, $A = (x \times y)$, where x represents the x-axis coordinate of the grid area and y represents the y-axis coordinate of the grid area. Each IoT node is assigned to a specific area of the grid as shown in Figure 4. The location of each sensor node is represented by the coordinates of its grid area (x_i, y_i).

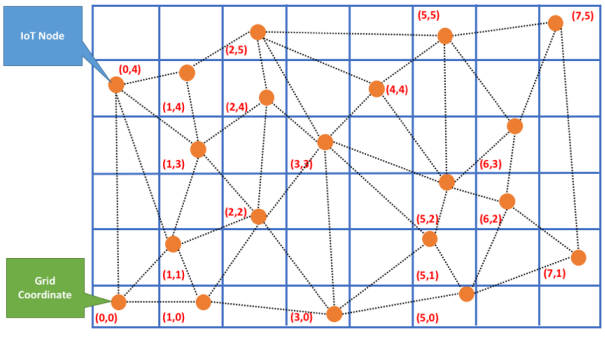


Figure 4: Cartesian grid layout used to identify location of IoT nodes in a two-dimensional space.

For communication model, HGN uses a single-hop mechanism for data transmission from IoT node to the sink. The use of *autosend* approach has been selected following the work by Saha and Bajcsy [13], to minimize error due to the loss of packets during data transmission. HGN event processing scheme involves minimal transmission of event data from IoT nodes to the sink, due to the ability for the front-end processing.

4 Integrating Blockchain in Distributed Event Monitoring

In this section, we present an overview of blockchain and its integration with the distributed event monitoring framework on IoT network.

4.1 Blockchain Technology

Blockchain provides a capability for verification of credentials and transactions using a distributed ledger mechanism. Such capability is of interest, especially when considering its integration with event monitoring system in IoT network.

Implementation of blockchain is not strictly limited to cryptocurrencies. In fact, there are many differences between cryptocurrency implementation and blockchain for industrial applications [14], including immutability and provenance. Blockchain has also been a point of interest in the field of shared economy applications (e.g. Uber, AirBnB) [15]. Nevertheless, past implementations of blockchain come with several limitations, including slow block generation period as mentioned in the Ethereum White Paper [16] and lack of loop implementation (Turing incomplete) [17]. Blockchain enables a decentralized mechanism for verification that allow participating nodes in the network to validate unique transactions through a distributed ledger (block). Figure 5 shows the decentralization approach in blockchain, in comparison with existing centralized approach.

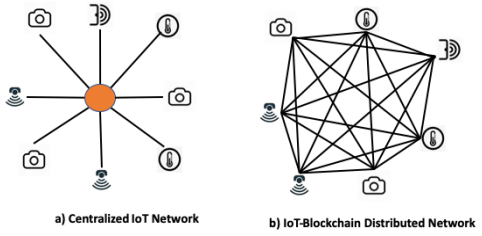


Figure 5: A comparison between conventional IoT network with IoT-Blockchain distributed network.

One of the important requirements for IoT integration on blockchain network is such that it requires fast validation and verification to be performed on blockchain network. Existing blockchain network such as Ethereum [18] has made this possible. Ethereum offers a smart contract deployment, as shown in Figure 6. It captures details of transactions including the value and its corresponding state. Developers can write a program that runs on top of Ethereum. With Ethereum as a blockchain platform, we can configure IoT devices to connect to this main network. This infrastructure would allow authentication of IoT devices through public key

infrastructure. IoT devices can utilize Ethereum platform for updating their action and behavior.

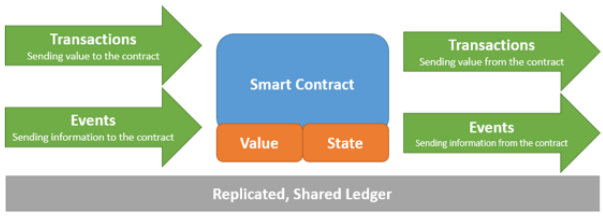


Figure 6: Smart contract deployment on Ethereum platform that captures details about transactions and events.

4.2 Device Synchronization

One of the important issues to be addressed in IoT network implementation within the IoT-Blockchain environment is device synchronization. As the number of devices increased, the need for proper synchronization mechanism is important. A work presented by Huh et al [19], demonstrated the synchronization of IoT devices on Ethereum blockchain network can be obtained using smart contract feature, which specify certain action and behavior of IoT devices. The smart contract can be used to place a code to indicate the detection of event by the IoT device.

In this paper, we demonstrate the use of blockchain network for the purpose of event data monitoring and verification. Blockchain offers immutability to the event data being captured from IoT network. Within the proposed model, sensor data are converted into activation patterns as shown in Figure 7.

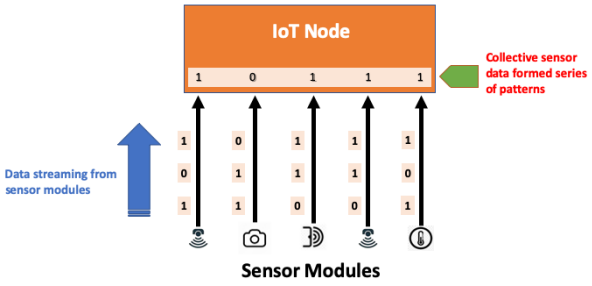


Figure 7: Inputs from sensor modules are treated as a pattern, to be processed by the IoT node.

From the temporal perspective, this sensor data patterns vary across time. Our aim is to capture unique event data to be stored and verified on blockchain network. HGN distributed pattern recognition scheme is used to detect unique event, prior to storage and verification on blockchain network. In this work, similar data patterns would be treated as non-event and will be discarded. The hypothesis for this work is such that the block generation could be minimized accordingly.

5 Proposed Infrastructure

In this section, we present our proposed IoT-blockchain with distributed pattern recognition framework for event detection and monitoring. In addition, this section will also include details on the simulation works and corresponding results involving forest fire detection using HGN distributed pattern recognition scheme.

5.1 System Model

The proposed event monitoring infrastructure incorporates a distributed pattern recognition on IoT- blockchain network. Hierarchical Graph Neuron (HGN) has been selected as the distributed pattern recognition algorithm to be used in our model. In this implementation, assumption is made as such that each IoT node is capable of performing event data collection. Each IoT node is linked to a blockchain-edge (BC-Edge) node that hold a distributed ledger, as part of the blockchain network. Figure 8 shows the proposed infrastructure for HGN deployment in IoT-blockchain network.

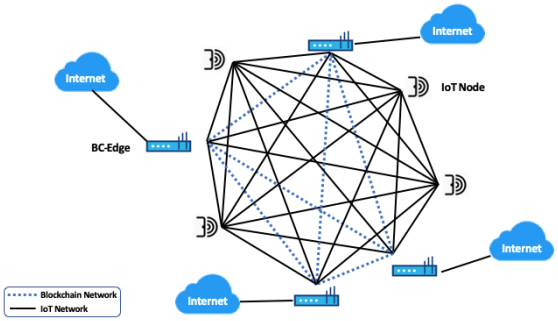


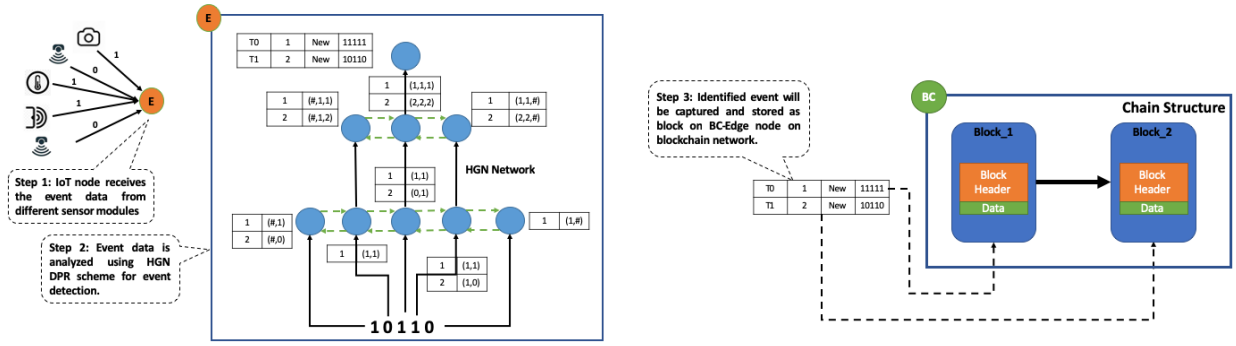
Figure 8: IoT-Blockchain network infrastructure for distributed event monitoring.

Note that the proposed scheme considered the use of multi-sensory IoT devices. Event monitoring is basically carried out at the device level, in which each device is capable of performing event detection using the HGN distributed pattern recognition algorithm. Figure 9 shows the detail event detection and verification procedures on IoT-blockchain network.

The network configuration used in this work is based upon the premise that all the nodes acquired communication capability with low latency in a full-mesh network structure.

In the proposed design, an IoT node N will establish its identity, and broadcast it to the BC-Edge nodes, B. For a given specific time interval T, the IoT nodes will communicate the event data (result from event activation using HGN scheme). This will then be processed by the BC-Edge node as event data for T. The analysis produces an output in the form of combined pattern index, event status, and event pattern as shown in the examples in Table 1. The event status contains information on the event, whether it is classified as new or recalled. Using this approach, only newly-recorded event will be stored and verified in the blockchain network.

As shown in Table 1, at T_2, event detected by IoT node *X02* is recalled and will be discarded. Similarly, at T_3, event



(a) Event data monitoring and detection.

(b) Unique event storage and verification on blockchain network.

Figure 9: Event monitoring with HGN distributed pattern recognition on integrated IoT-blockchain network.

Table 1: Outputs recorded by BC-Edge node, retrieved from the IoT nodes. Note that only unique event pattern will be stored in the blockchain network

Timestamp	IoT Node ID	Event Pattern	Index	Status
T_1	X01	10101	1	new
T_1	X02	11111	1	new
T_2	X01	10111	2	new
T_2	X02	11111	1	recall
T_3	X01	10111	2	recall
T_3	X02	11110	2	new

recorded by *X01* is recalled. Only unique event pattern will be captured and stored in the distributed ledger.

5.2 Event Monitoring Simulation

To demonstrate our proposed distributed event monitoring scheme, we designed a simulation work on event data analysis using the HGN algorithm. The analysis was conducted on the data related to forest fire, which contains 517 records as reported by Cortez and Morais [20]. Each event pattern detected was stored and verified using the SHA-256 secure hash algorithm on blockchain network.

We extracted five important features that contribute towards possibility of fire ignition, namely the ISI index, temperature, relative humidity (RH), wind and rain. For the purpose of simulation, the dataset is transformed into a binary representation using an interactive binning method. The range of values for each bin was specified according to the potential event (binary value 1) and non-event (binary value 0) as shown in the Table 2.

Table 2: Data values in different range used to identify each parameter as event.

Parameter	Data Range 1	Data Range 2
ISI	$x \geq 42.075$	$x \leq 14.025$
Temperature	$x \geq 23.325$	$x \leq 7.775$
Relative Humidity (RH)	$x \geq 63.75$	$x \leq 21.25$
Wind	$x \geq 6.75$	$x \leq 2.25$
Rain	$x \geq 4.8$	$x \leq 1.6$

Our assumption is such that the event data is continuous and recorded in a timely manner from T_0 till T_{516}, respectively.

5.3 Results and Discussion

In this section, we present the results of our simulation on event pattern recognition using the Hierarchical Graph Neuron (HGN) approach. The proposed scheme capable of detecting 19 unique events from the total of 517 event records. Consequently, the number of data blocks generated is reduced by 97%. This was made possible by identifying unique event patterns from the dataset. Figure 10 shows the event count distribution based on the 19 unique events detected using the proposed HGN distributed pattern recognition scheme.

Table 3 shows samples of event data for each corresponding unique event detected using our proposed HGN recognition scheme.

Apart from implementing the event data classification using HGN approach, we extended our analysis to examine the accuracy of our proposed recognition scheme on the simulated dataset. Based on the observation made, there were 40 event data that were misclassified as false positive (duplication of indices). As a result, the recognition accuracy for HGN in this simulation is at the rate of 92.3% with error rate of 7.7%. Further details on the misclassification can be seen in [21].

The simulation work that has been carried out demonstrate a high recall accuracy for HGN distributed pattern recognition implementation. The error rate for the event classification obtained is considerably low, as it only indicates value around 7.7%.In addition, the results prove that the number of unique events detected is significantly reduced. This is critical for ensuring efficient data block generation

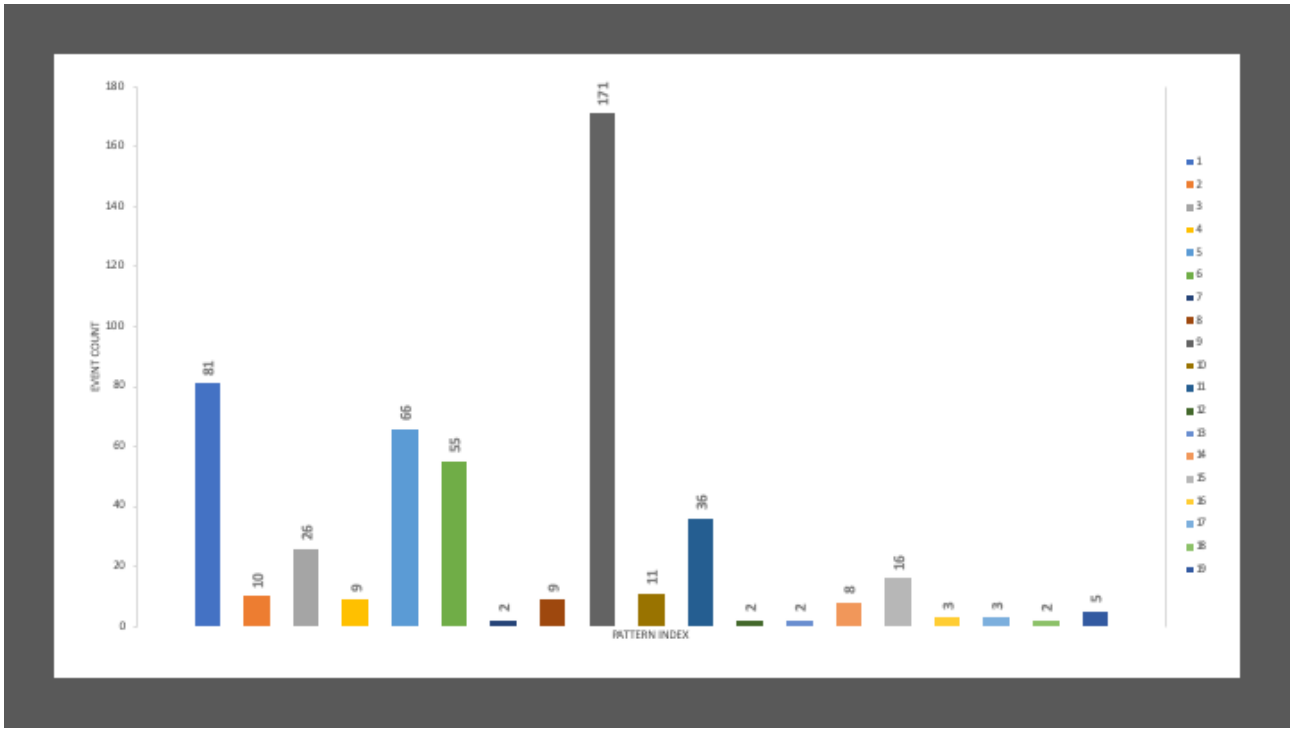


Figure 10: Unique event patterns detected using HGN distributed pattern recognition algorithm from the total of 517 event data recorded.

Table 3: Samples of forest fire event data that has been classified as unique event using HGN distributed pattern recognition algorithm.

Event ID	ISI Index	Temperature (°C)	Rel. Humidity (RH) (%)	Wind (km per hour)	Rain (mm per m^2)
1	5.1	8.2	51	6.7	0
2	7	21.3	42	2.2	0
3	5.8	23.4	22	2.7	0
4	6.2	12.9	74	4.9	0
5	17	20.1	40	4	0
6	7	21.6	33	2.2	0
7	8.9	18.4	42	6.7	0
8	15.9	25.9	24	4	0
9	10.7	10.3	74	2.2	0
10	7.8	17.4	24	5.4	0
11	3.8	15.2	51	2.7	0
12	8.1	29.6	27	2.7	0
13	6.7	18.4	25	3.1	0
14	14.3	19.1	53	2.7	0
15	14.1	33.1	25	4	0
16	10.8	26.4	35	2.7	0
17	14	30.8	30	4.9	0
18	17.7	26.4	34	3.6	0
19	14.3	27.3	63	4.9	6.4

within the blockchain network.

5.4 Future Work

Based on the simulation work that has been carried out, it was clearly indicated that the HGN distributed pattern recognition scheme is able to provide accurate classification of unique events. Our future work is basically to fully integrate the event classification module with blockchain network for event data storage and verification. The use of open-source hyperledger fabric platform [22] under the Linux Foundation Projects will be considered.

We intend to expand our research into effective scheme for data storage on blockchain. As the amount of data arises from events monitored on IoT network, the need for efficient data storage mechanism is important. At a preliminary stage, we propose a two-type storage mechanism, comprising event digest and event data. Event digest will be permanently stored in the blockchain network in the form of recorded blocks (on-chain), while the event data will be stored in conventional databases (off-chain). In implementing this scheme, it is expected that the storage requirement for event data could be further reduced.

In addition to the storage optimization, we would also aim to study on different consensus mechanisms for data verification on blockchain network. These include the proof-of-work (PoW) and proof-of-stake (PoS). The consensus mechanism enables peers and users within the network to verify the authenticity of the data that is stored on the blockchain network. Results of the study would assist us in implementing more efficient IoT-blockchain infrastructure for event monitoring on ioT network.

6 Conclusions

In this paper, we presented our on-going work on distributed event monitoring using a distributed pattern recognition algorithm on integrated IoT-Blockchain network. The proposed scheme implements a HGN recognition algorithm that performs a distributed event detection. Accuracy of HGN pattern recognition may improves the efficiency of distributed ledger storage mechanism in blockchain network. This enables unique event data to be stored and verified in a distributed environment. The simulation work being carried out shows that HGN algorithm exerts high recall accuracy with one-cycle learning. Furthermore, the algorithm is highly-suitable for in-network deployment.

HGN recognition approach also minimizes the needs for storing large amount of continuous event data on blockchain network. The detection mechanism in the proposed scheme also consider the entire state of the IoT network, rather than examining individual sensor input values.

For our future development, we intend to look into several aspects of IoT-blockchain integration, including the efficiency and effectiveness of the proposed scheme, through the scalability factor of the system in implementing multi-heterogeneous IoT-blockchain network. Apart from this, different consensus mechanisms in blockchain technology will be reviewed to analyze their effectiveness in the proposed IoT-blockchain scheme.

References

[1] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, 2016.

[2] Mohamed Baqer, Asad I Khan, and Zubair A Baig. Implementing a graph neuron array for pattern recognition within unstructured wireless sensor networks. In *International Conference on Embedded and Ubiquitous Computing*, pages 208–217. Springer, 2005.

[3] Anang Hudaya, Muhamad Amin, Nazrul Muhaimin Ahmad, and Subarmaniam Kannan. Integrating distributed pattern recognition technique for event monitoring within the iot-blockchain network. In *2018 International Conference on Intelligent and Advanced System (ICIAS)*, pages 1–6. IEEE, 2018.

[4] Benny B Nasution and Asad I Khan. A hierarchical graph neuron scheme for real-time pattern recognition. *IEEE Transactions on Neural Networks*, 19(2):212–229, 2008.

[5] Kangheng Wu, Xiaokang Xiong, Bert W. Leung, Jihyoun Park, and Zhibin Lei. Event processing of monitoring data of large hi-tech manufacturing equipment: Debs grand challenge. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, DEBS '12, pages 387–392, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1315-5. doi: 10.1145/2335484.2335535. URL `http://doi.acm.org/10.1145/2335484.2335535`.

[6] Mian Ahmad Jan, Priyadarsi Nanda, Xiangjian He, and Ren Ping Liu. A sybil attack detection scheme for a forest wildfire monitoring application. *Future Generation Computer Systems*, 80:613 – 626, 2018. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2016.05.034. URL `http://www.sciencedirect.com/science/article/pii/S0167739X16301522`.

[7] Graham Cormode. Continuous distributed monitoring: a short survey. In *Proceedings of the first international workshop on algorithms and models for distributed event processing*, pages 1–10. ACM, 2011.

[8] Amir H Basirat and Asad I Khan. Graph neuron and hierarchical graph neuron, novel approaches toward real time pattern recognition in wireless sensor networks. In *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, pages 404–409. ACM, 2009.

[9] Asad I Khan and Patrik Mihailescu. Parallel pattern recognition computations within a wireless sensor network. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 777–780. IEEE, 2004.

[10] Simon D Levy and Ross Gayler. Vector symbolic architectures: A new building material for artificial general intelligence. In *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, pages 414–418. IOS Press, 2008.

[11] Ram Kumar, Vlasios Tsiatsis, and Mani B Srivastava. Computation hierarchy for in-network processing. In *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*, pages 68–77. ACM, 2003.

[12] A. H. M. Amin and A. I. Khan. Parallel pattern recognition using a single-cycle learning approach within wireless sensor networks. In *2008 Ninth International Conference on Parallel and Distributed Computing, Applications and Technologies*, pages 305–308, Dec 2008. doi: 10.1109/PDCAT.2008.47.

[13] Sunayana Saha, Peter Bajcsy, et al. System design issues in single-hop wireless sensor networks. *Proc. of 2nd IASTED ICCIIT 2003*, 2003.

[14] Gideon Greenspan. Four genuine blockchain use cases. *MultiChain [blog]*, 10, 2016.

[15] Steve Huckle, Rituparna Bhattacharya, Martin White, and Natalia Beloff. Internet of things, blockchain and shared economy applications. *Procedia computer science*, 98:461–466, 2016.

[16] Vitalik Buterin et al. Ethereum white paper. *GitHub repository*, pages 22–23, 2013.

[17] Vitalik Buterin. Toward a 12-second block time. *Ethereum Blog*, 2014.

[18] Gavin Wood et al. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151:1–32, 2014.

[19] Seyoung Huh, Sangrae Cho, and Soohyung Kim. Managing iot devices using blockchain platform. In *2017 19th international conference on advanced communication technology (ICACT)*, pages 464–467. IEEE, 2017.

[20] Paulo Cortez and Aníbal de Jesus Raimundo Morais. A data mining approach to predict forest fires using meteorological data. 2007.

[21] Anang Hudaya Muhamad Amin, Sujni Paul, Fred N Kiwanuka, Imtiaz Ahmad Akhtar, et al. Improving event monitoring in iot network using an integrated blockchain-distributed pattern recognition scheme. In *International Congress on Blockchain and Applications*, pages 134–144. Springer, 2019.

[22] Hyperledger fabric - hyperledger. `https://www.hyperledger.org/projects/fabric`. Accessed: 2019-07-12.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 265-270 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

A Novel Strategy For Prompt Small Cell Deployment In Heterogeneous Networks

Dorathy Abonyi*

Department of Electrical and Electronic Engineering, Enugu State University of Science and Technology, Nigeria

ARTICLEINFO

Article history:
Received: 27 May, 2019
Accepted: 11 July, 2019
Online: 08 August, 2019

Keywords:
Heterogeneous networks
RSSI
Direction-of-arrival estimation

ABSTRACT

Popularity and affordability of smart phones and other data hungry devices add exponentially to the traffic demand of existing cellular networks. Cell densification and small cell deployment over existing macrocell has been identified as an effective solution to high traffic demand predicted for future wireless networks like 5G. Small cells are deployed over existing macrocells in a Heterogeneous network to offload traffic, ensure balanced load and good quality of service in the network. To achieve the purpose of HetNet, small cells are deployed in locations of high user concentration. Operators are required to be able to identify these locations for optimum small cell deployment. Presently, this decision is based on long term traffic data which create latency in small cell deployment especially in areas of unexpected hotspots. The work presented in this paper is a new strategy that can practically help operators to promptly identify locations for small cell deployment based on user cluster. This strategy involves a monitoring system that can be easily incorporated in a mobile cellular network base transceiver station (BTS) to monitor the coverage area and periodically identify positions for prompt and optimum small-cell deployment in HetNets. The monitoring system is a two element array system that uses RSS data to resolve DoA, range, user cluster and identify positions for small cell deployment. At any point in time, the network operator can visualize cluster locations relative to the base station and is able to identify hotspots. When small cells like micro-, pico-, or femto-cells are promptly deployed in areas of higher user concentration, high data rates and good quality of service in the network are maintained at all times. The network coverage area is divided into a series of azimuthal and range sectors, and users are located into the sectors where they belong. The strategy was simulated in MATLAB environment for WiFi networks with thirty users and results indicate 100% *accurate DoA estimation and MAE in range estimation of* 6*m for a known environment. Optimum positions for small cell deployment based on user clusters were correctly determined. Simulation was validated with experiment and result indicate close relationship between experiment and simulation with MAE difference in range estimation of* 1*m.*

1 Introduction

The exponential growth in data traffic demand on cellular networks and the promising solution offered by heterogeneous network and small cell deployment has triggered a lot of research interest in this area. It is obvious that homogeneous networks involving macro- only cells are insufficient in solving the capacity requirement of the present and future networks like 5G. This is due to high cost of macrocell BTS and the unavailability of cell sites especially in city centers. Small cells like micro-, pico- and femtocells are therefore deployed within the existing macrocell and integrated as a heterogeneous network. These small cells are cheaper to implement and site acquisition is also easier to get since even an existing electric pole or building roof is a potential site. Deploying small cells within existing macrocell gives rise to a big umbrella cell with a number of small cells, each positioned to serve a smaller coverage area within the big

*Department of Electrical and Electronic Engineering, Enugu State University of Science and Technology, Nigeria, +2348037576332 & abonyi.dorathy@esut.edu.ng

cell as shown in Figure 1.

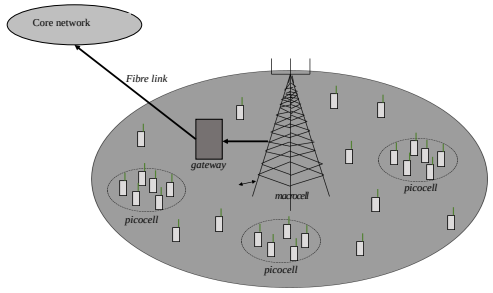


Figure 1: Heterogeneous network showing an umbrella macrocell with deployed picocells forming a 2-tier HetNet

Three main aim of heterogeneous network are (1) To offload traffic from macrocell, (2) Ensure balanced load in the network and (3) Ensure good quality of service to users. To achieve these objectives, small cells are deployed in a well targeted manager in most cases within locations of high user concentration (HUC). This therefore requires the network operators to be able to identify locations of HUC. Presently, these locations are determined using collected traffic data over a period of time and locations with high traffic demand is taken as HUC and position for small cell deployment. This approach does not give opportunity for prompt traffic offload, balanced load and good QoS using small cells especially in unexpected hotspot cases. This paper is an extension of the work originally presented in International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) [1] which introduced the system based on simulation only. The work presented in this paper presents the application of the previous work as a strategy for prompt identification of optimum position for small cell deployment. Both simulation and experimental validation are presented to demonstrate the applicability of proposed strategy.

Strategies for small cell deployment can broadly be classified as Random Deployment Strategy (RDS) and Deterministic Deployment Strategy (DDS). RDS in most In RDS, small cells are randomly deployed within the macrocell while in DDS, the position for small cell deployment is determined based on some standing factors. Network performance for DDS and RDS was compared in [2] and their result indicate that DDS is a better option. Well targeted small cells has been shown to enhance QoS, offload traffic from macro-cell [3], ensure balance network load [4], improve energy saving capability and generally optimize performance [5] of the network. DDS can further be classified into uniformly distributed strategy (UDS), Cell edge strategy (CES) and user aware strategy (UAS). In UDS, a predetermined pattern of small cell deployment is followed for all macrocells owned my the network operator [6]. For CES approach, Small cells are arranged around the macrocell edge [7]. In UAS, locations of users are considered for small cell deployment based on usage or population. Normally, these data are obtained from long term traffic information [8] which does not provide prompt response to capacity requirement of hotspots especially when they are unexpected. Hotspots are locations of high demand characterized by high user concentration. A mobile small cell deployment strategy for unexpected hotspots which will ensure prompt response to traffic was proposed in [9] but with the assumption that network operators are able to promptly identify hotspot locations. To the best of our knowledge, there is no existing small cell deployment strategy with hotspot identification which is the contribution of this paper.

2 Problem Statement and Proposed Strategy

Considering a macrocell of a cellular network with 360° coverage. Looking at only 180 degrees coverage of this network as illustrated in Figure2 with base station (BS) located as indicated with red dot. Assuming there are so many users causing traffic overload on the existing macrocell and low quality of service (QoS) resulting to customer dissatisfaction. Some locations in this network are characterized with obvious clusters of users within the cell causing an overload on the macrocell base transciever station servicing this area. The strategy proposes a monitoring system to be deployed on the BS tower to monitor the coverage area so that it divides the area into azimuthal and range sectors as shown in Figure 2. By locating users into the individual azimuth-range classes, positions with HUC can be identified for small cell deployment. With blind or random placement of small cells in this network, the uncertainty of achieving the objective of HetNet will be high but with a monitoring system, it is almost certain that optimal placement of small cell based on user concentration will be achieved.

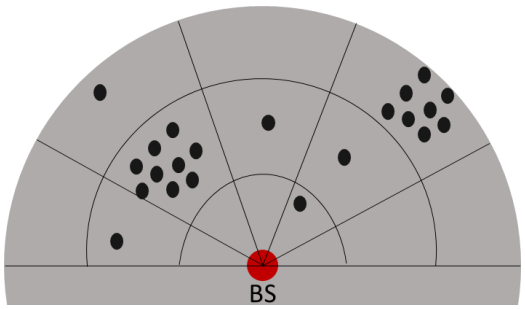


Figure 2: The proposed strategy for prompt small cell deployment

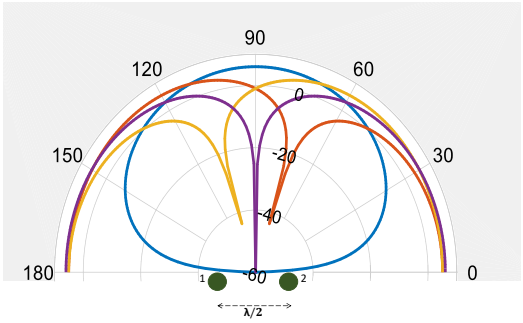


Figure 3: The proposed strategy for prompt small cell deployment

Assuming a two element array system is used for monitoring so that a directional main beam is produced on az-

imuth and steered to other azimuth locations as shown in Figure 3. This creates azimuth locations differentiated by main beam positions for angle of arrival localization of network users. By sharing the network coverage into range classes relative to the BS and estimating user range, users are located into different azimuth-range classes and users clusters can be identified.

3 User Cluster Localization and Optimum Position for Small Cell Deployment

To localize users by cluster, the system first estimates the DoA of each user. This is based on the azimuth position from where the maximum RSS was calculated when compared with measured RSS from other beam steered azimuth positions. It should be noted that the estimated DoA is not a precise estimate but considering the application, only a sector estimate is required. Second the system estimates range of each user using a range estimation model that is based on RSS-distance relationship. By fitting a curve in least square sense, a mathematical model that is applied for user range estimation is obtained. Combining estimated DoA and range, users are classified into created azimuth-range classes of the network.

3.1 Simulation

To test the proposed strategy, thirty users were randomly distributed within a simulated network coverage area as shown in Figure 4. The aim is to determine locations of user cluster and optimum positions for small cell deployment based on number of users in each cluster.

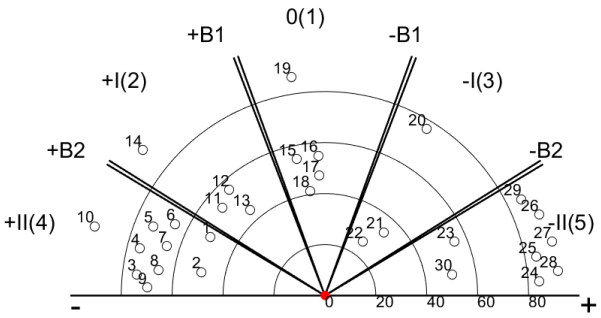


Figure 4: Actual location of users before localization

Assuming the network channel to be FSPL environment, the system acquires data for localization by calculating user RSS from all four main azimuth positions. Figure 5 is a plot of calculated RSS from all four main beam positions. From the calculated RSS data, it can be seen that different RSS values are measured from each azimuth position with a unique maximum value for each user. Using the calculated RSS, user DoA is calculated based on the azimuth position with highest calculated RSS for each user. Estimated DoA result of each user is shown in Figure 6.

Comparing the result of Figure 6 with the actual user locations of Figure 4, it can be seen that all users were correctly located to the azimuth position where they belong. Considering the accuracy of DoA estimation as the ability to locate users into each of the created azimuth sectors, this result gives 100% DoA estimation accuracy.

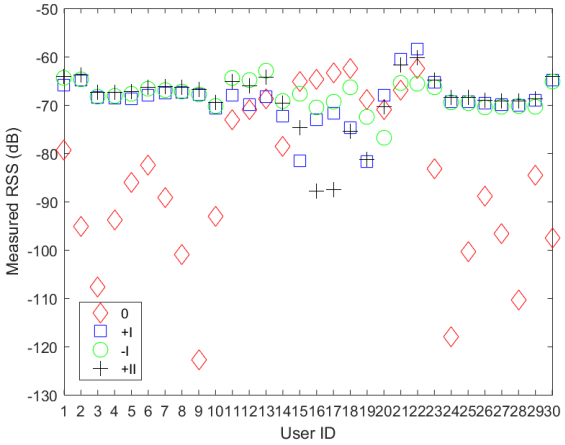


Figure 5: System simulated RSS from user devices in four main azimuth positions showing possibility of classifying DoA based on maximum RSS

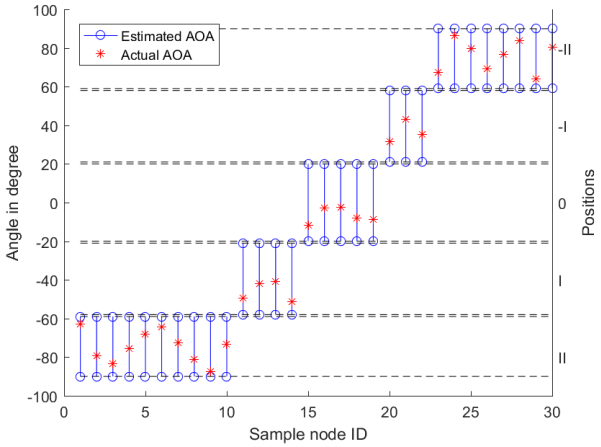


Figure 6: Estimated user DoA showing correct azimuth position localization of all users in the Network

To test the range estimation, four different environments, FSPL, HATA, ITU and WINNER defined by their empirical pathloss models of Equations 1, 2, 3 and 4 respectively were considered.

$$PL_{FSPL} = -27.5 + 20\log(f) + 20log(d) \quad (1)$$

$$PL_{HATA} = -35.4 + 20\log(f) + 26log(d) \quad (2)$$

$$PL_{ITU} = -28 + 20\log(f) + 20log(d) \quad (3)$$

$$PL_{WINNER} = -29.8 + 20\log(f) + 21.5\log(d) \quad (4)$$

where f is radio wave frequency in MHz, d is distance between BTS and user in meters. These environments were modeled and range estimation model developed for each environment. Using the measured RSS from FSPL environment, the four range estimation models for FSPL, HATA, ITU and WINNER were applied to estimate user range. Figure 7 shows the actual user range with the result of estimated range using these range models. To determine which of

the network location area is the optimum position for small cell deployment, users are classified into each of the sectors based on the estimated DoA and range. By counting the number of users in each class, user population per class is determined. Based on a preset threshold of number of users for small cell deployment depending on the type of small cell for the network, the optimum position for small cell deployment is determined. Assuming a micro-picocell network where a picocell is deployed in positions with atleast six users, optimum positions for small cell deployment in the simulated network are obtained as shown in Figure 8.

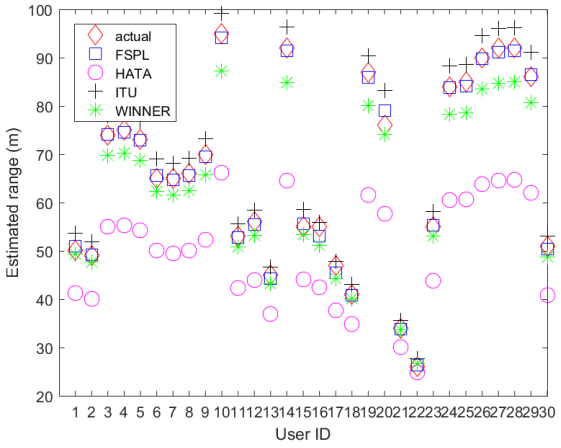


Figure 7: Estimated range of users using all four environment range estimation models for calculated RSS in FSPL environment

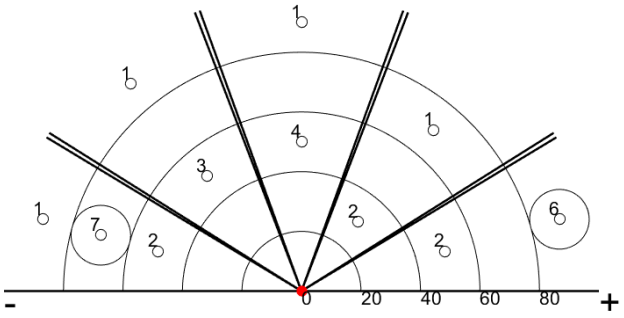


Figure 8: User cluster localization and optimum small cell deployment position identification

The range estimation result of Figure 7 gives a mean absolute error (MAE) of approximately 6m, 16m, 10m and 7m using FSPL, HATA, ITU and WINNER range models respectively. FSPL environment has given the least error in estimation because RSS was calculated in FSPL environment. This means that if the environment is known, the system is able to determine range estimation up to an accuracy of 6m but if the environment is unknown, the error in range estimation can increase up to 16m. By visually observing the result of Figure 7 and considering range estimated using FSPL range estimation model, it can be seen that users with ID numbers 1 to 10 and 23 to 30 located at azimuth positions +II and -II respectively were located with error in range estimation lying between 0 and 1m. Users with ID 11 to 14 and 20 to 22 located at azimuth positions +I and -I respectively were located with range estimation error between 1m and 37m. This is an indication that in a switched beam based localization systems, range estimation is azimuth position dependent. This will further be investigated in future work. Result of Figure 8 shows that user cluster and positions for small cell deployment can be determined.

3.2 Experimental Validation

Experiment was performed outdoor in the open field space of Ponderosa park in Sheffield, UK using a network of WiFi nodes as users. Figure 9 is a picture of the experimental setup. Figure 10 shows a model of the experimental environment with deployed users in their actual positions.



Figure 9: Outdoor experimental setup showing some of the distributed users on red and deployed observer system on yellow

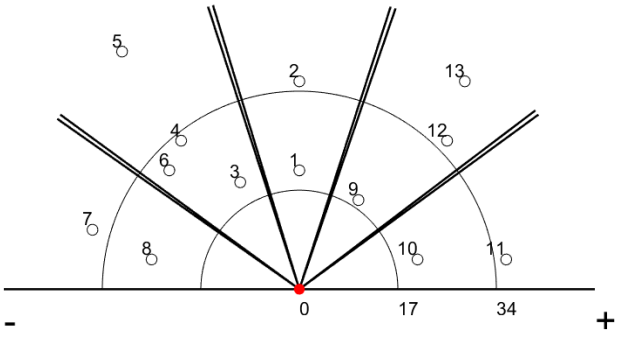


Figure 10: A model of the actual user position before localization

Ten RSS of these users were measured from a reference point where the monitoring system was positioned. Mean of the experimental measured RSS for each user from four main azimuth positions were obtained and plotted as shown in Figure 11. Using the measured RSS, DoA of users were estimated and result shown in Figure 12. Using the same measured RSS, user range were estimated using four range models representing four different environment scenarios. Result of Figure 13 shows actual, ITU simulated and experimental estimated range of users using all environment models. Next is to locate user clusters and identify locations for small cell deployment based on set threshold. Since only few (thirteen) users were used in this experiment, it was assumed that locations with up to two users are qualified for small cell deployment. Due to the small experimental area, coverage area was shared into three range sectors. Combining the estimated DoA and range, users are classified into

clusters by counting the number of users in each azimuth-range sectors of the coverage area. Figure 14 is result of experimental localized user clusters.

4 Result Analysis and Discussion

Comparing the actual user azimuth positions and the DoA estimation result of Figure 12, it can be seen that all users were correctly localized to their azimuth positions. Result has shown 100% accuracy for the used sample of users which validates the simulation DoA estimation result.

Result of range estimation shown in Figure 13 gives a mean absolute error (MAE) between actual and experiment of 2.51m, 3.56m, 2.19m and 2.54m for FSPL, HATA, ITU and WINNER open space range models respectively. This shows that the environment of RSS measurement is closely related to ITU model than other used range models. Comparing simulation and experiment, MAE in range estimation between experiment and simulation in ITU environment is 1.36m. This means that the mean deviation of experiment from simulation is approximately 1m and 2m from actual range. This shows that range estimation is achievable with manageable error levels. Apart from HATA which has a model that is well deviated from others, all other models are on the average, about 2.5m close to actual range. Deviation of experiment from simulation is just approximately 1m which means that this strategy is applicable and would work as modeled in simulation with a mean difference of $\leq$ 1.5m of range estimation in most known environments.

Comparing result of Figures 14 with the actual user positions, four small cell deployment locations were identified instead of two. This is due to over estimation of two of the users as a result of noise in measurement. This error may have been caused by the drizzling weather condition during this experiment. Also with few users and small number of users for small cell deployment as used here, this is expected but with a practical higher number of users required for small cell deployment in a real cellular network, this discrepancies will be reduced. The possibility of user cluster and small cell deployment position identification is promising with the results obtained from this experiment.

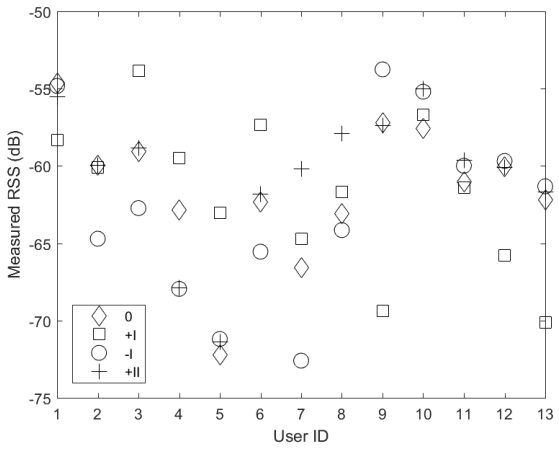


Figure 11: Measured RSS from user devices at all azimuth positions

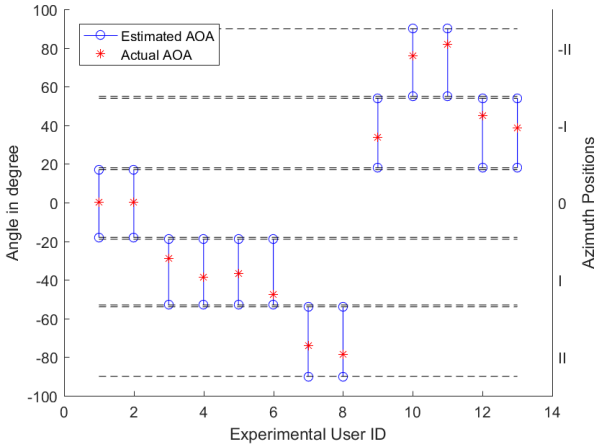


Figure 12: Experimental estimated DoA of users

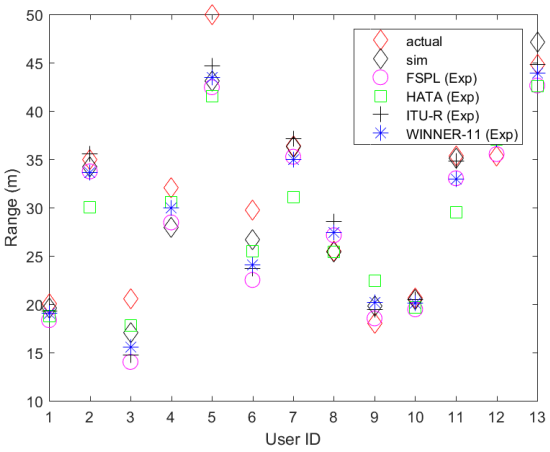


Figure 13: Experimental range estimation showing actual, simulated and experimental estimate

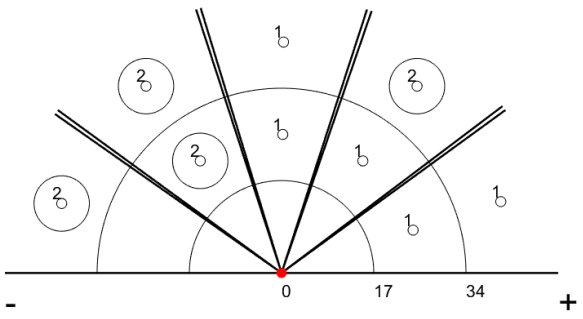


Figure 14: Experimental user cluster localization and optimum small cell deployment position identification

5 Conclusion

The work presented in this paper describes a novel and practical strategy for prompt small cell deployment in a heterogeneous network. This strategy involves using a monitoring system to periodically monitor the network coverage area of a wireless network and identify locations for small cell deployment based on user cluster localization. The entire

network coverage area is shared into azimuth and range locations for user cluster localization. By applying DoA and range estimation, user relative position is estimated. Number of users in each class is determined to identify optimum positions for small cell deployment. Both simulation and experimental results have shown that this strategy is applicable and user clusters can be correctly identified with minimal error.

Conflict of Interest The authors declare that there is no conflict of interest.

Acknowledgment I wish to acknowledge Enugu State University of Science and Technology and the Federal Government of Nigeria through Tertiary Education Trust Fund (TETF) for sponsoring this research.

References

[1] D. Abonyi and J. Rigelsford, "A System for Optimizing Small-Cell Deployment in 2-Tier HetNets," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, sep 2018, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8514981/

[2] G. Su, L. Li, X. Lin, and H. Wang, "On the optimal small cell deployment for energy-efficient heterogeneous cellular networks," in *2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, jul 2014, pp. 172–175. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6876775

[3] A. Nika, A. Ismail, B. Y. Zhao, S. Gaito, G. P. Rossi, and H. Zheng, "Understanding data hotspots in cellular networks," in *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. IEEE, aug 2014, pp. 70–76. [Online]. Available: http://ieeexplore.ieee.org/document/6928662/

[4] M. H. Qutqut, H. Abou-zeid, H. S. Hassanein, A. M. Rashwan, and F. M. Al-Turjman, "Dynamic small cell placement strategies for LTE Heterogeneous Networks," in *2014 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, jun 2014, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/6912536/

[5] S. Landstrom, H. Murai, and A. Simonsson, "Deployment Aspects of LTE Pico Nodes," in *2011 IEEE International Conference on Communications Workshops (ICC)*. IEEE, jun 2011, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/document/5963602/

[6] Y. Park, J. Heo, H. Kim, H. Wang, S. Choi, T. Yu, and D. Hong, "Effective Small Cell Deployment with Interference and Traffic Consideration," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE, sep 2014, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6965951

[7] J. Hoadley and P. Maveddat, "Enabling small cell deployment with HetNet," *IEEE Wireless Communications*, vol. 19, no. 2, pp. 4–5, apr 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6189405/

[8] I. Bahceci, "On femto-cell deployment strategies for randomly distributed hotspots in cellular networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, apr 2014, pp. 2330–2335. [Online]. Available: http://ieeexplore.ieee.org/document/6952713/

[9] N. E, X. Chu, and J. Zhang, "Mobile Small-Cell Deployment Strategy for Hot Spot in Existing Heterogeneous Networks," in *2015 IEEE Globecom Workshops (GC Wkshps)*. IEEE, dec 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7414060/

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 276-284 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Localization of Emerging Leakages in Water Distribution Systems: A Complex Networks Approach

Matteo Nicolini[*,1,2]

[1]*Polytechnic Department of Engineering and Architecture, University of Udine, 33100 Udine, Italy*

[2]*Idrostudi, AREA Science Park, 34149 Padriciano, Trieste, Italy*

ARTICLE INFO

Article history:
Received: 12 May, 2019
Accepted: 20 July, 2019
Online: 07 August, 2019

Keywords:
Correlation
Degree centrality
Pressure gage
Water loss

ABSTRACT

Water distribution networks are infrastructural systems designed for providing potable water to consumers. In these last decades, the importance of assessing and identifying emerging leakages has become a primary issue, because of the high level of water loss characterizing such systems worldwide. In this paper, a new approach aimed at the prompt localization of leakages occurring in water distribution systems is introduced. The methodology relies on the analysis of real-time pressure measurements and on Complex Networks Theory. Starting from a collection of nodes representing the locations of pressure sensors, links of a virtual, complex network are created on the basis of the values assumed by correlation coefficients between pressure measurements: if such values are above a given threshold, relevant nodes are considered to be connected to each other. In this way, information about the structure and topology of the complex network is easily derived. In particular, the degree centrality of the nodes is a key parameter allowing to identify the position of a leakage. The paper first analyzes a well-known literature example, and then proves the high reliability of the methodology for a real water distribution system.

1. Introduction

Water distribution systems (WDSs) are strategic infrastructures for the transport and the delivery of potable water to various types of customers [1]. However, the deterioration of ageing components (especially pipes and pumps), the rapid growth of urbanization and the statutory and contractual quality standards that have to be guaranteed to customers are playing a fundamental role in the decision-making process, especially for the increasing costs due to the operational management. In particular, water loss is a widespread problem in all the countries of the world: for a well-managed and controlled WDS, the level of water leakage can be less than 10% [2], but percentages of 40-50% are not uncommon even in developed countries [3,4].

Water utilities have to continuously monitor and control the functioning of the system: besides financial aspects and issues related to the interruption of service, energetic costs and environmental impacts are a major concern, making water loss identification and reduction one of the most challenging tasks [5].

[*]Matteo Nicolini, via del Cotonificio, 114, +39 0432 558742;
Email: matteo.nicolini@uniud.it

In this paper, a new approach for the early identification of emerging leakages in a WDS is presented. The novelty of the methodology resides in the analysis of pressure measurements through Complex Networks Theory [6]. The reliability of the method increases with the number of installed pressure gages, up to the ideal situation of having one sensor at every node of the WDS.

In this work, the assumption that pressure data are available at every junction of the system is made (a simulation model has actually been used in order to calculate them). In a real situation, such signals arrive from pressure gages properly installed in the field, and correlations between them should provide the information about the emergence of leakages to be detected. The aim of the study is to show the capability of the methodology of localizing a leakage since its first formation, provided the above hypotheses are satisfied. However, the methodology is still valuable even in real cases, characterized by a limited amount of measure points: in such situations, the zone with the most likely presence of a leakage can be identified.

Anyway, the proposed approach does not intend to solve the problem of precisely localizing a leakage in a real WDS. Instead,

the methodology represents a valuable alternative to other prelocalization techniques, or a useful tool to be adopted in conjunction with other approaches.

The rest of the paper is organized as follows: section 2 contains a literature review, section 3 describes the methodology, which is applied to a literature system in section 4; section 5 shows the results obtained from the application to a real WDS, and section 6 draws some concluding remarks.

2. Literature Review

Leakages in WDSs are usually classified into two main categories [7]: bursts, which are characterized by sudden activation and high flows, and background leakages, which do not surface and are low, background flows that can persist for years without being detected.

Since bursts can be identified by instrumentation (sometimes they are visible on the ground), they are repaired in a short period of time, leading to a small amount of water lost. In this context, several technologies are available: step-tests, noise correlators, gas-injections, acoustic sensing through conduit pigging, and others: see [8] for a review.

Instead, background leakages can determine huge volumes of water lost, unless some dedicated leakage identification and repairing activities are performed. In these last years, many water utilities have implemented pressure management, an approach consisting in the introduction and regulation of some pressure reducing valves (PRVs) with the aim to control the piezometric surface. The selective reduction of pressure in a WDS (typically during the nighttime) may result in considerable water and energy savings [9,10].

Other approaches rely on proactive identification of water leakages, in order to keep the level of water loss always under control. To this end, District Metered Areas (DMAs) are created, consisting in the subdivision of a WDS in continuously monitored portions [11]. The installed instruments (flow and pressure gages) allow to provide a nearly real-time estimation of water loss on a 24-hour basis, by analyzing the minimum night flow (MNF) occurring in the lowest consumption interval, that is, between 2:00 a.m and 4:00 a.m.. Such minimum demand conditions determine the maximum values of pressure in the system, and hence the highest values of leakage.

A totally different approach relies on transient methods, consisting in high-frequency analysis of pressures transients in a WDS after some surge has been created [12]. Others rely on a transient network simulation model, usually very difficult to calibrate [13].

In these last years, the rapid improvement in sensor technology and in data transfer and communication systems has greatly enhanced the activity of real-time control of water losses, although many problems remain unsolved: first of all, the possibility of identifying a leakage since its first formation [14].

Several researchers have been focused on the comparison between actual field measurements and the results obtained by a calibrated numerical model [15–17]. In such cases, the key-point is the level of accuracy of the simulation model, which can hardly be optimal, since it should also contain information on the leakages to be discovered. In particular [18] have introduced a methodology for leak-detection and localization coupled with demand calibration.

More recently, Complex Networks Theory (CNT) has received increasing attention for the comprehension of a wide spectrum of real systems, ranging from physical infrastructures to social communities [19,20]. Successful examples include functional (correlation) network approaches [21], to infer hidden statistical inter-relationships between macroscopic regions of the human brain [22] or the Earth's climate system [23], difficult to uncover with traditional non-linear time series analysis techniques [24].

The application of CNT to the design and operation of WDSs has attracted a growing number of researchers, because its inherent capability of unveiling hidden properties, not grasped by traditional analyses or modelling approaches [25]. CNT has been adopted for evaluating the topological characteristics and the resilience of a system [26], or for expansion strategies [27]. Several authors investigated vulnerability-related issues, like node vulnerability under cascading failures [28–30], spectral methods to establish vulnerability areas [31,32], or for evaluating robustness under random or intentional attacks [33]. Other studies focused on the analysis of the formation of isolated communities [34], on the segmentation of WDSs for the identification of District Metered Areas (DMAs) using general metrics and modularity [35,36]. The optimal sampling design has also been addressed with the modularity concept [37] and through a combination of classical optimization and CNT [38]. More recently, Complex Networks Theory has been adopted for a systematic classification of WDSs [39], and for their optimal design through a tradeoff between network cost and reliability, measured in terms of flow entropy [40].

All such studies have been mainly focused on the topological aspects of WDSs, deriving their main characteristics from the analysis of the real system. Actually, many hidden properties may be uncovered looking at functional or correlation ties between nodes, especially when simulation or field data are available.

3. Methodological Approach

Starting from a given WDS whose topological, geometrical and hydraulic characteristics are known, the simulation model of the system can be built. In this work, the software Epanet has been adopted, being an open-source and a standard toolkit for such kind of numerical analyses [41].

The two sets of equations the software solves at the generic time t_k are the continuity equation for each junction i:

$$\sum_j Q_{ij}(t_k) - D_i(t_k) = 0 \qquad (1)$$

and the flow-headloss relationship in every pipe connecting junctions i and j:

$$H_i(t_k) - H_j(t_k) = h_{ij}(t_k) = rQ_{ij}^n(t_k) + mQ_{ij}^2(t_k) \qquad (2)$$

where Q_{ij} is the flow in pipe ij, D_i is the demand at node i, H_i and H_j are nodal heads at junctions i and j, h_{ij} is the headloss between nodes i and j; r, n and m are, respectively, the resistance coefficient, the flow exponent and the minor loss coefficient. For each time t_k,

starting from known heads at the fixed grade junctions (typically, reservoirs or tanks), the software uses the gradient method [42] in order to solve the set of equations (1) and (2) for determining all the heads and flows. Once the head is known at each junction, the pressure is directly calculated as the difference between head and the elevation of that junction (the velocity head at each junction is neglected).

The simulation model allows to determine the variability of pressure, provided a user demand pattern is defined. A pressure 'signal', $p_i(t)$, may be associated at each junction i, and the correlation coefficient c_{ij} between every pair of pressures signals at junctions i and j can be calculated. In this paper, the Pearson correlation coefficient has been adopted; it is given by the following expression:

$$c_{ij} = \frac{\sum_{k=1}^{T}(p_i(t_k)-\bar{p}_i)(p_j(t_k)-\bar{p}_j)}{\sqrt{\sum_{k=1}^{T}(p_i(t_k)-\bar{p}_i)^2 \cdot \sum_{k=1}^{T}(p_j(t_k)-\bar{p}_j)^2}} \qquad (3)$$

in which t_k represents the discretized time of the numerical module, T is the number of time steps of the simulated time horizon, and $\bar{p}_i$ is the average pressure at the i-th junction.

The temporal variability of pressure depends mainly on the fluctuation of the customers' demand: the higher the request of water, the lower is the pressure. Usually, distribution systems supplying water to residential areas are subjected to an almost spatially-uniform user demand. In this way, the temporal variability of pressure is very similar at all junctions, and the correlation coefficient c_{ij} is very close to the value of one. In other words, there is a strong linear relationship between pressure at junctions, properly described by the (linear) Pearson correlation coefficient.

However, the presence of many tanks or control devices, such as pressure regulating valves, inverter of pumps and other facilities, may introduce non-linearities in the relationships between pressures, and the correlation coefficient may not represent an appropriate measure. The more general case, which will be evaluated in future research, can be analyzed through non-linear correlation parameters or, more simply, looking for subsets of the considered time horizon for which linear correlations exist between pressure signals.

The methodology starts by creating a similarity matrix in which every element c_{ij} is the Pearson correlation coefficient between real-time measurements of pressure sensors (or pressures calculated by the simulation model) at i and j.

From the similarity matrix, an adjacency matrix representing a virtual, undirected complex network may be built (Figure 1). In this network, the nodes are the points of measurement and the links are created if the correlation coefficient between pressure signals is high enough. In other words, if c_{ij} is above a chosen threshold, θ, the related element a_{ij} of the adjacency matrix is 1, and zero otherwise (elements in the diagonal are set to zero).

In an ideal system with no leakages and characterized by a spatial uniform demand pattern, that is, the same typology of customers' consumption, there is a strong correlation between pressure signals, and the related complex network is heavily connected (it is a complete graph). As an example, Figure 1 (left) shows the situation in which four sensors are installed at junctions 3, 9, 28 and 31. In the case of no leakage, the similarity matrix presents very high values of correlation coefficients, and the adjacency matrix is that typical of a complete graph (Figure 1, top right).

The formation of a leakage starts to 'break' correlations among the node nearest to the leakage and the others, inducing a progressive link removal, depending on the amount of water loss. This is due to the fact that only the time-varying flow out of the leakage is a function of pressure, with different dynamics from customer's demand.

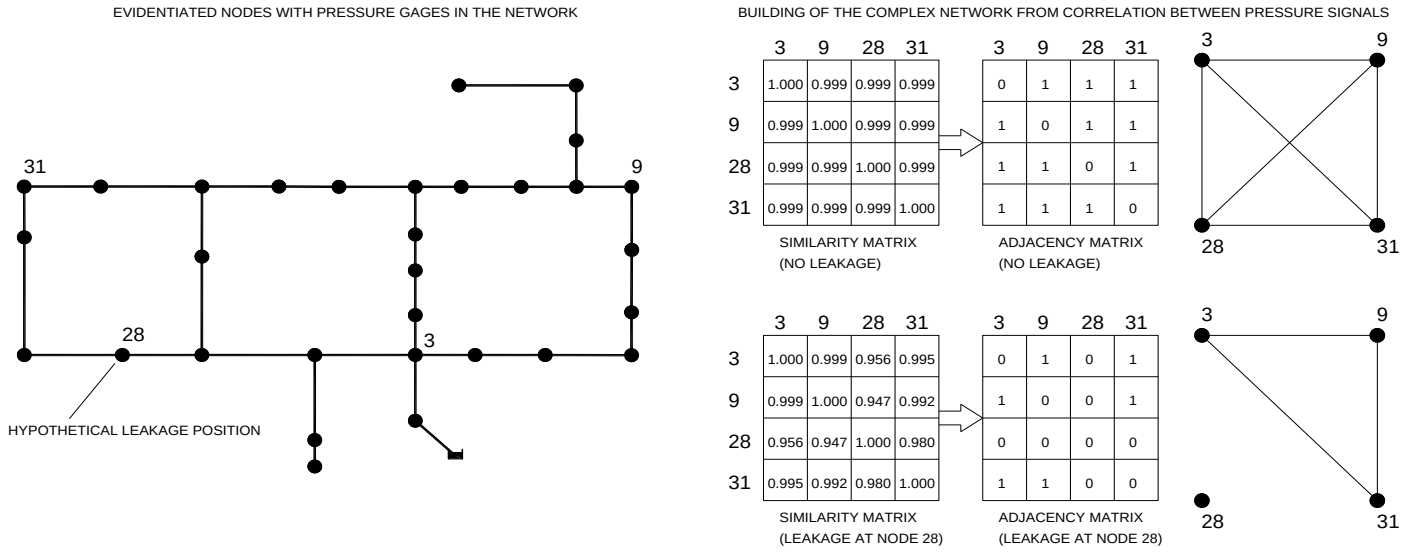


Figure 1: Difference between similarity matrix, made by Pearson correlation coefficients, c_{ij}, and adjacency matrix, in which the value of 0 or 1 assumed by each element depends on the threshold chosen for c_{ij} (in the example θ = 0.990). In this case, the size of matrices is less than the number of junctions of the water distribution system, being equal to the number of installed pressure gages (indicated on the left of the figure). The hypothetical position of a leakage is also shown, together with its impact on the matrices and the resulting virtual network.

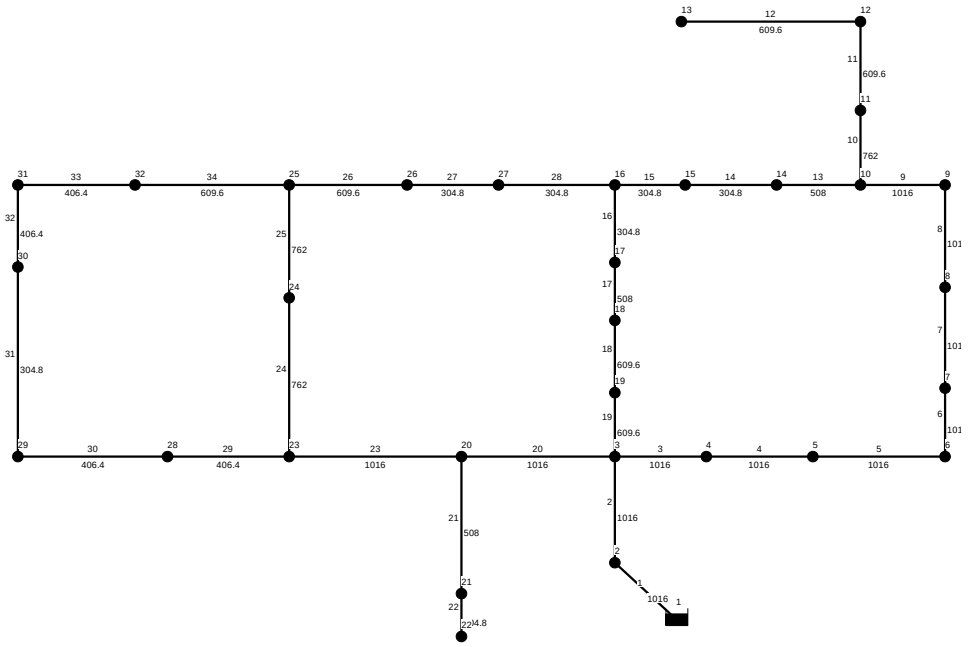


Figure 2: *Hanoi WDS adopted for the analyses. Link number and diameter (in mm) are indicated along the pipes*.

If a leakage is formed at the *i*-th junction of a WDS, the total discharge $Q_i(t_k)$ outflowing at time t_k equals the nodal demand, and can be expressed as the sum of consumer's request and pressure-dependent water loss:

$$D_i(t_k) = Q_i(t_k) = \alpha(t_k) q_i + c[p_i(t_k)]^u \qquad (4)$$

in which q_i is the average demand at node i, $\alpha(t_k)$ is the multiplier coefficient characterizing the demand pattern at time t_k (Figure 3), $p_i(t_k)$ is the pressure at node i, and c and u are the leakage coefficient and the leakage exponent: they determine, respectively, the leakage entity and its dependence on the pressure (for an ideal leakage of circular shape on a steel pipe, $u = 0.5$).

Figure 1 (bottom right) shows the example of a leakage present at junction 28: it is easily seen that the values of the correlation coefficient in the similarity matrix decrease when the position with the leakage is involved in the calculation, giving rise to a less connected network.

The parameter adopted for the identification of the junction in which a leakage is forming is the degree centrality: the degree of a node represents the number of its nearest neighbors (in other words, the number of its connections). For an undirected network (as in this case) the column vector of node degrees' **k** is given by (the symbol **1** represents an all-one column vector):

$$\mathbf{k} = (\mathbf{1}^T \mathbf{A})^T = \mathbf{A}\,\mathbf{1} \qquad (5)$$

in which **A** is the adjacency matrix [17]. Thus, the junction in which the leak is emerging is the one characterized by the lowest correlation with respect to all the others, and the related node in the complex network has the lowest degree centrality.

4. Application to a Literature System

The methodology has been applied to the Hanoi WDS (Vietnam), since it represents a well-known literature example (Figure 2): it was introduced for the first time by [43] and then analyzed by many authors [44]. The model of the main system consists of 32 junctions and 34 pipes, organized in three loops (see [43] for the data). The numerical simulation model has been used to artificially create pressure signals at every junction of the main system, where a pressure sensor is considered to be available.

Figure 3 shows the pattern assumed for customers' demand: its variability is typical of residential consumption, with two peaks during the day, respectively in the morning and in the evening, and a minimum during the nighttime. The time resolution adopted is 5 minutes.

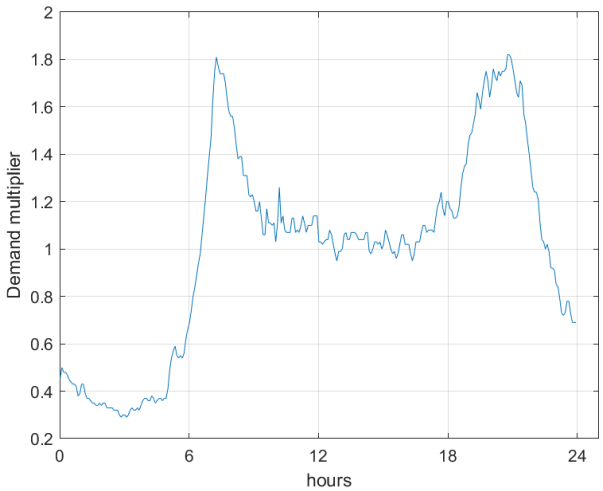


Figure 3: Temporal variability of demand multiplier $\alpha(t_k)$ in (4), describing customers' demand.

In Figure 3, the typical oscillations due to the stochastic nature of demand are evident. In the present paper, the simple case of uniform spatial distribution of demand pattern has been assumed. Due to the importance of correctly simulating users' demand variability, future research will focus on such issue: to this end, the

approach adopted by [45,46] appears the most suitable for real-time analyses. The time variability of pressure in the case of no leakage in the system is shown in Figure 4.

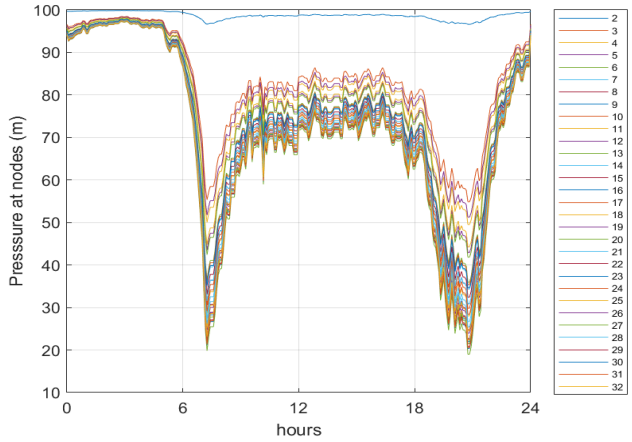


Figure 4: Pressure variability over 24 hours for the junctions of Hanoi WDS.

The results have been obtained by simulations performed with Epanet software [39], with a hydraulic time step of 5 minutes and a Hazen-Williams roughness coefficient of 130 for all pipes.

The correlation between signals is evident, and may be confirmed by the plot of Figure 5, where the relationship between pairs of nodes keeping junction 2 in abscissa are shown. Similar results may be plotted for other pairs of junctions. Such behavior confirms the validity of linearity assumption between pressures

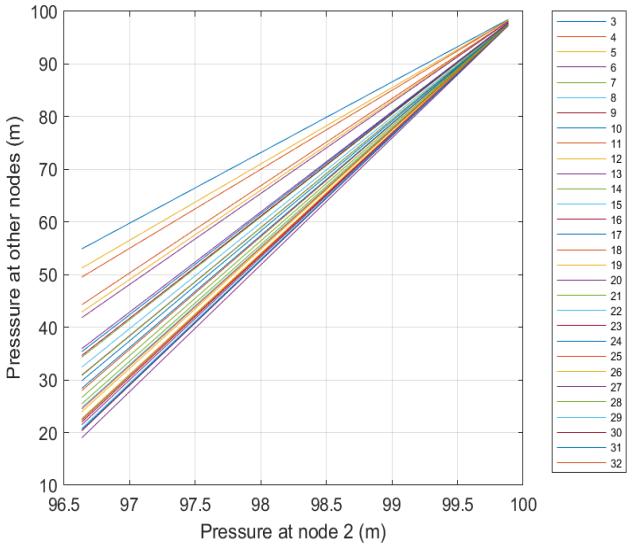


Figure 5: Correlation plot for pairs of pressures between junction 2 and others

Figure 6 is a heatmap plot of the global results obtained with reference to one leakage in all possible positions, and characterized by an average outflow of 5 l/s, which may appear to be large, but actually is a relatively small value if compared to the user demand at junctions (100 l/s on average). For each column, representing an assumed position of a leakage, the figure shows the degree centrality of the nodes on a blue-tone basis, i.e., the darkest the color, the bigger is the degree of a node.

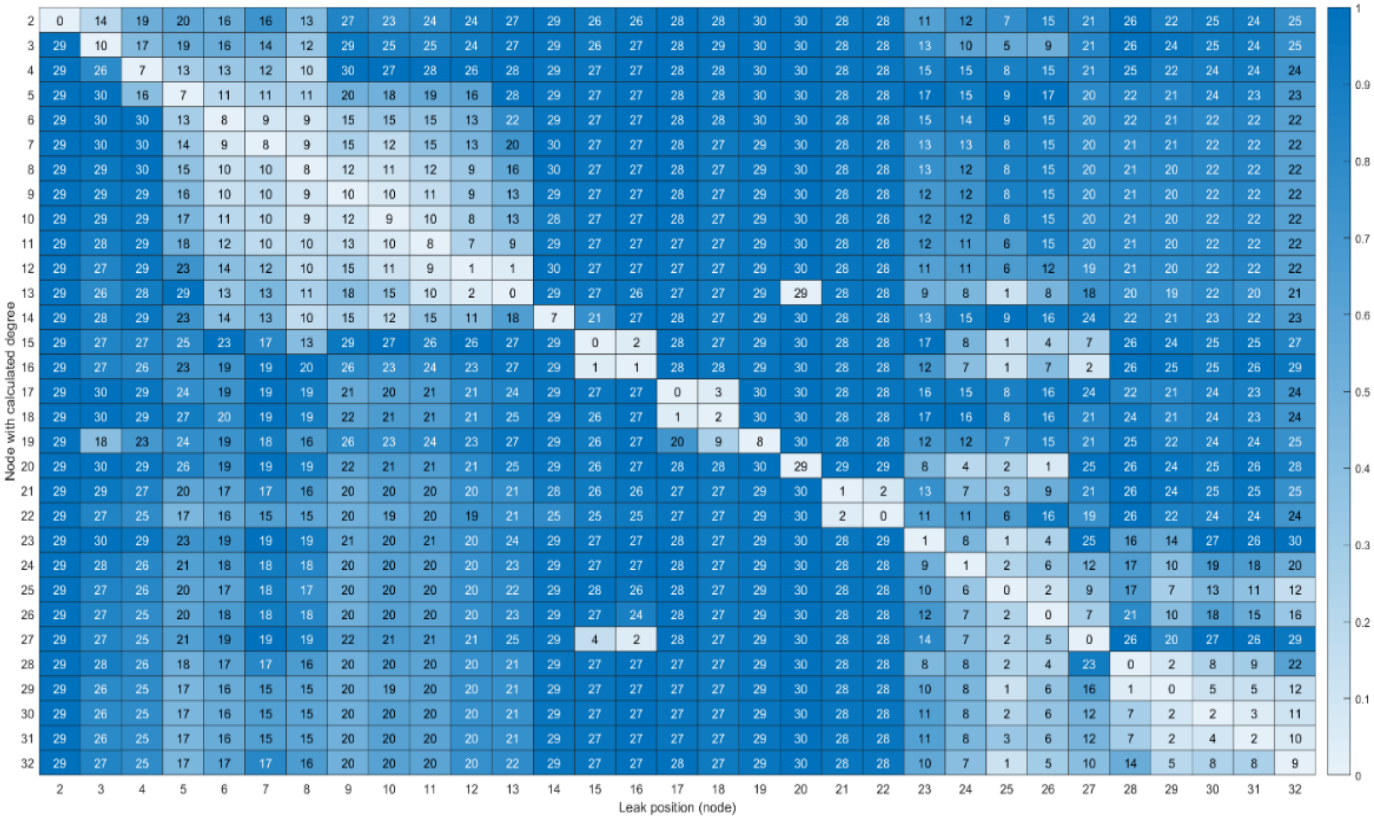


Figure 6: Heatmap representation of degree centralities when varying the position of one leakage of 5 l/s in Hanoi WDS. Number in columns are the node degrees for a fixed leakage position

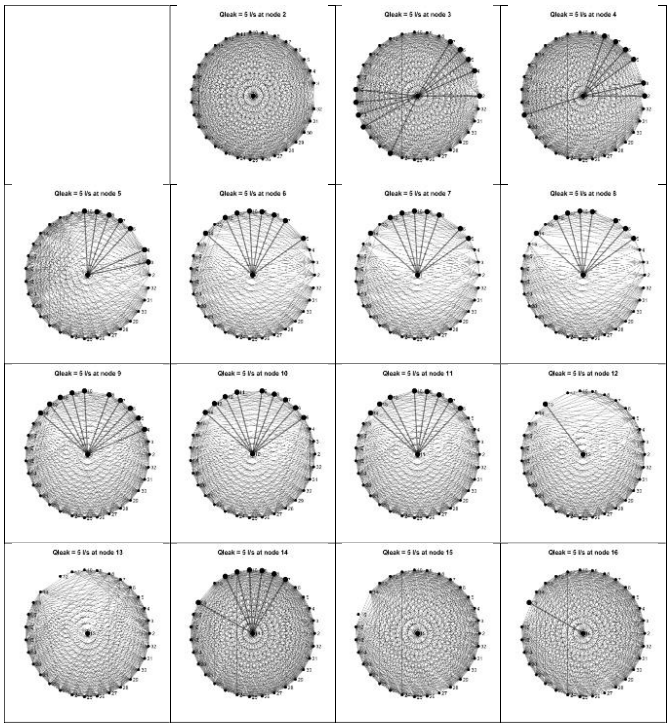


Figure 7: Round plots of virtual correlation networks obtained when varying leakage position (nodes 2 to 16). The junction where the leakage is activated is drawn at the center of each plot. Thick lines are the links connecting the central node to others, while other connections are indicated with thin lines.

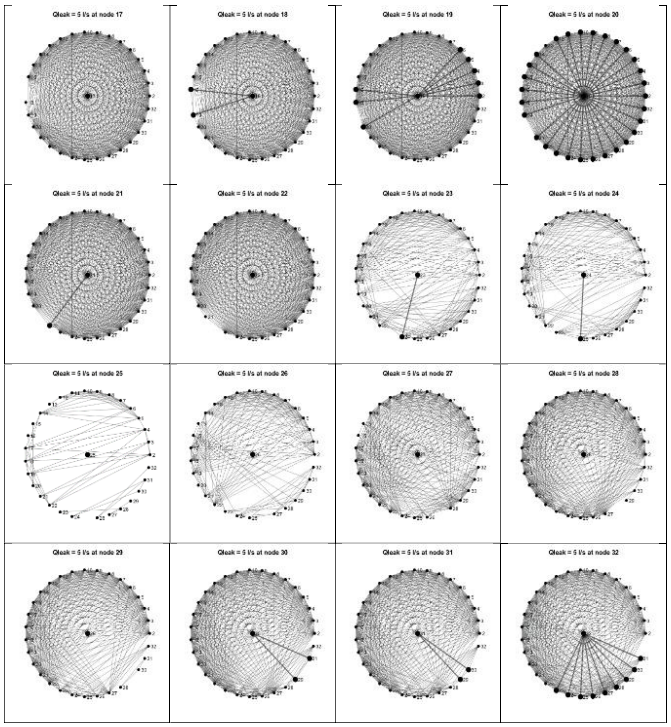


Figure 8: Round plots of virtual correlation networks obtained when varying leakage position (nodes 17 to 32). The junction where the leakage is activated is drawn at the center of each plot. Thick lines are the links connecting the central node to others, while other connections are indicated with thin lines.

In this way, the lightest cell represents (for each column) the node with the least degree centrality, that is, that affected by the leakage. This is confirmed by the number indicated in each cell, giving the degree centrality. Only simulations with one leakage at a time have been considered.It is evident that, in all the cases analyzed, the node with the smallest degree is that one in which the leakage is present.

Figure 7 and Figure 8 show round plots of the virtual networks in which the node where the leakage is activated is drawn at the center. Thick lines represent the connections the central node has with others, while thin lines indicate those between other nodes.

For each plot, the number of thick lines is the degree centrality of the node with the leakage. Such figures provide a pictorial representation of the absence of connections for the central node due to the presence of a leakage. In other words, such node is the most 'isolated'

5. Application to a real system

In order to test the reliability of the methodology, it has been applied to the western district of Tobruk city, Libya (Figure 9, left).



Figure 9: Left: layout of the city of Tobruk (Libya), and of the western district WDS (right), where the junctions characterized by the highest frequency values reported in Figure 10 are also indicated.

It covers a surface of 8 km^2 and serves a population of 54600 inhabitants. The average flow supplied is 240 l/s. The model of the system includes all the details of the network, and is made of 1139 junctions and 1621 pipes (Figure 9, right).

In this case, the performance of the methodology has been evaluated as the percentage of success in identifying the junction affected by the leakage (or its neighbors, according to a predefined tolerance of 200 m, representing the average length of the pipes), varying several parameters like threshold, θ, leakage coefficient, *c*, and leakage exponent, *u* (4).

Table 1 reports the results obtained for a tolerance distance of 200 m (in other words, it is considered a good result whenever the junction with the smallest degree centrality falls within 200 meters from that in which the leakage is activated). It can be observed that, independently of the parameters assumed, the rate of success of the methodology is around 90%. This is due to the fact that there are some junctions having the minimum value of degree centrality when leakages are activated at their surroundings.

To this end, the results in terms of frequency of occurrence of the nodes have been analyzed: Figure 10 shows several cases obtained with two values of the leakage exponent (u = 0.5; 1.0) and five values for the leakage coefficient (c = 0.1,…0.5), which are typically encountered in real situations.

Table 1: Performance indicator (PI) giving the percentage of success of the methodology when applied to Tobruk WDS, with varying threshold, θ, leakage coefficient, *c*, and exponent, *u*. (The threshold number, θ, indicated as 1-10^{-4} means 0.9999.)

θ	*c*	*u*	PI (%)
1-10^{-4}	0.5	0.5	90.34
1-10^{-6}	0.5	0.5	75.94
1-10^{-7}	0.1	0.5	86.65
1-10^{-6}	0.1	0.5	90.61
1-10^{-6}	0.1	1.0	90.61
1-10^{-5}	0.1	0.5	88.85
1-10^{-5}	0.2	0.5	90.17
1-10^{-5}	0.3	0.5	90.87
1-10^{-5}	0.4	0.5	91.83
1-10^{-5}	0.5	0.5	92.01
1-10^{-5}	0.1	1.0	88.85
1-10^{-5}	0.2	1.0	90.17
1-10^{-5}	0.3	1.0	90.87
1-10^{-5}	0.4	1.0	91.83
1-10^{-5}	0.5	1.0	92.01

In an ideal situation, all nodes should be characterized by a frequency of one, and the frequency plots would be uniform. However, from Figure 10 it can be observed that some of them exhibit very high frequency values, indicating that their minimum degree centrality occurs not only when a leakage is present there, but also when it is activated at several other surrounding junctions. When analyzing in more detail the nodes characterized by the highest values of frequency, it can be noted that they represent particular situations of 'bottlenecks' of the WDS (Figure 9, right). In these cases, their minimum degree centrality hinders some other non-linear property not captured by the linear Pearson correlation coefficient, which is still under investigation.

6. Discussion and Concluding Remarks

The paper has presented the results of a new methodology introduced for the localization of leakages emerging in water distribution systems. The main advantage of the approach is that it relies only on the measurements of pressure and on cross-correlations between signals, thus avoiding the need of comparing such data with a 'reference' (and not well defined) scenario given by a simulation model, whose optimal calibration is generally very difficult to attain. In this paper, the simulation model has been adopted with the only purpose of generating the pressure signals.

The novelty of the research resides in the fact that it analyzes pressure measurements through Complex Networks Theory, its reliability increasing with the number of installed pressure sensors. However, the proposed approach does not represent a final solution to the problem of leakage identification, but it has to be considered a further support in the process of leakage prelocalization, that is, a fast method to assess the integrity of a WDS and, eventually, to identify areas where prompt interventions should be planned.

Two test cases have been considered: the first, representing a well-known literature system, proved the best performance of the methodology in precisely localizing the node where the leakage is emerging (which is characterized by the least degree centrality with respect to all the others); the second, a real-world network, showed a high rate of success (around 90%). The difference between the results obtained can be ascribed to the fact that, in the first case, only the main transport pipes are included in the model (in other words, the model simulates the principal 'skeleton' of the water distribution system). In the second case, given by a real WDS, all the pipes have been modeled, and hence the flows in the links may arrange in such a way that the conditions of minimum degree centrality do not always occur at the junction where the leakage is present, but rather in some other surrounding nodes. Actually, this is not a problem in real situations, since water managers are interested in quickly identifying the area where the leakage is forming.

Other issues still under investigations are represented by the influence on the spatial distribution of demand pattern and the presence of several reservoirs and tanks, which actually limit the performance of the methodology. In any case, the current trend in WDS management is to divide such systems in DMA (District Metered Areas), for a better control of water leakages: thus, each DMA (which is often supplied by at most one reservoir, or may be modeled in such a way through appropriate boundary conditions) can be considered as a single system and the proposed methodology applied accordingly.

The requirement of linear correlation between pressure variability is fundamental for the success of the methodology. However it is only a necessary condition, as proved by the second test case, for which the linearity in correlation is not sufficient to always guarantee the minimum degree centrality at the node with the leakage, hindering some other non-linear property not captured by the linear Pearson correlation coefficient. Also this issue is currently under investigation.

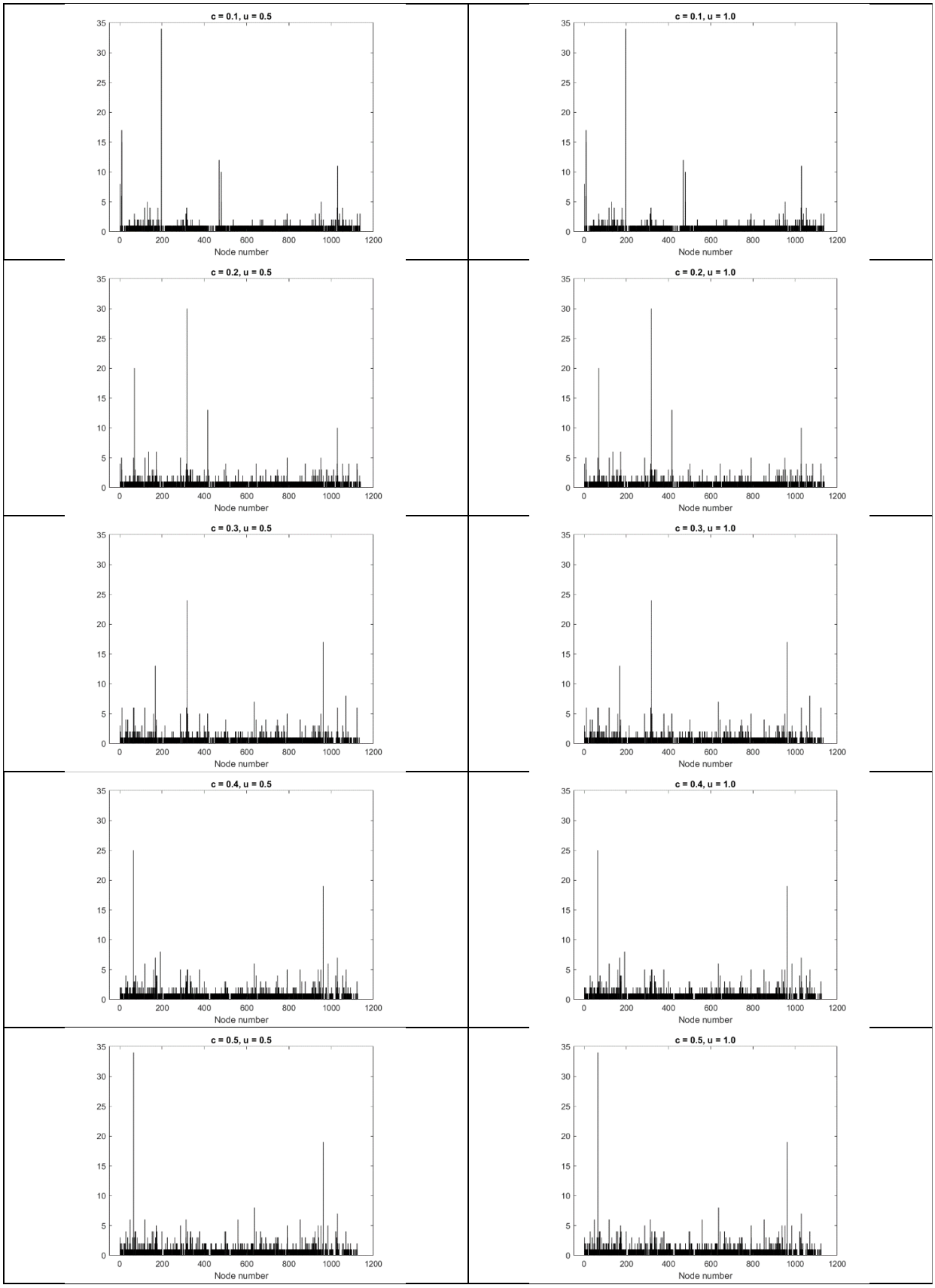


Figure 10: Frequency plots of junction occurrences in minimum degree centrality when varying leakage position and for different values of leakage coefficient, *c*, and exponent, *u*. Correlation threshold of $1\text{-}10^{-5}$ has been assumed.

Acknowledgment

The Author would like to thank the anonymous Reviewers for their comments and suggestions, which resulted in a considerable improvement of the paper.

References

[1] P. Swamee, A. Sharma, Design of Water Supply Pipe Networks, John Wiley & Sons, 2008.
[2] R. Beuken, C. Lavooij, A. Bosch, P. Schaap, Low leakage in the Netherlands confirmed, in Proceedings of the Water Distribution Systems Analysis Symposium 2008 (WDSA 2008), Kruger National Park, South Africa, 2008. DOI: 10.1061/40941(247)174
[3] A. Lambert, International Report: Water Losses Management and Techniques, Water Science and Technology: Water Supply, 2(4), 2002, 1–20. https://doi.org/10.2166/ws.2002.0115
[4] R.S. McKenzie, Z. Siqalaba, W. Wegelin, The State of Non-Revenue Water in South Africa; Report No. TT522/12, Water Research Commission: Lynnwood Manor, Pretoria, South Africa, 2012.
[5] A.F. Colombo, B.W. Carney, "Energy and costs of leaky pipes: toward comprehensive picture", J. Water Resour. Plann. Manage., 128(6), 441–450, 2002. DOI: 10.1061/(ASCE)0733-9496(2002)128:6(441)
[6] M. Nicolini, Leakage Identification in Water Distribution Systems with a Complex Networks Approach, in Proceedings of the 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII 2018), Jeju Island, South Korea, 2018. DOI: 10.1109/ICKII.2018.8569182
[7] J. Thornton, R. Sturm, G. Kunkel, Water Loss Control, 2nd Ed, McGraw-Hill, New York, 2008.
[8] R. Puust, Z. Kapelan, D.A. Savic, T. Koppel, "A review of methods for leakage management in pipe networks", Urban Water Journal, 7(1), 25–45, 2010. https://doi.org/10.1080/15730621003610878
[9] M. Nicolini, L. Zovatto, "Optimal location and control of pressure reducing valves in water networks", J. Water Resour. Plann. Manage., 135(3), 178–187, 2009. DOI: 10.1061/(ASCE)0733-9496(2009)135:3(178)H.
[10] M. Nicolini, C. Giacomello, K. Deb, "Calibration and optimal leakage management for a real water distribution network", J. Water Resour. Plann. Manage., 137(1), 134–142, 2011. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000087
[11] S. Alvisi, M. Franchini, "A procedure for the design of district metered areas in water distribution systems", Procedia Engineering, 70, 41–50, 2014. https://doi.org/10.1016/j.proeng.2014.02.006
[12] B. Brunone, M. Ferrante, "Detecting leaks in pressurized pipes by means of transients", J. of Hydraulic Research, 39(5), 539–547, 2001. https://doi.org/10.1080/00221686.2001.9628278
[13] D. Misiunas, J. Vitkovsky, G. Olsson, A. Simpson, M. Lambert, "Pipeline break detection using pressure transient monitoring", J. Water Resour. Plann. Manage., 131(4), 316–325, 2005. DOI: 10.1061/(ASCE)0733-9496(2005)131:4(316)
[14] M. Nicolini, A. Patriarca, Model Calibration and System Simulation from Real Time Monitoring of Water Distribution Networks, in IEEE 3rd Int.l Conf. on Computer Research and Development (ICCRD 2011), Shangai, China, 2011. DOI: 10.1109/ICCRD.2011.5763972
[15] R. Perez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, A. Peralta, "Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks", Control Engineering Practice, 19(10), 1157–1167, 2011. https://doi.org/10.1016/j.conengprac.2011.06.004
[16] M. V. Casillas Ponce, L. E. Garza Castañón, V. Puig Cayuela, "Model-based leak detection and location in water distribution networks considering an extended horizon analysis of pressure sensitivities", Journal of Hydroinformatics, 16(3), 649–670, 2013. https://doi.org/10.2166/hydro.2013.019
[17] D. Perfido, T. Messervey, C. Zanotti, M. Raciti, A. Costa, Automated leak detection system for the improvement of water network management, in 3rd Int.l Electronic Conf. on Sensors and Applications, 2016. https://doi.org/10.3390/ecsa-3-S5002
[18] G. Sanz, R. Perez, Z. Kapelan, D. Savic, "Leak detection and localization through demand components calibration", J. Water Resour. Plann. Manage., 142(2), 2016. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000592
[19] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, 2010.
[20] A.L. Barabási, "The network takeover", Nature Physics, 8, 14–16, 2012.
[21] R. Donner, E. Hernández-García, E. Ser-Giacomi, "Introduction to Focus Issue: Complex network perspective on flow systems", Chaos, 27, 2017. https://doi.org/10.1063/1.4979129
[22] C. Zhou, L. Zemanová, G. Zamora-López, C. C. Hilgetag, J. Kurths, "Structure-function relationship in complex brain networks expressed by hierarchical synchronization", New J. Phys. 9:178, 2007. doi:10.1088/1367-2630/9/6/178
[23] L. C. Carpi, P. Saco, O. Rosso, M. Gómez Ravetti, "Structural evolution of the Tropical Pacific climate network", Eur. Phys. J. B, 85:389, 2012. DOI: 10.1140/epjb/e2012-30413-7
[24] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd Ed., Cambridge University Press, 2004.
[25] O. Giustolisi, A. Simone, L. Ridolfi, "Network structure classification and features of water distribution systems", Water Resources Research, 53(4), 3407–3423, 2017. https://doi.org/10.1002/2016WR020071
[26] A. Yazdani, P. Jeffrey, "A complex network approach to robustness and vulnerability of spatially organized water distribution networks". arXiv Prepr. arXiv1008.1770, 2010.
[27] A. Yazdani, R. A. Otoo, P. Jeffrey, "Resilience enhancing expansion strategies for water distribution systems: A network theory approach", Environ. Model. Softw., 26(12), 1574–1582, 2011. DOI:10.1016/j.envsoft.2011.07.016
[28] Q. Shuang, M. Zhang, Y. Yuan, "Node vulnerability of water distribution networks under cascading failures", Reliab. Eng. Syst. Saf., 124(C), 132–141, 2014. DOI: 10.1016/j.ress.2013.12.002
[29] L. Berardi, R. Ugarelli, J. Rostum, O. Giustolisi, "Assessing mechanical vulnerability in water distribution networks under multiple failures", Water Resour. Res., 50(3), 2586–2599, 2014. https://doi.org/10.1002/2013WR014770
[30] F. Chaoqi, W. Ying, W. Xiaoyang, "Research on complex networks' repairing characteristics due to cascading failure", Physica A, 482, 317–324, 2017. DOI: 10.1016/j.physa.2017.04.086
[31] J.A. Gutiérrez-Pérez, M. Herrera, R. Pérez-García, E. Ramos-Martínez, "Application of graph-spectral methods in the vulnerability assessment of water supply networks", Math. Comput. Model., 57(7-8), 1853–1859, 2013. https://doi.org/10.1016/j.mcm.2011.12.008
[32] J.A. Gutiérrez-Pérez, M. Herrera, J. Izquierdo, R. Pérez-García, An approach based on ranking elements to form supply clusters in Water Supply Networks as a support to vulnerability assessment, in International Environmental Modelling and Software Society (iEMSs) Conference, Leipzig, Germany, 2012.
[33] A. Yazdani, P. Jeffrey, "A note on measurement of network vulnerability under random and intentional attacks", 2010. arXiv:1006.2791
[34] N. Sheng, Y.W. Jia, Z. Xu, S.L. Ho, C.W. Kan, "A complex network based model for detecting isolated communities in water distribution networks", Chaos, 23(4), 2013. doi: 10.1063/1.4823803
[35] O. Giustolisi, L. Ridolfi, "A novel infrastructure modularity index for the segmentation of water distribution networks", Water Resour. Res., 50(10), 7648–7661, 2014. https://doi.org/10.1002/2014WR016067
[36] O. Giustolisi, L. Ridolfi, L. Berardi, "General metrics for segmenting infrastructure networks", Journal of Hydroinformatics, 17(4), 505–517, 2015. DOI: 10.2166/hydro.2015.102
[37] A. Simone, O. Giustolisi, D.B. Laucelli, "A proposal of optimal sampling design using a modularity strategy", Water Resources Research, 52(8), 6171–6185, 2016. https://doi.org/10.1002/2016WR018944
[38] R. Nazempour, M.A.S. Monfared, E. Zio, "A complex network theory approach for optimizing contamination warning sensor location in water distribution networks", arXiv preprint arXiv:1601.02155, 2016
[39] C. Giudicianni, A. Di Nardo, M. Di Natale, R. Greco, G.F. Santonastaso, A. Scala, "Topological Taxonomy of Water Distribution Networks", Water, 10(4), 444, 1–19, 2018. https://doi.org/10.3390/w10040444
[40] G.F. Santonastaso, A. Di Nardo, M. Di Natale, C. Giudicianni, R. Greco, "Scaling-Laws of Flow Entropy with Topological Metrics of Water Distribution Networks", Entropy, 20(2), 95, 1–15, 2018. https://doi.org/10.3390/e20020095
[41] L. A. Rossman, EPANET 2 Users Manual, USEPA, Cincinnati, OH, 2000.
[42] E. Todini, S. Pilati, A gradient method for the analysis of pipe networks, in Int.l Conf. on Computer Applications for Water Supply and Distribution, Leicester Polytechnic, UK, 1987.
[43] O. Fujiwara, D. B. Khang, "A two-phase decomposition method for optimal design of looped water distribution networks", Water Resources Research, 26(4), 539–549, 1990. https://doi.org/10.1029/WR026i004p00539
[44] A. Marchi, G. Dandy, A. Wilkins, H. Rohrlach, "A methodology for comparing evolutionary algorithms for the optimization of water distribution systems, J. Water Resour. Plann. Manage., 140(1), 786–798, 2014. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000321
[45] E. Creaco, G. Pezzinga, D. Savic, "On the choice of the demand and hydraulic modeling approach to WDN real-time simulation", Water Resources Research, 53(7), 6159–6177, 2017. https://doi.org/10.1002/2016WR020104
[46] E. Creaco, E.J.M. Blokker, S.G. Buchberger, "Models for generating household water demand pulses: literature review and comparison", J. Water Resour. Plann. Manage., 143(6), 2017. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000763

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 285-291 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

DSTATCOM-Fuel Cell System on Radial Low-voltage Distribution Network for Mitigating Voltage Rise Caused by High Penetration of Photovoltaic Systems

Dylon Hao Cheng Lam, Jianhui Wong*, Yun Seng Lim, Jimmy Chin Yong Hee

Department of Electrical and Electronic Engineering, Universiti Tunku Abdul Rahman, Selangor, Malaysia

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 27 July, 2019
Online: 07 August, 2019

Keywords:
Grid-connected
DSTATCOM
Fuel cell
Distribution Network
Voltage rise

ABSTRACT

The Malaysian government has recently announced to increase the share of renewable energy generation from 2% to 20% of the total electricity consumption by year 2030. However, integrating a significant level of renewable energy sources on the distribution network without proper coordination will cause adverse effects that leads to unnecessary operation costs to the stakeholders. Various solutions including energy storage system are proven to be an effective way to overcome the issues caused by the integration of renewable energy sources. However, energy storage system using batteries have limitations such as limited life cycle and environmental hazards. As such, alternative source such as fuel cell which is well known for its zero pollutant can be considered. In this paper, a distributed static synchronous compensator integrated with a hydrogen system (DSTATCOM-Fuel cell) is proposed. The hydrogen system consists of fuel cell and electrolyzer to perform power exchange within the grid for regulating the voltage magnitude at 1 p.u. Various case studies have been conducted to validate the efficacy of this proposed combination and the results have shown that the DSTATCOM-Fuel cell can effectively mitigates the voltage rise up to 12.3% while maintaining high penetration of the photovoltaic systems.

1. Introduction

This paper is an extension of work originally presented in the International Conference on Smart Grid and Clean Energy 2018 [1]. Malaysia is expected to receive significant growth of the grid-connected renewable energy sources as the government has recently set a new target to commit on increasing the renewable energy generation from the current 2% to 20% of the total electricity consumption by year 2030 [2]. However, the growth of renewable energy sources on the electrical networks without proper coordination can create technical challenges that lead to the malfunction of power electronic devices, reduction of network efficiency, and improper operation of power protection system which gives high operating cost impact to the stakeholders. Curtailing and limits the amounts of grid-connected renewable energy have always been applied in order to avoid the deterioration of electrical power quality [3] despite these measures can greatly reduce the generation of clean energy and affect the commitment to increase the share of renewable energy generation [4]. Many solutions have been research, and realized that integrating large-scaled energy storage system by manipulating the real and reactive power exchange within the system can be used to regulate the network voltage [5–9]. Besides, energy storage system integrated with STATCOM are used to enhance the transient stability [10–14], improve voltage sags [15], smoothen wind farm power output and terminal voltage by incorporates real and reactive power control ability [16–18]. Although these combinations have proven useful, however, batteries have few disadvantages such as limited life cycle, voltage and current limitation and environmental hazards [19–21]. Therefore, more exploration on new alternative solutions and their corresponding control strategies become obligatory. Unlike the other batteries, fuel cells is an environmental friendly and well known for its zero pollutant, high efficiency, modularity, high power and reliable [22] which has the capability of high charge and discharge cycles. In this paper, a distributed static synchronous compensator integrated with a hydrogen system (DSTATCOM-Fuel cell) is proposed for active and reactive power compensation. In this combination, the DSTATCOM regulates reactive power output to maintain the voltage at 1 p.u. with the hydrogen system operates through the discharging and charging of the fuel cell and electrolyzer respectively to regulate active power based on the PV power output and voltage magnitude. Various case studies are conducted

*Jianhui Wong, Universiti Tunku Abdul Rahman, wongjh@utar.edu.my

to validate the performance of the proposed solution to mitigate the voltage rise issues while allowing high penetration from grid-connected PV systems.

2. Typical Malaysian Low-voltage Distribution Network in Radial Topology

The paper begins with a preliminary study to investigate the impact on system voltage with the integration of grid-connected PV systems on the radial distribution network. It is followed by an efficacy analysis using MATLAB Simulink to validate the performance of the proposed DSTATCOM-Fuel cell system to regulate the system voltage at 1 p.u while maximize the capacity of the grid-connected PV systems on the distribution network.

2.1. Modeling Electrical Network

A typical Malaysian low-voltage distribution network for small sub-urban area as shown in Figure 1. This network can be categorized into three sections according to its voltage level; (i) 132 kV incoming supply, (ii) 33 kV distribution substations, (iii) 11 kV distribution substations. The 132 kV incoming supply is the sub-transmission network of the national grid, the voltage is then stepped down to 33 kV by the 132/33 kV power transformers installed in the 33 kV distribution substations. The 33 kV is further stepped down to 11 kV using the 33/11 kV power transformer to a suitable voltage level for utilization. In this network, feeder 4 being the longest feeder will be modelled and the proposed DSTATCOM-Fuel cell system is integrated on feeder 4 to validate the efficacy.

There are 11 distribution substations across the radial feeder 4 with each of them are rated at 11/0.4 kV. The distance between each substation is indicated in Figure 1. The primary sides of the 11/0.4 kV transformers are labelled as 1, 2, 3 … and 11, while the secondary sides of the 11/0.4 kV transformers are labelled as 1a, 2a, 3a … and 11a. Table 1 and Table 2 show the technical specifications of the power transformer and conductor respectively. The locations of the distribution substations, the typical load characteristics and photovoltaic system power output were collected from the utility company.

A three-phase PV system is placed at the terminals 1a, 2a, and 3a in case 1 and 9a, 10a, and 11a in case 2 to investigate the impact of PV systems towards voltage rise on the network. Equation 1 shows sizing of the PV system is based on the penetration levels as follows:

$$PL = \frac{E_{PV}}{E_{load}} \times 100\% \quad (1)$$

where,

PL is the penetration level of the PV system (%)
E_{PV} is the total PV energy yield for the day (kWh)
E_{load} is the daily energy consumption at the load (kWh)

Based on the PV power output profile and load profile used in this study, it is determined that for 100% PV penetration level, the PV system should be sized at 500 kW. Thus, the PV system capacity for 60%, 70%, 80%, and 90% PV penetration are 300 kW, 350 kW, 400 kW, and 450 kW respectively.

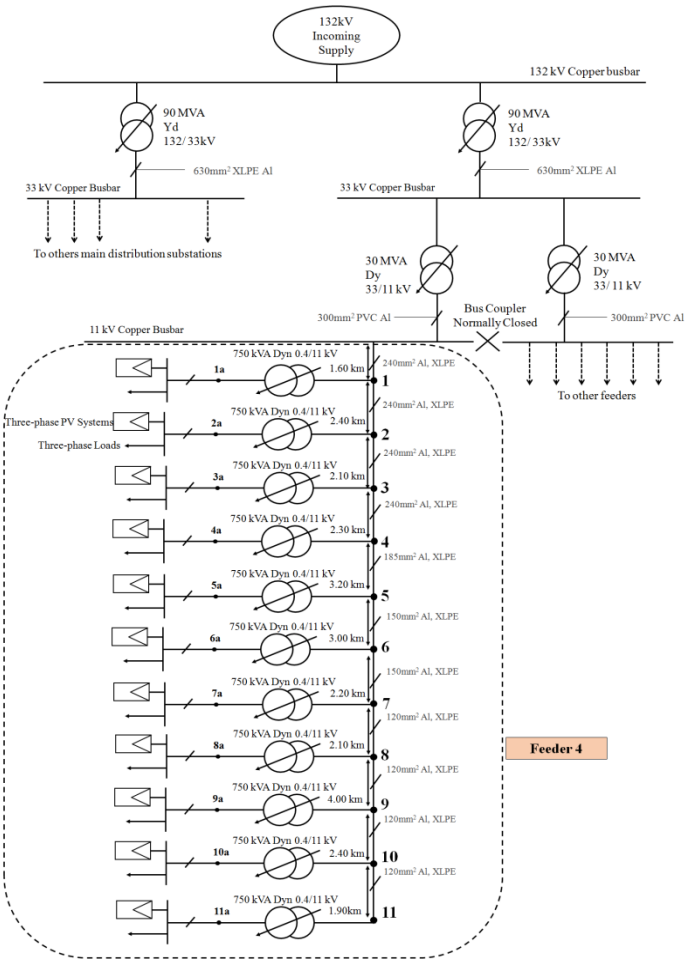


Figure 1: Typical Malaysian low-voltage distribution network in a sub-urban area.

Table 1: Technical Specification of the oil-immersed power transformers

Descriptions	Voltage Level (kV)	Rated Power (MVA)	Vector Group	Percent Reactance (%X)
Three-phase power transformers	132/33	90.00	Yd11	13.50%
	33/ 11	30.00	Dyn11	10.00%
	11/ 0.4	0.75	DYn11	4.75%

Table 2: Technical Specification of the conductors

Voltage Level (kV)	Cross-sectional Area (mm2)	Material	Insulator	R (Ω/km)	X_L (Ω/km)
0.4	300	Al	PVC	0.1300	0.0720
11.0	120	Al	XLPE	0.2990	0.0873
	150	Al	XLPE	0.2645	0.1603
	185	Cu	PILC	0.1950	0.0829
	240	Al	XLPE	0.1610	0.1524
33.0	630	Al	XLPE	0.0627	0.1070

2.2. Modeling DSTATCOM-Fuel Cell System

The proposed DSTATCOM-Fuel cell system is shown in Figure 2. The system consists of a DSTATCOM with an integrated hydrogen system. The hydrogen system consists of a fuel cell, an

electrolyzer, and a hydrogen storage tank. The hydrogen system is connected to the DC capacitor of the DSTATCOM through a boost converter which regulates the DC voltage from the hydrogen system to the desired level. The DSTATCOM consists of a DC capacitor, a three-phase pulse-width-modulated voltage source converter (VSC), and a coupling transformer. The coupling transformer connects the VSC to the network and functions to provide galvanic isolation and voltage regulation at the point of common coupling (PCC). The DSTATCOM, fuel cell, and electrolyzer are rated at 500 kVAr, 500 kW, and 500 kW respectively to match the PV system capacity at 100% PV penetration.

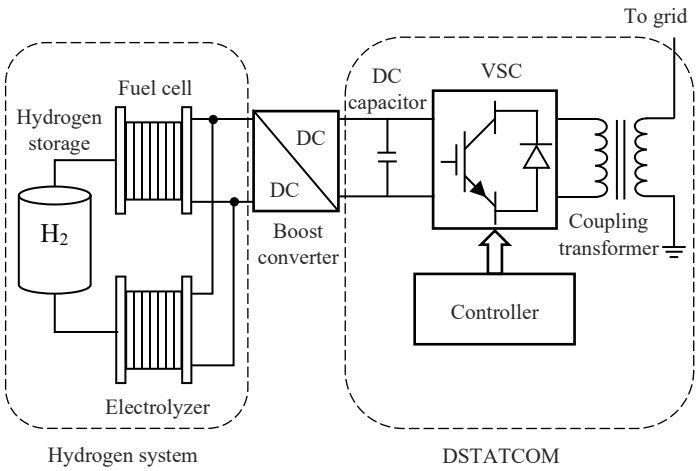


Figure 2: Schematic diagram of DSTATCOM-Fuel cell

The proton membrane exchange fuel cell (PEMFC) is used for the hydrogen system as it has a low operating temperature of 60-80°C, a high electrical efficiency range of 40-50%, a fast start up time, and high power density which makes it suitable for use in distribution networks [23,24]. A 500 kW PEMFC is modelled based on the NedStack PS6 datasheet. The parameters of the 500 kW PEMFC stack is shown in Table 3. The voltage-current (V-I) and power-current (P-I) curve of the fuel cell stack is shown in Figure 3.

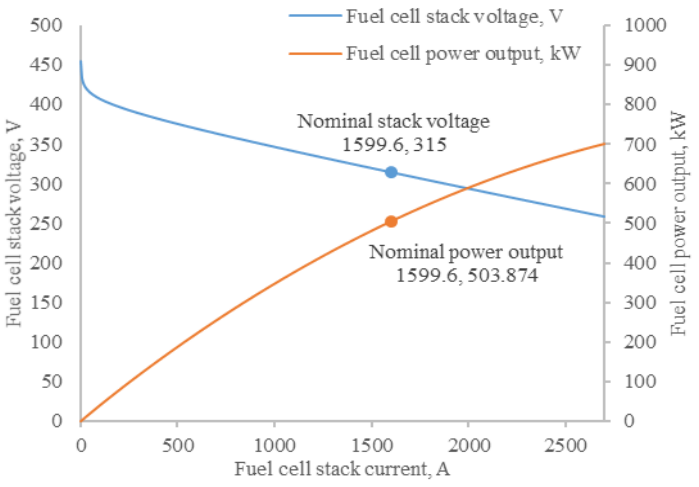


Figure 3: V-I curve and P-I curve of the fuel cell stack.

Table 3: Technical specification of the PEMFC stack

Rated capacity	504 kW
Nominal voltage	315 V
Nominal current	1333 A
Number of cells connected in series	455
Number of strings connected in parallel	12
Total number of cells in the fuel cell stack	5460

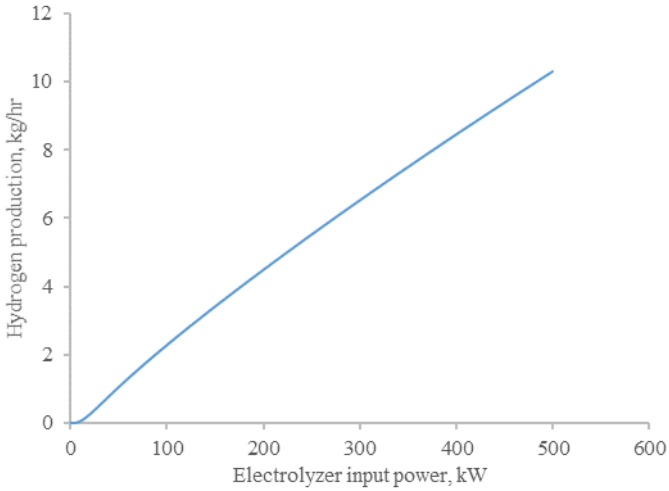


Figure 4: Hydrogen production of electrolyzer stack.

An alkaline electrolyzer stack is modelled based on the Photovoltaik, Elecktrolyse, Brennstoffzelle Und Systemtechnik (PHOEBUS) plant installed at Julich, Germany as reported in [25]. In order to create a 500 kW electrolyzer system, 19 electrolyzer stacks each rated at 26 kW are modelled and connected in parallel. The voltage-current relationship and hydrogen production of the electrolyzer system are modelled as described in [26,27]. The hydrogen produced by the electrolyzer is stored into a hydrogen storage tank without compression. The hydrogen storage tank capacity is 30 kg of hydrogen. The hydrogen production of the electrolyzer model is shown in Figure 4.

2.3. Control Strategies for Voltage Regulation

Figure 5 shows the control scheme for the DSTATCOM-Fuel cell. The objective of the controller is to mitigate voltage rise issue on the network by active and reactive power compensation. The voltage at the PCC, V_{PCC}, is measured and DSTATCOM performs reactive power compensation to regulate the voltage at 1 p.u. The voltage rise on the network occurs mainly due to high PV penetration. Hence, controller will initiate the hydrogen system charging operation when the PV power output, P_{PV}, is higher than the load consumption, P_{load} and the hydrogen tank level is less than 100%. If P_{PV} is less than P_{load}, the control will check if V_{PCC} is below the statutory limit of 1.1 p.u. If yes, the control will initiate the hydrogen system discharging operation if the hydrogen tank level is more than 20%. A minimum threshold of 20% for the hydrogen tank level is set to ensure a safe operation of the fuel cell.

If V_{PCC} exceeds the 1.1 p.u. limit, the control will stop the charging and discharging of the hydrogen system and only reactive power compensation will be used to regulate the voltage.

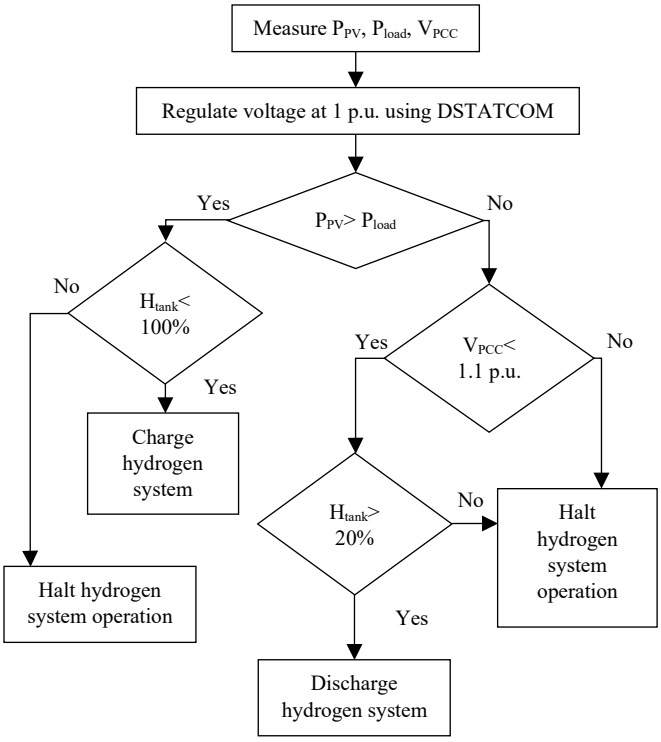


Figure 5: Flow chart of the DSTATCOM-Fuel cell controller.

3. Simulation Results

3.1. Impact of PV system allocation on voltage rise severity:

In this case study, the impact of PV system allocation on the system voltage is studied. Two cases are considered; i) Case 1 where a PV system is connected at terminals 1a, 2a, and 3a respectively, ii) Case 2 where a PV system is connected at terminals 9a, 10a, and 11a respectively. The PV power output profile used in this case study is the average hourly PV power output measured at Selangor, Malaysia during the month of April, 2018. The network is simulated for a period of 24 hours to determine the maximum voltage rise that occur within one full day.

Figure 6 shows the voltage magnitude of the load terminals 1a, 2a, and 3a for case 1. Figure 7 shows the voltage magnitude of the load terminals 9a, 10a, and 11a for case 2. In both cases, the maximum voltage is measured when there is 0%, 60%, 70%, 80%, 90%, and 100% PV. The voltage magnitudes shown in these two figures are the maximum voltage measured within the 24-hour period.

In case 1, the maximum voltage at terminal 1a, 2a, and 3a violates the 1.1 p.u. statutory limit when PV penetration exceeds 70%, 80% and 90% respectively. The maximum voltages at 1a, 2a, and 3a when PV penetration is 100% are 1.12 p.u., 1,12 p.u. and 1.11 p.u. respectively. It is noticed that in this case, the voltage rise severity reduces at terminals further away from the 11 kV main substation.

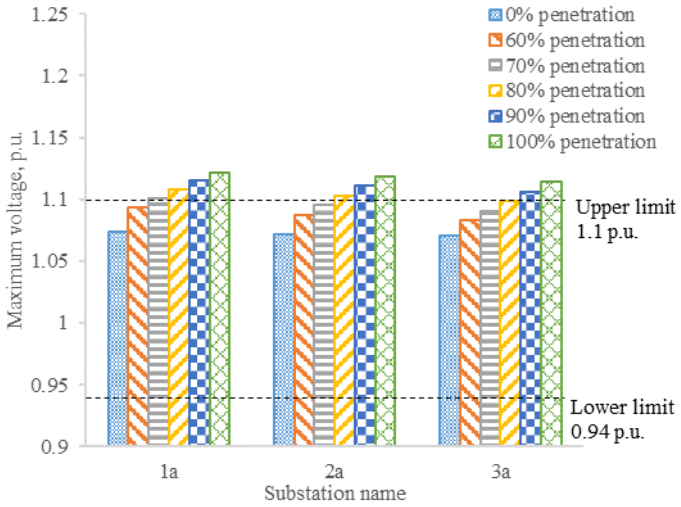


Figure 6: Maximum voltage magnitude of the load terminals when the PV system is connected at the terminals 1a, 2a, and 3a.

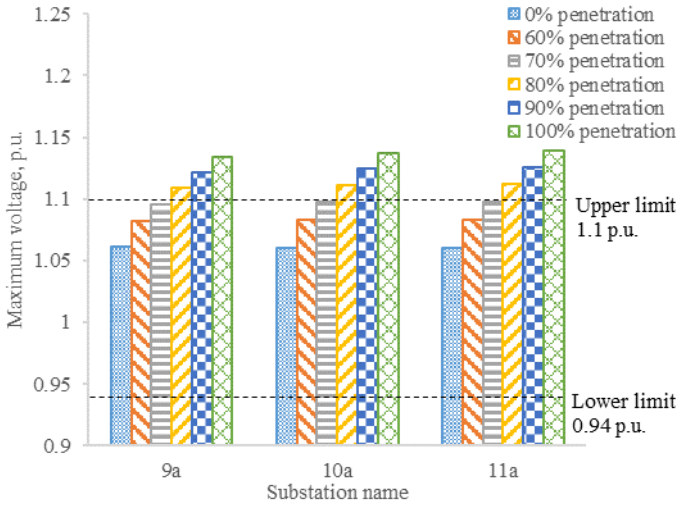


Figure 7: Maximum voltage magnitude of the load terminals when the PV system is connected at the terminals 9a, 10a, and 11a.

In case 2, the maximum voltage at terminal 9a, 10a, and 11a violates the 1.1 p.u. statutory limit when PV penetration exceeds 80%. The maximum voltages at 9a, 10a, and 11a when PV penetration is 100% are 1.13 p.u., 1.14 p.u., and 1.14 p.u. respectively. It is noticed that in this case, the voltage rise severity increases at terminals further away from the 11 kV main substation.

Comparing case 1 and case 2, it is noticed that when several PV systems are placed near the 11 kV substation, the voltage rise impact is greatest at the terminal located nearest to the main substation. Conversely, when multiple PV systems are placed at the end of the load terminals, the voltage rise is most severe at the terminal located furthest from the main substation.

Overall, it is observed that the voltage rise is more severe when multiple PV systems are placed at terminals furthest to the 11 kV main substation. The voltage increase as PV penetration increases is also more significant when the PV system is connected to the last three terminals compared to the first three terminals.

3.2. Efficacy Analysis for DSTATCOM-Fuel Cell System:

In order to determine the effectiveness of the proposed system for mitigating voltage rise caused by PV systems, a DSTATCOM-Fuel cell system is connected to terminals 1a, 2a, and 3a in case 1 and to terminals 9a, 10a, and 11a in case 2.

Figure 8 and Figure 9 show the comparison of the voltage rise for both cases when the PV system penetration level is 100%. It is observed that in both cases, the maximum voltage is maintained at 1 p.u when DSTATCOM-Fuel cell is connected. The regulation of voltage at the terminals with a connected PV system also affects the voltage at the other terminals. This effect is more significant in case 2 where the voltage at the neighboring terminals is reduced by as much as 2%. Overall, by using DSTATCOM-Fuel cell voltage rise can be reduced by as much as 10.7% in case 1 and 12.3% in case 2.

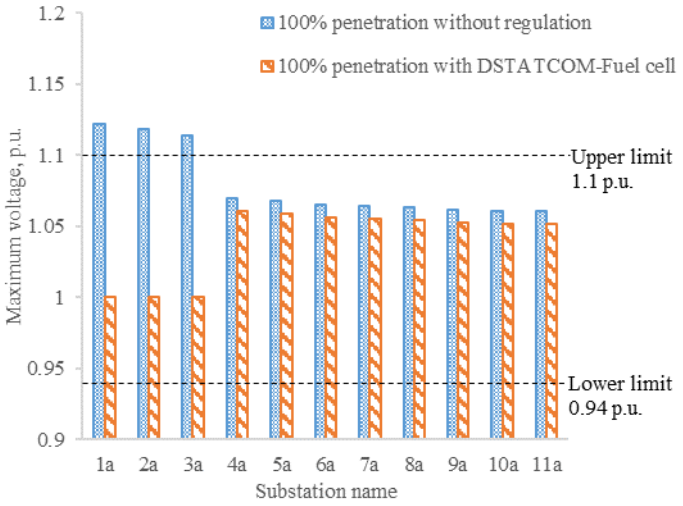


Figure 8: Comparison of voltage rise at 100% with and without regulation by DSTATCOM-Fuel cell in case 1.

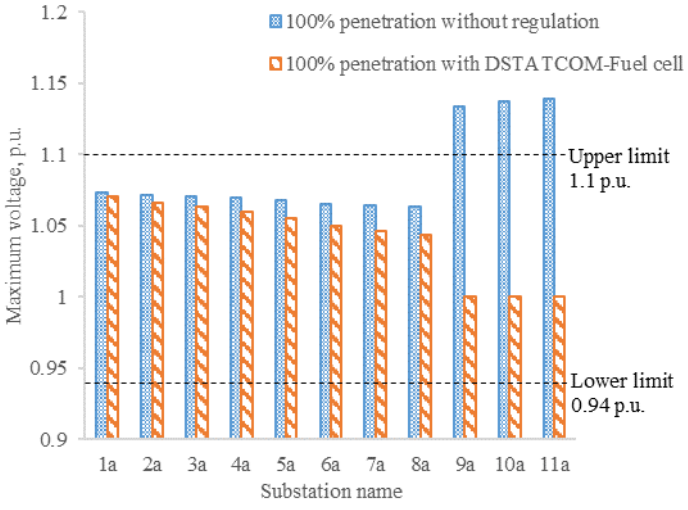


Figure 9: Comparison of voltage rise at 100% with and without regulation by DSTATCOM-Fuel cell in case 2.

Figure 10 shows the regulated power output at terminal 11a in case 3 after active power compensation by the DSTATCOM-Fuel cell system relative to the PV power output and load consumption. In order to mitigate voltage rise issues caused by reverse power flow, a combination of active and reactive power compensation is used depending on the PV power output and hydrogen tank level. When the PV power output is less than the load consumption, the hydrogen system in the DSTATCOM-Fuel cell will supply active power to the network to match the load consumption if the voltage does not exceed the statutory limit of 1.1 p.u. and the hydrogen tank level is not less than 20%. When the PV power output exceeds the load consumption, the hydrogen system charging operation is initiated if the hydrogen tank level is not full and the excess power is absorbed by the hydrogen system and converted to hydrogen.

It is observed that from 3:45:00 to 7:15:00 and 21:30:00 to 0:00:00, the regulated power is 0 kW. This is due to the hydrogen tank level being below 20%. Hence, the hydrogen system does not carry out the discharging operation during this period.

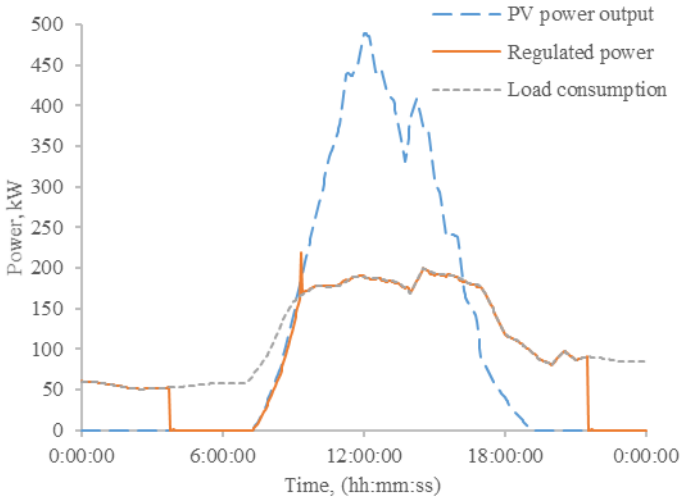


Figure 10: Regulated power output relative to PV power output and load consumption at terminal 11a in case 2.

Figure 11 shows the voltage levels throughout the day at terminal 11a in case 3. It is observed that there is significant voltage rise up to 1.14 p.u. which coincides with the peak output of the PV system. When the DSTATCOM-Fuel cell system is connected, the voltage is regulated at 1 p.u. throughout the day.

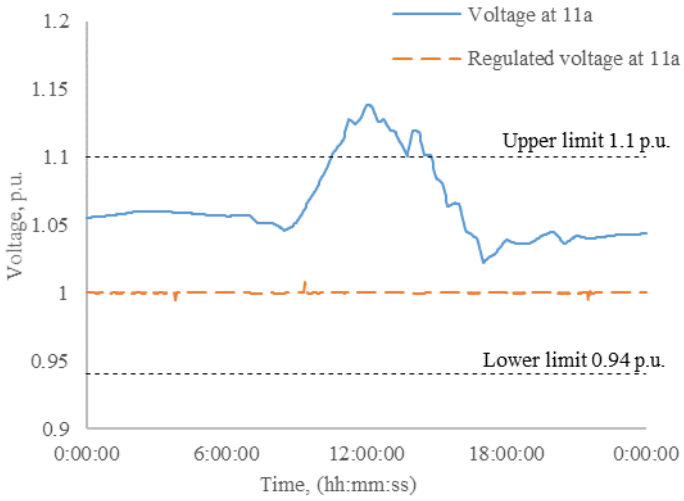


Figure 11: Regulated and unregulated voltage using DSTATCOM-Fuel cell at terminal 11a when a photovoltaic system generates 100% of the total consumed power.

Figure 12 shows the active power compensation by the fuel cell and electrolyzer and the reactive power compensation of the DSTATCOM. It is noticed that the DSTATCOM is constantly absorbing reactive power to reduce voltage level and regulated the voltage at 1 p.u. The hydrogen system charging and discharging operation depends on the PV power output. When the PV power output is high and exceeds the load consumption, the electrolyzer absorbs active power and converts it to hydrogen. When the PV power output is low and is less than the load consumption, the fuel cell supplies active power if the voltage magnitude is within the statutory limit of 1.1 p.u. It is noticed that from 3:45:00 to 7:15:00 and 21:30:00 to 0:00:00, the fuel cell does not supply any active power due to low hydrogen tank levels.

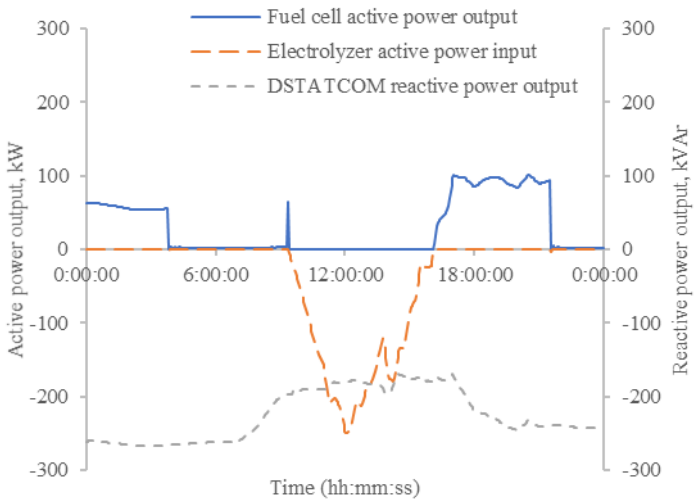


Figure 12: DSTATCOM-Fuel cell active and reactive power output at terminal 11a.

Figure 13 shows the hydrogen tank levels throughout the day. For this study, a starting level of 50%, which is 15 kg of hydrogen, has been assumed in this simulation. It is noticed that when the hydrogen tank level reaches 20%, the discharging operation of the system is turned off. Hence, only a minimal amount of hydrogen is consumed by the fuel cell to maintain the DC voltage in the DSTATCOM.

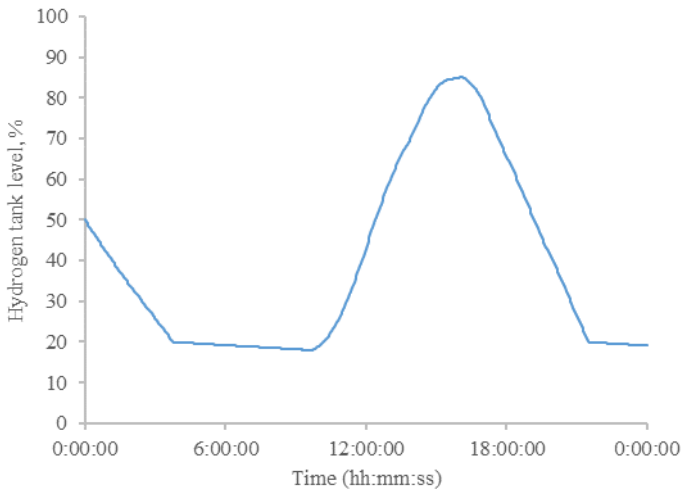


Figure 13: Hydrogen tank levels throughout the day.

4. Conclusion

Integrating large amount of renewable energy sources without suppressing the power quality always has been a challenge to the utility company. The voltage rise issues become severe when large amount of PV systems are installed across the end of a long feeder in a radial distribution network. This paper presented a DSTATCOM-Fuel cell system to perform power exchange within the grid for mitigating the voltage rise issues without curtailing the high penetration of PV systems. Excess power generated from the PV system will be converted into hydrogen and store in a hydrogen tank via an electrolyzer. In the event of high load consumption during low renewable energy generation, fuel cell will generate electricity using the hydrogen stored in the tank. Various case studies are conducted in this paper to validate the efficacy of the proposed DSTATCOM-Fuel cell system and the results have shown that this combination can mitigate the voltage rise issues up to 12.3%, maintaining the voltage at 1 p.u. while allowing high penetration of PV systems presence in the network.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS) under the grant number FRGS/1/2017/TK04/UTAR/02/2 and Universiti Tunku Abdul Rahman (UTAR) Research Fund.

References

[1] J. C. Y. Hee, J. Wong, Y. S. Lim, J. H. Tang, "Assessment on Various Allocations of Energy Storages Systems on Radial Distribution Network for Maximum PV Systems Penetration" in 2018 Int. Conf. Smart Grid Clean Energy Technol. ICSGCE, 2018. https://doi.org/10.1109/ICSGCE.2018.8556803.

[2] N. S. Eusoff, "Malaysia sets new goal of 20% clean energy generation by 2030" The Edge, 2018. https://www.theedgemarkets.com/article/malaysia-sets-new-goal-18-clean-energy-generation-2030.

[3] A. Shafiu, T. Bopp, I. Chilvers, G. Strbac, "Active management and protection of distribution networks with distributed generation" in IEEE Power Eng. Soc. Gen. Meet. 2004, 2004. https://doi.org/10.1109/PES.2004.1373011.

[4] L. Kane, G. Ault, "A review and analysis of renewable energy curtailment schemes and Principles of Access: Transitioning towards business as usual" Energy Policy, 72, 67–77, 2014. https://doi.org/10.1016/j.enpol.2014.04.010.

[5] T. Mahto, V. Mukherjee, "Energy storage systems for mitigating the variability of isolated hybrid power system" Renew. Sustain. Energy Rev., 51(Supplement C), 1564–1577, 2015. https://doi.org/10.1016/j.rser.2015.07.012.

[6] M. S. Ali, M. M. Haque, P. Wolfs, "A review of topological ordering based voltage rise mitigation methods for LV distribution networks with high levels of photovoltaic penetration" Renew. Sustain. Energy Rev., 103, 463–476, 2019. https://doi.org/10.1016/j.rser.2018.12.049.

[7] P. Ariyaratna, K. M. Muttaqi, D. Sutanto, "A novel control strategy to mitigate slow and fast fluctuations of the voltage profile at common coupling Point of rooftop solar PV unit with an integrated hybrid energy storage system" J. Energy Storage, 20, 409–417, 2018. https://doi.org/10.1016/j.est.2018.10.016.

[8] F. H. M. Rafi, M. J. Hossain, J. Lu, "Hierarchical controls selection based on PV penetrations for voltage rise mitigation in a LV distribution network" Int. J. Electr. Power Energy Syst., 81, 123–139, 2016. https://doi.org/10.1016/j.ijepes.2016.02.013.

[9] E. L. Ratnam, S. R. Weller, "Receding horizon optimization-based approaches to managing supply voltages and power flows in a distribution grid with battery storage co-located with solar PV" Appl. Energy, 210, 1017–1026, 2018. https://doi.org/10.1016/j.apenergy.2017.08.163.

[10] Z. Yang, C. Shen, L. Zhang, M. L. Crow, S. Atcitty, "Integration of a StatCom and battery energy storage" IEEE Trans. Power Syst., 16(2), 254–260, 2001. https://doi.org/10.1109/59.918295.
[11] M. A. Rahman, M. H. Shawon, S. K. Ghosh, S. K. Nath, "Transient stability enhancement of fuel cell system using D-STATCOM" in 2014 3rd Int. Conf. Dev. Renew. Energy Technol. ICDRET, 2014. https://doi.org/10.1109/ICDRET.2014.6861663.
[12] M. Stella, M. A. G. Ezra, A. Peer Fathima, C. Khang Jiunn, "Research on the efficacy of unified Statcom-Fuel cells in improving the transient stability of power systems" Int. J. Hydrog. Energy, 41(3), 1944–1957, 2016. https://doi.org/10.1016/j.ijhydene.2015.11.130.
[13] S. Morris, M. Ezra, Y.-S. Lim, "Multi-machine power transmission system stabilization using MPSO based neuro-fuzzy hybrid controller for STATCOM/BESS" Int. Rev. Electr. Eng., 7, 2012.
[14] M. A. Bhaskar, D. Sarathkumar, M. Anand, "Transient stability enhancement by using fuel cell as STATCOM" in 2014 Int. Conf. Electron. Commun. Syst. ICECS, 2014. https://doi.org/10.1109/ECS.2014.6892553.
[15] A. Hajizadeh, M. A. Golkar, "Control of hybrid fuel cell/energy storage distributed generation system against voltage sag" Int. J. Electr. Power Energy Syst., 32(5), 488–497, 2010. https://doi.org/10.1016/j.ijepes.2009.09.015.
[16] S. M. Muyeen, R. Takahashi, T. Murata, J. Tamura, Mohd. H. Ali, "Application of STATCOM/BESS for wind power smoothening and hydrogen generation" Electr. Power Syst. Res., 79(2), 365–373, 2009. https://doi.org/10.1016/j.epsr.2008.07.007.
[17] S. W. Mohod, M. V. Aware, "A STATCOM-Control Scheme for Grid Connected Wind Energy System for Power Quality Improvement" IEEE Syst. J., 4(3), 346–352, 2010. https://doi.org/10.1109/JSYST.2010.2052943.
[18] S. W. Mohod, S. M. Hatwar, M. V. Aware, "Grid Support with Variable Speed Wind Energy System and Battery Storage for Power Quality" Energy Procedia, 12, 1032–1041, 2011. https://doi.org/10.1016/j.egypro.2011.10.135.
[19] A. Lenka, A. Papachen, A. P. Fathima, "Three Area AGC of An Interconnected Power System with Superconducting Magnetic Energy Storage" Int. J. Ind. Electron. Electr. Eng., 2, 54–58, 2014.
[20] A. A. K. Arani, H. Karami, G. Gharehpetian, M. Akhavanhejazi, "Review of Flywheel Energy Storage Systems structures and applications in power systems and microgrids" Renew. Sustain. Energy Rev., 69, 2017. https://doi.org/10.1016/j.rser.2016.11.166.
[21] E. MacA. Gray, C. J. Webb, J. Andrews, B. Shabani, P. J. Tsai, S. L. I. Chan, "Hydrogen storage for off-grid power supply" Int. J. Hydrog. Energy, 36(1), 654–663, 2011. https://doi.org/10.1016/j.ijhydene.2010.09.051.
[22] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, "Prognostics and Health Management of PEMFC – State of the art and remaining challenges" Int. J. Hydrog. Energy, 38(35), 15307–15317, 2013. https://doi.org/10.1016/j.ijhydene.2013.09.051.
[23] E. G. T. Services, Fuel Cell Handbook, Seventh Edition, 2004.
[24] H. R. Ellamla, I. Staffell, P. Bujlo, B. G. Pollet, S. Pasupathi, "Current status of fuel cell based combined heat and power systems for residential sector" J. Power Sources, 293(Supplement C), 312–328, 2015. https://doi.org/10.1016/j.jpowsour.2015.05.050.
[25] H. Barthels, W. A. Brocke, K. Bonhoff, H. G. Groehn, G. Heuts, M. Lennartz, H. Mai, J. Mergel, L. Schmid, P. Ritzenhoff, "Phoebus-Jülich: An autonomous energy supply system comprising photovoltaics, electrolytic hydrogen, fuel cell" Int. J. Hydrog. Energy, 23(4), 295–301, 1998. https://doi.org/10.1016/S0360-3199(97)00055-4.
[26] M. H. Nehrir, C. Wang, Modeling and Control of Fuel Cells: Distributed Generation Applications, John Wiley & Sons, 2009.
[27] Ø. Ulleberg, "Modeling of advanced alkaline electrolyzers: a system simulation approach" Int. J. Hydrog. Energy, 28(1), 21–33, 2003. https://doi.org/10.1016/S0360-3199(02)00033-2.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 299-304 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Conversion of 2D to 3D Technique for Monocular Images using Papilio One

Priya Arokia*, Anupama Patil

E&TC department, DYPIEMR, Pune, 411 044, India

ARTICLE INFO

Article history:
Received: 26 December, 2018
Accepted: 24 March, 2019
Online: 10 April, 2019

Keywords:
Monocular image
Depth
Scenic feature

ABSTRACT

A 3D image adds realism in viewing experience and can assist in simplifying the graphical displays. A Third dimension supplement to the input can improve pattern recognition, and can be used for 3D scene reconstruction and robot navigation. Recently popularity of 3D hardware is also increased which makes it a hot topic. The production of content as 3D is not matching with its need so there is scope of improvement of these 3D contents. Monocular cues give profundity data when seeing a scene with one eye. When a spectator moves, the evident relative movement of a few stationary articles against a foundation gives indicates about their relative separation. Depth estimation from monocular cues is a difficult task because single image lacks prior information like depth information, motion information etc. In Depth using scene features depth is estimated by exploring the features like shape, edges, color, texture and as well as an analysis of the environment of the scene that are of interest with respect to the target. Different objects have different hue and value and hence color is useful for depth estimation. Shape and texture provides disparity which is used to estimate depth. The main problem in converting a 2 dimensional to 3 dimensional images using single image is that it lacks information required for reconstruction in 3D data. While doing conversion by taking different cues or combination of multiple cues from scene conversion has been done e.g. structure form shape, motion, defocus etc. But such methods work for restricted scenarios not for global scenes. For instance, outdoor algorithms worked poor for indoor algorithms. Here we have implemented automatic conversion of 2 dimensional to 3 dimensional images using monocular image which can convert global images in visually comfortable 3D image.

1. Introduction

In this era 3D supporting hardware's are increasing, but the demand of content in 3D and its availability does not go hand in hand. Still 3D contents are surpassed by the 2D contents. So there is a requirement of 3D data and obvious many researchers are already working on this to close this gap in the future. One solution of direct taking 3D using multi-view method is available but it is costly and already there is large amount of 2D data is available; it will be costly to create newly 3D contents. To work on this problem, there should be the technique which convert large amount of 2D available data into 3D. The converted data should have comfortable visual quality and should not time consuming.

The available 2D data is a monocular data, which is taken from using only single view. The main problem in conversion of 2 dimension to 3 dimension using single image is that it lacks information required for reconstruction in 3D data. As discussed earlier in 2 dimension to 3 dimension conversion two approaches are available i.e. automatic and semi-automatic. Semiautomatic method gives good results but this method is time consuming and costly. Real time implementation of automatic method is preferable as semi-automatic method requires human intervention and is also known as a most trustable tool in extracting the depth from a particular scene. Disparity-pixels extracting algorithm and the camera quality determines the accuracy of the results.

2D image depth estimation is usually done in steps. A monochromatic-picture is oulined by depth where a distant separation from a camera is demonstrated by a low intensity and neaby separation is shown by a high power. i.e. it deals with luminance intensity [1]. Thus depth map, as afunction of the image co-ordinates, presents the depth in corresspondence to the camera position of a point in the visualscene. Estimating depth from a single picture or video quite exciting. As depth is an

*Priya Arokia,VIIT Research scholar,9923240728,Email: prinnu@yahoo.com

important cue of a scene which is lost in image acquisition. Because of this reason and several application needs this information, hence various methods are proposed to extract depth.

Depth Image Based Rendering (DIBR) needs a flow of mono-scopic images and a subsequent flow of depth map images that gives details of per pixel depth. Knowing the depth of each point in an input picture, rendering of any close-by perspective into a sample picture can be done by anticipating the pixels of the first picture to their appropriate 3-D areas and re-anticipating them onto the world-picture-plane. In this manner, DIBR grants the formation of novel pictures, utilizing data from the depth maps, as though they were caught with a camera from various perspectives. A favorable position of the DIBR approach is the higher capability with which coding of the depth maps can be done, than two scenes of common pictures, consequently lessening the transfer speed required for transmission. Rendering an image /video has been well realized and there are also very many algorithms available for producing quality and standard images.

In the left image the closer objects to the camera are located in the right while in the right image they are located in the left. The objects farther away from the camera appears in the same place in both the images. The object displacement in the two images is the disparity. The higher depth is indicated by a smaller disparity and a lower depth by a higher disparity. This is why pixel matching between the two images is essential. Output classification into dense and sparse is done on the stereo correspondence algorithms. An optimum mix of speed and accuracy is obtained by applying method of segments & edges.

Real time applications like robotics require dense output. Since our focus was oriented towards algorithms for real time application, the focus was on ‘dense stereo correspondense’. According to [2] dense matching algorithms are divided as local and global matching. The local one is either area based or feature based. Since the disparity is determined chiefly by the intensity values of a particular window, area-based local method applies variable windows. The window size, depending on the output can be changed such that a dense output is obtained to suit most of the real time applications. The feature extraction on which the feature based methods rely on provides a sparse output. However, they are quite fast. The accurate global methods, also called the energy based methods are rather very slow and computanionally expensive.

The object/features mapped in both right and left images is assumed to have the same intensity and the algorithms would be implemented. The illumination effects on the objects, with different view points of the cameras, may at times create a diffrence in the intensity values and invalidate these assumptions. Although the two cameras under consideration are completely tuned and both the images are simltaneously captured, stereo vision normally fails in non-ideal lighting. The reason is that the orientation or the pose of the cameras with respect to the light source may result in the variation of the light intensity of the captured images from the respective view points.

Pixel matching is then the whole idea. For this matching the pixels in the left images are to be looked for in the right images. The search becomes difficult if by chance the variations are both in the x and the Y co-ordinates, since the images are captured by to cameras from two different view points.. Image rectification, nothing but the process of horizontally aligning the epipolar lines of the two cameras is deployed to avoid this. Unidirectional alignment is done by means of linear transformations like rotate, skew and translate.

Authors have proposed an automatic 2D to 3D conversion algorithm rooted on multiple depth indications. Three distintive methods were considered for depth generation and consequently 2-D scene one depth generation method was carriedout [3]. The limitation of this method is latency. A few researchers have presented automatic pathway of learning the 3-D scene structure. In this pathway, an easier algorithm is recommended that learns the scene depth from a huge store of image + depth pairs. The produced anaglyph images provides a decent 3-D perception. However, they are notabsolutely devoid of distortions.

As image or video is having its attributes at a pixel level that are learned by a point transformation, it is applied to a monocular image.. The limitation of this system is cannot apply same transformation to images with potentially different global 3D scene structure.The same authors have also presented method to automatically convert a monoscopic video into stereo for 3D visualization. They are taking plausible automatically extracted depth maps at every frame and presented a framework using temporal information for improved and time-coherent depth when multiple frames are available. They have mixed indoor and outdoor images; hence indoor images work poor for outdoor images and vice-versa. This system is complex.

Guttmann proposed semi-automatic user defined strokes corresponding to a rough estimate of the depth values are defined for the image of interest in the scene[4]. This particular system finds the depth values for the rest of the image, and produces a depth map in order to create stereoscopic 3D image pairs. This can be further extended to automatic method.

This paper proposes the use of image rectification after the calibration of the camera.

The remainder of this brief is arranged as follows. Section II provides the process block diagram and its explanation. Section III presents the simulation results, Section IV subjective analysis and Section V draws the conclusion.

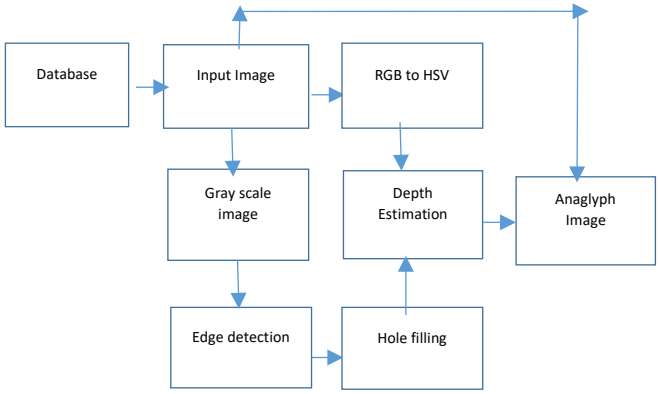


Figure 1.Block diagram of process

2. Process Block Diagram

2.1. Process flow

In 2D-to-3D conversion, input is a RGB monocular color image. Bit depth gives number of bits per pixel and here bit depth should be 24 as the input expected is a RGB image. As an input one can give image of file formats JPEG, BMP and PNG. Figure 1 is the block diagram of the whole process.

The input image is taken from database. Four datasets are used i.e. Make3D dataset [5], NYU dataset [6], Middlebury dataset [7] and our own dataset. Make3D dataset is of outdoor images. It consists of 534 images, each is having 1704X272 dimensions.

The NYU-depth v2 dataset is involved video grouping from assortment scenes taken indoor captured by both RGB and Depth cameras of Microsoft Kinect. This dataset consists of indoor scenes like basements, bedrooms, bathrooms, bookstores, cafe, classroom, dining rooms, kitchens, living rooms, offices etc. Middlebury dataset is one of the well-known dataset, which comprised of stereo images of various scenes along with its ground truth. We also have created indoor dataset of 22 images each image is of 4608X3456 using Sony Device [8]. Input image is given from one of this dataset. For computational efficiency original input images are re-sized.

2.2. RGB to HSV Conversion

The Papilio one is an expandable development board having a "Xilinx Spartan 3E FPGA chip", which is powerful and is in open-source. The whole conversion process happens in Papilio one.

The conversion process of RGB to HSV is shown in Figure 2. 2D query image is given from PC through MATLAB to Papilio one and the development board is used for conversion.From HSV value (V) attribute is considered here to estimate depth and hence a color model from RGB to HSV conversion is done.

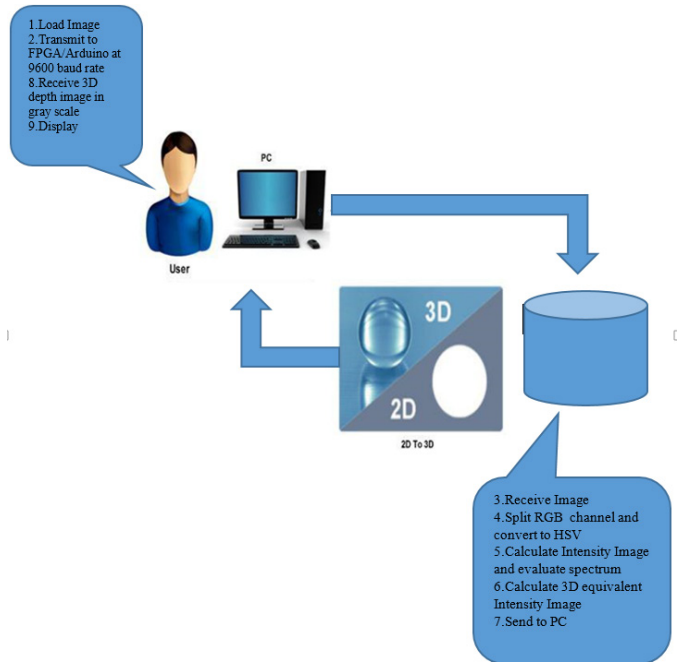


Figure 2. Block diagram of process

Since Papilio One is used for conversion and it has memory constraint ,instead of whole image being sent at a time to the kit, serially pixels are sent to Papilio One for computation one row after another. HSV image is then generated in MATLAB from these processed pixels.

2.3. Edge detection

As per Figure 1 the next step is to detect the edge. Here the input images are converted to gray scale for proper edge detection. Edge detection is the technique of identifying those points in the digitized image which causes a drastic change in the splendor of the image or that causes discontinuities. The edges are identified by rapid change in frequency. The objective is to capture these much needed rapid variations in the properties of the world. Edge detection is done by applying masks on the input which transforms the image into output image highlighting the sharp edges in the object. There are several masks available like Canny, Prewitt, Sobel, Roberts etc. which extracts the sharpness in the images.

The Canny detection is a popular edge detection method. It basically locates the maxima of gradient-I in the neighborhood. Gaussian channel derivative is used to compute the gradient. All the four steps [9] are used to detect the edges of the image. The strategy utilizes two edges, to distinguish solid and frail edges, and incorporates the powerless edges in the yield just on the off chance that they are associated with solid edges. Gaussian channel is connected to evacuate any clamor show in a picture, and more inclined to identify genuine powerless edges [10].

Object should be detected properly, because if it is not then the next operation of hole filling will not be performed properly. Morphological dilation operation is added before the hole-filling to improve the object detection. Pixels are added to the object boundaries during dilation and the number of pixels added is a function of the shape & size of the structuring elements deployed for processing the image.

2.4. Hole filling and depth estimation

The images under go morphological operations viz., dilation and after dilation operation, the detected objects are filled using MATLAB imfill function. For this filling operation first labeling is done using `bwlabel' function which generates a matrix same size as that of the input matrix containing labels for all connected objects in the input matrix.

The input "n" can either be a four connected or eight connected objects with a default value of eight. The pixels with 0 labelling coressponds to the background and the labelling from 1 through 8 coressponds to each of the coneected objects. The ouput gives the number of the connected objects in the iput image. Along with bwlabel function n is used to return vectors of indices for the pixels that make up a specific object. Then that specific object is cropped for filling new depth values using `imcrop' function. This function creates an interactive crop image. From that cropped object maximum and minimum values are calculated. For the same coordinates, values of V (intensity) retrieved from FPGA are stored in array. From the array maximum and minimum values of V are searched to calculate the range. 255/range = multiValue. So for the same coordinates the

new value which gives depth is calculated using following equation:

$$val=((imgHSV(yy,xx,3))-minVal)*multiVal \quad (1)$$

where, val is new depth value of depth, (imgHSV(yy,xx,3) gives values of V for that particular pixel, minVal is minimum value from array and multivalue .

2.5. Anaglyph Image Generation

So far depth map image is estimated, now using this depth map and 2D input image anaglyph image is generated. The anaglyph image is generated by using the difference value of each pixel from the depth map. Each pixel is shifted by the corresponding difference value in an input image. The parallax is given by

$$parallax(x,y)=C[1-depth(x,y)]/128 \quad (2)$$

where C is the maximum parallax. Anaglyph image is got by shifting each pixel to the right by parallax(x,y)/2.

Algorithm:

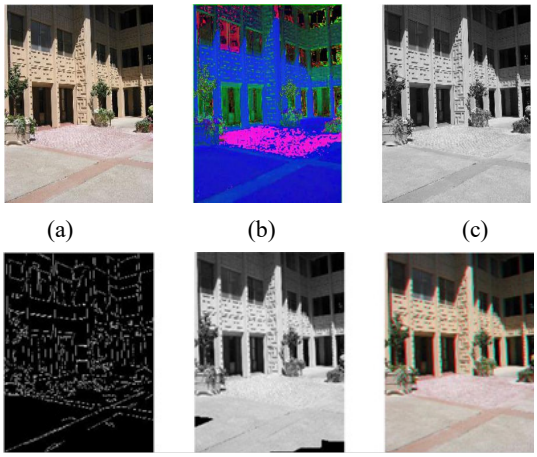
- Load input 2D image from database.
- Send input images pixel serially to FPGA to perform RGB to HSV conversion.
- Display HSV image on PC.
- Convert input image to grayscale image.
- Perform Canny edge detection on grayscale image.
- Display edge detection output.
- Perform dilation operation on edge detection output.
- Perform hole filling operation.
- Label images for object detection.
- Crop the detected objects.
- From the HSV image get the values of V for the same coordinates of the object.
- Assign depth value to new pixel using equation 1 and display.
- Generate anaglyph images and display.

3. Simulation Results

The algorithms we used work only for restricted scenario since cues are not considered. Proposed algorithm is tested for indoor image database, outdoor images database and database created by our own. Figure 3 is the output of various stages of proposed block diagram.

Input is from Make3D dataset of 500X 500 resolution. The first dataset used is a Make3D dataset made up of out-door images which has the depth fields that are acquired by a laser finder. For computational efficiency, the images are resized and the resolutions of the Make3D images are 1704 x 2272. All the stages of the block diagram are given as the output. Figure 3 (f) gives the 3d output which has to be viewed by the 3d glasses.

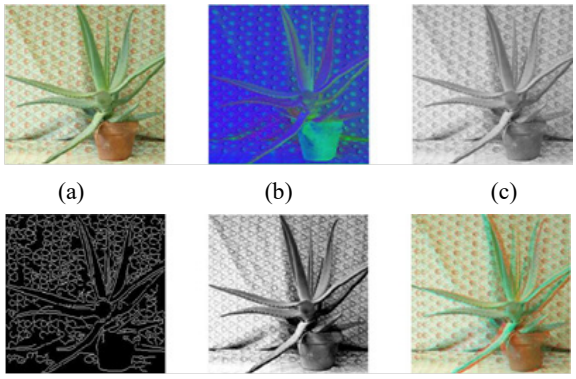
Figure 4 is the output of various stages and Input is from Middlebury dataset of 300X 259 resolution an in-door image, for computational efficiency, the images are re-sized. Figure 4 (f) gives the 3d output which has to be viewed by the 3d glasses.



(a) (b) (c)

(d) (e) (f)

Figure 3: (a) Query (b) HSV (c) Grayscale (d) Edge Detected (e) Depth (f) Anaglyph .

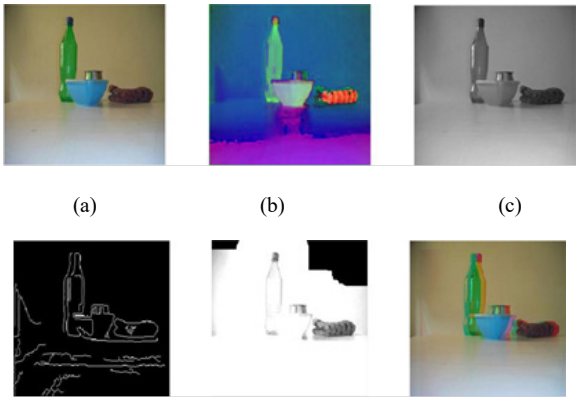


(a) (b) (c)

(d) (e) (f)

Figure 4: (a) Query (b) HSV (c) Grayscale (d) Edge Detected (e) Depth (f) Anaglyph.

We continued the experiment with our own database with dataset of 300 X 225 resolution, indoor images. The quality of the output was good as that of the Middlebury.



(a) (b) (c)

(d) (e) (f)

Figure 5: (a) Query (b) HSV (c) Grayscale (d) Edge Detected (e) Depth (f) Anaglyph

Following is the analysis of the images in terms of computing time efficiency of Make 3D dataset. The algorithm was executed in MATLAB and in FPGA and the table shows the time taken of the images for execution.

Table 1: Analysis of Images from Make3D Dataset

Size of input image	Execution time	
	MATLAB	FPGA
100*100	1.08ms	15min 26s
200*200	1.76ms	1hr 00min
300*300	9.05ms	2hr 26min
400*400	37.41ms	4hr 06min
500*500	122.65ms	-
600*600	274.75ms	-
700*700	505.38ms	-
800*800	888.00ms	-
900*900	1451.07ms	-
1000*1000	2196.97ms	-

The time taken by MATLAB to execute the complete process using a single core processor is in milliseconds and for FPGA the time taken is exponentially increasing with increase in the size of the input image. For any real time processing MATLAB would give a time efficient output and the future scope is to extend this algorithm work for multi-core processors.

4. Subjective analysis

Various objective quality metrics like MSE and PSNR are available. This can be used for objective analysis of 2D but not suitable for the 2-D to 3-D contents [11]. The generated depth-map is a pseudo-depth map and is not a real depth-map. So to evaluate the visual quality of output of proposed algorithm it is compared with other algorithms [12]. Figure 6 shows the form shared with people to mark the quality of the images shared with them to analyze. It shows the credit score in percentage depending on the visual perception of the viewer.

Figure 7 shows the subjective analysis of people. For subjective analysis survey from 50 people with normal or correct-to-normal visual acuity and stereo acuity for visual comfort of output. We have shown them output of our algorithm and for the same images output of other two methods. Images are taken from standard dataset.

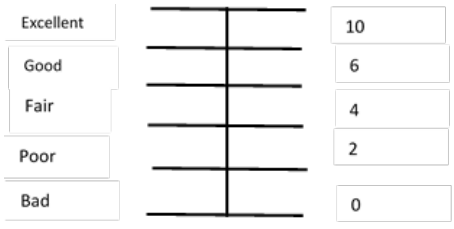


Figure 6: Rating scales used for assessment

Table 2 shows subjective evaluation, where M1 is our method, M2 is global method and M3 is Make3d algorithm. People has given good scale to our algorithm while excellent scale to other to algorithms. The scale on which people has given rating is shown in Figure 6. From the survey its found that they felt proposed methods output good, while of other two method's excellent. The subjective analysis also shows that proposed algorithm provides comfortable viewing experience.

Table 2: Sample Analysis of Images from Make3D Dataset

viewer count	Image1			Image 2			Image 3			Image 4		
	M 1	M 2	M 3	M 1	M 2	M 3	M 1	M2	M 3	M 1	M2	M 3
1	8	10	10	8	10	10	8	10	8	6	10	8
2	6	10	10	6	10	10	8	10	10	6	10	10
3	8	10	10	6	10	10	8	10	8	6	10	8
4	8	10	10	8	10	10	8	10	10	8	10	10
5	8	10	10	8	10	10	8	10	10	6	10	10
6	8	10	10	8	10	10	8	10	8	8	10	8
7	8	10	10	8	10	10	8	10	8	8	10	10
8	8	10	10	8	10	10	8	10	8	8	10	10
9	8	10	10	8	10	10	8	10	8	8	10	8
10	10	8	8	10	8	8	8	10	8	8	10	8
total	80	98	98	80	98	98	80	100	86	72	100	90

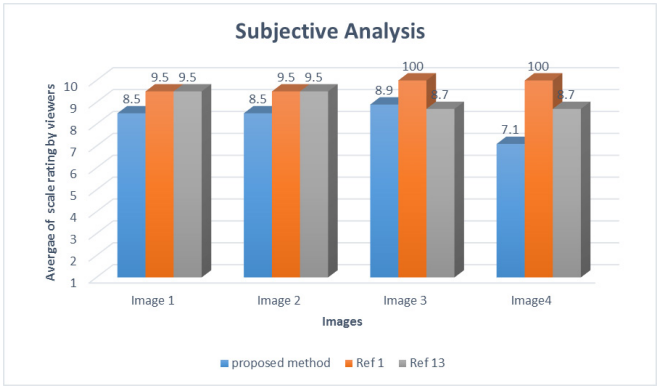


Figure 7: Graph of Subjective Analysis



(a) (b) (c)

Figure 8: output images (a) proposed method (b) global method (c) Make3d

5. Conclusion

In current scenario, 2D-to-3D conversion is a very active research area as it gives more lifelike viewing experience and hence its popularity is increasing. As there are many hardware available which support 3D and hence there is an urgent need of 3D contents. Though there are distinct methods introduced by researchers but these method works for restricted scenes.

In this paper 2D-to-3D image conversion was implemented which can take global image as an input. The work comes under automatic approach of conversion, in which algorithm itself does the whole conversion. Using RGB to HSV conversion along with simple MATLAB operations 2D image is converted into 3D. The method was tested on three distinct standard databases and on our own database. For output analysis subjective evaluation was done on 10 people for visual comfort of implemented systems output with the output of other two methods. People rated output as good from scaling range as visual comfort.

The work can be further extended by working on video. Here conversion from RGB to HSV step is implemented on starter kit of FPGA; it may be possible to implement whole system on a computational efficient FPGA.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors hereby express their gratitude to Dr. D Y. Patil Institute of Engineering, Research and Management, Akurdi for facilitating this research. We are highly privileged to have got the much needed support and assistance from almost each and every people of the institution for the successful outcome of this work. Special thanks, to all the teaching staffs of the institution, for their constant encouragement and guidance which has helped us in successfully completing this work. .

References

[1] R. Phan and D. Androutsos " Literature survey on recent methods for 2d to 3d video conversion" Multimedia Image and Video Processing, Second Edition, pages 691to716, 2012.

[2] J. Konrad, Meng Wang, P. Ishwar, Chen Wu, and D. Mukharjee. "Learning-based, automatic 2d-to-3d image and video conversion". Image Processing, IEEE Transactions on, 22(9):3485-3496, 2013. https://doi:10.1109 / TIP. 2013.2270375.

[3] Pan Ji, Lianghao Wang, Dong-Xiao Li, and Ming Zhang. "An automatic 2d to 3d conversion algorithm using multi-depth cues", In Audio, Language and Image Processing (ICALIP), 2012 International Conference on, pages 546-550, 2012.

[4] M. Guttmann, L. Wolf, and D. Cohen-Or. "Semiautomatic stereo extraction from video footage" Computer Vision, 2009 IEEE 12th International Conference on, pages 136-142, 2009.

[5] Make3D, Convert your still image into 3D model, (2015, June 20). http://make3d.cs.cornell.edu/data.html.

[6] NYU Depth Dataset V2, (2016, Jan 11) , https://cs.nyu.edu/ ~ silberman / datasets /nyu_depth_v2.html

[7] Middlebury Stereo Datasets, (2013, Nov 14), http:// vision.middlebury.edu /stereo/data/

[8] Welcome to Sony Support , (2015, Dec 24), https:// www.sony.co.uk /electronics /support

[9] Zhao Xu1, Xu Baojie, Wu Guoxin1 "Canny edge detection based on Open CV" 2017 IEEE 13th International Conference on Electronic Measurement & Instruments ICEMI', pages 53-56, 2017.

[10] Janusz Konrad, Meng Wang, and Prakash Ishwar," 2D-to-3D Image Conversion by Learning Depth from Examples", Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference, 2012. https://doi.org/ 10.1109/CVPRW.2012.6238903.

[11] Mani B. Fard, Mani B. Fard, Ulug Bayazit, Ulug Bayazit, "Automatic 2D-to-3D video conversion by monocular depth cues fusion and utilizing human face landmarks", Proc. SPIE 9067, Sixth International Conference on Machine Vision (ICMV 2013), 90670B, 2013. https://doi.org/10.1117/12.2049802.

[12] Saxena, A. and Min Sun and Ng, A.Y. Make3D:" Learning 3D Scene Structure from a Single Still Image". Pattern Analysis and Machine Intelligence, IEEE Transactions, pages 824-840,2009.https://doi.org/ 10.1109/TPAMI.2008.132.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 300-307 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Investigating The Detection of Intention Signal During Different Exercise Protocols in Robot-Assisted Hand Movement of Stroke Patients and Healthy Subjects Using EEG-BCI System

Maryam Butt*,[1], Golshah Naghdy[1], Fazel Naghdy[1], Geoffrey Murray[2], Haiping Du[1]

[1] *School of Electrical Computer and Telecommunications Engineering, University of Wollongong (UOW), 2522, Australia*

[2] *School of Medicine, University of Wollongong (UOW), 2522, Australia*

Article history:
Received: 17 May, 2019
Accepted: 20 July, 2019
Online: 07 August, 2019

Keywords:
Intention detection
EEG
Brain-computer interfaces
MRCP
AMADEO hand robot
Support vector machines
Stroke

A B S T R A C T

Improving the hand motor skills in post-stroke patients through rehabilitation based on movement intention derived signals from the brain in conjunction with robot-assistive technologies are explored. The experimental work is conducted using Electroencephalogram based Brain-Computer Interface (EEG-BCI) system and the AMADEO hand rehabilitation robotic device. Two protocols using visual-cues and then using a 2-Dimensional (2D) interactive game is presented on a computer screen to healthy subjects as well as post-stroke patients performing the hand movements. The movement intention signals during hand movement are detected through the Support Vector Machine (SVM) classifier. The intent signals produced at six distinct electrodes are investigated to determine electrodes contributing most to the SVM classifier's performance. Overall, the game protocol shows better classification results for both healthy and stroke patients compared to the visual-cues protocol. FC3 is found to be the most consistent electrode site for the detection of the motor intention of the hand for both protocols. In the experimental work, average classification accuracy for the visual-cues protocol of 67.56% for healthy subjects and 56.24% for stroke patients were obtained. For the game protocol, the classifier accuracy produced for healthy participants was 79.7% and for the post-stroke patients was 66.64%. The results confirm that the intention signal is more pronounced during more engaging activities, such as playing games, for both healthy and stroke subjects. Therefore, the effectiveness of rehabilitation therapy for post-stroke patients could be significantly enhanced using interactive and engaging exercise protocols.

1. Introduction

This paper is an extension of work presented in 18th IEEE International Conference on Bioinformatics and Bioengineering (BIBE) [1]. Stroke is the main cause of prolonged disability among adults [2]. The most common impairment occurs when a stroke victim has a motor loss of limb(s) on one side of the body causing difficulty with walking and ability to perform activities of daily living. There are many types of rehabilitation therapies currently being practiced, and others being investigated. One of the therapies types being investigated and coming into clinical practice is Robot-Assisted therapy. Advancement in robotic technology over the last decade has led to increasing interest in this type of therapy. For example, Ang et al. [3] used the MIT-Manus robot for a randomized control study with 26 stroke patients to restore their arm movement. Similarly, a haptic knob robot is being used for arm rehabilitation of stroke survivors with positive outcomes [4]. A state-of-the-art robotic device called AMADEO, designed for fine finger motor skill improvement was tested on eight cortical stroke patients and resulted in a 35% increase in their hand movement after multi-session training [5]. All these strategies of post-stroke rehabilitation depend on some neurological adaption that occurs in the patient's brain to restore the impaired function of the limb. The brain's ability to adapt to new neural changes, even in adulthood, is called the neuroplasticity or brain plasticity [6].

Brain-Computer Interface (BCI) has an ability to utilize neuroplasticity mechanisms to improve motor function of post-stroke patients during their rehabilitation process [7]. Enhanced motor intention improves motor execution which in turn is fed back to the rehabilitation process. The mechanism of neuroplasticity reinforces the neural pathways which control the movement, leading to functional and motor recovery of the

*Corresponding Author: Maryam Butt, mb077@uowmail.edu.au

patients [8]. Brain signals that have been used in BCI systems include Electroencephalogram (EEG), Magnetoencephalography (MEG), ElectroCorticoGraphy (ECoG), as well as functional Near-InfraRed Spectroscopy (fNIRS). Amongst these modalities, EEG is the most common modality employed to drive various BCI systems [9]. It is popular because it is not an invasive approach to record the brain's electrical activities, and because EEG signals have the highest resolution in time-domain [10].

Many studies demonstrate applications of EEG-BCI systems to detect movement intention for various upper-limbs, such as movements of the arm [11-15], elbow [16], [17] and wrist [18] and hand [19-23] for post-stroke rehabilitation.

A self-paced reaching arm movement with the help of a haptic device has been studied in [11] and [12]. In a single-trial EEG protocol, the intention of reaching movement was detected about 500 ms before the actual limb movement [11]. In the follow-up study, the movement directions were decoded up to 76% accuracy for healthy subjects and up to 47% accuracy for stroke patients performed with their impaired arm [12]. Ibanez et al. [13] recruited six healthy subjects and six stroke patients to detect the movement onset of voluntary arm reaching action. The results showed that the True Positive Rate (TPR) of the classifier for healthy participants was 74.5±13.8% while for stroke patients, TPR was 82.2±10.4%. In another study, four chronic post-stroke patients participated in eight training sessions based on arm reaching actions [14]. The intention for the movement was decoded using an EEG-BCI system which then triggered Functional Electrical Stimulation (FES) for further assistance to perform the task. The authors stated that their protocol correctly classified about 66% of movements with an average detection latency of 112±278 ms. Moreover, clinical tests and kinematics results proved the feasibility of the designed intervention for post-stroke recovery.

Among many classifiers, Support Vector Machine (SVM) classifier is commonly used in intent detection of limb movement. For example, Frisoli et al. [15] proposed a gaze-based robot-assisted arm movement protocol in which the real-time movement of a robotic device was controlled by the intended signal of subjects. A Motor Imagery (MI) based EEG-BCI system was used to detect the intention signal through SVM. The authors reported a classification accuracy rate of 89.4±5% for robot-based movement mode. Similarly, Hortal et al. [16] measured the classification performance of SVM by designing a hybrid EEG-BCI system composed of an exoskeleton device and FES for elbow movement. The hybrid BCI system was tested using two training protocols which were MI-based training protocol and a training protocol based on motor intention detection. In the first training protocol, the authors achieved a classification accuracy of about 83% for healthy subjects and 65.3% for stroke patients, while in the second training protocol, the accuracy achieved was approximately 77% and 71.6% for healthy and patient groups, respectively.

Asynchronous EEG-BCI systems are also being deployed for motor intent detection of upper limbs. The triggering of a robotic device through intention signal was studied by Bhagat et al. [17] for elbow joint movement using an asynchronous EEG-BCI robot-assisted system. The motor intent signal of four chronic stroke patients was detected which then triggered an exoskeleton device called MAHI-EXO-II to encourage and guide the active participation of subjects. The intention signal was detected on average −367±328 ms before the actual motor execution. Bai et al. [18] also developed an asynchronous BCI protocol for real-time and online prediction of the self-paced movement of the wrist. Seven healthy subjects were recruited in the study to perform an extension of their wrist whenever they wanted. The authors successfully predicted the voluntary movement at approximately 0.6 s before the real execution of the movement.

The latest applications of EEG-BCI systems have been in detecting intention for movement of the hand, which has a greater complexity of movement than other upper limbs. The hand occupies the largest cortical representation in the motor cortex of the brain [24] as well as playing the most vital role in our daily life activities. Muralidharan et al. [19] distinguished resting state from the extension of fingers by using an open-loop EEG-BCI system. They detected attempted finger extension by post-stroke patients at about 200 to 600 ms before the actual movement onset. Only one EEG channel has been used by Jochumsen et al. [20] to detect motor intent signal during palmar grasp action using handgrip dynamometer. The signal was then decoded into speed and force. They showed that approximately 75% of the movements were detected 100 ms before movement onset, and approximately 60% of task-related movements were accurately decoded into the force and speed levels according to the performed task movements. In a separate study, Jochumsen et al. [21] also distinguish between the intention for three types of grasp tasks (palmar, pinch and lateral) achieving accuracy of 79%, 76%, and 63% respectively during 1-class, 2-class and 3-class classification problems. Ofner et al. [22] classified several arm movements which include hand-opening, hand-closing, forearm supination, forearm pronation, elbow extension as well as elbow flexion. The authors state that one movement can be classified from another with 55% classification accuracy. This research group achieved the classification accuracy of 93% while classifying grasping actions of hand from its rest state compared to the classification results stated by Jochumsen et al. in [21] [23].

In this study, we combine detection of motor intention signals using EEG-BCI system with robot assistive technologies to improve the effectiveness of the hand motor skills in post-stroke patients. Our first stage of research based on an EEG-BCI system and the AMADEO finger-hand rehabilitation robotic device (Tyromotion GmbH, Graz, Austria) is reported. The experimental setup is designed for classification of movement intention of hand-closing versus rest state. It consists of two distinct protocols, which we refer to as protocol A and protocol B. Protocol A, termed as visual-cues protocol, consists of hand-opening and hand-closing pictures which are presented to subjects. In protocol B subjects played an AMADEO game called Shoot-out. We utilize SVM, which is the most common supervised learning algorithm used for these classification problems. Intention signal produced at different single electrodes' positions during hand movement is analyzed to determine the electrode that is most consistent for the accuracy of SVM classifier.

The rest of the paper is organized as follows. In section 2, the experimental platform and experimental set up are introduced. In section 3, the pre-processing of the signals produced in the experimental work and the classifications method, the classification performance metrics are described. In section 4, the

results of the validation of the approach on stroke patients and benchmarking against the healthy subjects are presented. Some conclusions are finally drawn in section 5.

2. Experimental Design and Platform

2.1. Participants

The participants in the experimental work consisted of healthy subjects and post-stroke patients. The healthy subjects consist of four male participants with a mean age of 28 years and no history of any neurological disorder. Two post-stroke patients both right hand dominant were initially assessed through two commonly used clinical tests, namely the Motor Assessment Scale (MAS) and Fugl-Meyer Assessment (FMA) scale. Table I shows each patient's demographic details as well as their clinical tests scores. The Ethics committee at the University of Wollongong approved all methods and procedures performed in this experiment (Ethics application number: 2014/400). All participants provided written informed consent before this study commenced.

2.2. AMADEO Finger-Hand Rehabilitation Device

The AMADEO finger-hand rehabilitation device is an end-effector robot-assistive system specifically developed for hand movement recovery of post-stroke patients as shown in Figure 1 [25]. It has five Degrees-of-Freedom (DoF) allowing both passive and active movements of fingers as well as the thumb. The robot can generate various patterns of fingers and thumb movements as well as subjects can interact with the robotic device through 2D games.



(A) (B)

Figure 1: (A) AMADEO Finger-Hand Rehabilitation Unit (B) Hand-Arm Adjustment Support

2.3. Acquisition of EEG Signal

The EEG signals were acquired by deploying a 32 electrodes Ag/AgCl Quick-Cap (Compumedics-Neuroscan) in accordance to the 10-20 system for positioning electrodes. The electrode positioning diagram of the Quick-Cap is given in Figure 2. The Grael 4K EEG amplifier was employed in this study which has been set to 2048 Hz of sampling frequency. Of the 32 channels, the FPz electrode was used as the ground electrode and the CPz was set as the reference electrode. Thirty remaining electrodes were used for the acquisition of brain signal. In addition, the two electrodes were placed below the left eye and on the supraorbital ridge to record vertical eye movements and eye-blinks. Moreover, two other electrodes placed over the outer canthus of both eyes were used to monitor horizontal movements of eyes.

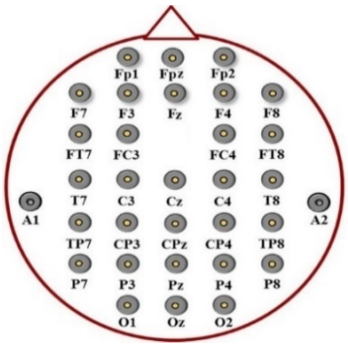


Figure 2: Quick-Cap Electrodes Positions

2.4. Experimental Setup

Participants were seated upright on a comfortable chair with their right arms attached to the AMADEO hand-arm adjustment support unit. During protocol A, each participant was trained to focus on visual-cues displayed on a computer screen. Visual-cues displaying hand-opening and hand-closing pictures to alert subjects to perform these specific hand movements. The hand-closing pictures were displayed every 5 s, followed by the hand-opening pictures 1 s later with a 4 s waiting period. This resulted in 5 s gap between any two hand-closing visual-cues. Each set comprised 23 trials of hand movements during this protocol.

Of the games available on the AMADEO system, the 'Shoot-out' was chosen for use in protocol B. The playing screen of Shoot-out game is shown in Figure 3. In this game, the subject closes their hand to shoot the drum coming out at equal time intervals. Subjects have up to 23 trials of the hand movements in each block.



Figure 3: AMADEO Shoot-out Game

Table 1: Stroke Patients' Details and Clinical Test Scores

Gender	Age	Stroke type	Time since stroke (months)	Lesion location	Paretic Hand	MAS Hand Score (0-6)	FMA Hand Score (0-14)
Female	64	Ischemic	7	Left pons	Right	2	9
Male	60	Ischemic	4	Left pons & left frontal regions	Right	2	8

All participants performed 6 blocks consisting of 23 trials (6 x 23 = 138) for both protocols. At each hand-closing event, manual event markers were sent to the CURRY 8 (Compumedics, Neuroscan) software which was used for EEG acquisition. Protocol B was the Shoot-out game of AMADEO, which was more interesting and interactive compared to protocol A. We explored the hypothesis in this study that whether a stronger motor intention signal can be produced and detected through SVM during protocol B than protocol A.

3. Signal Processing for Motor Intention Detection

The intention signal during hand-closing movements was detected by offline analysis of the data for both healthy and stroke subjects. The movement intention of a limb produces special EEG patterns in specific parts of the brain and such slow event-related potential is called Movement Related Cortical Potential (MRCP). It appears in the delta frequency band of EEG signal as a direct-current shifts up to 2 s prior to cue-based as well as self-initiated movements [26]. When the person performs the required motor action, the MRCP disappears. The continuous EEG signal at the selected electrodes is divided into two types of epochs. The first, containing MRCP signal is called Move epoch and the second, which does not have an MRCP signal, is named as No-Move epoch. Different time-domain features of MRCP signal are extracted from both these epochs. SVM algorithm is then employed to classify Move and No-Move epochs and, therefore, detecting intention signal of the hand movement.

3.1. Pre-processing

The first step in pre-processing is to filter the EEG signal for the frequency range between 0.1-1 Hz [17] as this narrow range of delta band best captures the anticipatory based MRCP [27]. This band-pass filtering was applied in three steps: first, a high-pass filter with cut-off frequency (f_c) of 0.1 Hz was applied; second, for re-referencing of signals, a Common Average Reference (CAR) filter was applied to all signals; and third, a low-pass filter with f_c of 1 Hz was applied. Both low-pass and high-pass filters were 4th order Butterworth causal filters. After filtering, the signals were down-sampled from 2048 Hz to 20 Hz to increase computational efficiency [17]. The CURRY 8 (Compumedics-Neuroscan) software was used to perform all the aforementioned pre-processing steps.

3.2. Epoch Extraction & Channel Selection

For further processing, the data were imported into a MATLAB toolbox known as 'EEGLAB'. Generally, MRCP signal does not exist 2 s before the onset of movement, all Move epochs were extracted from continuous EEG data between -2 s to 0 s with 0 s indicating the instance of actual motor execution. Between 0.5 s to 2.5 s, No-Move epochs were extracted when subjects opened or relaxed their hands. Therefore, epoch length was fixed at 2 s for both epochs. All Move epochs were inspected visually for artifacts, for instances eye-blinks, head movement, and other movement-related artifacts, and then these were removed from the data.

The EEG electrodes located over the left hemisphere of the brain were only chosen because the right-hand movement was considered in this experiment. The electrodes selected were C3, FC3, CP3, Cz, T7 and a Short Laplacian (SL) channel calculated using the C3-(FC3+Cz+CP3+T7)/4 formula [28]. In literature, the C3 channel is most commonly used for right-hand motor intention detection using the MRCP signal. The four neighboring electrodes of the C3 channel and their linear combination were also investigated to determine the best electrode choice for each protocol. It is possible that the negative peak of the MRCP exists before -1.5 s is due to artifact presence. Therefore, every Move epoch was again visually inspected and such Move epochs and their counter-part No-Move epochs were deleted to get the final processed data which was then used to extract features of MRCP signal.

3.3. Feature Extraction & Classification

The two time-domain features from both the Move and No-Move epochs were extracted. These features were the negative peak and the slope of the MRCP signal. The Move epochs have prominent negative peak and slope of MRCP signal compared to No-Move epochs. Based on these time-domain features, the SVM classifier differentiates between both epochs. These were plotted and outliers were removed (along with their counter-class features) before applying the SVM. This prevented biasing the results of SVM due to outliers. In each protocol, an average of 10±2 epochs was rejected per subject. The Move class was marked with a label of '1' and the No-Move class was marked with a label of '0'. Depending on these input features, the SVM classified between class 1 and 0. Three-fourths of the dataset was utilized for training data while one-fourth was used as test data for classification.

3.4. Evaluation of SVM Classifier Performance

The SVM classifier's performance was determined based on classification accuracy percentage, True Positive Rate (TPR)–also known as sensitivity and True Negative Rate (TNR)–also called specificity. They were calculated using the relationships given in (1), (2) and (3) where True Positive, False Positive, True Negative and False Negative were abbreviated as TP, FP, TN, and FN respectively.

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$TPR=TP/(TP+FN) \quad (2)$$

$$TNR=TN/(TN+FP) \quad (3)$$

Receiver Operating Characteristics (ROC) curve was analyzed to determine the classifier's performance during protocol A and B. Moreover, Area Under the Curve (AUC) for the ROC curve was calculated for both protocols. In the end, all these performance metrics were compared to find out which protocol helps better to maintain stronger motor intention level in both healthy subjects and stroke patients and therefore, obtaining better classification results. In addition, these results also could show which electrode was the best choice for intent detection during protocol A and protocol B.

4. Results & Discussion

The experimental work demonstrated the utility of the AMADEO device for rehabilitation of post-stroke patients, while also performing the motor intention detection using the described SVM algorithm.

Six electrodes which include C3, FC3, CP3, Cz, T7, and SL were chosen to determine the importance of electrode selection in intent detection during each protocol. Performance metrics were calculated after the SVM application on test data, including classification accuracy, TPR, TNR and AUC for ROC.

4.1. Intent Detection of Stroke Patients

The performance of the SVM algorithm for detection of hand motor intention signal produced by the stroke patients is studied in this sub-section. Table II presents the accuracy of the SVM classifier at all six selected electrodes for protocols A and B. For protocol A, the FC3 electrode shows the maximum accuracy of the SVM classifier of about 72%. Whereas, the same electrode shows classification accuracy of 89% for intent detection when stroke patients perform the hand movement during protocol B. This supports our hypothesis that during protocol B, subjects are more involved in performing hand movement and are likely to have greater classification accuracy as compared to protocol A. The same trend is demonstrated by the results of C3, CP3, T7, and SL electrodes. However, for electrode Cz, protocol B is showing lower classifier's accuracy than that for protocol A. The list of electrodes based on the classification accuracies from high-to-low for protocol A is FC3, Cz, SL, C3, T7, and CP3. A similar list for the protocol B is FC3, SL, C3, CP3, T7, and Cz. It is clear from the classifier accuracy results that for both protocols, FC3 shows the maximum classifier performance. However, for protocol B the classifier's accuracy is higher than that for protocol A at all electrode sites except Cz.

4.2. Intent Detection of Healthy Subjects

The results of the classifier accuracy for healthy subjects for protocols A and B are presented in Table III. The analysis of data for healthy subjects shows that protocol B has better accuracy of SVM classifier than protocol A, except for the channels Cz and T7. With respect to the results acquired for FC3, protocol B has an accuracy of SVM classifier of 98% while the classification accuracy of protocol A is 86%. From Table III, it is clear that the classifier's accuracy varies in accordance with the selected electrode. For protocol A, the electrodes can be listed as FC3, Cz, C3, T7, SL, CP3 electrodes can be ranked in order of their effectiveness in determining the intent signal. Similarly, for protocol B, the ranking is FC3, C3, SL, CP3, Cz, T7.

Table 2: Accuracy of SVM Classifier for Post-Stroke Patients

Channel	SVM Classifier's Accuracy (%)	
	Protocol A	Protocol B
C3	52.88	81.01
CP3	26.25	60.16
FC3	72.22	88.89
Cz	70.45	34.56
T7	49.29	51.43
SL	66.33	83.78

Table 3: Accuracy of SVM Classifier for Healthy Subjects

Channel	SVM Classifier's Accuracy (%)	
	Protocol A	Protocol B
C3	64.58	92.86
CP3	57.7	76.3
FC3	86.25	98.0
Cz	73.81	70.0
T7	63	57.69
SL	60.0	83.33

When the Cz electrode from stroke patients' data as well as the Cz and T7 electrodes from healthy subjects' data were chosen, protocol A had slightly better classification accuracy than protocol B. There could be many possible reasons for this apparent contradiction. The artifact removal is performed using visual inspection method so it might be possible that the data at Cz and T7 electrodes still contain some artifacts in the test data used for protocol B that made the classifier accuracy for protocol B less than protocol A. Moreover, the outlier values removal form Cz and T7 electrodes might cause loss of large data points for protocol B and due to fewer test data values, the performance of protocol B has deteriorated.

4.3. Other Performance Metrics of the SVM Classifier

The accuracy alone cannot justify the performance of the classifier, therefore, sensitivity (TPR) and specificity (TNR) parameters should be also considered. The TPR and TNR values show whether the extracted features in Move and No-Move classes are distinct enough to be classified accurately by SVM. Figure 4 shows bar graph representations of TPR and TNR using stroke patients' data from protocol A at all six electrodes. Similarly, Figure 5 shows TPR and TNR values at all selected electrodes for protocol B. For protocol A, Figure 4 shows that

TNR for CP3, FC3, Cz, and SL is significantly lower than its corresponding TPR. On the other hand, in the case of T7, TNR is slightly higher and slightly lower for C3. However, for protocol B, TNR is higher than TPR for each choice of electrode except for CP3 and T7 as shown in Figure 5.

Similarly, TPR and TNR values using data of healthy subjects for both protocols is shown in Figure 6 and Figure 7. In the case of protocol A, Figure 6 demonstrates that TPR obtained for C3, Cz and FC3 is higher than its corresponding TNR, whereas, for CP3, T7, and SL, the reverse is the case. Figure 7 shows that TPR, for all six electrodes, is higher than its corresponding TNR for protocol B indicating that the Move class contains the adequate features to correctly detect motor intention signal.

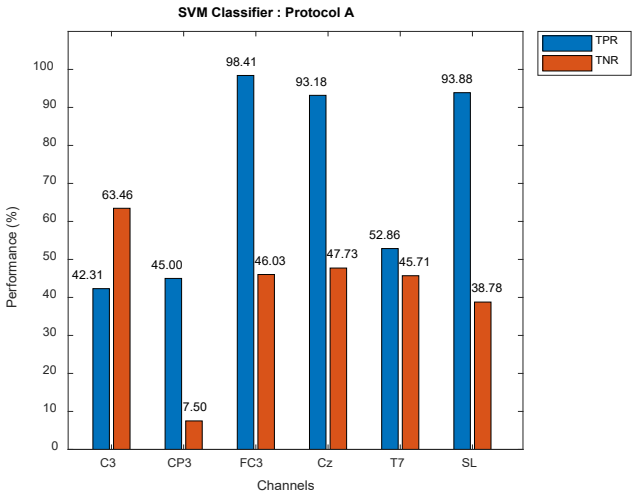


Figure 4: Protocol A Stroke Patients Data

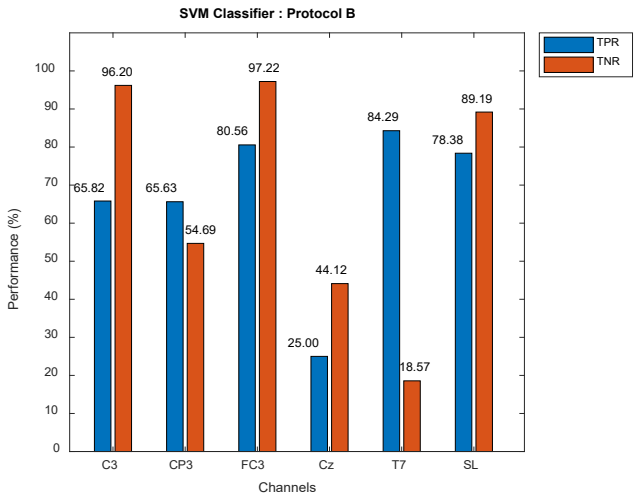


Figure 5: Protocol B Stroke Patients Data

ROC curve is another way of comparing the performance of the SVM classifier for protocols A and B. Figure 8 shows the ROC curve for the FC3 channel, the most appropriate channel for intent detection of hand movement for this experiment, for healthy participants which shows that protocol B has superior performance.

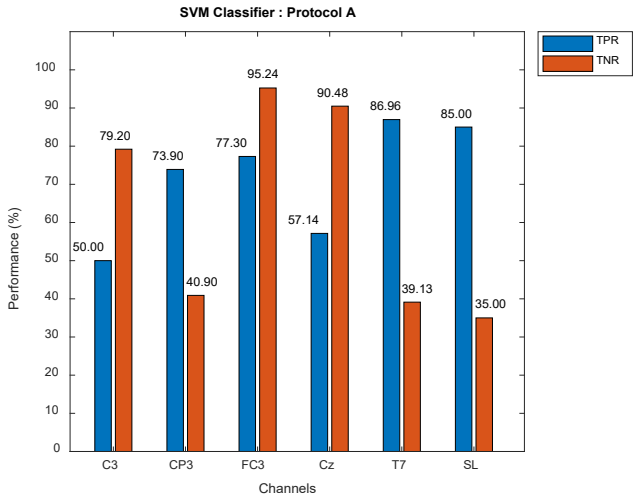


Figure 6: Protocol A Healthy Subject Data

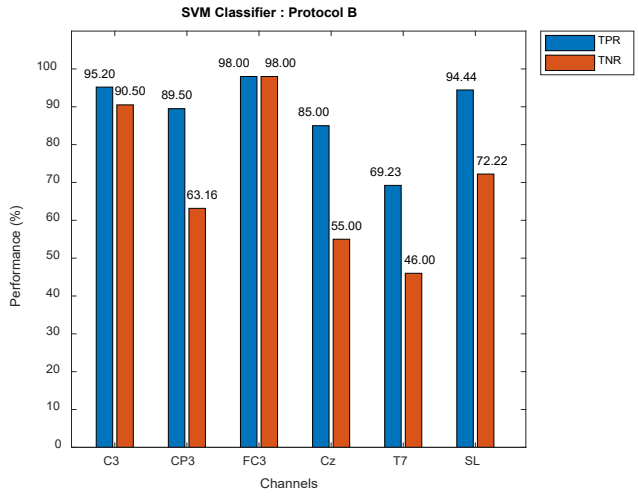


Figure 7: Protocol B Healthy Subject Data

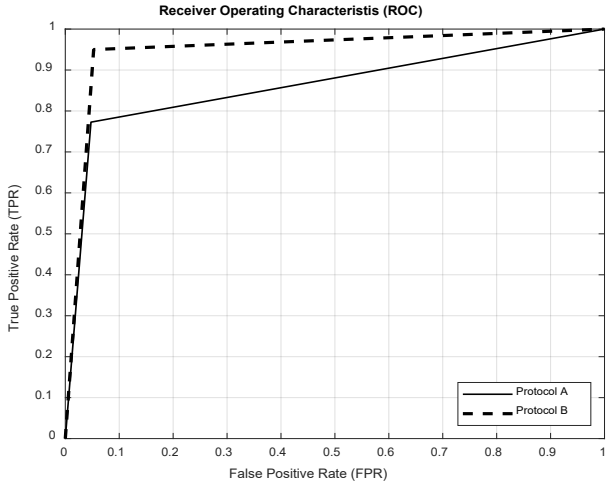


Figure 8: ROC Curve Using FC3 Channel

The Area Under the Curve (AUC) for ROC is also an important factor in assessing the classifier performance. The excellent classifier has an AUC value of 1 whereas 0.5 value

shows its random guessing behavior. The AUC values for FC3 channel for both healthy subjects and stroke patients are listed in Table IV. It is found that AUC for protocol B is greater than protocol A for both healthy subjects and stroke patients which was the expected outcome. Moreover, AUC value for protocol B in the healthy group is ~0.95 which is approximately equal to ideal AUC value for the classifier.

Table 4: Area Under the ROC Curve for FC3 Channel

Group	AUC for FC3 Channel	
	Protocol A	Protocol B
Stroke	0.7222	0.8889
Healthy	0.8626	0.9487

5. Conclusion

The utilization of EEG-BCI system together with AMADEO robotic rehabilitation system for the functional recovery of the hand of the post-stroke patients was reported as the first stage of our work. The movement intention for both healthy participants as well as post-stroke patients was detected using SVM classifier. Two protocols that were based on simple visual-cues and the AMADEO 2D game were used to detect the hand-closing movement intention vs the resting state. This was followed by an evaluation of the SVM classifier performance by selecting different single electrodes. The average classifier accuracy of 67.56% for the visual-cue protocol and 79.7% for the gaming protocol was achieved for healthy subjects. Similarly, for stroke patients, the classifier accuracy obtained was 56.24% and 66.64% for the visual-cue and the game protocols respectively. The results demonstrated that the 2D games, for example, AMADEO Shoot-out were better activities in retaining concentration compared to static pictures using visual-cues. We conclude that gaming scenarios are preferable for robot-based rehabilitation exercises to promote the active participation of the patients. The FC3 electrode was found to be the best single electrode for the designed experiment with both protocols. Both TPR and TNR were important in determining the performance of the classifier, as these parameters could indicate whether good results of the classifier were obtained either by detecting positive class or negative class accurately. In addition to TPR and TNR, it was shown that the ROC plot and AUC parameters of the classifier could also be used as the performance metric evaluators for the classifier.

Conflict of Interest

There is no conflict of interest among authors.

Acknowledgment

This work is performed by joint support by The University of Wollongong and Higher Education Commission (HEC) Pakistan.

References

[1] M. Butt, G. Naghdy, F. Naghdy, G. Murray, and H. Du, "Investigating Electrode Sites for Intention Detection During Robot Based Hand Movement Using EEG-BCI System," in 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), 2018, pp. 177-180: IEEE.

[2] J. Adamson, A. Beswick, and S. Ebrahim, "Is stroke the most common cause of disability?," Journal of stroke and cerebrovascular diseases, vol. 13, no. 4, pp. 171-177, 2004.

[3] K. K. Ang et al., "A Randomized Controlled Trial of EEG-Based Motor Imagery Brain-Computer Interface Robotic Rehabilitation for Stroke," Clinical EEG and Neuroscience, Article vol. 46, no. 4, pp. 310-320, 2015.

[4] K. K. Ang et al., "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: Results of a three-armed randomized controlled trial for chronic stroke," Frontiers in Neuroengineering, Article vol. 7, no. JUL, 2014, Art. no. 30.

[5] X. Huang, F. Naghdy, H. Du, G. Naghdy, and G. Murray, "Design of adaptive control and virtual reality-based fine hand motion rehabilitation system and its effects in subacute stroke patients," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 6, no. 6, pp. 678-686, 2018.

[6] B. B. Johansson, "Brain plasticity and stroke rehabilitation: the Willis lecture," Stroke, vol. 31, no. 1, pp. 223-230, 2000.

[7] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," Nature Reviews Neurology, vol. 12, no. 9, pp. 513-525, 2016.

[8] M. Grosse-Wentrup, D. Mattia, and K. Oweiss, "Using brain–computer interfaces to induce neural plasticity and restore function," Journal of neural engineering, vol. 8, no. 2, p. 025004, 2011.

[9] H.-J. Hwang, S. Kim, S. Choi, and C.-H. Im, "EEG-based brain-computer interfaces: a thorough literature survey," International Journal of Human-Computer Interaction, vol. 29, no. 12, pp. 814-826, 2013.

[10] J. S. Kumar and P. Bhuvaneswari, "Analysis of Electroencephalography (EEG) signals and its categorization–a study," Procedia engineering, vol. 38, pp. 2525-2536, 2012.

[11] E. Lew, R. Chavarriaga, S. Silvoni, and R. Millan Jdel, "Detection of self-paced reaching movement intention from EEG signals," (in eng), Front Neuroeng, vol. 5, p. 13, 2012.

[12] E. Y. L. Lew, R. Chavarriaga, S. Silvoni, and J. D. Millan, "Single trial prediction of self-paced reaching directions from EEG signals," (in English), Frontiers in Neuroscience, Article vol. 8, p. 13, Aug 2014, Art. no. 222.

[13] J. Ibanez et al., "Detection of the onset of upper-limb movements based on the combined analysis of changes in the sensorimotor rhythms and slow cortical potentials," (in English), Journal of Neural Engineering, Article vol. 11, no. 5, p. 12, Oct 2014, Art. no. 056009.

[14] J. Ibáñez et al., "Low latency estimation of motor intentions to assist reaching movements along multiple sessions in chronic stroke patients: a feasibility study," Frontiers in neuroscience, vol. 11, p. 126, 2017.

[15] A. Frisoli et al., "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1169-1179, 2012.

[16] E. Hortal, D. Planelles, F. Resquin, J. M. Climent, J. M. Azorin, and J. L. Pons, "Using a brain-machine interface to control a hybrid upper limb exoskeleton during rehabilitation of patients with neurological conditions," Journal of Neuroengineering and Rehabilitation, vol. 12, Oct 2015, Art. no. 92.

[17] N. A. Bhagat et al., "Design and optimization of an EEG-based brain machine interface (BMI) to an upper-limb exoskeleton for stroke survivors," Frontiers in Neuroscience, Article vol. 10, no. MAR, 2016, Art. no. 122.

[18] O. Bai et al., "Prediction of human voluntary movement before it occurs," Clinical Neurophysiology, vol. 122, no. 2, pp. 364-372, 2011.

[19] A. Muralidharan, J. Chae, and D. Taylor, "Extracting attempted hand movements from EEGs in people with complete hand paralysis following stroke," Frontiers in neuroscience, vol. 5, p. 39, 2011.

[20] M. Jochumsen, I. K. Niazi, D. Taylor, D. Farina, and K. Dremstrup, "Detecting and classifying movement-related cortical potentials associated with hand movements in healthy subjects and stroke patients from single-electrode, single-trial EEG," Journal of neural engineering, vol. 12, no. 5, p. 056013, 2015.

[21] M. Jochumsen, I. K. Niazi, K. Dremstrup, and E. N. Kamavuako, "Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation," Medical & biological engineering & computing, vol. 54, no. 10, pp. 1491-1501, 2016.

[22] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, "Upper limb movements can be decoded from the time-domain of low-frequency EEG," PloS one, vol. 12, no. 8, p. e0182578, 2017.

[23] A. Schwarz, P. Ofner, J. Pereira, A. I. Sburlea, and G. R. Müller-Putz, "Decoding natural reach-and-grasp actions from human EEG," Journal of neural engineering, vol. 15, no. 1, p. 016005, 2017.

[24] A. Nakamura et al., "Somatosensory Homunculus as Drawn by MEG," NeuroImage, vol. 7, no. 4, pp. 377-386, 1998.
[25] Z. Yue, X. Zhang, and J. Wang, "Hand rehabilitation robotics on post stroke motor recovery," Behavioural neurology, vol. , 2017.
[26] H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?," Clinical neurophysiology, vol. 117, no. 11, pp. 2341-2356, 2006.
[27] G. Garipelli, R. Chavarriaga, and J. d. R. Millán, "Single trial recognition of anticipatory slow cortical potentials: the role of spatio-spectral filtering," in 2011 5th International IEEE/EMBS Conference on Neural Engineering, 2011, pp. 408-411: IEEE.
[28] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," Electroencephalography and clinical Neurophysiology, vol. 103, no. 3, pp. 386-394, 1997.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 300-305 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Resonator Influence Simulation of Designed Close-Open Standing Wave Thermoacoustic Engine

Sugiyanto[*,1,2], Samsul Kamal[1], Joko Waluyo[1], Adhika Widyaparaga[1]

[1]*Department of Mechanical and Industrial Engineering, Faculty of Engineering, Gadjah Mada University, 55281, Indonesia*

[2]*Department of Mechanical Engineering, Vocational School, Gadjah Mada University, 55281, Indonesia*

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 15 September, 2019
Online: 15 October, 2019

Keywords:
SWTE
Length of Resonator Tube
Acoustic Power

ABSTRACT

Thermo-acoustic technology is very potential to be applied to convert heat into another energy source. The complete thermoacoustic engine construction requires a resonator connected to a core where there is a heat exchanger. The thermoacoustic core has a function as an acoustic power generator while the resonator functions as a storage of acoustic power which will shift the amplitude or phase of oscillating pressure or volume flow rate. Experimental testing was carried out to obtain the performance of a quarter wave length of standing wave thermoacoustic engine (SWTE). The resonator is made of a 2 inch Sch 40 stainless steel pipe with a length of 390 mm. The results of this experiment showed that the measured pressure amplitude was 4800 kPa with a working frequency of 138 HZ and produced an acoustic power of 22.85 watts. The parameters of this experimental test are used for the simulation program. The acoustic power generated from the simulation is 20.13 watts, 18.67 watts and 17.82 watts respectively for the length of the resonator 390 mm, 780 mm and 1170 mm.

1. Introduction

Thermoacoustic is a technology related to heat transfer to pressure oscillation, so the use of the term "Thermo" comes from the heat transfer process while "acoustic" comes from pressure oscillating in the working fluid. This term was first introduced by [1] who was credited with laying a theoretical foundation for the field of thermoacoustic research. Thermoacoustic technology is more interesting to develop because of its superiority compared to traditional energy conversion system technologies [2]. The advantage lies in thermoacoustic devices that have immovable components and simple construction (basically pipelines), making such systems very reliable and requires low construction and maintenance costs.

Thermoacoustic devices classified into thermoacoustic engine (TAE) and thermoacoustic refrigerators (TAR) depending on the practical application of the thermoacoustic effects. TAE turns thermal energy into acoustic energy. The thermal energy comes from a significant temperature difference derived from a heat exchanger that applied to two ends of porous material. TAR converts acoustic energy into temperature differences utilizing acoustic waves imposed along with porous material. This process causes temperature differences that occur in the two heat exchangers [3].

The main components of a TAE or TAR are a resonator tube, a porous material (stack) and two heat exchangers. A standing wave type produces a half wavelength or a quarter wavelength which generate in the resonator. The thermoacoustic phenomenon takes place in porous material when the temperature gradient that is worn along the porous material interacts with the oscillation of sound waves. The heat exchanger is responsible for transferring heat in and out of the thermoacoustic device at the desired temperature, thereby maintaining the temperature gradient applied to the porous material [4].

Swift [5] built and tested this SWTE using a 5-inch diameter stainless steel pipe with electric heating. The working fluid used is Helium at an average pressure of 13.8 bar. The result is an acoustic power that can be generated at 630 W with an efficiency of heat conversion to acoustic power of 9% and a Carnot efficiency of 13%. Thermoacoustic prime mover (TAPM) which has a resonator length of 128 cm, uses a stack of stainless-steel mesh numbered 14, two heat exchangers and water at atmospheric pressure as gas acting inside the resonator. Successfully designed, built, and evaluated by [6].

[*]Corresponding Author: Sugiyanto, Department of Mechanical Engineering, Vocational School, Gadjah Mada University, 55281, Yogyakarta, Indonesia. Tel : +628112502021, Email: sugiyanto_t@ugm.ac.id

There are two types of acoustic waves generated by the thermoacoustic engine, standing waves, and traveling waves. Both of these wave types depend on the type of resonator used. Thermoacoustic engines require high-temperature gradients to produce sound. The temperature gradient itself occurs in part often called the stack (for a traveling wave thermoacoustic engine called a regenerator), placed between the hot and cold sides of the thermoacoustic engine. Thermoacoustic engine research mostly is done on the standing wave type due to its ease of configuration, which only consists of a resonator pipe and the thermoacoustic core which is directly connected to the resonator and placed near the other end. This thermoacoustic core is a heat exchanger (HHX), stack, and cold heat exchanger (CHX).

The influence of the length of the resonator pipe on frequency is carried out by [7] who investigate the performance of prime mover thermoacoustic traveling waves at a small onset temperature difference, which is obtained by mounting the stack in the right position on the horizontal resonator pipe.

The resonator tube is one of the critical components of a thermoacoustic engine. Many studies have been conducted on the effect of resonance tube on thermoacoustic performance, most of them focused their research on resonance tube length which is considered the dominant factor that determines the frequency of the system

The effect of the resonance tube on a standing wave thermoacoustic engine is researched by simulating a linear thermoacoustic. In this simulation, the SWTE is connected to a pulse tube refrigeration to obtain the temperature gradient on the core side. The results of this simulation calculation show that the prolongation of the resonance tube can cause a decrease in the working frequency [8].

Hariharan et al. [9] conducted a study with variations in stack geometry and resonator length to determine the SWTE performance. In this study, variations in the length of the resonator carried out with lengths of 200, 300, 400, and 500 mm. Variation of the length of the resonator influences the pressure amplitude, wave frequency, and acoustic power so that it will also affect its efficiency.

The TAE model whose resonator length can be regulated was examined by [10] which consisted of 103 mm ceramic honeycomb stack, volume buffers, and shell-tube type heat exchangers. The TAE performance was evaluated using six different types of stack. The length of the resonators used were 50 mm, 100 mm, and 150 mm placed in three different places. The length of the resonator that can be set in this study aims to choose the best working frequency so that the TAE can work optimally and produce the highest acoustic power.

The influences of the tube diameter of traveling wave thermoacoustic are not sensitive; hence, the resonance frequency of the radial dimension is always ignored. Therefore, the way to adjust the working frequency is to vary the length of the resonator tube. However, for a small thermoacoustic system, the influence of the radial dimension cannot be neglected [11].

The resonator is one of the key components of a thermoacoustic engine because it's save acoustic power. Resonator supply the instances and compliances that the working frequency is related to it. The unwanted eddying or non-linear pressure variations should be avoided by using a straight tubular resonator pipe without step, misalignment, and sudden transition. However, A straight, uniform-diameter resonator is seldom the best design [12].

This paper is an extension of work originally presented in the 2018 4th International Conference on Science and Technology (ICST), by changing some variations but still focusing on simulations as a way to analyze the performance of this SWTE. The author has conducted research on the effects of resonator diameter on SWTE performance by simulations using DeltaEC. The results show that the simulations with resonator diameter variations from the close-open SWTE have no significant effect [13].

Experiment testing was conducted beforehand to get the results of the parameters used. The results of this experiment will be used to validate the results of the simulation. The parameters of the simulation results that have been confirmed with settings in the SWTE experiment are then used as SWTE simulation parameters in variation length of the resonator. Simulation of an SWTE is easy to do using DeltaEC . This program was made by Dr. Bill Ward and Dr. Gregori Swift from Los Alamos National Laboratories (LANL) [14].

2. Theoritical basis

The quarter wave length resonator is a tube closed at one end and open at the other. This resonator can be explained in a lumped element model consisting of compliance in series with inertance [15].

The working frequency is related to the inertance (L) and compliance C) of the component in the thermoacoustic engine, and they can be described by the equation [12]:

$$L = \frac{\rho_m l}{A} \tag{1}$$

$$C = \frac{V}{\gamma P_m} \tag{2}$$

Where A and l are the sectional area and length of the resonator tube, respectively. ρ_m is average density and P_m is the pressure of the gas. V and γ are the volume compliance and the specific heat, respectively.

The acoustic intensity I for the measured sound wave was calculated using the modified two sensor method [16]:

$$I = \frac{1}{8\omega\rho_m}\left\{\mathrm{Im}[H]\left(|P_A|^2 - |P_B|^2\right) + 2\,\mathrm{Re}[H]\,|P_A||P_B|\sin\theta\right\} \tag{3}$$

H is defined by:

$$H = \frac{kF}{\cos\left(k\frac{\Delta x}{2}\right)\sin\left(k\frac{\Delta x}{2}\right)} \tag{4}$$

Where *Re* [ ] is the real component and *Im* [ ] is the imaginary components. P_A and P_B are acoustic pressure waves measured by pressure sensors mounted close together along the resonator, and $\theta = \arg [P_A/P_B]$ represents phase lead of pressure wave P_A relative to P_B, k is complex wave number and Δx is the difference in distance between the two pressure sensors.

F is a complex factor calculated by:

$$F = 1 - \frac{2J_1\left(i^{3/2}\sqrt{2}r_0/\delta\right)}{\left(i^{3/2}\sqrt{2}r_0/\delta\right)J_0\left(i^{3/2}\sqrt{2}r_0/\delta\right)} \quad (5)$$

And

$$k = -ik_0\sqrt{\frac{J_0\left(i^{3/2}\sqrt{2}r_0/\delta\right)}{J_2\left(i^{3/2}\sqrt{2}r_0/\delta\right)}}\sqrt{\gamma + (\gamma-1)\frac{J_2\left(i^{3/2}\sqrt{2\sigma}r_0/\delta\right)}{J_0\left(i^{3/2}\sqrt{2\sigma}r_0/\delta\right)}} \quad (6)$$

Where γ and σ is notation of the specific heat ratio and Prandtl number, respectively. J_n represent the n^{th} order complex of Bessel function, k_0 is the wave number, and r_0 is the radius of duct.

The results of the acoustic intensity calculation (I) are then used to calculate the acoustic power $(\dot{E})$ in the following equation:

$$\dot{E} = \dot{I}A \quad (7)$$

3. Methodology

This research uses an SWTE that is open at the end of the resonator but closed at the end adjacent to the thermoacoustic core as seen in Figure 1. Resonators are made of 2-inch Sch. 40 stainless steel pipes with a length of 390 mm. Two heat exchanger devices used are made of copper both for hot heat exchangers (HHX) and cold heat exchangers (CHX). Both of these heat exchangers flank the stack made of stainless-steel wire mesh, with a porosity of 88.26%. The heat source for heating the HHX comes from 6 pcs glow plug from an 11 VDC diesel engine, and CHX is cooled using water.

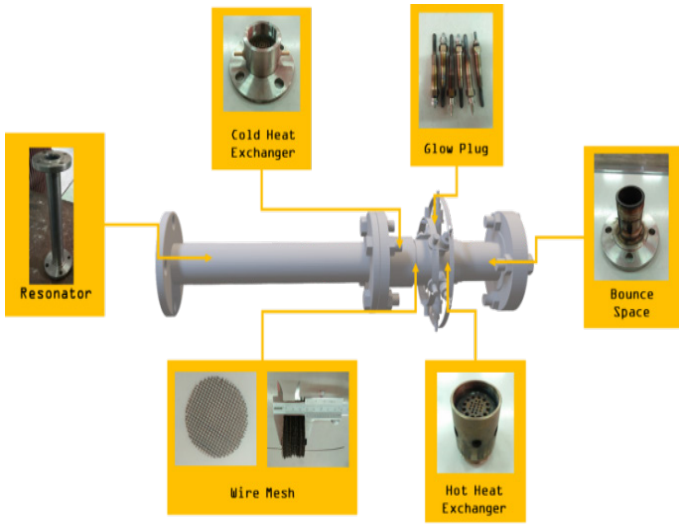


Figure 1: The main components of a close-open thermoacoustic engine

The first step is experimental testing to find out the acoustic intensity of SWTE with a resonator length of 390 mm. In this experimental test, temperatures were measured using two type-K thermocouples placed in HHX (hot side) and CHX (cold side) connect to National Instrument 9234 module. Two M102A05 PCB pressure sensors are mounted on the resonator with a distance between them of 50 mm, used to measure acoustic pressure with readings using the National Instrument module 9234 as well. As a heat source, 6 glow plugs from diesel engines are used which are connected to the transformer battery charger. The set-up of this experimental test is shown in Figure 2.

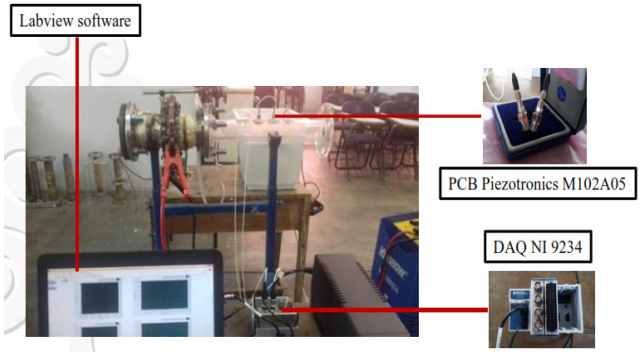


Figure 2: SWTE experiment set-up

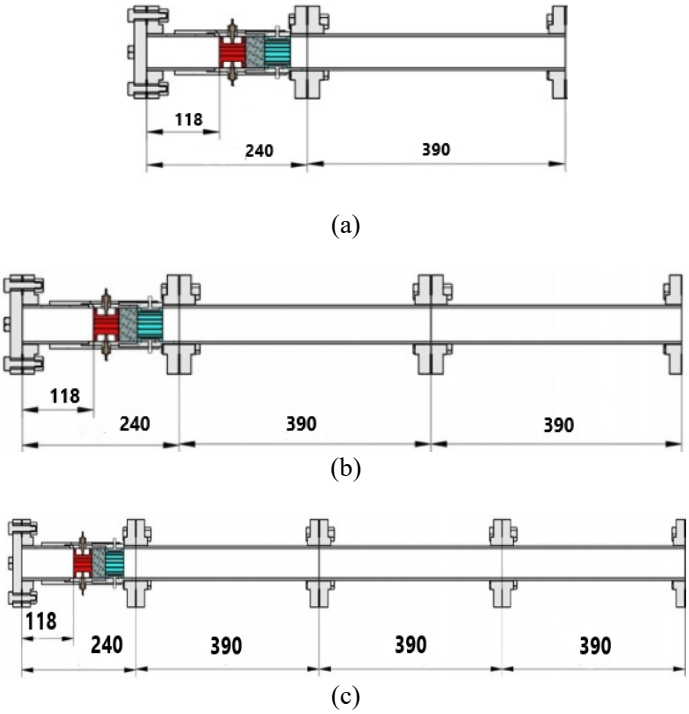


Figure 3: SWTE simulation set-up with a variation of resonator length; a. 390 mm , b. 780 mm , c. 1170 mm length of resonator tube.

The second step is simulation with DeltaEC software to find out the performance of SWTE. The first simulation uses experimental test parameters at a 390 mm resonator length. The experimental data in the first step of the study were used to validate the simulation results in the second step. After the two results between the experiment and the simulation are validated, a simulation program is further developed to predict the performance of the SWTE at variations in the length of 780 mm and 1170 mm as seen in Figure 3.

The display of measurement results using NI 9234 data logger as a simulation parameter that was read by Labview is shown in Figure 4. at point 1 and around 3600 Pa, 138 Hz at point 2. The measurement data from these two pressure sensors are then processed to get the acoustic intensity and acoustic power using equations 3 and 7. The final result for this experiment is the acoustic power of 22,853 watts.

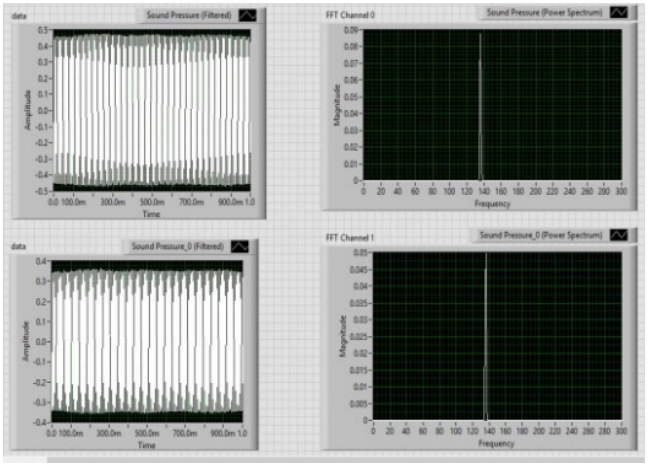


Figure 4: Pressure Amplitude and frequency of SWTE ekperiment result.

4. Result and Analysis

The result of program execution is a schematic SWTE construction diagram with segments of the components used, as presented in Figure 5. It is in a closed on one side state expressed by the SURFACE segment. The other components are Excess pipe, which is stated by DUCT segment, HHX and CHX components are expressed by HX segment, STKSCREEN segments express wire mesh stack, and resonator tube is expressed in DUCT segment.

The DeltaEC program shown in Figure 6 begins by entering the initial parameters as used in the experiment testing, followed by settings for each segment used. In Segment 6 which is the Duct segment for the resonator, a variation of the resonator length of 390 mm is included so that the total power and the acoustic power generated by the SWTE system. The simulation program that has been successfully executed to obtain the performance of SWTE is then used to simulate the performance of SWTE at the length of other resonators. This simulation is done by changing the program simulation parameter at segment DUCT.

The acoustic power that was successfully generated from the simulation for a 390 mm resonator length was 20,13 watts. The difference in acoustic power between experimental results and simulation results is 3%, so this difference is still acceptable because the value of the deviation is even smaller than 10% [12].

The distribution of acoustic power from the system can be seen from the graph in Figure 7 as a result of program execution. The red line shows the total energy on the device while the blue line is the acoustic power. In the segments 1 and 2, which are bounce space areas, there is no power generation. Acoustic power starts to be generated in the core area of SWTE, and the highest is in the field of stack wire mesh which is almost 30 watts. Momentary out of CHX this power goes down and will tend to be constant along the resonator tube which is around 20 watts. The total power of this system is 680 watts, so the efficiency of SWTE with a resonator length of 390 m is 2.94%.

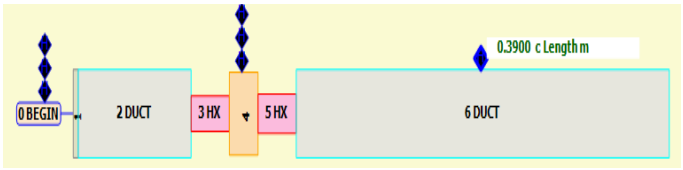


Figure 5: Schematic diagram result of simulation for 390 mm resonator length.

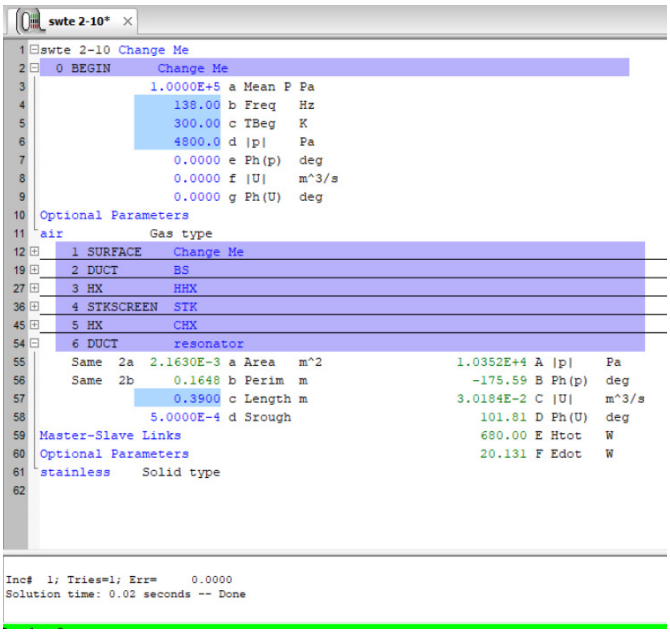


Figure 6: Result simulation of power at 390 mm of resonator length

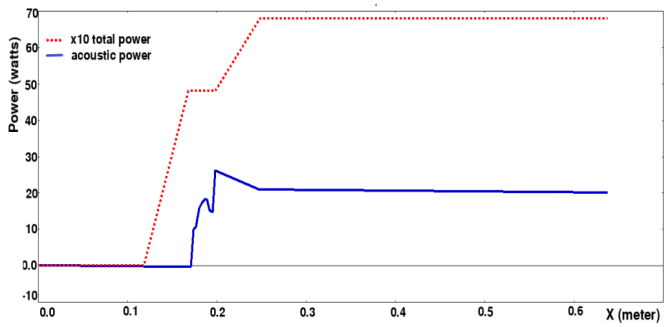


Figure 7: Power distribution of SWTE use 390 mm of resonator length.

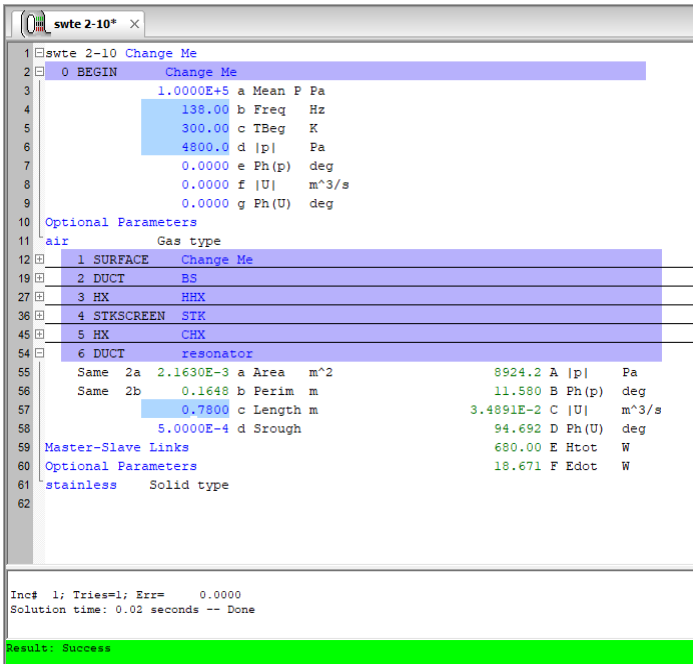


Figure 8: Result simulation of power output at 780 mm of resonator length

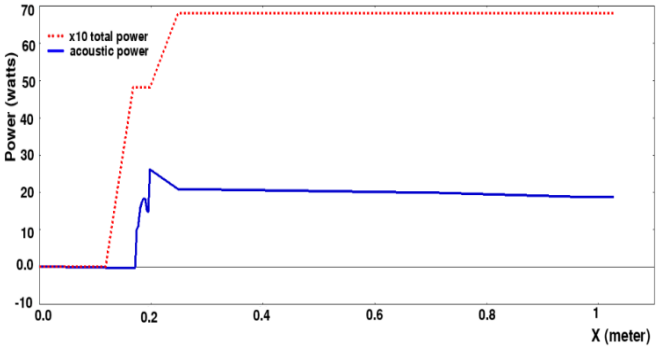


Figure 9: Power distribution of SWTE use 780 mm of resonator length.

Figure 8 to 11 presents acoustic power generated for a resonator length of 780 mm and 1170 mm, respectively. The simulation with a 780 mm resonator length is done by changing the segment 6 in the program by liquefying the other parameters in a fixed condition, as shown in Figure 8. The output power for this condition is 18.67 watts for acoustic power and 680 watts for total power. There was no change in total power due to variations in this second simulation; no other parameters were changed. The efficiency gained from this conditioning system is 2.7%.

As shown in Figure 9, for the length of resonator tube 780 mm at a position of 0 to 0.18 m there is no power generation because a bounce space occupies this position as an air reservoir. At intervals of $0.18 \leqslant x \leqslant 0.2$ m, which is the position of the HHX and stack, the acoustic power starts to rise and reaches the highest value shortly after leaving the stack, which is 26.12 watts. CHX occupies the area at $0.2 \leqslant x \leqslant 0.4$ m intervals, resulting in acoustic power of 20.96 watts. This value is lower than when the stack exits due to water cooling on the CHX side. This value will tend to be constant when passing through the resonator at a distance of $0.4 \leqslant x \leqslant 1.2$ m, which is equal to 18.67 watts.

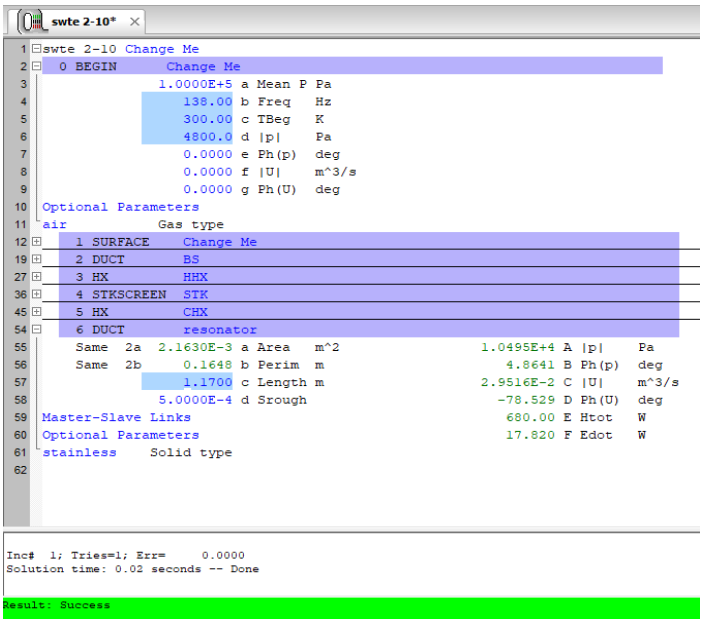


Figure 10: Result simulation of power output at 1170 mm of resonator length

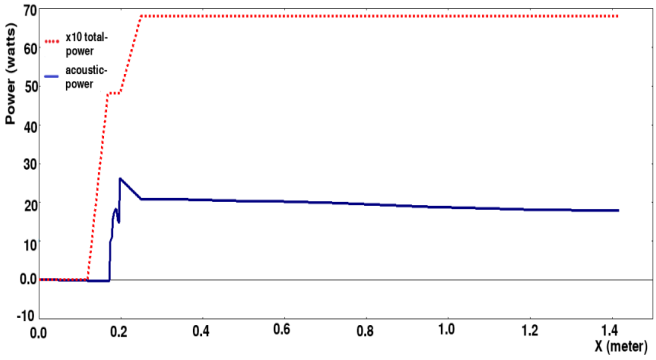


Figure 11: Power distribution of SWTE use 1170 mm of resonator length

The simulation results for the variation of the resonator length of 1170 mm are presented in Figure 10. Like the two previous simulations, segment 6 of the resonator length is changed to obtain information on the amount of output power generated by this system. The results can be seen that the total power produced is still constant at 680 watts because there is no change in the thermoacoustic core parameters. The acoustic power generated is smaller by 17.82 watts so that the efficiency of the system as a result of this third simulation is 2.62%.

The acoustic power distribution as presented in Figure 11 has the same trend with the simulation results for the variation of the resonator lengths of 390 mm and 780 m, where on the bounce space side the end is closed, there is no power generated, the core has the highest value and the resonator will tend to go down until the side open end on this system.

Based on the simulation results for SWTE with variations in the length of the resonator, the results show that the longer the resonator tube, the output power of the system will decrease. The prediction from this simulation is following the research conducted by [9] and [10], namely that the length of the resonator will affect the acoustic power so that it will also affect its efficiency.

The way to increase the standing type of performance can be done by changing to the co-axial type without changing the length of the resonator. Takeyama [17] has tested it by inserting an inner tube into the resonator to the bounce space area. The inner tube used is smaller than the diameter of the resonator tube, as seen in Figure 12. Estimation of co-axial performance can be done by simulations using DeltaEC, such as research by [18] which conducted this simulation to optimize co-axial traveling-wave thermoacoustic cooler powered by the standing-wave thermoacoustic engine.

The simulation results of the co-axial type of thermoacoustic engine which the inner tube diameter of 25 mm and 580 mm length presented in Figure 13. The acoustic power is 30 watts. There was an increase of about 10 watts compared to the SWTE type result in both simulation and experimental. The total energy that can be achieved by co-axial is about 100 watts, so the efficiency of this system is around 3%, slightly higher than the SWTE.

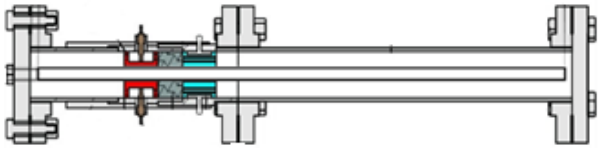


Figure 12: Co-axial thermoacoustic engine contruction.

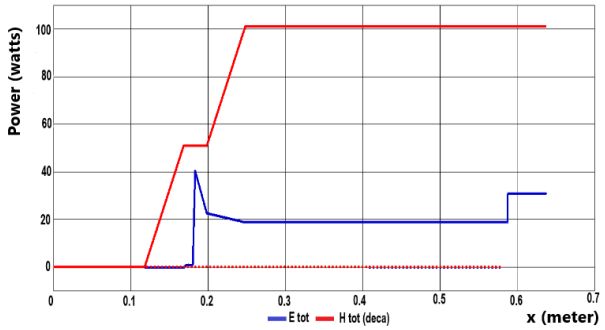


Figure 13: Power distribution of co-axial use 25 mm inner tube diameter

A summary of the comparison of research results in this paper is presented in Table 1

Table 1: Comparison of test results with variations in resonator length

Methods	Length of Resonator (mm)	Acoustic Power (watts)
Experiment	390	22.85
Simulation	390	20.13
Simulation	780	18.67
Simulation	1170	17.82
Simulation (co-axial)	390	30.00

5. Conclusion

Simulations to predict the performance of SWTE open-close have been carried out. Its application for a 390 mm resonator length successfully made, and the results match the experimental data. Execution of other resonator lengths, namely 780 mm and 1170 mm, states that a prolonged resonator reduces the acoustic power produced by this device. Another method for increasing engine performance without changing the length of the resonator is by making the device co-axial type.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The author would like to thank the Gadjah Mada University Higher Education Research Scheme, which has supported part of this research.

References

[1] N. Rott, “Thermoacoustics,” Adv. Appl. Mech., vol. 20, pp. 135–175, 1980. https://doi.org/10.1016/S0065-2156(08)70233-3

[2] S. L. Garrett, “Thermoacoustic Refrigeration,” Naval Postgraduate School, Breckenridge Hilton, Breckenride CO, 1993.

[3] G. . Swift, “Thermoacoustic engines,” J. Acoust. Soc. Am., vol. 84, no. 4, pp. 1145–1180, 1988. https://doi.org/10.1121/1.396617

[4] H. Babaei and K. Siddiqui, “Design and optimization of thermoacoustic devices,” Energy Convers. Manag., vol. 49, no. 12, pp. 3585–3598, 2008. https://doi.org/10.1016/j.enconman.2008.07.002

[5] G. W. Swift, “Analysis and performance of a large thermoacoustic engine,” J. Acoust. Soc. Am., vol. 92, no. 3, pp. 1551–1563, 1992. https://doi.org/10.1121/1.403896

[6] I. Setiawan, P. Murti, W. N. Achmadin, A. B. S. Utomo, and M. Nohtomi, “Design, construction and evaluation of a standing wave thermoacoustic prime mover,” in AIP Conference Proceedings, 2016, vol. 1717. https://doi.org/10.1063/1.4943482

[7] T. Jin, B. Zhang, K. Tang, R. Bao, and G. Chen, “Experimental observation on a small-scale thermoacoustic prime mover,” J. Zhejiang Univ. A, vol. 8, no. 2, pp. 205–209, 2007. https://doi.org/10.1631/jzus.2007.a0205

[8] K. Tang, G. B. Chen, T. Jin, R. Bao, B. Kong, and L. M. Qiu, “Influence of resonance tube length on performance of thermoacoustically driven pulse tube refrigerator,” Cryogenics (Guildf)., vol. 45, no. 3, pp. 185–191, 2005. https://doi.org/10.1016/j.cryogenics.2004.10.002

[9] N. M. Hariharan, P. Sivashanmugam, and S. Kasthurirengan, “Influence of stack geometry and resonator length on the performance of thermoacoustic engine,” Appl. Acoust., vol. 73, no. 10, pp. 1052–1058, 2012. https://doi.org/10.1016/j.apacoust.2012.05.003

[10] A. C. Alcock, L. K. Tartibu, and T. C. Jen, “Experimental investigation of an adjustable standing wave thermoacoustic engine,” Heat Mass Transf. und Stoffuebertragung, vol. 55, no. 3, pp. 877–890, 2018. https://doi.org/10.1007/s00231-018-2469-1

[11] G. Zhou, Q. Li, Z. Li, and Q. Li, “Influence of resonator diameter on a miniature thermoacoustic Stirling heat engine,” Chinese Sci. Bull., vol. 53, no. 1, pp. 145–154, 2008. https://doi.org/10.1007/s11434-008-0019-9

[12] G. . Swift, “Thermoacoustics : A Unifying Perspective for Some Engines and Refrigerators,” Leb. und-Technologie, vol. 35, no. 7, p. 636, 2002. https://doi.org/10.1016/s0023-6438(02)90911-4

[13] Sugiyanto, S. Kamal, J. Waluyo, and A. Widyaparaga, “Simulation of Close-Open Standing Wave Thermoacoustic Engine Toward Variation of Resonator Diameter,” in 4th International Conference on Science and Technology, ICST 2018, 2018, vol. 1, pp. 1–5. https://doi.org/10.1109/ICSTC.2018.8528657

[14] B. Ward, J. Clark, and G. Swift, Design environment for low-amplitude thermoacoustic energy conversion (DeltaEC), vol. 122, no. 5. 2007. https://doi.org/10.1121/1.2942768

[15] T. D. Rossing, Springer Handbook of Acoustic. LLC New York: Springer Science+Business Media, 2007.

[16] T. Biwa, Y. Tashiro, H. Nomura, Y. Ueda, and T. Yazaki, “Experimental verification of a two-sensor acoustic intensity measurement in lossy ducts,” J. Acoust. Soc. Am., vol. 124, no. 3, pp. 1584–1590, 2008. https://doi.org/10.1121/1.2953311

[17] Y. Takeyama, S. ichi Sakamoto, and Y. Watanabe, “Study on the setting position of a prime mover in the coaxial-type thermoacoustic cooling system: Comparison with the straight-tube-type thermoacoustic system,” Jpn. J. Appl. Phys., vol. 57, no. 7, 2018. https://doi.org/10.7567/JJAP.57.07LE14

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 2, 305-313 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Noise Cancellation Algorithm Based on Air- and Bone-Conducted Speech Signals by Considering an Unscented Transformation Method

Hisako Orimoto*, Akira ikuta

Prefectural University of Hiroshima, Faculty of Management and Information System, 734-8338, Japan

ARTICLEINFO

Article history:
Received: 22 February, 2019
Accepted: 10 April, 2019
Online: 12 April, 2019

Keywords:
Noise Cancellation
Speech Signal
Air- and Bone-conducted
UT Method

ABSTRACT

Noise control is essential when applying speech recognition in noisy environments such as factories. In this study, a signal processing for noise cancellation is proposed by using a noise-insensitive bone-conducted speech signal together with an air-conducted speech signal. The speech signal is generally expressed by a nonlinear model. The extended Kalman filter is very famous as a state estimation method for nonlinear systems. However, this filter needs a linearized approximation model for the nonlinear systems. By using the sample point called Sigma point, the unscented Kalman filter (UKF) can be applied to the nonlinear system model without linear approximation. In this study, new type method is proposed based on the UKF. Although UKF considers Gaussian noise, an extended UKF considering non-Gaussian noise is proposed. A noise cancellation method is derived by use of air- and bone-conducted speech signals. The validity of this method is investigated by using both conducted speech signals measured in a noisy real environment.

1 Introduction

Recently, speech recognition systems are used in car navigation systems and smart speakers etc.. However, speech recognition can not be performed effectively in circumstances with heavy noises. Therefore, some countermeasures against surrounding noise are indispensable in such situations.

Kalman filter has been applied to many noisy circumstances as a noise cancellation method for speech signal [1],[2],[3]. This filter assumes a linear model subject to white and Gaussian noise as the system equation and the observation equation [4],[5]. Though, the extended Kalman filter (EKF) [6] can be applied to nonlinear systems, linear approximation models of nonlinear systems are required. Therefore, many improvements are necessary for the noise cancellation method to apply it to actual speech signal processing. From the above viewpoint, in our preciously reported study, a noise cancellation method was proposed by using air- and bone-conducted speech signals in the situation contaminated by non-Gaussian and non-white noises [7].

However, since the calculation of the expansion coefficients in the previous algorithm was very complicated, a simplified method is required.

On the other hand, the unscented Kalman filter (UKF) by use of unscented transformation (UT) method can be applied to nonlinear system[9]. The UT method is a technique for calculating the statistics of a random variable that has been nonlinearly transformed. The set of samples on so-called sigma points(σ-points) are chosen so that they capture the specific properties of the underlying distribution. Therefore, this method can be applied to arbitrary nonlinear systems. In our previous study, a noise cancellation method based on only air conducted speech signal has been proposed by applying the UT method [8].

In this study, a new noise cancellation algorithm based on air- and bone-conducted speech signals is proposed by considering the UT method. The relationship between air-conducted speech signal and backgrand noise is expressed as an additive model based on the additive property of sound pressure. However, propagation mechanism of bone-conducted speech signal is complicated and has to be considered as an unknown system in general. Therefore, a system model including unknown parameters is introduced in this study. More specifically, the sample points obtained by using the UT method are introduced. The noise cancellation algorithm is derived by use of an expansion expression of Bayes' theorem. This method can be considered non-Gaussian properties of noises and nonlinear correlation information between the speech signal and observation. Furthermore, the validity of the proposed method is experimentally confirmed by applying it to real speech signal with noises.

*Corresponding Author: Hisako Orimoto, Address, +81 82 251 9804 & orimoto@pu-hiroshima.ac.jp

https://dx.doi.org/10.25046/aj040239

2 Theory

2.1 Modeling of Air- and Bone-Conducted Speech Signals

We consider the original speech signal x_k, observed air-conducted speech signal y_k and bone-conducted speech signal z_k at discrete time k. The observation y_k is contaminated by a surrounding noise v_k. According to the additive property of sound pressure, the following relationship can be established.

$$y_k = x_k + v_k, \tag{1}$$

where the mean and variance of v_k are known. In order to derive the propagation model of the bone-conducted speech signal, the correlation information between x_k and z_k is required.However, it is difficult to obtain prior information on the unknown speech signal x_k. In this study, a new adaptive algorithm for noise cancellation is proposed by introducing a propagation model with unknown parameters between x_k and z_k as the bone-conducted speech signal model for z_k:

$$z_k = a_k x_k + b_k w_k, \tag{2}$$

where w_k is a random noise ($mean : 0, variance 1$) and a_k and b_k are unknown parameters.

2.2 Estimation Method Combined Bayes' Theorem with UT Method

The conditional joint probability distribution of the specific signal x_k and the unknown parameters a_k and b_k is expressed by using expansion expression of Bayes' theorem[10]. To simplify the derivation process of the estimation algorithm, σ-points and the weighting coefficients are introduced in expansion coefficients.

The conditional probability distribution of x_k, a_k and b_k is expressed as

$$\begin{aligned}
&P(x_k, a_k, b_k | Y_k, Z_k) \\
&= P_0(x_k | Y_{k-1}, Z_{k-1}) P_0(a_k | Y_{k-1}, Z_{k-1}) P_0(b_k | Y_{k-1}, Z_{k-1}) \\
&\quad \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{lmnst} \\
&\quad \varphi_l^{(1)}(x_k) \varphi_m^{(2)}(a_k) \varphi_n^{(3)}(b_k) \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k) \\
&\quad / \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k),
\end{aligned} \tag{3}$$

with

$$A_{lmnst} \equiv \left\langle \varphi_l^{(1)}(x_k) \varphi_m^{(2)}(a_k) \varphi_n^{(3)}(b_k) \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k) | Y_{k-1}, Z_{k-1} \right\rangle, \tag{4}$$

where $Y_k (= \{y_1, y_2, ..., y_k\})$ and $Z_k (= \{z_1, z_2, ..., z_k\})$ are sets of air- and bone-conducted speech signal data up to time k. The above five functions $\varphi_l^{(1)}(x_k)$, $\varphi_m^{(2)}(a_k)$, $\varphi_n^{(3)}(b_k)$, $\varphi_s^{(4)}(y_k)$ and $\varphi_t^{(5)}(z_K)$ are orthonormal polynomials of degrees l, m, n, s and t with weighting functions $P_0(x_k \mid Y_{k-1}, Z_{k-1})$, $P_0(a_k \mid Y_{k-1}, Z_{k-1})$, $P_0(b_k \mid Y_{k-1}, Z_{k-1})$, $P_0(y_k \mid Y_{k-1}, Z_{k-1})$, $P_0(z_k \mid Y_{k-1}, Z_{k-1})$, which can be chosen as the probability functions describing the dominant part of the fluctuation. As the examples of standard probability functions, Gaussian distribution is adopted:

$$\begin{cases}
P_0(x_k \mid Y_{k-1}, Z_{k-1}) = N(x_k; x_k^*, \Gamma_{x_k}), \\
P_0(a_k \mid Y_{k-1}, Z_{k-1}) = N(a_k; a_k^*, \Gamma_{a_k}), \\
P_0(b_k \mid Y_{k-1}, Z_{k-1}) = N(b_k; b_k^*, \Gamma_{b_k}). \\
P_0(y_k \mid Y_{k-1}, Z_{k-1}) = N(y_k; y_k^*, \Omega_k), \\
P_0(z_k \mid Y_{k-1}, Z_{k-1}) = N(z_k; z_k^*, \Phi_k),
\end{cases} \tag{5}$$

with

$$N(x; \mu, \Gamma) = \frac{1}{\sqrt{2\pi\Gamma}} e^{\left\{ -\frac{(x-\mu)^2}{2\Gamma} \right\}},$$

$$\begin{aligned}
&x_k^* \equiv \langle x_k \mid Y_{k-1}, Z_{k-1} \rangle, \\
&\Gamma_{x_k} \equiv \left\langle (x_k - x_k^*)^2 \mid Y_{k-1}, Z_{k-1} \right\rangle, \\
&a_k^* \equiv \langle a_k \mid Y_{k-1}, Z_{k-1} \rangle, \\
&\Gamma_{a_k} \equiv \left\langle (a_k - a_k^*)^2 \mid Y_{k-1}, Z_{k-1} \right\rangle, \\
&b_k^* \equiv \langle b_k \mid Y_{k-1}, Z_{k-1} \rangle, \\
&\Gamma_{b_k} \equiv \left\langle (b_k - b_k^*)^2 \mid Y_{k-1}, Z_{k-1} \right\rangle, \\
&y_k^* \equiv \langle y_k \mid Y_{k-1}, Z_{k-1} \rangle, \\
&\Omega_k \equiv \left\langle (y_k - y_k^*)^2 \mid Y_{k-1}, Z_{k-1} \right\rangle, \\
&z_k^* \equiv \langle z_k \mid Y_{k-1}, Z_{k-1} \rangle, \\
&\Phi_k \equiv \left\langle (z_k - z_k^*)^2 \mid Y_{k-1}, Z_{k-1} \right\rangle.
\end{aligned} \tag{6}$$

The orthonormal polynomials[11] with five weighting probability distributions in Eq. (5) are then specified as

$$\varphi_l^{(1)}(x_k) = \frac{1}{\sqrt{l!}} H_l\left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}} \right), \tag{7}$$

$$\varphi_m^{(2)}(a_k) = \frac{1}{\sqrt{m!}} H_m\left(\frac{a_k - a_k^*}{\sqrt{\Gamma_{a_k}}} \right), \tag{8}$$

$$\varphi_n^{(3)}(b_k) = \frac{1}{\sqrt{n!}} H_n\left(\frac{b_k - b_k^*}{\sqrt{\Gamma_{\beta_k}}} \right), \tag{9}$$

$$\varphi_s^{(4)}(y_k) = \frac{1}{\sqrt{s!}} H_s\left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right), \tag{10}$$

$$\varphi_t^{(5)}(z_k) = \frac{1}{\sqrt{t!}} H_t\left(\frac{z_k - z_k^*}{\sqrt{\Phi_k}} \right). \tag{11}$$

The estimates for mean and variance (i.e., conditional mean and variance) of x_k, a_k, b_k, which are the first and second order statistics, can be expressed as follows:

$$\begin{aligned}
&\hat{x}_k \equiv \langle x_k \mid Y_k, Z_k \rangle \\
&= \frac{\sum_{l=0}^{1} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{l00st} c_{1l} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)},
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\hat{a}_k \equiv \langle a_k \mid Y_k, Z_k \rangle \\
&= \frac{\sum_{m=0}^{1} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{0m0st} d_{1m} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)},
\end{aligned} \tag{13}$$

$$\hat{b}_k \equiv \langle b_k \mid Y_k, Z_k \rangle = \frac{\sum_{n=0}^{1}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{00nst} e_{1n} \varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{000st}\varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}, \quad (14)$$

$$P_{x_k} \equiv \left\langle (x_k - \hat{x}_k)^2 \mid Y_k, Z_k \right\rangle = \frac{\sum_{l=0}^{2}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{l00st} c_{2l} \varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{000st}\varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}, \quad (15)$$

$$P_{a_k} \equiv \left\langle (a_k - \hat{a}_k)^2 \mid Y_k, Z_k \right\rangle = \frac{\sum_{m=0}^{2}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{0m0st} d_{2m} \varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{000st}\varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}, \quad (16)$$

$$P_{b_k} \equiv \left\langle (b_k - \hat{b}_k)^2 \mid Y_k, Z_k \right\rangle = \frac{\sum_{n=0}^{2}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{00nst} e_{2n} \varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{000st}\varphi_s^{(4)}(y_k)\varphi_t^{(5)}(z_k)}. \quad (17)$$

Where coefficients are appropriate constant satisfying the following equality:

$$\begin{aligned} x_k &= \textstyle\sum_{l=0}^{1} c_{1l}\varphi_l^{(1)}(x_k), (x_k - \hat{x}_k)^2 = \sum_{l=0}^{2} c_{2l}\varphi_l^{(1)}(x_k), \\ a_k &= \textstyle\sum_{m=0}^{1} d_{1m}\varphi_m^{(2)}(a_k), (a_k - \hat{a}_k)^2 = \sum_{m=0}^{2} d_{2m}\varphi_m^{(2)}(a_k), \\ b_k &= \textstyle\sum_{n=0}^{1} e_{1n}\varphi_n^{(2)}(b_k), (b_k - \hat{b}_k)^2 = \sum_{n=0}^{2} e_{2n}\varphi_n^{(3)}(b_k). \end{aligned} \quad (18)$$

By using the UT method, the expansion coefficients defined by (4) can be realized for arbitrary nonlinear systems. When the UT method is applied to approximate the means and variances of x_k, a_k, b_k, y_k and z_k, the σ-points $x_k^{*(i)}$, $a_k^{*(i)}$, $b_k^{*(i)}$, $y_k^{*(i)}$ and $z_k^{*(i)}$ are obtained as sample points, as follows:

$$\begin{aligned} x_k^{*(0)} &= x_k^*, \\ x_k^{*(1)} &= x_k^* + \sqrt{(1+\lambda)\Gamma_{x_k}}, x_k^{*(2)} = x_k^* - \sqrt{(1+\lambda)\Gamma_{x_k}}, \\ a_k^{*(0)} &= a_k^*, \\ a_k^{*(1)} &= a_k^* + \sqrt{(1+\lambda)\Gamma_{a_k}}, a_k^{*(2)} = a_k^* - \sqrt{(1+\lambda)\Gamma_{a_k}}, \\ b_k^{*(0)} &= b_k^*, \\ b_k^{*(1)} &= b_k^* + \sqrt{(1+\lambda)\Gamma_{b_k}}, b_k^{*(2)} = b_k^* - \sqrt{(1+\lambda)\Gamma_{b_k}}, \\ y_k^{*(0)} &= y_k^*, \\ y_k^{*(1)} &= y_k^* + \sqrt{(1+\lambda)\Omega_k}, y_k^{*(2)} = y_k^* - \sqrt{(1+\lambda)\Omega_k}, \\ z_k^{*(0)} &= z_k^*, \\ z_k^{*(1)} &= z_k^* + \sqrt{(1+\lambda)\Phi_k}, z_k^{*(2)} = z_k^* - \sqrt{(1+\lambda)\Phi_k}, \end{aligned} \quad (19)$$

The σ-points are decided so as to obtain the approximately same mean and variance as original variables. Where λ is a regulation parameter. The weights to be used are obtained as follows:

$$\begin{aligned} W^{(0)} &= \frac{\lambda}{(1+\lambda)}, \\ W^{(1)} &= \frac{1}{2(1+\lambda)}, \\ W^{(2)} &= \frac{1}{2(1+\lambda)}. \end{aligned} \quad (20)$$

Here, the weighing coefficients $W^{(i)}$ have to satisfy the normalization constraint.

$$\textstyle\sum_{i=0}^{2} W^{(i)} = 1. \quad (21)$$

Each expansion coefficient A_{lmnst} defined by (4) is obtained specifically by substituting σ-points $x_k^{*(i)}$, $a_k^{*(i)}$, $b_k^{*(i)}$, $y_k^{*(i)}$ and $z_k^{*(i)}$ into the conditional expectation of x_k, y_k and z_k.

$$\begin{aligned} A_{10010} &= \tfrac{1}{\sqrt{\Gamma_{x_k}}\sqrt{\Omega_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*)(y_k^{*(i)} - y_k^*), \\ A_{10020} &= \tfrac{1}{\sqrt{2}\sqrt{\Gamma_{x_k}\Omega_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*)\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}, \\ A_{20010} &= \tfrac{1}{\sqrt{2\Gamma_{x_k}}\sqrt{\Omega_k}} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\}(y_k^{*(i)} - y_k^*), \\ A_{20020} &= \tfrac{1}{2\Gamma_{x_k}\Omega_k} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\} \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}, \\ A_{10001} &= \tfrac{1}{\sqrt{\Gamma_{x_k}}\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*)(z_k^{*(i)} - z_k^*), \\ A_{10002} &= \tfrac{1}{\sqrt{2}\sqrt{\Gamma_{x_k}\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*)\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \\ A_{20001} &= \tfrac{1}{\sqrt{2\Gamma_{x_k}}\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\}(z_k^{*(i)} - z_k^*), \\ A_{20002} &= \tfrac{1}{2\Gamma_{x_k}\Phi_k} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\}\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \\ A_{10011} &= \tfrac{1}{\sqrt{\Gamma_{x_k}}\sqrt{\Omega_k}\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*)(y_k^{*(i)} - y_k^*)(z_k^{*(i)} - z_k^*), \\ A_{20011} &= \tfrac{1}{\sqrt{2\Gamma_{x_k}}\sqrt{\Omega_k}\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\} \cdot(y_k^{*(i)} - y_k^*)(z_k^{*(i)} - z_k^*), \\ A_{10021} &= \tfrac{1}{\sqrt{2}\sqrt{\Gamma_{x_k}\Omega_k}\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*) \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}(z_k^{*(i)} - z_k^*), \\ A_{10012} &= \tfrac{1}{\sqrt{2}\sqrt{\Gamma_{x_k}}\sqrt{\Omega_k\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*) \cdot(y_k^{*(i)} - y_k^*)\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \\ A_{20012} &= \tfrac{1}{2\Gamma_{x_k}\sqrt{\Omega_k\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\} \cdot(y_k^{*(i)} - y_k^*)\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \\ A_{10022} &= \tfrac{1}{2\sqrt{\Gamma_{x_k}}\Omega_k\Phi_k} \textstyle\sum_{i=0}^{2} W^{(i)}(x_k^{*(i)} - x_k^*) \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \\ A_{20021} &= \tfrac{1}{2\Gamma_{x_k}\Omega_k\sqrt{\Phi_k}} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\} \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}(z_k^{*(i)} - z_k^*), \\ A_{20022} &= \tfrac{1}{2\sqrt{2}\Gamma_{x_k}\Omega_k\Phi_k} \textstyle\sum_{i=0}^{2} W^{(i)}\left\{(x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k}\right\} \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}, \end{aligned} \quad (22)$$

with

$A_{00000} = 1, A_{10000} = A_{00010} = A_{00001} = A_{20000} = A_{00020} = A_{00002} = 0.$

Furthermore, expansion coefficients of a_k, y_k and z_k are

expressed as follows:

$$
\begin{aligned}
A_{01010} &= \frac{1}{\sqrt{\Gamma_{a_k}}\sqrt{\Omega_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)(y_k^{*(i)} - y_k^*),\\
A_{01020} &= \frac{1}{\sqrt{2}\sqrt{\Gamma_{a_k}\Omega_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\},\\
A_{02010} &= \frac{1}{\sqrt{2}\Gamma_{a_k}\sqrt{\Omega_k}} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}(y_k^{*(i)} - y_k^*),\\
A_{02020} &= \frac{1}{2\Gamma_{a_k}\Omega_k} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\},\\
A_{01001} &= \frac{1}{\sqrt{\Gamma_{a_k}}\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)(z_k^{*(i)} - z_k^*),\\
A_{01002} &= \frac{1}{\sqrt{2}\sqrt{\Gamma_{a_k}\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\},\\
A_{02001} &= \frac{1}{\sqrt{2}\Gamma_{a_k}\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}(z_k^{*(i)} - z_k^*),\\
A_{02002} &= \frac{1}{2\Gamma_{a_k}\Phi_k} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\},\\
A_{01011} &= \frac{1}{\sqrt{\Gamma_{a_k}}\sqrt{\Omega_k}\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)(y_k^{*(i)} - y_k^*)(z_k^{*(i)} - z_k^*),\\
A_{02011} &= \frac{1}{\sqrt{2}\Gamma_{a_k}\sqrt{\Omega_k}\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot(y_k^{*(i)} - y_k^*)(z_k^{*(i)} - z_k^*),\\
A_{01021} &= \frac{1}{\sqrt{2}\sqrt{\Gamma_{a_k}\Omega_k}\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)\\
&\qquad \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}(z_k^{*(i)} - z_k^*),\\
A_{01012} &= \frac{1}{\sqrt{2}\sqrt{\Gamma_{a_k}}\sqrt{\Omega_k\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)(y_k^{*(i)} - y_k^*)\\
&\qquad \cdot\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\},\\
A_{02012} &= \frac{1}{2\Gamma_{a_k}\sqrt{\Omega_k\Phi_k}} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot(y_k^{*(i)} - y_k^*)\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\},\\
A_{01022} &= \frac{1}{2\sqrt{\Gamma_{a_k}\Omega_k\Phi_k}} \sum_{i=0}^{2} w^{(i)}(a_k^{*(i)} - a_k^*)\\
&\qquad \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\},\\
A_{02021} &= \frac{1}{2\Gamma_{a_k}\Omega_k\sqrt{\Phi_k}} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}(z_k^{*(i)} - z_k^*),\\
A_{02022} &= \frac{1}{2\sqrt{2}\Gamma_{a_k}\Omega_k\Phi_k} \sum_{i=0}^{2} w^{(i)}\left\{(a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k}\right\}\\
&\qquad \cdot\left\{(y_k^{*(i)} - y_k^*)^2 - \Omega_k\right\}\left\{(z_k^{*(i)} - z_k^*)^2 - \Phi_k\right\}.
\end{aligned}
\tag{23}
$$

The expansion coefficients of b_k, y_k and z_k ($A_{00110}, A_{00120}, A_{00210}$, A_{00220}, $A_{00101}, A_{00102}, A_{00201}$, A_{00202}, $A_{00111}, A_{00211}, A_{00121}$, A_{00112}, $A_{00212}, A_{00122}, A_{00221}$, A_{00222}) are calculated through the same manners.

After substituting (1) (2) into the definition of four parameters y_k^*, Ω_k, z_k^* and Φ_k in (6), the following expressions can be derived.

$$y_k^* = x_k^* + \langle v_k \rangle, \tag{24}$$

$$\Omega_k = \Gamma_{x_k} + \left\langle (v_k - \langle v_k \rangle)^2 \right\rangle, \tag{25}$$

$$z_k^* = a_k^* x_k^*, \tag{26}$$

$$\Phi_k = (\Gamma_{a_k} + a_k^{*2})(\Gamma_{x_k} + x_k^{*2}) + (\Gamma_{b_k} + b_k^{*2})\left\langle w_k^2 \right\rangle - a_k^{*2} x_k^{*2}. \tag{27}$$

In order to derive the predicted values of the speech signal x_k and the unknown parameters a_k, b_k, the time transition of the speech signal x_k is expressed as follows.

$$x_{k+1} = F x_k + G u_k, \tag{28}$$

where, u_k is a random input with mean 0 and variance 1. Parameters F and G are calculated from time correlation information of x_k and x_{k+1}:

$$F = \frac{\langle x_{k+1} x_k \rangle}{\langle x_k \rangle^2}, \quad G = \sqrt{(1 - F^2)\left\langle x_k^2 \right\rangle}. \tag{29}$$

Therefore, x_{k+1}^* and $\Gamma_{x_{k+1}}$ can be expressed as follows:

$$x_{k+1}^* = F\hat{x}_k + G\langle u_k \rangle, \tag{30}$$

$$\Gamma_{x_{k+1}} = F^2 P_{x_k} + G^2\left\langle (u_k - \langle u_k \rangle)^2 \right\rangle, \tag{31}$$

Since the parameters a_k and b_k are constants, time transition models are introduced for the recursive estimation.

$$a_{k+1} = a_k, b_{k+1} = b_k. \tag{32}$$

By using these relationships, the predictions are given as follows,

$$a_{k+1}^* = \hat{a}_k, \ \Gamma_{a_{k+1}} = P_{a_k}, \tag{33}$$

$$b_{k+1}^* = \hat{b}_k, \ \Gamma_{b_{k+1}} = P_{b_k}. \tag{34}$$

The state estimation algorithm with expansion coefficient A_{lmnst} reflecting linear and nonlinear correlation information among variables and statistics of non-Gaussian noise is completed.

3 Experiment

In order to confirm the validity of the proposed noise cancellation algorithm, we compared it with the method using only the air-conducted speech signal. The compared method was derived by considering the following conditional probability distribution.

$$
\begin{aligned}
&P(x_k, a_k, b_k \,|\, Y_k)\\
&= P_0(x_k \,|\, Y_{k-1}) P_0(a_k \,|\, Y_{k-1}) P_0(b_k \,|\, Y_{k-1})\\
&\quad \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{s=0}^{\infty} A_{lmns}\\
&\quad \varphi_l^{(1)}(x_k)\varphi_m^{(2)}(a_k)\varphi_n^{(3)}(b_k)\varphi_s^{(4)}(y_k)\\
&\quad / \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{000s}\varphi_s^{(4)}(y_k).
\end{aligned}
\tag{35}
$$

Based on (35), the estimates of mean and variance of x_k are derived as follow:

$$
\begin{aligned}
\hat{x}_k &\equiv \langle x_k \mid Y_k \rangle\\
&= \frac{\sum_{l=0}^{1}\sum_{s=0}^{\infty} A_{l00s} c_{1l} \varphi_s^{(4)}(y_k)}{\sum_{s=0}^{\infty} A_{000s}\varphi_s^{(4)}(y_k)},
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
P_{x_k} &\equiv \left\langle (x_k - \hat{x}_k)^2 \mid Y_k \right\rangle\\
&= \frac{\sum_{l=0}^{2}\sum_{s=0}^{\infty} A_{l00s} c_{2l} \varphi_s^{(4)}(y_k)}{\sum_{s=0}^{\infty} A_{000s}\varphi_s^{(4)}(y_k)}.
\end{aligned}
\tag{37}
$$

Male and female speech signals were used in the experiment. The speech signal data were measured in the anechoic

chamber in the acoustic laboratory. The observed speech signal are contaminated with the white noise, the pink noise and the machine noise respectively. The spectra of these noises are shown in Figures 1-3. Furthermore, the observation data of air-conduced speech signal were created by mixing noises with speech signal on a computer.

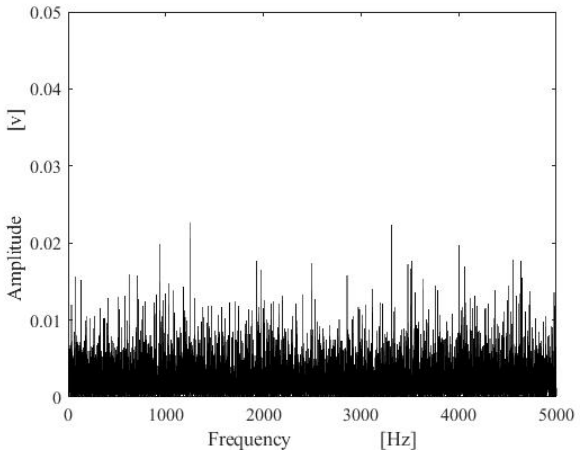


Figure 1: Spectrum of white noise.

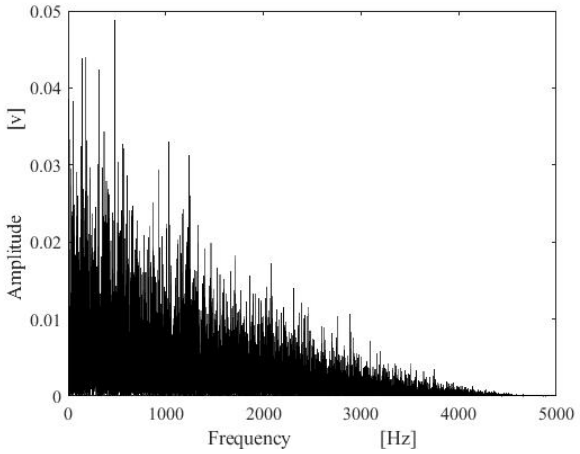


Figure 2: Spectrum of pink noise.

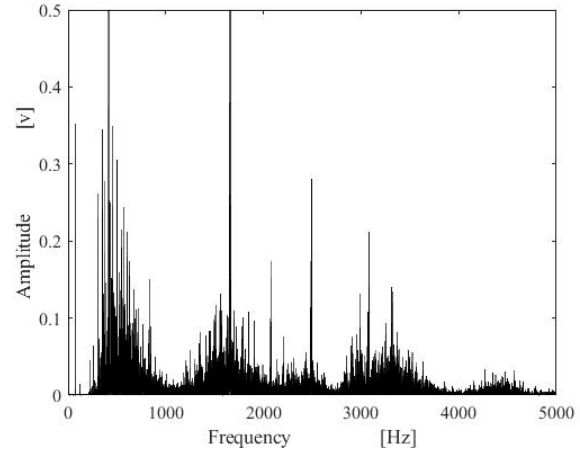


Figure 3: Spectrum of machine noise.

Table 1 shows the specifications of the personal computer for signal processing in the experiment. The signal processing time for speech signal of about 3.5 seconds in length was from 0.5 to 0.8 seconds.

Table 1: The specifications of the personal computer.

	Specification
PC	Dell Vostro 3650
CPU	Intel Core i7-6700 @ 3.40GHz
MEMORY	8.00G
OS	Win 10 Pro 64bit

As an evaluation method of estimation result, the Root Mean Square Error (RMSE) and Performance Evaluation Index (PEI) are adopted.

$$RMSE = \sqrt{\frac{\Sigma(x_k - \hat{x}_k)^2}{N}}, \tag{38}$$

$$PEI = 10\log_{10}\frac{\Sigma x_k^2}{\Sigma(x_k - \hat{x}_k)^2} \quad [dB]. \tag{39}$$

As the RMS Error is smaller value, the better estimation result is obtained. On the other hand, the larger the PEI is, the better the estimation is. Table 2 and Table 3 show the results for the male speech signal and a female speech signal respectively. In the cases of lower noises, almost the same estimation results are obtained in the proposed method and the compared method. On the other hand, in the case of higher noises, the proposed method obtains better results than the compared method. Furthermore, in comparison with our previous method [7] with complicated algorithm, almost the same accurate estimation results are obtained as the proposed method. Therefore, the superiority of the proposed method adopting the simplified algorithm could be confirmed.

Table 2: Comparisons of RMSE and PEI for a male speech signal.

(a) white noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0160	6.8438	0.0163	6.7057
1/2	0.0225	3.8693	0.0269	2.3286
1/3	0.0267	2.4111	0.0450	-2.1282
1/4	0.0330	0.5615	0.0838	-7.5301
1/5	0.0480	-2.6982	0.1338	-11.6004

(b) pink noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0192	5.2693	0.0207	4.6124
1/2	0.0281	1.9414	0.0345	0.1846
1/3	0.0322	0.7773	0.0626	-4.9946
1/4	0.0419	-1.5062	0.0823	-7.3776
1/5	0.0669	-5.5803	0.1334	-11.5703

(c) machine noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0238	3.4024	0.0272	2.2381
1/2	0.0313	1.0116	0.0489	-2.8538
1/3	0.0399	-1.0931	0.0767	-6.7613
1/4	0.0586	-4.4310	0.1080	-9.7397
1/5	0.0686	-5.8003	0.1409	-12.0497

Some of the waveform summarized in Tables 2 and 3 are shown in Figures 4-23. Figures 4 and 14 show the original speech signals of male and female, respectively. Figures 5, 8, 11, 15, 18 and 21 show the speech signals contaminated by noises with amplitude of 3 times larger than the original signals. The estimated results by using of proposed method are shown in Figures 6, 9, 12, 16, 19 and 22. On the other hand, the comparison results are shown in Figures 7, 10, 13, 17, 20 and 23. In the cases of contaminated by white noise and pink noise, the better results are obtained by the proposed method than the compared method.

Table 3: Comparisons of RMSE and PEI for a female speech signal.

(a) white noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0134	4.8382	0.0104	7.0371
1/2	0.0159	3.3213	0.0160	3.2779
1/3	0.0180	2.2390	0.0195	1.5406
1/4	0.0198	1.4447	0.0273	-1.3608
1/5	0.0260	-0.9548	0.0457	-5.8342

(b) pink noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0136	4.656	0.0128	5.2366
1/2	0.0176	2.4293	0.0211	0.8777
1/3	0.0231	0.0790	0.0271	-1.297
1/4	0.0241	-0.2875	0.0374	-4.1092
1/5	0.0264	-1.0874	0.0510	-6.8009

(c) machine noise

	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0147	3.9846	0.0168	2.8276
1/2	0.0207	1.0503	0.0237	-0.1545
1/3	0.0293	-1.9655	0.0499	-6.6083
1/4	0.0456	-5.8243	0.0553	-7.5024
1/5	0.0425	-5.2075	0.0660	-9.0326

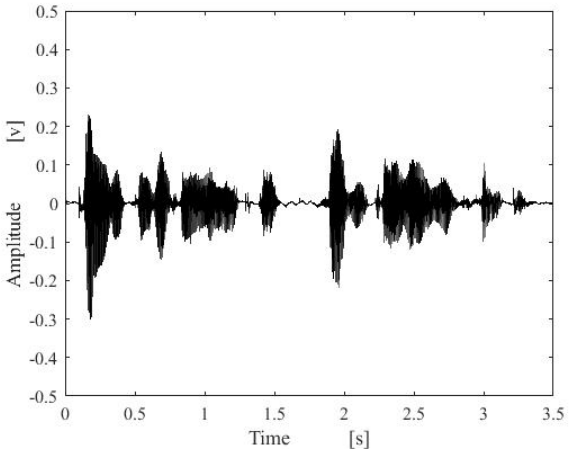


Figure 4: Original male speech signal

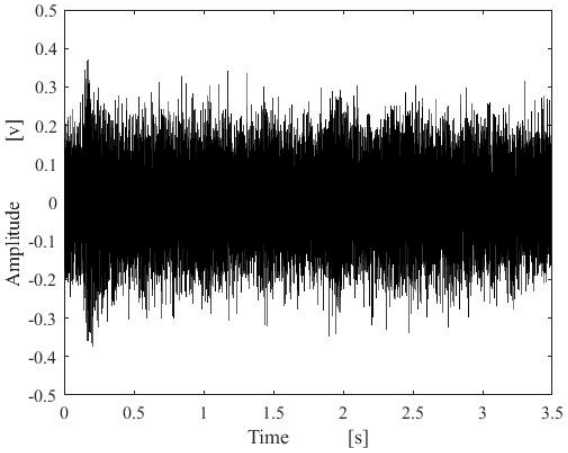


Figure 5: Male speech signal containing a white noise

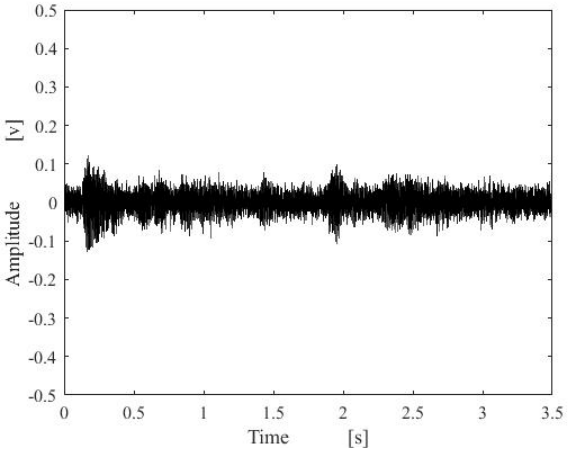


Figure 6: Estimated results by using proposed method.

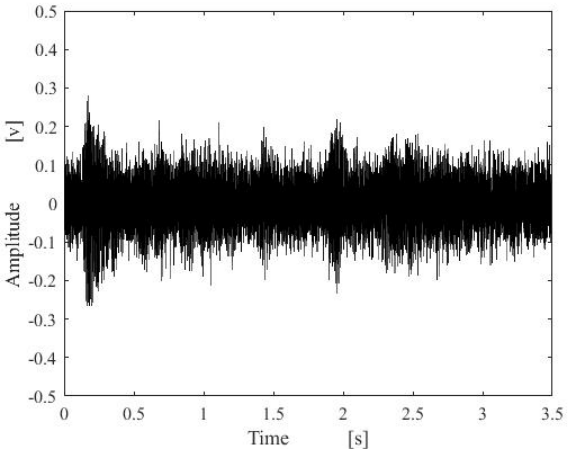


Figure 7: Estimated results by using compared method.

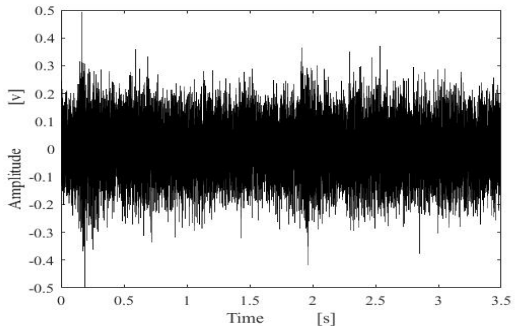


Figure 8: Male speech signal containing a pink noise

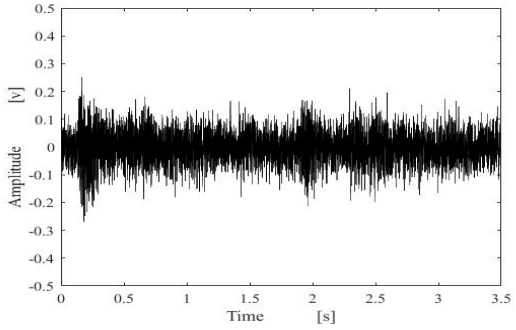


Figure 9: Estimated results by using proposed method.

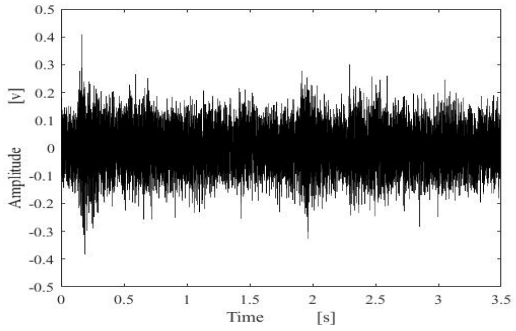


Figure 10: Estimated results by using compared method.

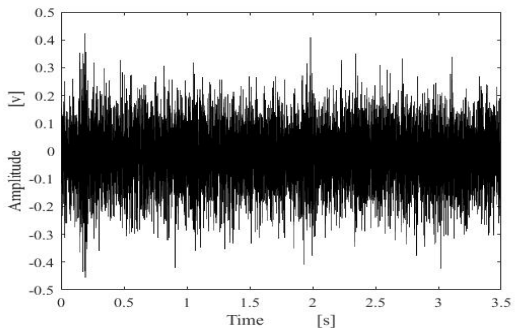


Figure 11: Male speech signal containing a machine noise

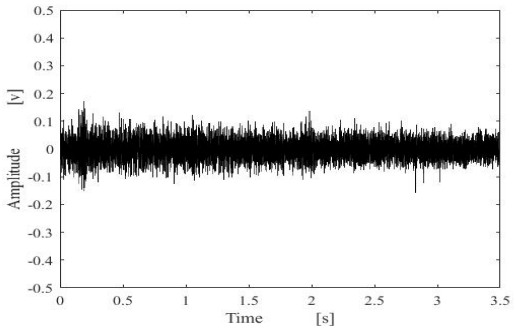


Figure 12: Estimated results by using proposed method.

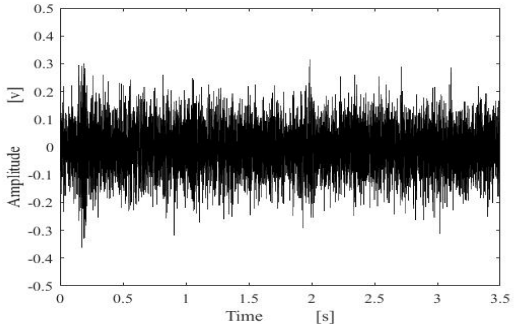


Figure 13: Estimated results by using compared method.

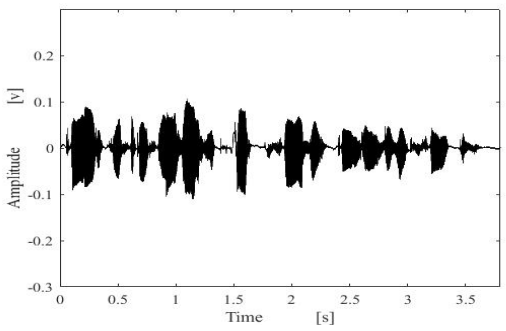


Figure 14: Original female speech signal

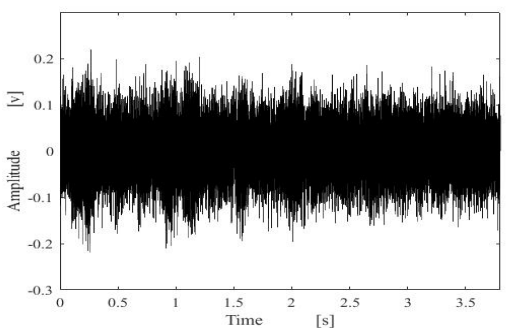


Figure 15: Female speech signal containing a white noise

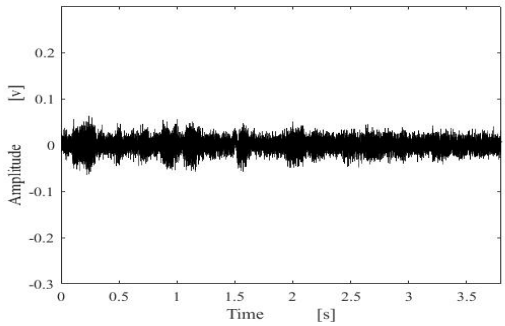


Figure 16: Estimated results by using proposed method.

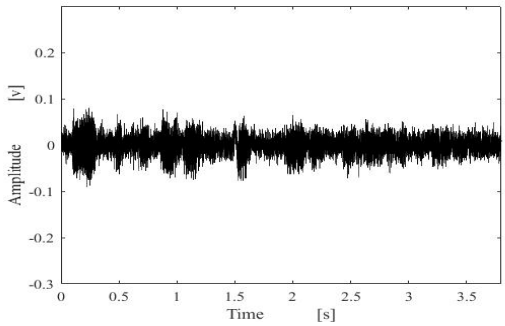


Figure 17: Estimated results by using compared method.

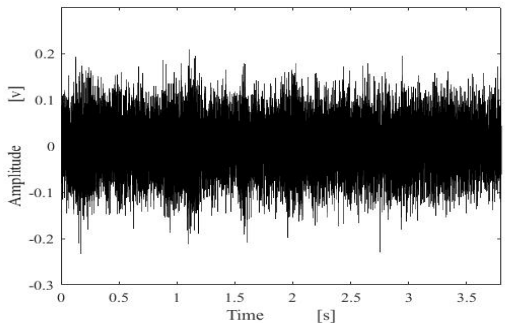


Figure 18: Female speech signal containing a pink noise

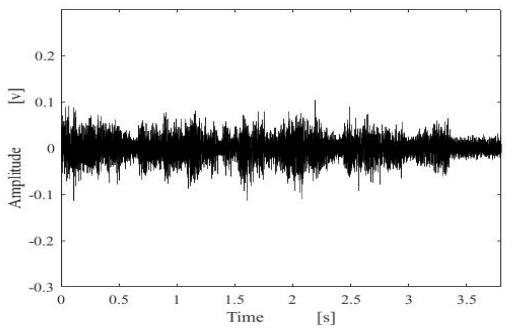


Figure 19: Estimated results by using proposed method.

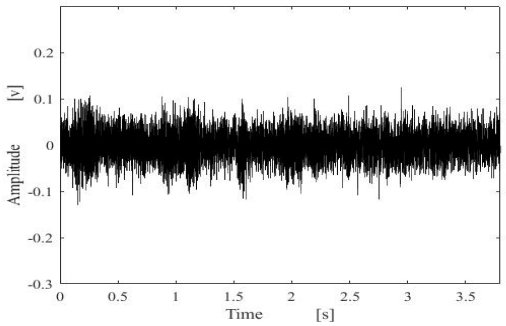


Figure 20: Estimated results by using compared method.

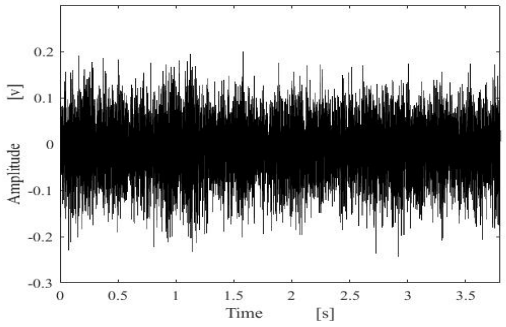


Figure 21: Female speech signal containing a machine noise

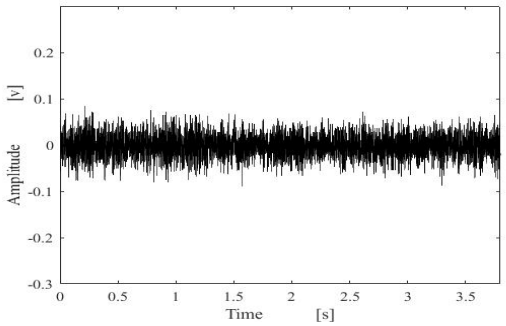


Figure 22: Estimated results by using proposed method.

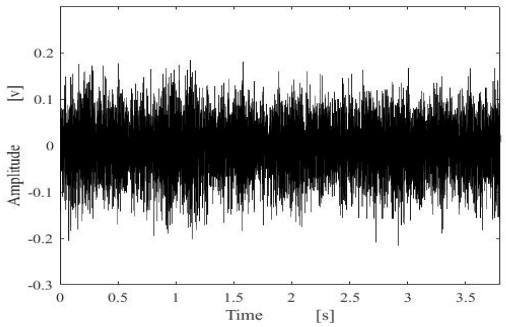


Figure 23: Estimated results by using compared method.

4 Conclusion

In this paper, a new method to suppress noise for speech signal has been proposed, which is applicable to actual en-

vironment with non-Gaussian and non-white noises. The aim of the proposed method is to improve the accuracy of estimation by using air- and bone-conducted speech signals.

The proposed method considered σ-points of not only x_k but also unknown parameter a_k, b_k and observation values y_k, z_k. Moreover, this study has proposed a method including the higher order correlation information between σ-points. Our algorithm has been realized by utilizing the Bayes' theorem as the fundamental principle of estimation and UT method using σ-points. Application of our algorithm has been made to real speech signal contaminated by noises. It has been revealed by experiments that better estimation results could be obtained by the proposed algorithm as compared with the method without using bone-conducted speech signal. However, we have not tried to apply the proposed algorithm to real speech recognition by use of a voice recognition software. Therefore, by applying the algorithm to speech recognition system, the effectiveness of the theory has to be confirmed experimentally.

The proposed approach is quite different from those traditional standard techniques. However, we are still in an early stage of development, and a number of practical problems are yet to be investigated in the future. These include: (i) Introduction of a realistic nonlinear model expressing the actual propagation characteristics of bone-conducted speech signal instead of the simple model in (2). (ii) Consideration of higher order expansion coefficients A_{lmnst} $(l, m, n, s, t \geqq 3)$, in the estimation algorithm. (iii) Selection of an optimal point to put the sensor to measure the bone-conducted speech signal.

References

[1] M. Gabrea, E. Grivel, and M. Najim, "A single microphone Kalman filter-based noise canceller", IEEE Signal Process. Lett., 6 (3), 55-57, 1999.

[2] W. Kim, and H. Ko, gNoise variance estimation for Kalman filtering of noisy speech", IEICE Trans. Inf. and Syst., E84-D (1), 155-160, 2001.

[3] N. Tanabe, T. Furukawa, and S. Tsuji, "Robust noise suppression algorithm with the Kalman filter theory for white and colored disturbance", IEICE Trans. Fundamentals, E91-A (3), 818-829, 2008.

[4] R. E. Kalman, "A new approach to linear filtering and prediction problems", Trans. ASME, Series, D, J. Basic Engineering, 82 (1), 35-45, 1960.

[5] R. E. Kalman and R. S. Buch, "New results in linear filtering and prediction theory", Trans. ASME Series D, J. Basic Engineering, 83 (1), 95-108, 1961.

[6] H. J. Kushner, "Approximations to optimal nonlinear filter", IEEE Trans. on Automatic Control, 12 (5), 546-556, 1967.

[7] A.Ikuta, H.Orimoto and G. Gallagher, "Noise suppression method by jointly using bone- and air-conducted speech signals", Noise Control Engr. J. 66 (6), 472-488, 2018.

[8] H.Orimoto and A.Ikuta, "Signal processing for Noise cancellation method of speech signal by using an extension type UKF", Proceedings of SIGNAL PROCESSING algorithms, architectures, arrangements, and applications (SPA), 304-309, 2018.

[9] James V. Candy, Bayesian Signal Processing Classical, Modern, and Particle Filtering Methods, Wiley-IEEE Press, 2009.

[10] M. Ohta and H. Yamada, "New Methodological Trials of Dynamical State Estimation for the Noise and Vibration Environmental System", Acustica, 55 (4), 199-212, 1984.

[11] M. Ohta and T. Koizumi, "General statistical treatment of the response of a non-linear rectifying device to a stationary random input", IEEE Trans. Inf. Theory, 14 (4), 595-598, (1968).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 314-322 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

The Role of Information Technology Human Capability in the Implementation of Information Technology Governance (ITG): A Systematic Literature Review on Malaysian Organizations

Nor Aziati Abdul Hamid[*,1], Chin Wei Liew[1], Nor Hazana Abdullah[2], Siti Sarah Omar[3]

[1]*Faculty of Technology Management and Business, University Tun Hussein Onn Malaysia, 86400, Malaysia*

[2]*Department of Technology Management, FPTP, University Tun Hussein Onn Malaysia, 86400, Malaysia*

[3]*Department of Business, FPTP, University Tun Hussein Onn Malaysia, 86400, Malaysia*

ARTICLE INFO

Article history:
Received: 13 May, 2019
Accepted: 08 July, 2019
Online: 07 August, 2019

Keywords:
Information technology governance (ITG)
Information technology human capability
Systematic literature review
Governance issues
Malaysia organizations

ABSTRACT

This study reviewed articles on information technology governance (ITG) in the Malaysian context and categorized the issues found in ITG implementation according to organizational-related and human-related issues. It further analysed the impact and role of information technology (IT) human capability in ITG implementation in developing countries like Malaysia, which is still in the infancy stage of ITG. This study found that most organizations in Malaysia are not only facing organizational-related issues but also human-related issues as well. However, there are insufficient studies on how implementation issues between organization and humans are interrelated. The roles of IT human capability in ITG implementation has been heavily underrated as most studies focused on ITG mechanisms and the impact of ITG on firm performance. Issues highlighted in past studies were summarized and used to explain the roles of IT human capability in ITG implementation. The findings revealed that the number of issues reported in the human-related category was slightly more than organizational-related issues. This study explains the role of IT human capability in ITG and reduces the ambiguity of IT human capability in ITG for future research activities.

1. Introduction

Since the beginning of the 21st century, organizations worldwide have been focusing on integrating Information Technology (IT) into organizational business processes. IT has emerged as one of the most important tools for organizations to retain and improve their competitive advantage. Rapid development in IT has enabled organizations to streamline their operations by improving organizational agility and information flow throughout the entire organization. Strategic alignment between business & IT strategies has become one of the main focuses of organizations in order to ensure that IT tools can be fully utilized to support business decisions. However, organizations have encountered more IT investment failures than expected. According to Aziz *et al*., [1], the number of IT failures has been growing since the last decade. Despite all the attention given, the implementation of IT governance (ITG) has not been easy, especially for developing country.

*Corresponding Author: Nor Aziati Abdul Hamid, aziati@uthm.edu.my

2. Problem Statement

The ITG concept has become increasingly well-known to worldwide organizations due to the importance of good IT governance that is constantly developing at a rapid pace in support of business decisions. The ITG concept has garnered increased attention from scholars and practitioners around the globe but the concept is still relatively new to Malaysia. There are numerous ITG frameworks that are readily available for adoption, such as Control Objectives for Information and Related Technologies (COBIT 5.0) and Information Technology Infrastructure Library (ITIL v3). The ITG concept has been constantly evolving as the COBIT and ITIL frameworks have already seen multiple updated versions up till now. Organizations have been trying to adopt various practices and frameworks for governance purposes. However, the implementation of ITG is a difficult process especially for organizations in developing country like Malaysia as there are various barriers to the adoption of ITG practices [2]. In addition, the number of studies on factors that inhibit the positive implementation of ITG has been less than ideal, and even

worse for developing countries. ITG awareness and practices are severely lacking in the Asian region [3]. For a developing country like Malaysia, where ITG implementation is still at the infancy stage, more studies should be conducted to increase the understanding of ITG. Lacks of ITG awareness, guidelines and practices have also been reported concerning the public sector in Tanzania [4]. Therefore, this study attempted to examine and categorize ITG issues that are consistently faced by Malaysian organizations. This study also explained why most issues are interrelated and emphasized on how IT human capability can mitigate ITG implementation issues.

3. Research Context

3.1. Information Technology Governance (ITG)

Due to the massive amount of attention given to the field of ITG, researchers and experts have given various definitions and expectations of ITG. One of the most accepted definition of ITG is "the distribution of decision rights and accountability in order to induce positive behaviour or patterns in managing and using IT resources" [5]. ITG is also defined as the "utilization of organizational capacity by the board, executives and IT management team to formulate and implement IT strategy as well as achieve strategic alignment between business and IT" [6]. ITGI defined ITG as an integral part of corporate governance aided by adequate leadership, structures and processes to ensure IT sustainability and act as an extension of an organization's strategies and objectives [7].

3.2. Information Technology Capability

Although IT capability has seen various definitions in previous studies, the definitions have been centred on "the ability to manipulate IT resources and costs" [8]. The definition was extended by stating that IT resources have been combined to support business-related decisions [9]. Both definitions referred to the ability to utilize and manipulate IT-related resources and other complementary resources to support businesses in value-added ways [10]. Aside from controlling IT resources, IT capability is specifically referred to as the ability to affect organizational goals through the control of IT-related costs [11].

IT capability is placed in between technical capabilities and IT-related organizational capabilities. However, most of the classifications included three of the following, namely IT infrastructure capabilities, IT human capabilities, and IT-enabled intangibles [8, 10-13]. IT infrastructure capabilities is defined as the ability to provide sustainable infrastructure services to support business-related processes or IT-related tangible assets that include IT systems, gadgets, and various tools and applications that form the very base of IT infrastructure and shared throughout the firm [8,12,14]. IT human capabilities refer to the combination of technical and managerial skills of IT professionals in solving and attending various organizational needs [8,12]. IT-enabled intangibles are referred to as intangibles made available due to the utilization of IT resources. These intangibles include knowledge assets, synergy, relationship assets, responsiveness and others [8,10]

3.3. ITG Issues in Malaysia

Since the sources of information for ITG implementation issues were found to be lacking in the review process, data collected for this study included those from studies concerning various sectors, such as public, financial and higher education sectors. Most organizations in Malaysia are still in the infancy stage of ITG implementation, as reported by [2]; hence, the factor of sector diversity is expected to have minimal effect on the barriers to initially accept the basic implementation of ITG. This study has classified issues faced by organizations that implement ITG into two main categories. The first category is organizational-related, and the second category is human-related. Further details on the categorization and issues found will be discussed in the section on findings.

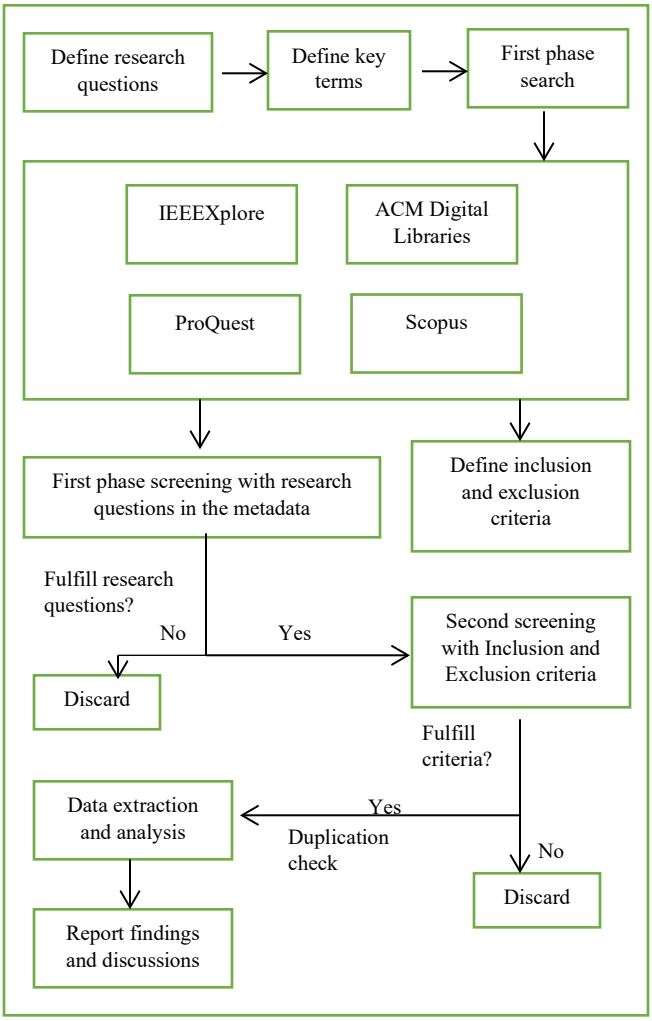


Figure 1: Systematic Literature Review Process

4. Systematic Literature Review

The methodology used in this study was the systematic literature review method. The guidelines for a systematic literature review were provided by Kitchenham [15]. This type of review is mostly used when the study intends to identify any research gap or interpret any area of interest in a span of time. This methodology helps the researcher to be better positioned for future research activities according to the research gaps found in the review process. The steps undertaken for systematic literature review in this study are as follows: i) Define the research questions; ii) Define key terms for searching the database; iii) First phase literature search and selection process; iv) Define

inclusion and exclusion criteria for the second phase selection process; v) Data extraction and analysis; and vi) Report findings and discussions. Figure 1 illustrates the systematic literature review process undertaken in this study.

The first step in the systematic literature review is to define the research questions for the literature screening process involving the metadata in the first phase. The research questions set for literature screening are as follows:

i) Is the article related to information technology governance?

ii) Are there any primary findings on the implementation of information technology governance?

iii) Did the study take place in developing countries or Malaysia?

The literature searches were conducted on 6 databases, which were IEEEXplore, ACM Digital Libraries, Scopus, Science Direct, Springer Link, and ProQuest. The initial search began in November 2018 and was updated in June 2019. Some of the terms used for literature searches only in the metadata are "information technology governance", "information and communication technology governance", "IT governance", and "ICT governance". Thus, if the initial search presented 500 or more hits, the term "Malaysia" is added to the search results, if available, or in the full text to limit the search results to the context of this study. Moreover, if the results yield less than 50 hits, the newly added term "Malaysia" is deleted, and the previous results are considered. As for the literature search in Science Direct, there were 1,297 hits after the term was added. The results were then filtered according to articles with open accessibility. The results of the literature searches are listed below:

i) IEEEXplore – 454

ii) ACM Digital Libraries – 158

iii) SCOPUS – 97

iv) ProQuest – 25

Next, initial screenings were carried out by referring to the research questions. After the initial screenings, the breakdown in the number of qualified articles were 16 articles from IEEEXplore, 7 from ACM Digital Libraries, 47 from SCOPUS, and 1 from the ProQuest database. The inclusion and exclusion criteria for the second phase of screening were then determined, as shown in Table 1 below.

Table 1: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Paper written in English	Paper not written in English
Full accessibility	Partial/No accessibility
Findings related to organizations in Malaysia	Findings unrelated to organizations in Malaysia
Findings reveal IT governance implementation issues and causes	Findings did not reveal IT governance implementation issues
Primary study	Non primary study

In the second phase, the criteria were used in the literature selection process after reading the full text. Finally, 11 articles were selected, and after removing redundant articles, 10 articles had fulfilled the criteria, and 2 additional articles outside the database were found from previous studies and referred by related authors in the qualified articles. A total of 12 articles had qualified for this study. As the study focused only in the context of Malaysian organizations, the number of qualified articles was low. Figure 2 illustrates the number of articles published per year.

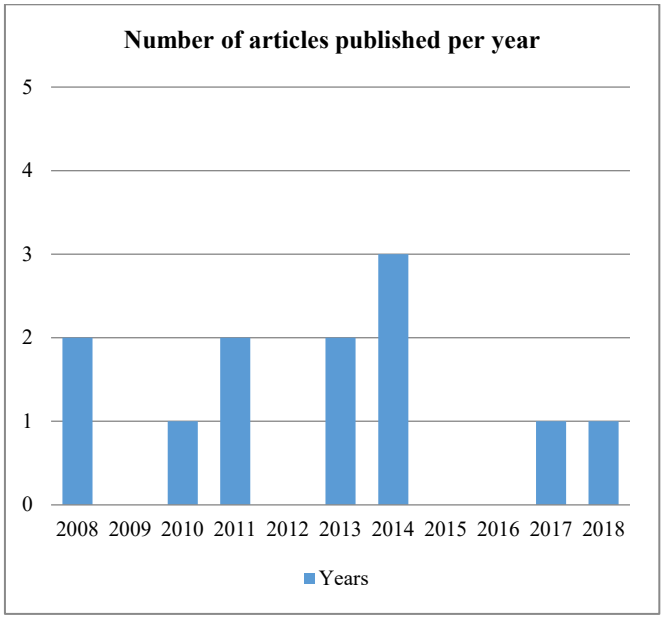


Figure 2: Number of articles published per year

As shown in Figure 2, the number of articles published per year is unusually low, especially after 2014. This might be because most studies on ITG had focused on the benefits and impact of ITG on firm performance instead of barriers to ITG implementation. As reported in [16], most organizations that participated in the study had a low ITG maturity level. Therefore, this study attempted to summarize issues that consistently inhibited the process of ITG implementations and adoption, as well as to propose potential solutions for future studies.

Table 2 reveals that most of the articles reviewed had both organizational and human-related issues in ITG implementation with the exception of 2 papers where one focused solely on IT practitioners and the other focused more on structural issues, such as the lack of defined roles and responsibilities. Human-related issues were mentioned 21 times and organizational-related issues were mentioned 18 times in all 12 articles. Table 3 shows the issues that were found in each category, while Table 4 shows the classifications of key issues found in articles after the reviewing process.

4.1. Organizational-related Issues

The issues in this category are regarded as issues faced during ITG implementation due to organizational decisions pertaining to the structure, guidelines and practices, as well as the support from top management. Organizational issues are related to top-down decision making where structures, responsibility and accountability, policy, guidelines and practices as well as

direction and goals are determined by the top management to be implemented by the entire organization. Hence, organizational-related issues are mostly related to the upper level of an organization. The subcategories proposed by this study pertaining to organizational-related issues are structure, guidelines and practices and top management.

4.1.1. Structure

Structure issues defined in this study refer to issues developed due to lack of assignation in accountability and responsibility in ITG. Structural issues are one of the most commonly found issues faced by newcomers to ITG implementations [4, 17-20]. As ITG is still a relatively new concept in Malaysia, organizations often struggle to develop appropriate structures or hierarchy of power in terms of ITG implementation. More often, a IT steering committee and IT strategy committee should be established by involving the Chief Information Officer (CIO). It was reported that structural issues were not functioning well as the CIO did not understand his/her role and responsibility in ITG [18]. Without an appropriate structure in place, there would be lack of transparency in the governance process and even difficulties in justifying IT-based decisions [19]. In addition, lack of structure can lead to communication issues [2]. Organizations will find it difficult to refer to the personnel in-charge as no one will assume the responsibility for related issues without a structure in place [18]. Similar issues were reported as there was a lack of a platform for management to discuss IT-related issues [21].

4.1.2. Guidelines and Practices

Guidelines and practices issues arise due to the lack of ITG guidelines and practices in the organization. Guidelines and practices are complementary factors to a reliable governance structure, which also contributes to the creation of transparency for justifying the ITG implementation processes [19]. The lack of strategic guidelines for IT funding in public universities was reported in [18]. It was revealed that IT projects that were initially approved or agreed by the IT committee were rejected eventually due to financial issues.

Miscommunication and improper guidelines led to a lack of coordination between departments and facilities regarding the utilization of IT resources. Guidelines and practices together with a reliable ITG structure prevent information disparity between the expectations of the CIO and implementation on the ground [21]. A study on Malaysia's public sector revealed that 76.19% of organizations had a low ITG maturity level. Findings also showed that IT managers expected at least an average level of ITG practice despite the lack of proper mechanisms and exposure to ITG practices [16].

4.1.3. Top Management

ITG should be driven and initiated from the organization's top level [19]. Top management support is classified as an organizational issue because top level executives are responsible for deciding organizational initiatives for moving forward in the right direction. Successful ITG implementation often depends on the top management's initiative to offer sufficient support. In addition, ITG implementation is also influenced by internal organizational cultures and politics. It is important for top management to recognize the needs for ITG and assign appropriate accountability according to initiatives, beginning from the top of the organization [3].

Lack of support and commitment from top management is one of the current issues faced by Malaysia organizations when implementing ITG [2, 22]. Similar situations where the top management was lackadaisical in supporting ITG was also reported [1], which then resulted in a lack of adequate resources to manage assigned controls as the top management willingness to incorporate ITG into the organization's future strategy is highly dependent on the recognition and knowledge about the importance of ITG.

Table 2: Key Findings

No.	Authors	Context	Key findings/Issues					
			①	②	③	④	⑤	⑥
1	[17]	Malaysia Ministry of Education	√	√		√	√	
2	[18]	Malaysia public university	√	√		√	√	√
3	[3]	Malaysian senior management		√	√		√	
4	[1]	Malaysian organizations			√	√		
5	[2]	Malaysian organizations	√	√	√	√	√	
6	[25]	Malaysian IT practitioners				√	√	
7	[20]	Malaysian universities	√					
8	[21]	Malaysian public universities	√	√		√	√	√
9	[26]	Malaysian public sector	√	√			√	
10	[19]	Malaysian public sector	√					√
11	[16]	Malaysian public sector		√		√	√	√
12	[22]	Malaysian financial sector			√	√	√	

Legend	
Organizational-related	Human-related
① Structure ② Guidelines and Practices ③ Top Management	④ Competencies ⑤ Awareness and Understanding ⑥ Leadership

Table 3: ITG Issues According to Categories

No.	Key Findings/ Issues					
	Organizational-related			Human-related		
	①	②	③	④	⑤	⑥
1	Lack of steering committee	Lack of documented practices		IT skills at basic level	Lack of understanding	
2	Communication breakdown and lack of commitment	Lack of approach for funding strategy		Lack of skills and knowledge in CIO and staffs	Lack of CIO with adequate ITG understanding	Lack of CIO presence in ITG and lack of IT initiative
3		Lack of exposure in ITG practices	IT accountability for top management		Recognition and knowledge from top management required	
4			Lack of top management support	Lack of adequate human resource		
5	ITG affected by culture and politics	Communication issues	Lack of top management support	Lack of time, training, knowledge and skills	Lack of ITG awareness	
6				ITG requires continuous commitment	ITG requires special knowledge	
7	Weak structure and lack of clear roles and responsibilities					
8	Communication breakdown	Lack of coordination and alignment		Lack of capable human resources	ITG responsibilities placed on IT only	Lack of CIO presence and effort to create awareness
9	Lack of governance in collaboration	Lack of governance in collaboration			Lack of understanding towards the importance of ITG	
10	Lack of accountability and transparency	Lack of justification on expense and resources	ITG should be driven from top management			ITG requires leadership
11		Lack of exposure on proper guidelines and mechanisms		Insufficient proficient IT personnel	Lack of consistent understanding on ITG	Lack of official directive
12			Lack of top management support	Complexity in ITG practices		

Table 4: Classification of Key Issues

Key Findings/Issues					
Organizational-related			Human-related		
Structure	Guidelines and Practices	Top Management	Competencies	Awareness and Understanding	Leadership
• Lack of adequate structure and accountability	• Lack of documented practices and guidelines • Lack of guidelines for collaborations • Lack of exposure	• Lack of top management support • Lack of top management accountability	• Lack of adequate IT-related skills • Lack of adequate managerial skills • Lack of human resources	• Lack of awareness and understanding on ITG • Lack of recognition of ITG	• Lack of role in CIO • Lack of IT initiative

4.2. Human-related Issues

Human-related issues refer to issues originating from human resources caused by the lack of quality, competency, and understanding required for managing and completing tasks required by organizations. Specifically, human-related issues in this study refer to barriers to ITG implementation due to insufficient personnel with quality to manage and govern IT-related resources and issues. In human-related issues, the personnel included in this study were the CIO and other personnel involved in ITG from the middle to the top management level. Three subcategories in human-related issues are competencies, awareness and understanding as well as leadership quality.

4.2.1. Competencies

Competency issues are often overlooked by organizations since most of the attention is placed on ITG frameworks and practices. Competency in this study refers to the ability of ITG personnel to manage and complete assigned responsibilities and practices related to ITG. The terms competency and capability are often used interchangeably due to how closely their interpretations are. Competence is defined as having the ability to utilize skills, knowledge and the capacity to satisfy or achieve current needs. Capability is defined as qualities, abilities, and the potential to develop accordingly to meet future needs by adapting to changes in a dynamic environment while generating new knowledge for improving working performance [23]. Therefore, in order to fit in the context of this study, where current ITG issues can be safely regarded as current ITG needs, the term competency is used in this subcategory. Most senior management and CIOs lack knowledge and experience in ITG because ITG is a relatively new concept compared to working skills and experience [1,3,17-18]. In addition, it was reported that the complexity of ITG practices was the highest ranked barrier from the technical aspect [22]. The prerequisites for ITG personnel have been elevated as ITG committees are required to plan, manage, and monitor IT initiatives. ITG personnel also need to master non-technical skills such as business strategic planning, leadership, and communication [16]. Hence, due to the increased prerequisites for ITG personnel, organizations have been lacking human resources equipped with proficient skills in technical and non- technical aspects to resolve ITG issues and challenges.

4.2.2. Awareness and Understanding

As mentioned previously, top and senior management need to recognize the need for ITG to ensure smooth top-down implementation [3]. Such recognition requires adequate awareness and understanding about ITG. Awareness and understanding issues refer to issues that occur due to the lack of experience and understanding on the potential benefits of ITG in organizations. Awareness can elevate performance and allow accurate understanding of situations that can further lead to better decision making by leaders [3]. The authors had further revealed 3 level of ITG awareness in reference to Endsley's theory of awareness [24]. The first level of awareness includes gathering perceptual information from the surrounding and focusing on elements that are beneficial and relevant. One example of this is to understand the availability of resources, infrastructure, and skills required to achieve current tasks. The second level of awareness includes integrating perceptual information with experiences and knowledge to assess the significance of the context, which is the role of ITG in organizations [25]. The third level of awareness includes projecting future elements by anticipating the changes that can be brought to the operational environment. Examples of third level awareness are management of IT resources that meet organizational needs, lead the initiative in addressing IT issues, ensure continuous alignment between business and IT, and leverage IT potentials to enhance business growth. It was found that top management perceives that there is no need for ITG as there is a lack of awareness and knowledge [2]. There were no efforts by the top management to create awareness about ITG adoption [21]. Instead, all the issues were thought to be an IT-only responsibility, which further leads to a disoriented alignment between strategic business and IT planning. Business executives did not understand the strategic value of IT as it was thought to be insignificant [26]. Business executives failed to understand IT issues and there was a severe lack of collaboration between business and IT. IT managers assumed a high-level of maturity for ITG but in reality, 76.19% of the organizations were still in position of low ITG maturity [16]. This could be due to different levels of understanding regarding the definitions of ITG and ITG itself. Most implementations were highly dependent on instructions, and there was a lack of initiative to actively seek improvement. All the issues mentioned above indicate that most

organizations in Malaysia still lack ITG awareness and understanding.

4.2.3. Leadership

Leadership quality in this study refers to the ability of top and senior management to spearhead the IT initiative in ITG implementation. As most of the organizations are still at the infancy stage of ITG, lower levels of the organization hierarchy need to be guided by top management. This requires top management to lead using the IT initiative and provide direction from the highest level of the organization [19]. For top management to lead, adequate knowledge about ITG is required. Therefore, a lot of issues arise because even the top management has a low understanding of ITG itself [18]. There is a possibility of a lack of official directives concerning ITG, which causes the lack of collaboration between business and IT [16]. Top management should always initiate the active involvement of stakeholders in ITG, while providing sufficient guides and empowering key personnel with expectations and support. As most personnel lack ITG knowledge and experience, top management plays a major role in ensuring the continuity and sustainability of ITG. Presence of the CIO was insignificant because the responsibilities were not up to a respectable standard [21]. There was no effort to create awareness and liaison between business and IT. Hence, the IT department assumed all responsibility for issues identified. Leadership in ITG is especially vital for organizations at the infancy stage as the top management is responsible for setting the guidelines and directions for organizations to move forward. Newcomers to ITG need explicit guidance or they would just become simple rule followers instead of skilled personnel actively seeking continuous innovation and improvement.

5. Role of IT Human Capability

IT human capability mentioned in this study refers to technical capabilities, business capabilities and behavioural capabilities [12]. Technical capabilities refer to technical skills and knowledge in a specific area of expertise. Business capabilities refer to the IT personnel's ability to understand and manage overall business processes and environments. Behavioural capabilities refer to the IT personnel's interpersonal ability to interact effectively with others. The results show that both organizational-related and human-related issues are major factors that help organizations to implement ITG in Malaysia. It is undeniable that both issues need unwavering attention from organizations in order to implement ITG and move it in the right direction.

ITG is more than just following rules and guidelines. ITG requires continuous commitment from organizations to ensure sustainability, and having competent human resources is one of the prerequisites for future sustainability, whether in business or ITG [25]. It has been stated that organizations might be governing IT with less than adequate prerequisites in terms of IT capabilities [10]. Development of ICT personnel has been one of the focuses of the Malaysian Strategic IT Plan 2011 [26]. However, findings indicate that there is a lack of effort or mechanisms for developing IT personnel in the area of ITG. Although issues have been categorized into organizational-related and human-related categories, both issues are interrelated as one issue can cause the other to surface.

One of the primary issues in organizational-related issues is the lack of adequate structure and accountability in ITG. The lack of organizational structure is not directly related to IT human capability, but it should be taken into consideration when forming the ITG structure. When forming the ITG structure, organizations should assess whether there are qualified personnel available and accountable for ITG implementation. It is undeniable that collaboration issues [21,26] have consistently emerged during ITG implementation. In the selection of qualified personnel for ITG, there is no doubt that at least certain levels of business and behavioural capability should be considered as ITG often requires collaborative efforts for strategic alignment purposes. This requires personnel to possess some level of business acumen and the capability to work well with other units when needed. It raises further questions on whether business or IT executives should be held accountable, as most business executives possess business capabilities without IT skills, and most IT executives possess IT capabilities without business skills. It does not mean that business executives are required to possess IT capabilities but at least they should be aware of the importance of ITG since IT executives are now required to extend their understanding of business knowledge [27]. It was also revealed that business and IT, which act in collaboration, can lead to ITG effectiveness and strategic alignment as knowledge sharing between business and IT executives lead to a better understanding of each other's environment [28-29]. For example, it has been revealed that a web-based system developed in a university aligns the entire existing or new program with the Malaysian Qualification Framework (MQF) [30]. Similar web-based systems can be designed to monitor and align ITG prerequisites pertaining to certain personnel involved. Training should be given for those who do not meet the prerequisite capabilities for ITG.

Guidelines and practices are some of the most common issues faced when implementing ITG. Guidelines and practices are available through frameworks like COBIT 5.0 and ITIl.v3. However, despite the availability of frameworks in the market, not all practices and mechanisms are suitable for every organization. One main purpose of guidelines and practices is to help and give direction to organizations intending to implement ITG. It is also important to practice the relevant mechanisms that help develop positive IT behaviours throughout the entire organization. One of the common issues with the availability of frameworks is the lack of instructions on how to perform tasks with the right mind-set. ITG is more than just simply following rules and guidelines. Organizations must also implement various ITG mechanisms that help to develop individuals with adequate capabilities so that ITG efforts are sustainable in the future. It is important for organizations to develop enough ITG awareness among IT personnel through various practices and mechanisms. As reported, organizations with awareness imbue governance-related mind-sets into daily routines and practices [3]. Personnel with the ability to actively lead and think using IT initiatives are better than rule followers. Guidelines and practices provide the basis for incorporating ITG into daily routines, but it depends on the personnel's capability to understand and manage tasks effectively. In addition, only qualified IT personnel who are aware of the importance of ITG will actively think of the processes and

guidelines that need to be improved and adapted to the changing business process. Therefore, IT human capability is important for the continuous improvement of ITG implementation in an organization.

Gaining top management support has been an issue for organizations in Malaysia as top management lack the understanding of the importance of ITG. Moreover, lack of top management support directly affects the outcome of ITG implementation as resources are often procured with the approval of the top management. It is important to convince the top management about the implementation of ITG in order to gain their support. It has been proven that new IT deployment is easier with firm commitment from the top management [30]. However, it is not easy to convince the top management as ITG is a long-term project that requires various resources to succeed. Therefore, having capable human resources that are knowledgeable in ITG can ensure the continuity and success of ITG implementations, especially in the future. The CIO plays an important role in convincing top management as the CIO often acts as a liaison between business and IT executives. The CIO who is capable of recognizing, understanding, and relaying the importance of ITG to the top management ensures better allocation of resources for a smoother implementation process.

Unlike organizational-related issues, human-related issues are directly related to IT human capability. As mentioned before, competencies refer to the ability to satisfy current needs with skills and knowledge. Therefore, organizations need to develop current competencies into capabilities in order to adapt to future challenges. Currently, ITG is presented as a brand-new challenge that requires human resources with proficient skillsets and capabilities for organizations in Malaysia. Most Malaysian organizations lack competent human resources for implementing ITG. The lack of competency generally includes IT as well as managerial skills and knowledge, which are part of technical and business capabilities. The issues are especially troubling for the public sector as most public sectors have a more or less fixed number of staffs. Therefore, a recruitment of additional experts or capable personnel was less likely to occur unless external expertise is specifically consulted. Hence, organizations need to develop human resources that are proficient in ITG supplemented by the help of various practices and mechanisms. However, competency issues consistently emerge from organizations suggesting that organizations are more focused on setting rules and guidelines rather than developing capable human resources that are supposedly accountable for ITG tasks. This creates the culture of passive governance where actions are only taken based on rules and instructions. As mentioned earlier, personnel involved in ITG should be equipped with business knowledge and ITG awareness. This can be achieved with job rotations and cross training between business and IT units [27]. A web-based system can be developed to record all qualifications and skills of related personnel. All training and rotations should be provided based on specific skills that are lacking. Qualified personnel with cumulated experience in the task can improve the efficiency of the process and are more inclined to offer and seek improvements for ITG.

Another factor of passive governance comes from the lack of awareness and understanding of ITG. ITG requires leaders with enough understanding and knowledge about the ITG concept to lead with IT initiative and active thinking in order to create a good culture of governance. Results revealed that the importance of ITG was severely underestimated by organizations. Few cases reported that senior management did not recognize the importance of ITG and felt that ITG was unnecessary [3,26]. There was also a case where ITG was thought to be only as an IT responsibility [21]. Issues like this clearly show that the roles of ITG in organizations were ambiguous to entire organizations, or at least to the top management and business executives. For ITG implementation to be successful, ITG awareness must be spread throughout the organization so that strategic alignment between business and IT can be achieved. Organizations must reach consensus on the role of ITG, and how it can affect organizations moving forward. As mentioned previously, awareness allows organizations to imbue certain mindset into daily practices. For this case, ITG awareness prevents the culture of passive governance and allows personnel to be actively involved in the governance process moving forward. In this case, awareness and understanding must be included as part of IT human capability as both play a big role in adapting to new challenges.

In this study, leadership is listed under behaviour capability since leadership is an essential ability to lead and interact with others. From the leadership aspect, the CIO is expected to be in the midst of ITG implementation. However, previous articles reveal that CIOs in Malaysian organizations lacked leadership qualities, especially in terms of IT initiative [16, 18-19]. In ITG's infancy stage, leadership plays one of the most important roles in ITG implementation as beginners in the organization need more guidance than ever. The CIO is seen by practitioners as the main driver in ITG [27]. However, the CIO should not be the only one to be held responsible for leadership issues. Anyone in the organization who is aware of the importance of ITG should be leading-by-example in order to create more ITG exposure. As mentioned previously, ITG should be driven from the top management downwards. Therefore, the top management must take responsibility in leading the IT initiative by recognizing the importance of instilling a successful ITG.

Finally, as mentioned by [10], the use of IT capability has been ambiguous in the ITG literature. Therefore, this study attempted to clear the ambiguity by exploring and explaining the role of IT human capability in ITG implementation. This study can contribute to future studies on the relation between ITG and IT human capability.

6. Conclusion

In conclusion, this study explained the interrelated relations between organizational-related and human-related issues by justifying the role of IT human capability. This study also clarified the ambiguity concerning the use of IT capability on ITG literature. Through systematic literature review, it has become clear that structure and guidelines help ITG to consistently move in the right direction. However, without specific capabilities, such as business, technical, and behavioural capabilities possessed by related personnel, ITG cannot be implemented effectively since most ITG implementation and practice requires collaboration from both business and IT units for the sake of strategic alignment. Therefore, it can be concluded that both organizational and human

factors are important for the consistency, efficiency and effectiveness of ITG. Additionally, researchers are encouraged to commit further studies on IT human capability and ITG implementation in Malaysia in order to compensate the lack of proven ITG studies in the Malaysian context. This study also attempted to further highlight the importance and role of IT human capability in the midst of ITG implementation and mechanisms in order to garner more attention toward human-related issues since developing countries like Malaysia are affected by issues such as the lack of qualified human resources. Therefore, organizations should pay equal attention to both organizational and human-related issues in order to create a well-built structure, framework, and environment that are well-suited for ITG implementation at the infancy stage.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Special thanks to University Tun Hussein Onn Malaysia (UTHM) for assisting in the completion of the paper.

References

[1] K. A. Aziz, M. B. Norhashim, E. M. Halim, "Information security and information technology governance: a Malaysian case study" International Journal of Management Practice, 4(4), 331, 2011. https://doi.org/10.1504/ijmp.2011.039204

[2] M. F. I. Othman, T. Chan, E. Foo, "IT Governance Adoption in Malaysia : A Preliminary Investigation" in 2011 22nd Australasian Conference on Information Systems (ACIS), Sydney, Australia http://aisel.aisnet.org/acis2011/69

[3] Y. M. Lin, N. H. Arshad, H. Haron, Y. B. Wah, M. Yusoff, A. Mohamed, "IT Governance Awareness and Practices: an Insight from Malaysian Senior Management Perspective" Journal of Law and Governance, 5(1), 2010. https://doi.org/10.15209/jbsge.v5i1.177

[4] E. N. Nfuka, L. Rusu, "Critical Success Factors for Effective IT Governance in the Public Sector Organizations in a Developing Country: The Case of Tanzania" in 2010 ECIS 18th European Conference on Information Systems, Pretoria, South Africa, 2010. https://doi.org/10.1080/1097198X.2013.10845642

[5] P. Weill, J. W. Ross, "It Governance on One Page" MIT Sloan Working Paper No. 4517-04; Research Working Paper No.349, 2004. http://dx.doi.org/10.2139/ssrn.664612

[6] W. V. Grembergen, S. D. Haes, "Exploring the relationship between IT governance practices and business/IT alignment through extreme case analysis in Belgian mid-to-large size financial enterprises" Journal of Enterprise Information Management, 22(5), 615-637, 2009. http://doi.org/10.1108/17410390910993563

[7] ITGI, Board Briefing on IT Governance, 2nd Edition, 2003.

[8] A. S. Bharadwaj, "A Resource-based perspective on information technology capability and firm performance: an empirical investigation" MIS Quarterly, 24(1), 169-196, 2000. http://dx.doi.org/10.2307/3250983

[9] V. Sambamurthy, W. Zmud, "Arrangements for information technology governance: A theory of multiple contingencies" MIS Quarterly, 23(2), 261-290, 1999. http://dx.doi.org/10.2307/249754

[10] K. Hiekkanen, J. Korhonen, E. Patricio, M. Helenius, J. Collin, "IT Governance, Decision-Making and IT Capabilities in 2013 Proceedings of The 9th European Conference on Management Leadership and Governance, Klagenfurt, Austria, 2013.

[11] Z. Aiqun, "An IT capability approach to informatization construction of higher education institutions" Procedia Computer Science, 131, 683–690, 2018. https://doi.org/10.1016/j.procs.2018.04.312

[12] D. Ko, D. Fink, "Information technology governance: An evaluation of the theory-practice gap" Corporate Governance: The international journal of business in society, 10(5), 662–674, 2010. https://doi.org/10.1108/14720701011085616

[13] J. S. Chen, H. T. Tsou, "Performance effects of IT capability, service process innovation, and the mediating role of customer service" Journal of Engineering and Technology Management, 29(1), 71-94, 2012. https://doi.org/10.1016/j.jengtecman.2011.09.007

[14] L. Fink, S. Neumann, "Gaining agility through IT personnel capabilities: The mediating role of IT infrastructure capabilities" Journal of the Association for Information System, 8(8), 440-462, 2007. https://dx.doi.org/10.17705/1jais.00135

[15] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Review in Software Engineering, Software Engineering Group, University of Durham, 2007.

[16] R. A. Razak, M. S. Zakaria, "The Information and Communication Technology Governance Maturity Level for Malaysian Public Sector" in Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing, Cambridge, United Kingdom, 2017. https://dx.doi.org/10.1145/3018896.3018967

[17] S. Ismail, R. A. Alias, A. A. Rahman, "IT Governance Implementation in the Malaysian Ministry of Education" in 2008 Postgraduate Annual Research Seminar (PARS '08). Johor Bahru, Malaysia, 2008.

[18] N. A. Ismail, "Information technology governance, funding and structure" Campus-Wide Information Systems, 25[3], 145-160, 2008. https://dx.doi.org/10.1108/10650740810886321

[19] R. A. Razak, M. S. Zakaria, "The critical success factors for effective ICT governance in Malaysian public sector : A Delphi study. International Journal of Economics and Management Engineering" 8(11), 2014. http://www.waset.org/publications/9999698

[20] A. R. Ahlan, B. A. Ajayi, H. Hussin, "Roles of Governance, Ethics and Integrity in Managing Information Technology Resources: Muslim Practitioner's View" in 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), Rabat, Morocco, 2013. https://doi.org/10.1109/ICT4M.2013.6518926

[21] A. R. Ahlan, Y. Arshad, and B. A. Ajayi, "IT governance in a Malaysian public institute of higher learning and intelligent decision making support system solution," Engineering and Management of IT-based Service System, 19(33), 2014. https://doi.org/10.1007/978-3-642-39928-2_2

[22] M. F. I. Othman, "Using analytical hierarchy process (AHP) to evaluate barriers in adopting formal IT governance practices" Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10, 35-40, 2018.

[23] R. Nagarajan, R. Prabhu, "Competence and capability- a new look" International Journal of Management (IJM), 6(6), 7-11, 2015.

[24] M. R. Endsley, "Measurement of situation awareness in dynamic systems." Human Factors, 37(1), 65-84, 1995. https://doi.org/10.1518/001872095779049499

[25] W. L. Teo, A. A. Manaf, P. L. F. Choong, "Perceived effectiveness of information technology governance initiatives among IT practitioners. International Journal of Engineering Business Management, 5(1), 1–9, 2013. https://doi.org/10.5772/56661

[26] J. Kaur, S. A. H. Bahri, "Implementation of Information Technology Governance in the Malaysian Public Sector Practice" in Pacific Asia Conference on Information Systems (PACIS), Chengdu, China, 2014. http://aisel.aisnet.org/pacis2014/119

[27] B. A. Ajayi, H. Hussin, "IT governance from practitioner's perspective: sharing the experience of a Malaysian university" Journal of Theoretical and Applied Information Technology, 88(2), 219-230, 2016.

[28] J. Kaur, N. Mohamed, A. R. Ahlan, "A Confirmatory Factor Analysis of the Information Technology Governance Effectiveness: Evidence from Malaysia" in 2011 International Conference on Research and Innovation in Information Systems, Kuala Lumpur, Malaysia, 2011. http://dx.doi.org/10.1109/ICRIIS.2011.6125666

[29] M. A. J. Sabegh, S. M. Motlagh, "The role and relevance of IT governance and IT capability in Business- IT alignment in medium and large companies" Business and Management Review, 2(6), 16-23, 2012.

[30] N. Musa, D. H. A. Ibrahim, N. A. Bolhassan, J. Abdullah, N. Kulathuramaiyer, M. N. Khairuddin, "An IT Governance Framework for Achieving the Development of Academic programme in Higher Institutions: A Case of Universiti Malaysia Sarawak (UNIMAS)" in The 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), Kuching, Malaysia, 2014. http://dx.doi.org/10.1109/ICT4M.2014.7020673

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 321-326 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Estimating Academic results from Trainees' Activities in Programming Exercises Using Four Types of Machine Learning

Toshiyasu Kato*[1], Yuki Terawaki[2], Yasushi Kodama[2], Teruhiko Unoki[3], Yasushi Kambayashi[1]

[1]*Nippon Institute of Technology, Saitama, Japan*

[2]*Hosei University, Tokyo, Japan*

[3]*Photron Limited, Tokyo, Japan*

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 09 October, 2019
Online: 22 October, 2019

Keywords:
Programming Exercises
Classroom Assistants
Learning analytics
Machine Learning
Grading estimate

ABSTRACT

Predicting trainees' final academic results in the early stage of programming class is a significant mission in the field of learning analytics. Performing exercises in programming class is hard and it takes a lot of time for trainees. For this reason, careful support with trainees are offered in many classes through classroom assistants (CAs). Even with CAs' assistances, managing a programming class is difficult. Because each trainee's coding activity is different from another's, even when each of them is solving the same exercise. Classroom assistants with little teaching experience have difficulty for understanding the unique features of trainee's coding activity. We have employed data mining to analyze trainees' coding activities to distinguish those various features. The objective of this research is to present such behavioral features of trainees to CAs to enrich their assistance for the trainees. In order to investigate the timing of guidance, we conjectured the academic results from the chronicle record of coding activities.

1. Introduction

This paper is an extension of the work originally presented in 2017 International Conference on Computational Science and Computational Intelligence (CSCI) [1].

In programming exercises, Individual trainee's coding activity varies from person to person, even when each trainee is answering the common exercise. The aim of this research is to display many coding features of trainees to classroom assistants (CAs). We conjectured the academic results to indicate the timing of instruction to the CAs.

Through the analysis of the logs of trainees by using the data mining techniques, we can find the learning pattern of each trainee. This analysis helps instructors to identify "at-risk" trainees [2]. Programming exercises in higher education institutions allow trainees to solve individual tasks. Since the skill levels of trainees differ, each of them has to work on tasks in an individual way [3]. Some trainees solve exercises easily, while others take much time. Therefore, many schools employ CAs to provide assistances for the trainees. CAs are expected not only answer questions of trainees but also to provide comprehensive assistances for the trainees with large assignments [4]. For this reason, a CA needs to understand who needs assistance, what kind of assistance the trainee needs, and under what circumstances assistance the trainee confusing [5].

Predicting trainees' final academic results in the early stage of programming class is a *significant mission* in the field of learning analytics. For instance, one can perform regression analysis for this purpose [6]. In order to catch the timing when the instructors intervene to the trainees who need help, instructors need to conjecture the trainees' future scores of the programming exercises. In order to provide such estimations, the authors have developed and used a unique programming exercise support system [7]. The system is called PRESS, and through the analysis of the outputs of PRESS, we have succeeded to identify the coding activities.

Trainees' progresses differ from each other, because of the difference of individual skill levels [3]. We have found that this skill level difference can be observed in the coding mode. Trainees who show slow progress have poor coding modes. PRESS demonstrates that such situations are common. Such trainees often

*Toshiyasu Kato, Nippon Institute of Technology 4-1 Gakuendai, Miyashiro-machi Minami-saitama-gun Saitama, 345-8501, Japan, katoto@nit.ac.jp

stop programming and become disgusted [8]. Therefore, it is important for instructors to identify what kind of trainees have programming exercises at an early stage [4].

We have conjectured the trainee's academic results in order to get the timing when CAs start assistance for the problematic trainees. According to the instructor's opinion, trainees who take a lot of time for exercises tend to have low academic results. In this paper, we give details of the results of evaluation experiments on this estimation. As a result, CAs became to comprehend the trainee's coding activities without having prior programming support experience.

2. Related Work

Several researchers have reported that coding activities has improved trainee's skills by assisting in debugging and coding methods. Ryan has built a model that enables trainees to improve their skills by analyzing debug logs and execution results [9]. Our study analyzes coding history in addition to debug logs.

Alex proposed a tutor on a computer to support trainee programming [10]. The tutor checks whether the trainees are on the correct trajectory and advises trainees in a wrong track. For instance, an instructor may give trainees a hint on how to reconstruct their programs. Our research aim is to improve trainees' practical skills by supporting coding methods.

There are past researches targeted at study how to support trainees through their learning chronicle data. In such studies, scientists have analyzed teaching materials using techniques based on statistical methods on Support Vector Machines (SVM). For instance, Yamada et al. detected contents on the Web pages that were difficult for trainees to comprehend by logistic regression [11]. In that research, they have created models with seven explanatory parameters, namely latent time, gaze point distance moved, gazing point movement speed, mouse distance moved, mouse speed, the amount of mouse wheel rotation, and correct identification. Yamada's evaluation experiment have achieved the identification rate of 81% by using SVM and time series data to improve coding activities, such as compilation and execution frequency.

Researches on foreseeing trainee's performance by machine learning (ML) algorithm that takes advantages of trainee's learning chronicle data as teaching data are as follows.

Guo, Zhang, Xu, Shi, and Yang applied deep neural network (DNN) to the data from 12,000 trainees in Chinese elementary and junior high schools to predict five academic results in high school entrance exams [12]. They found that deep learning can predict academic results with highest accuracy rate (Accuracy) among many ML methods.

Belachew and Gobena employed three types of ML algorithms, namely neural networks, Naive Bayes and support vector machines to predict results of 993 trainees at Wolkite University in Ethiopia [13]. They found that, as a result of comparison, Naive Bayes had the highest accuracy rate.

Ahmad, Ismail, and Aziz used three ML algorithms, namely Naive Bayes, decision trees, and rule bases to predict the performance of about 500 trainees in Malaysian [14]. As a result, they reported that the rule base has the highest accuracy rate.

Amrieh, Hamtini, and Aljarah used three ML algorithms, namely neural nets, decision trees, and Naive Bayes, to predict performance with data from 480 trainees at Jordanian University [15]. They found that decision trees and neural networks obtained the highest accuracy rates.

In these preceding researches, variety of ML algorithms are applied to trainee data, and researchers who tackled this problem demonstrated how and what way each of the ML methods succeeds to predict the final scores. Unfortunately, the data used for the evaluation are largely different from each other, and the data used in many studies are not open to the public. It is difficult for a third party to pursue further study through conducting a comparative experiments with a follow-up test or alternative method. We can only use them as reference data.

3. Analysis of Coding Activities

In this research, we have set up data on the coding activities of trainees in 2014 and 2015 classes analyzed in previous study [16, 17]. Table 1 displays the explanatory parameters used for the analysis. We have measured all of the data and then analyzed the analysis targets using PRESS. We have examined the following two approaches to the data.

- Unsupervised classification.
- Estimation of exercise-solving time.

Table 1: Explanatory parameters

Type of characteristics		
Time	Frequency	Error
Problem-Solving	Compilation	Number of errors
Compilation interval	Execution	Number of similar errors
Execution interval		

3.1. Unsupervised classification

We have used Ward's method, which is a representative hierarchical clustering [18] [19]. Clustering is a process that groups objects into classes of similar objects [20]. We have formed from

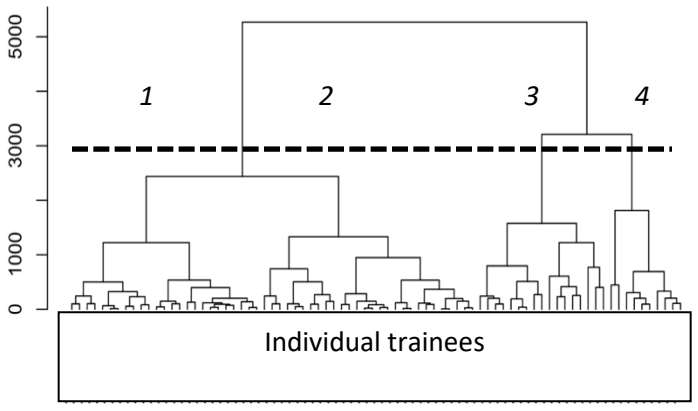


Figure 1: Clustering by Ward's method

Table 2: Explanatory parameters for each cluster

Cluster	Solving time [s]	Compilation interval [s]	Execution interval [s]	Compilation frequency	Execution frequency	Number of errors	Number of similar errors	Number of trainees
1	528	209	267	7	5	1	1	25
2	812	379	606	6	3	3	3	28
3	1405	590	1160	7	3	4	3	17
4	1506	228	337	18	13	5	4	10

the patterns (observations, data elements, or characteristic vectors) into clusters (groups or subsets) based on those localities and connectivity in the n-dimensional space as shown in Figure 1. Figure 1 demonstrates the cluster analysis of the seven explanatory parameters. Table 1 indicates the summary of explanatory parameters for each cluster in Figure 1.

We have perceived the following clusters:

- Cluster 1: A group solved the exercise in short time. In this group, trainees understand each problem they encounter while coding.
- Cluster 2: A group solved the exercise in short time. In this group, trainees are coding without understanding the details of their problems.
- Cluster 3: A group solved the exercise in long time. In this group, trainees are cursorily checking the error contents and cording without full comprehension of their problems.
- Cluster 4: A group solved the exercise in long time. In this group, trainees are cording without checking any error details.

3.2. Length estimation of time for solving exercises

The trainees in each cluster should receive cluster-specific instructions. Therefore, instructors and CAs need to identify the features of the trainees in the class and pay special attentions to those who take a long time to solve given exercises. The cluster analysis, however, cannot conjecture the duration for solving exercises in the future. For this reason, CAs cannot know which trainees need certain types of guidance at the beginning or before class. In order to ameliorate this hindrance, we used the 2014 class coding activities as training data and tried to predict the exercise-solving time in 2015 class activities. Table 3 displays the coding activity data for two years. The values of the explanatory parameters for each year are averaged values (excluding the number of trainees). These data are taken from two programming exercises of the same content each year.

We have used logistic regression to decide whether trainees' exercise-solving time would be long or short. Logistic regression is a typical statistical method that illustrates the relationship between two different series of data [21]. We can assume that one of the series of data as independent variables, and the other data are dependent variables. We assume that the output variable takes one value from a set of fixed values. This value corresponds to a class of the classification problem. The regression equation is displayed in Equation (1). p is the probability conjecture of the

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1\chi_1+\beta_2\chi_2+\cdots+\beta_k\chi_k)}} \quad (1)$$

Table 3: Coding activities in 2014 and 2015

	2014	2015
Number of trainees	85	80
Solving time (Problem) [s]	815	936
Compilation interval (Average) [s]	295	352
Execution interval (Average) [s]	513	584
Compile frequency	10	8
Execution frequency	5	5
Number of errors until correct answers	7	3
Number of similar errors that are most frequent	4	3
Number of trainees late submission	32	35
Number of exercise resolutions	40	41

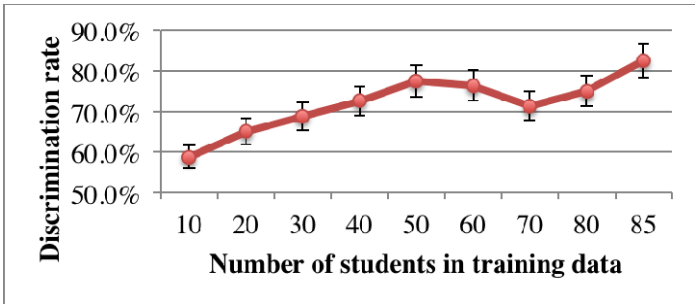


Figure 2: The number of trainees in the training data and discrimination rate

objective variable. In this research, we obtained the probability p of a long exercise-solving time. β_0 is a constant, β_1, β_2, ... β_k are regression coefficients, and $\chi_1, \chi_2, \ldots \chi_k$ are explanatory parameters.

We first applied training data to logistic regression to determine constant and regression coefficients. We used Tensor Flow for this calculation. Tensor Flow is a representative tool for DNN [22]. We used cross-entropy as a learning error function for the supervised classification problem and carried out optimization using the steepest descent method [23]. The objective variable for logistic regression must be represented in the 0/1 binary. We set the teaching data to 0 when it was longer than the mean exercise-solving time and to 1 if it was shorter than the mean exercise-solving time.

We created a model with the 2014 data, and then we applied the 2015 data to the model and conjectured the length of the exercise-solving time. The discrimination rate was 82.5%. The

parameters at this time were a learning rate of 0.015 and the number of learning iterations was 70,000. Figure 2 displays the number of trainees and the identification rate used in the training data for ML.

4. Estimating Academic results from Trainees' Activities

Now that we have the conjecture of the exercise-solving time, we next conjectured academic results to find the timing of instruction. In order to improve accuracy much further, we have used 2016 data in addition to the last two years (Table 3). Table 4 displays the coding activities for 2016 class and Table 5 displays the academic results by three years. The academic results are based on the final exam at the end of the semester.

The subjects (trainees) are sophomores in a Java programming class from our institute. Each year, the same instructors took the charge of the class, but the academic results are differed from year to year as a subjective evaluation of instructors has been added. For this reason, we have defined the estimation rate as follows along with the coincidence rate:

$$Estimation\ rate\ = \frac{Estimated\ academic\ results\ \pm 1}{Number\ of\ trainees}$$

Based on the data of these three years, we have conjectured the academic results through the following procedure.

1. Normalize the input data.
2. Decide the features.
3. Decide network configuration and parameters.
4. Conjecture the accuracy by creating a network model.
5. Check the estimation accuracy.
6. Repeat steps 3 to 5 to improve accuracy of the estimation.
7. Analyze by changing the hidden layer

Table 4: Coding activities in 2016

Number of trainees	107
Solving time (Exercise) [s]	639
Compile interval (Average) [s]	405
Execution interval (Average) [s]	139
Compile frequency	7
Execution frequency	2
Number of errors until correct answers	5
Number of similar errors that are most frequent	2
Number of trainees late submission	51
Number of exercise resolutions	38

Table 5: Academic results in programming exercises

Academic results	2014	2015	2016
1: Failing	17	1	5
2: Below pass	33	13	34
3: Pass	17	17	34
4: Very good	21	36	30
5: Excellent	2	15	7

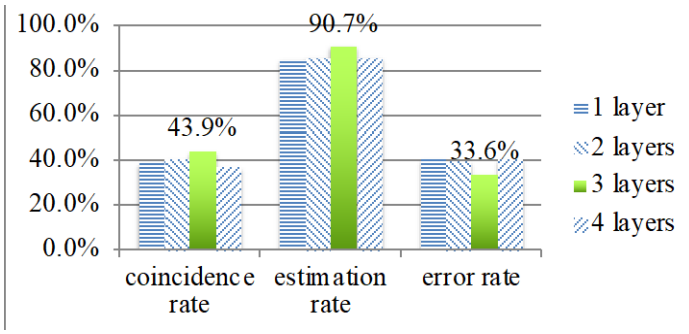


Figure 3: The coincidence rate, estimation rate, and error rate

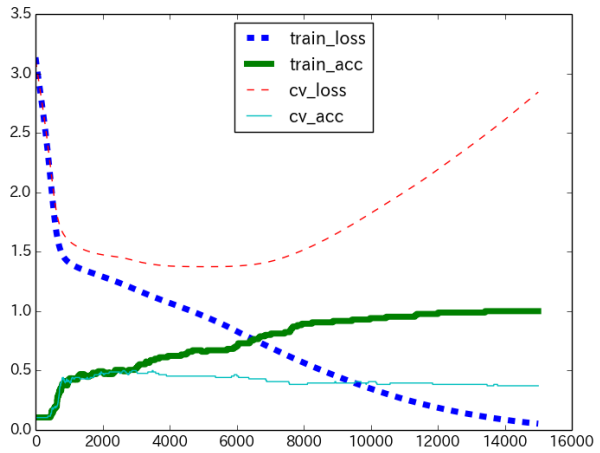


Figure 4: Loss and accuracy when the hidden layer is three layers

Figure 3 displays the coincidence rate, estimation rate, and error rate in the case of one, two, three and four hidden layers respectively. The error rate in Figure 3 is calculated by sum of the divergences between conjectured academic results and final academic results then divided by the number of trainees.

In addition to these parameters, we have chosen the following features to analyze the trainees' activities. They are duration for solving the exercises compile frequency, execution frequency, the number of similar errors that are most frequent, and the number of exercises they resolved. Through various machine-learning situations, we have found that the case of three hidden layers provides the most accurate estimation. Figure 4 displays the best case of our experiments; that has three hidden layers.

Table 6 displays the mean values of explanatory parameters. We have normalized them for each academic result. There was no trainee conjectured to earn the academic result of 5 (highest). In academic result, the number of exercises solved and the time for solving exercises are particularly related.

In addition to the experiments using DNN, we have performed performance prediction using three ML algorithms. The algorithms used in this experiment are Random Forest, SVM, and Naive Bayes. The objective of these experiments is to assess the performance of different ML algorithms.

With those ML algorithms, the purpose is also classification; the same as DNN. Also, we used the same data set for analysis as

Table 6: The mean values of feature quantities normalized for each academic result

Ac. result	Exercise-solving times	Compile times	Execution times	Similar Errors	Number of exercise resolutions
1	1.66	-0.40	-0.96	-0.78	-2.28
2	0.25	0.44	-0.44	0.47	-0.65
3	0.20	-0.77	-0.69	-0.68	-0.61
4	-0.96	-0.32	0.32	-0.30	1.01
5	-	-	-	-	-

Table 7: Estimation results

	Accuracy	Precision	Recall
DNN	0.70	0.41	0.46
Random Forest	0.62	0.34	0.37
SVM	0.58	0.32	0.42
Naive Bayesian	0.58	0.32	0.38

DNN. The reason for adopting these three types is that they were used for performance prediction in the previous researches [11] [13] [14] [15].

Table 7 displays the results of various data mining methods. The interpretation of precision and recall values here depends on each class and each academic result. This value is based on the average of all academic result classes. From these results, we can confirm that the academic results can be conjectured through analysis of coding activities by any MLs. DNN provides the most accurate predictions, but precision and recall values are low. From the perspective of accuracy, our method can find trainees who would fail from the class. However, since precision and recall are low, there is an exercise that the same method does not guarantee the same accuracy every year.

5. Discussion

We have conducted experiments of DNN with various combinations of explanatory parameters. We could not identify one single particular dominating explanatory parameter. We can conclude that all the explanatory parameters contribute almost equally to the accurate estimations. What we can state is that the DNN provides the highest values as displayed in Table 7.

For this data set, we can conclude that we can predict the performance of each trainee by any algorithms we select, even though the Deep Leaning provides the best results. However, precision and recall values are low even in DNN. The reason may be that the trainees are not the same every year and the final exam is different each year. The low value of precision and recall can be explained that the past trainee's patterns were not suitable for the target trainee's patterns. This can be expected to improve accuracy by increasing the number of trainee patterns each year.

Regarding the explanatory parameters, the most relevant activities to the academic results are the frequency of compilations and the number of similar errors. Trainees who display too frequent compilation and produce a number of similar errors are expected to receive low academic results. We can conclude that these trainees are doing programming without knowing what wrong they have done in the programming. However, there were some trainees with low academic results despite the small amount of compilations. There is a possibility that these trainees just copied their programs from other trainees. Instructor should intervene trainees those who compile too frequently in a short period of time. Also, we have observed that intervention is necessary for trainees who have not compiled for a long time. From these observation, we can conjecture that not only we can build a system that let instructors know which trainees need interventions but also we can construct a computerized automatic intervention system for trainees who need help.

We interpret the results in Table 6 and consider each group of trainees has the following features.

- Ac. result 1: A group solved the exercises in short period and produced a small number of executions. Some trainee in this group may be programming carefully, but in practice most of them are less motivated and not programming at all. This group of trainees may be waiting for their friends' solutions of the task.
- Ac. result 2: A group compiled the programs very frequently, causing the similar errors. This group of trainees may not investigate the cause of the error very well, and may just execute their program mindlessly. Because this type of trainees do not understand the cause of the error, the final exam scores also tend to be low.
- Ac. result 3: A group produced a small number of compilations and executions**.** In this group, trainees do not produce the same type of errors many times. Trainees are coding while they are investigating the cause of the error carefully. They are trying to understand every error they produced. Therefore, their final exam scores tend to be high.
- Ac. result 4: A group consumed very short period for solving the exercises. Trainees in this group solve many exercises. We observed trainees in this group can understand exercise sentences quickly and can program for solutions quickly. Therefore, their final exam scores tend to be very high.

The numbers of incorrect estimations are as follows. There were six cases where the conjectured academic result was 4 and the actual academic result was 2. Also, there were four cases where the conjectured academic result was 2 and the actual academic result was 4.

The reason why there were no trainees of academic result 5 in Table 6 may be that the number of exercises solved was small. Actually, we do not have enough training data with trainees with this highest academic result.

6. Conclusion

In this paper, we have analyzed the trainee's coding history to conjecture exercise-solving time and academic results. The objective of this estimation is to support CAs in order to assist trainees in programming exercises. We have carried out the analyses to classify trainees who perform programming exercises.

We have applied ML to construct a model that represents the coding activities of a programming class and then applied that model to the coding activities of trainees in the following year to conjecture the results from the data for the past two years. As the results, we have observed that the coincidence rate is 43.9%, and the plus or minus 1 estimation rate was 90.7%. In addition, in the analysis using four types of ML, we have found that DNN provides the most accurate estimation. However, since the values of precision and recall are low, it is necessary for us to investigate the coding history of the trainees of each year, and then improve our analysis.

As the data mining, we discovered that DNN, Random Forest, SVM and Naive Bayes can foresee the same results with sufficiently high accuracy by the same data. We clarified the results and found that we could conjecture the academic results even with a small amount of explanatory parameters.

As a future direction for research, we analyze each lesson more methodically so that we can conjecture the results at an early stage while maintaining the estimation rate high. We also plan to apply this method to other programming classes that use other programming languages than Java, such as Python.

Acknowledgment

This work was supported by Japan Society for Promotion of Science (JSPS), with the basic research program (C) (18K02910), and (A) 26245048, Grant-in-Aid for Scientific Research (KAKENHI).

References

[1] Kato, T., Kambayashi, Y., Terawaki, Y. & Kodama, Y., “Estimating Grades from Students' Behaviors in Programming Exercises Using Deep Learning”, In 2017 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, 1115-1119, 2017.

[2] Baradwaj, B. & Pal, S., “Mining Educational Data to Analyze Student’s Performance”, International Journal of Advanced Computer Science and Applications, 6, 2, 63-69, 2011.

[3] Horiguchi, S., Igaki, H., Inoue, A., Yamada M., Hoshi T. & Okada K., "Progress Management Metrics for Programming Education of HTML-based Learning Material", Journal of Information Processing Society of Japan., 53, 1, 61-71, 2012. (In Japanese)

[4] Sagisaka T. & Watanabe S., "Investigations of Beginners in Programming Course Based on Learning Strategies and Gradual Level Test, and Development of Support-Rules", Journal of Japanese Society for Information and Systems in Education., 26, 1, 5-15, 2009. (In Japanese)

[5] Yasuda K., Inoue A. & Ichimura S., "Programming Education System that can Share Problem-solving Processes between Students and Teaching Assistants", Journal of Information Processing Society of Japan., 53, 1, 81-89, 2012. (In Japanese)

[6] You, J. W., “Identifying significant indicators using LMS data to predict course achievement in online learning”, Internet and Higher Education, 29, 23-30, 2016.

[7] Kato T. & Ishikawa T., "Design and Evaluation of Support Functions of Course Management Systems for Assessing Learning Conditions in Programming Practicums", ICALT 2012, pp. 205-207, 2012.

[8] Igaki. H, Saito S., Inoue A., Nakamura R., & Kusumoto S., "Programming Process Visualization for Supporting Students in Programming Exercise", Journal of Information Processing Society of Japan, 54, 1, 2013. (In Japanese)

[9] Ryan C. & Michael C. L., "Debugging: from Novice to Expert", ACM SIGCSE Bulletin: 36, 1, 17-21, 2004.

[10] Alex G., Johan J. & Bastiaan H., "An Interactive Functional Programming Tutor", ITiCSE '12: Proceedings of the 17th. ACM, 250-255, 2012.

[11] Yamada T., Nakamichi N. & Matsui T., "Detection of Low Usability Web Pages with Logistic Regression", GN Workshop 2010, 57-62, 2010. (In Japanese)

[12] Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L., "Predicting Students Performance in Educational Data Mining", 2015 International Symposium on Educational Technology (ISET), doi:10.1109/iset.2015.33, 2015.

[13] Belachew, E. B., & Gobena, F. A., "Student Performance Prediction Model using Machine Learning Approach: The Case of Wolkite University", International Journal of Advanced Research in Computer Science and Software Engineering, 7, 2, 46-50. doi:10.23956/ijarcsse/v7i2/01219, 2017.

[14] Ahmad, F., Ismail, N. H., & Aziz, A. A., "The Prediction of Students Academic Performance using Classification Data Mining Techniques", Mathematical Sciences 9, 6415-6426. doi:10.12988/ams.2015.53289, 2015.

[15] Amrieh, E. A., Hamtini, T., & Aljarah, I., "Preprocessing and Analyzing Educational Data Set using X-API for Improving Students Performance", 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), doi:10.1109/aeect.2015.7360581, 2015.

[16] Kato T., Kambayashi Y. & Kodama Y., "Data Mining of Students' Behaviors in Programming Exercises", Smart Education and e-Learning 2016. 59, 121-133, 2016.

[17] Kato T., Kambayashi Y., Terawaki Y. & Kodama Y., " Analysis of Students' Behaviors in Programming Exercises using Deep Learning", Smart Education and e-Learning 2017. 38-47, 2017.

[18] Kamishima T., "A Survey of Recent Clustering Methods for Data Mining (Part 1): Try Clustering!", Journal of the Japanese Society for Artificial Intelligence, 18, 1, 59-65, 2003. (In Japanese)

[19] Michael R. A., "Cluster Analysis for Applications", Academic Press, 1973.

[20] Jain, A. K., Murty, M. N. & Flynn, P. J., "Data Clustering: A review", ACM Computing Surveys. 31, 3, 264-323, 1999.

[21] Uchida O., "Logistic Regression Analysis by SPSS", Ohmsha, 2011. (In Japanese)

[22] Tensor Flow, Google, https://www.tensorflow.org

[23] Ozawa Y., Kohno Y. & Takahashi T., "Generalization method with satisficing on ANN", The 29th Annual Conference of the Japanese Society for Artificial Intelligence, 29, 1-4, 2015. (In Japanese)

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 327-333 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Improving the Performance of Hadoop Framework Using Optimization Process in the Information Management

Ramachandran Ravi Sowmiyasree, Nachimuthu Maheswari*, Manickam Sivagami

School of Computing Science and Engineering,Vellore Institute of Technology, Chennai,India

ARTICLE INFO

Article history:
Received: 01 June, 2019
Accepted: 17 September, 2019
Online: 22 October, 2019

Keywords:
Bigdata
Hadoop
Performance
Map reduce
Hive
sqoop

ABSTRACT

Hadoop has certain issues that could be taken care to execute the job efficiently. These limitations are due to the locality of the data in the cluster, allocation of the jobs, scheduling of the tasks and resource allocations in Hadoop. Execution in the mapreduce remains a challenge in terms of efficiency. So, an improved Hadoop architecture that takes care of the computation time has been discussed. The improved architecture addresses the communication issues with the task trackers, inefficient clean up task, heartbeat function. Comparing with native Hadoop, the improved Hadoop reduces the total time taken for running the reducer tasks. The performance of the improved system using optimization serves better in the computation time.

1. Introduction

Big data has a huge volume, velocity and variety of information, which produces cost effective, new forms of information processing that helps in decision making, prediction and automation etc. Big data analytics is the process of analyzing large and varied data sets i.e., big data to discover hidden patterns, unknown correlations, trends, user preferences and other useful information that can help organizations to take better decisions in business. Parallel processing in data analytics has appeared as an interdisciplinary research area due to the nature and large size of data [1]. Pattern matching/mining or analysis need huge amount of complex data processing and computing [2].

Significant and reasonable allocation of resources are required to solve complicated problems [3]. It is very difficult to understand and process the data using traditional processing techniques. Big data parallel processing platforms has selected new possibilities to process the structured, semi-structured or unstructured data [4].

1.1 Problem Statement

Data localization and the resource allocation for the tasks are the challenges in the Hadoop. Effective and efficient resource allocation are the challenges in the map reduce framework. In order to implement an architecture with the data gathered from different sources of the systems as in the banking practice and various data processing supports as in the traditional global financial systems, an improvement in the computation time is required.

1.2 Motivation

Financial sectors and the banks are facing severe demands due to the growth of data processing. This happens not only from the improved regulatory requirements and an inconsistency in the data sources. The cost has to be reduced without compromising scalability and flexibility. In this scenario, the financial services industry shows tremendous interest in applying big data technologies to extract the significant value from the large amount of generated data. So, there is a requirement to improve the computation time of the process.

The proposed system addresses the issues in communicating with the task trackers, inefficient cleanup tasks, heartbeat function. Time taken for the reducers to complete the task has been reduced, which helps in better computation time.

The sections of the paper are divided as follows: Section 2 discussed some related works. In section 3, Hadoop map reduce

* Nachimuthu Maheswari, Email: maheswari.n@vit.ac.in

work flow and its performance is discussed. In section 4, the enhanced hadoop system is discussed. In section 5, the implementation phase is discussed. In section 6, the results are evaluated and discussed. In section 7, the conclusion is provided. Finally, in section 8, the future works are discussed.

2. Related Work

The improvement in Hadoop Mapreduce performance has suggested from different concerns or aspects. Many works have suggested in the improvement of the performance of mapreduce jobs and the hadoop, such as scheduling the jobs and computation time improvement. Scheduling the job and the execution time are considered as the important features of Hadoop [5] [6]. Various studies have given the information and have received the results using their ideas [7] [8]. Few people focus on the time of starting and ending process of map reduce job processing [9]. Issues in the system memory could be addressed to improve the performance of the overall system [10]. Apart from Hadoop, few studies utilize a distributed caching method to enhance the performance of Hadoop [11] [12]. Shm Streaming [13] proposes a streaming type of schema to give First In First Out queue as lock less which joins Hadoop and other programs. Hadoop spreads the redundant data into various nodes in various racks. This will be helpful in false tolerant issues. Various studies consider the improvements in data locality development for the betterment of hadoop [14]. Few, discuss on the data types for the betterment in the performance of Hadoop [15].

3. Hadoop Overview

Hadoop is an open source framework for distributed storage and processing. It gives solutions for big data processing and analysis. Hadoop has a file system that stores the data relevant to the applications. The interface is known as the Hadoop Distributed File System (HDFS). Hadoop Distributed File System shares the resources for data analysis. The major parts of Hadoop are map reduce and Hadoop Distributed File System. Other hand, moving the computing towards the data is less costlier than movement of data towards computing [16]. Hadoop allocates file system to store huge amount of data files across the nodes in the clusters.

In HDFS, the Hadoop cluster has two components, which are the masters- Name Node and the slaves- Data Nodes. In Hadoop cluster, name node is responsible for maintaining the file system. It helps in maintaining the data and sending the jobs to the corresponding data nodes responsible for the application data [17]. Job tracker works in the master node. Job Tracker takes care of the data related to the applications in the data nodes with the help of the task tracker for processing. Each task is running in the respective allotted slot in a data node, which has a fixed amount of map/reduce slots.

3.1 Mapreduce

Map reduce processes the job that splits the input data of the job into independent parts and stores the data in HDFS. During map reduce execution phase, multiple map tasks are processed in parallel. After the completion of the map tasks multiple reduce tasks are processed in parallel [18]. With respect to the applications the total number of allotted map tasks be different than that of reduce tasks. Data that are stored in HDFS and processed in map/reduce framework is the form of key and value pair. It has stored and used in the map/reduce tasks to determine the required results at the end of the job. The final results will be displayed in the key and value combination.

3.2 Mapreduce Work flow

HDFS stored the data required for the map reduce job and the data is distributed in to the blocks. Each mapper processed one block in the same time. Inside the mapper phase, the user can give the logic as per theproblem requirements. So that the map tasks runs on all the nodes in the cluster and computes the data parallelly that are stored in the blocks.

The mapper output is stored in the local disk as it is only the intermediate output. If it is written on HDFS, that will create unnecessary multiple copies as the HDFS always replicates the data. The mapper output is sorted/shuffled and submitted to the reducer. The shuffling/sorting has taken care by the internal process. Reducer phase is the next phase of processing and the user can specify his/her own statements. Input to the reducer is given by all the mappers. The final output is produced by the reducer and written on HDFS.

4. Improved Hadoop System

The focus is on transferring the relational data into HDFS using sqoop. The exported data is analysed using Hive by using Hive Query Language [19]. The data is partitioned and is fed into MapReduce for further processing. The MapReduce jobs are optimized due to the changes done in the configuration of Job in progress, Task in progress and Task tracker files of hadoop.

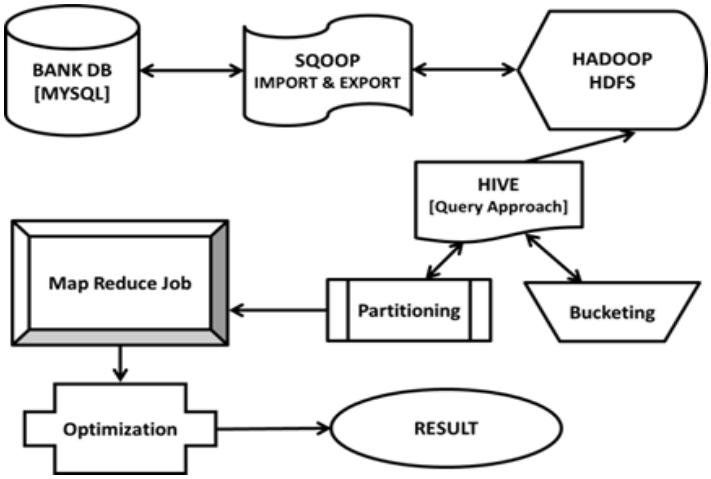


Figure.1 Diagrammatic representation of the proposed system

The bank MySql Database contains the required datasets. The data is exported to HDFS using Sqoop tool. Then Hive does the partition of the exported data, for which the results gets stored in HDFS. Then the Hive partitioned data is analysed and fetched from HDFS for MapReduce processing. The next step is Optimisation, which is like the heart of the work. Once the optimisation is over, the results are compared with that of the native hadoop to prove the working correctness of the work as

shown in Figure 1. The flow of the proposed system is explained below:

System Work flow

Step 1: Client requests for data sets. Bank data set is transferred from Mysql to HDFS using Sqoop.

Step 2: Hive extracts the data and partition the data set for map reduce programming.

Step 3: Changes performed in the Jobinprogress, Taskin progress, task tracker files.

Step 4: Name node initializes the job. Data fetched from the data node and performs the job.

Step 5: Performance of the results with respect to the computation time is compared with the native Hadoop.

4.1 Data Collection and Preprocessing

The Initial phase of the work is data collection. The article has a bank dataset that contains the account details, customer details and transaction details in MySql database.The transfer of the dataset into hadoop (HDFS), i.e, data migration is done using Sqoop tool. Sqoop supports for transferring data from relational databases to Hadoop.

Data from the three datasets (account,customer and transaction) are combined with its attributed and produced 21,215 records. The final combined dataset has the following attributes: account number, transaction type, amount, month and year.

In this phase, the dataset is fetched into hadoop (HDFS) using Sqoop Tool as in Figure 2. Using Sqoop, customized functionalities can be performed like fetching the particular column or fetching the dataset with specific condition and data can be stored in hadoop (HDFS).



Figure.2 Data preprocessing using sqoop

4.2 Data Analysis

In this phase, the dataset is analysed using HIVE tool which will be stored in hadoop (HDFS). For analyzing the dataset, HIVE uses HQL Language. Hive is chosen because it is a data warehouse by itself. Hive has the ability to work on top of an existing Hadoop cluster and provides SQL-like interface. The user can map the existing Sqoop tables to Hive and operate on them as shown in Figure3.

4.2.1 Partitioning

Hive arranges the tables into partitions. The records are partitioned based on the transaction type(with draw, transfer, deposit). Partitioning helps in dividing a table into related parts based on the values of partitioned columns as shown in Figure 4. Partitioning is useful in querying a portion of the data.

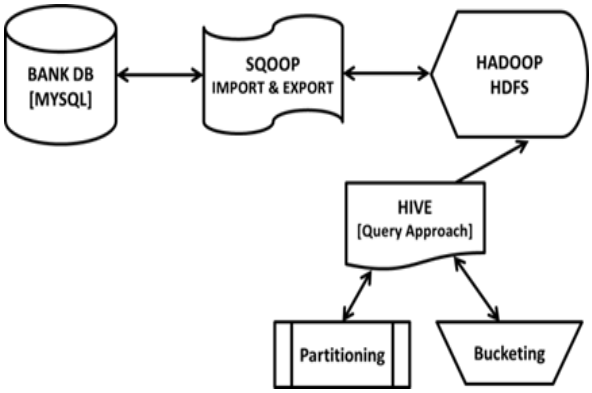


Figure.3 Data analysis using Hive

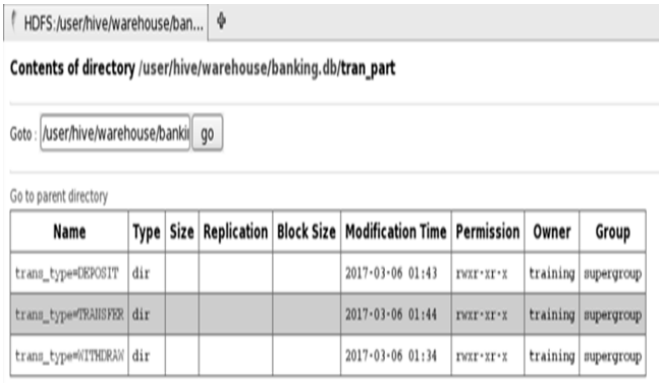


Figure.4 Partitioning the transaction data

The number of rows loaded and the time taken to load the partitioned data into HDFS is shown in Figure 5.

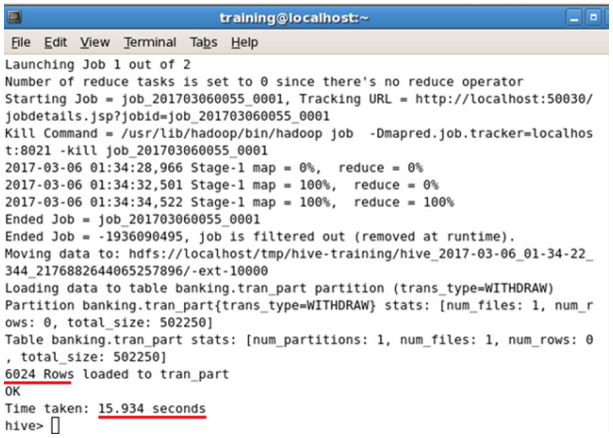


Figure.5 Loading partitioned data into HDFS

4.2.2 Bucketing

Bucketing decomposes data into manageable or equal parts as shown in Figure 6. The user can restrict the number of buckets to store the data. It provides faster query response. The number of required buckets can be given during the table creation. Loading of equal volume of data has to be done manually by programmers is a significant issue.

Choosing partitioned data over bucketed data

The size of the loaded partitioned data in the three partitions is less when compared to the size of the loaded bucketed data.

Bucketing works well when the field has high cardinality (number of possible values a field can have) and data is evenly distributed among buckets. Partitioning works best when the cardinality of the partitioning field is low.

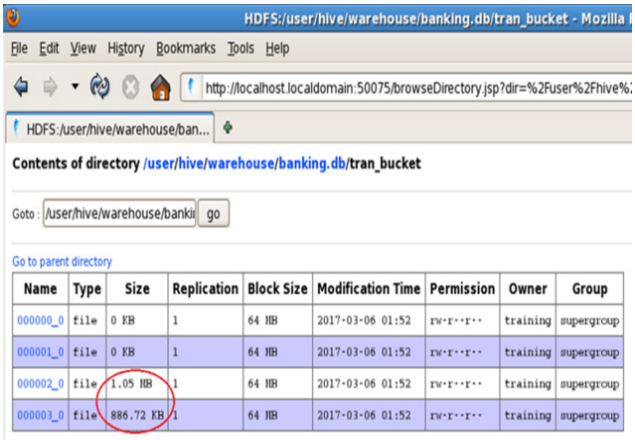


Figure.6 Bucketed data in HDFS

4.3 Data Processing

Data Processing is done using Hadoop Madpreduce. The partitioned data is fed into MapReduce for the processing. The main factor considered is the SLOTS_MILLIS_REDUCES[20]. It represents the total time taken for running the reduce tasks in milliseconds in the occupied slots. It also includes the tasks that where started speculatively as shown in Figure 7.

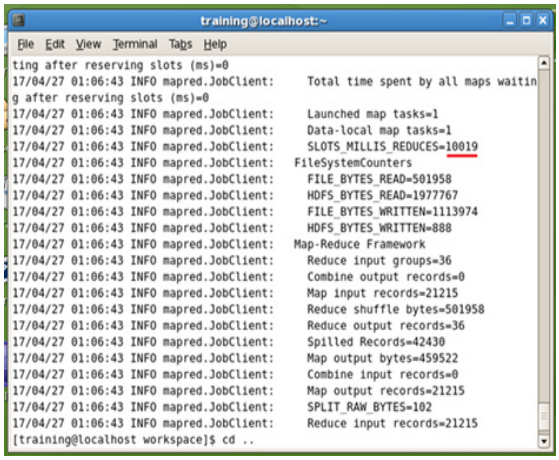


Figure.7 Running mapreduce job

4.4 Optimisation

Optimisation can be done to the native hadoop by changing the configuration factors of the system. There are three main files in hadoop[21], to which the changes are done to obtain an improved and an optimised version of hadoop. The files in hadoop to which the changes are made in order to obtain an improved version are

JobInProgress.java

JobInProgress maintains all the information in order to keep the job on the straight. It maintains its JobProfile and its updated JobStatus, along with a set of tables for doing bookkeeping of its tasks.

TaskInProgress.java

TaskInProgress (TIP) maintains all the information required for a task in the lifetime of its owning Job. A given task might be executed or re-executed, so level of indirection is needed above the running-id itself. A given TaskInProgress contains multiple taskids, zero or more of which might be executing at any one time, allowing speculative execution. A TIP allocates enough taskids to account for all the speculation and failures it will ever have to handle. TIP is dead when they are up.

TaskTracker.java

Task tracker is a process that initiates and monitors the map reduce tasks in the environment. It communicates the Job tracker for assigning the tasks and update the status.

4.4.1 JobInProgress

The changes made to the JobInProgress.java are as follows,

Contacting the tasktracker at once

In native hadoop, the task can contact the TT only after the whole task gets completed. So change is made so that, once the runJobSetupTask() works, the task can contact the TT then and there. It is beneficial as the TT is well informed of the on-going tasks and can also help to avoid speculations.

Launching Synchronised Cleanup

In native hadoop, Cleanup task is called once at the end of the task. The change made here is the introduction of Synchronised Cleanup task, which is launched before the cleanup task. It updates the last known cluster size then and there. It is highly beneficial in case of multiple tasks.

Later check of Cleanup

It checks whether the cleanup task is launched already or if the setup is not launched already. The later check is beneficial when the number of maps is zero.

4.4.2 TaskInProgress

The changes made to the TaskInProgress.java are as follows,

Skipping feature

Since completed reduces (for which the outputs go to hdfs) are not failed, this failure is noted only for completed maps, only if the particular tasked completes the particular map. However if the job is done, there is no need to manipulate completed maps. Reset the successful TaskId since no successful tasks can be concluded while running. There can be failures of tasks that are hosted on a machine that has not yet registered with restarted jobtracker. Therefore, the skipping feature recalculates the counts only if it is a genuine failure.

SetComplete function

When the TIP is complete, the other speculative subtasks will be closed when the owning tasktracker reports in and calls should close on the required object.

4.4.3 TaskTracker

The changes made to the TaskTracker.java are as follows,

Resetting heartbeat interval from the response

The name node and the data node communicate using Heartbeat messages. After a certain amount of time, if the Name Node does not receive any response from Data Node, then that particular Data Node is declared as dead. The normal heartbeat period is three seconds. If the name node is not receiving the heartbeat from a data node in ten minutes, then the name node decides the data node to be not in service and the data node's replicas are considered to be unavailable. The name node allocates the new replicas for the unavailable blocks in the other data nodes. The heartbeat message from the data node carries the information regarding the storage, amount of usage and transfer rate of data in the current state. Such details are helpful for the name nodes allocation of blocks and the decision making in load balancing.

By resetting heartbeat interval from the response, the TT knows the completed tasks. Once cleanup of the completed task is over, normal operation is resumed.

5. Implementation

5.1 Mapper Class

The map task is to process the input data. The input data is stored in the Hadoop Distributed File System (HDFS). The input file is passed to the map function and read line by line. The mapper processes the data and produces the intermediate data. The amount, transaction type, year and month are initialised according to the positions in the input data. The mapper class writes the year, month, transaction type and amount along with the sum into HDFS.

5.2 Reducer Class

This stage is the combination of the shuffle and the reduce stage. The Reducers job is to process the data that comes from the mapper. After processing, it produces a new collection of output, which will be stored in the HDFS.

5.3 Driver Class

The Driver class verifies the arguments from the command line. The input/output file details, job details are available in the command line. It assigns values for the job. The driver starts the execution from the main () method. In this method, a new configuration object and the Job are instantiated.

The job.waitForCompletion() helps to unveils the job perfectly. This driver code helps to wait for the job completion. The job status will be updated after the job completion. The true argument informs the framework to write verbose output to the controlling terminal of the job.

6. Results and Discussion

The work has done using bank data, which is analysed to identify the future trends in the sector. The improvement in the Hadoop performance is ensured by the factors such as computation time, time taken for running reduce tasks, contacting the task trackers at once, synchronised clean up, skipping feature, resetting the heartbeat interval.

6.1 Health of the file in HDFS after optimization

The fsck command in hadoop runs a HDFS file system checking utility. It is designed for reporting problems with various files. It checks health of hadoop file system. It produces a report that gives the complete health of the file system. HDFS is considered healthy if all the files have a minimum number of replicas as shown in Figure.8.

Table 1. Performance Analysis

Factors	Before Optimization(sec)	After Optimization(sec)
Job Duration	82	23
Map Time	60	3
Reduce Time	22	20

```
[training@localhost workspace]$ hadoop fsck /user/training/after_optimisation -f
iles -blocks -locations
FSCK started by training from /127.0.0.1 for path /user/training/after_optimisat
ion at Wed May 03 02:16:39 PDT 2017
/user/training/after_optimisation <dir>
/user/training/after_optimisation/_SUCCESS 0 bytes, 0 block(s):  OK

/user/training/after_optimisation/_logs <dir>
/user/training/after_optimisation/_logs/history <dir>
/user/training/after_optimisation/_logs/history/localhost_1490078156466_job_2017
03202335_0037_conf.xml 47417 bytes, 1 block(s):  OK
0. blk_-5486851598621712856_2862 len=47417 repl=1 [127.0.0.1:50010]

/user/training/after_optimisation/_logs/history/localhost_1490078156466_job_2017
03202335_0037_training_bank.jar 9391 bytes, 1 block(s):  OK
0. blk_6501123563346896786_2865 len=9391 repl=1 [127.0.0.1:50010]

/user/training/after_optimisation/part-r-00000 888 bytes, 1 block(s):  OK
0. blk_3551654415947294813_2864 len=888 repl=1 [127.0.0.1:50010]

Status: HEALTHY
 Total size:    57696 B
 Total dirs:    3
 Total files:   4
 Total blocks (validated):      3 (avg. block size 19232 B)
 Minimally replicated blocks:   3 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    1
 Average block replication:     1.0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Number of data-nodes:          1
 Number of racks:               1
FSCK ended at Wed May 03 02:16:39 PDT 2017 in 2 milliseconds


The filesystem under path '/user/training/after_optimisation' is HEALTHY
```

Figure.8 Health status of the result

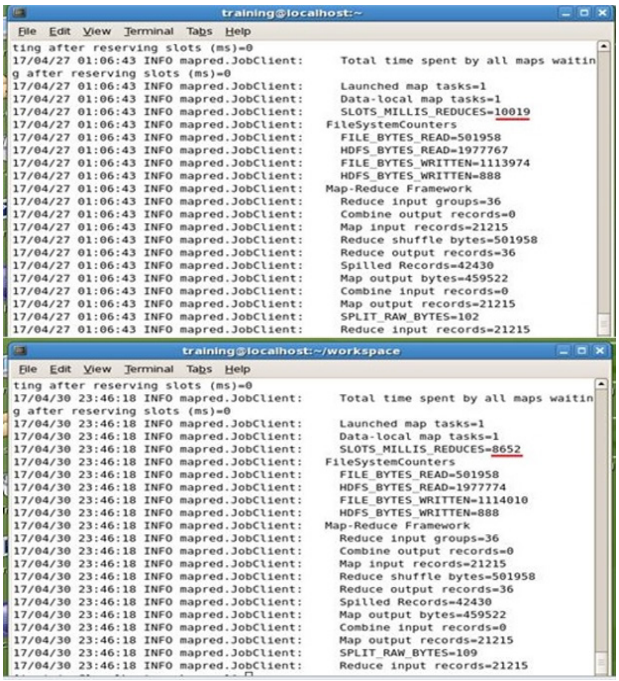


Figure.9 Performance - before and after optimization

SLOTS MILLIS REDUCES is the total time spent by all reduces in the occupied slots. Difference in the time taken for running reduces tasks before and after optimisation is noted to be 1367ms as shown in Figure 9. After optimization it takes less time.

6.2 Performance Analysis

The performance of the job for the same input is analysed before and after the optimisation is show in Table 1. Job duration, Map task time and Reduce task time are taken as the factors for the performance analysis. The time taken to complete the job is less after optimization is shown in Figure 10. The performance has evidently improved after the optimisation.

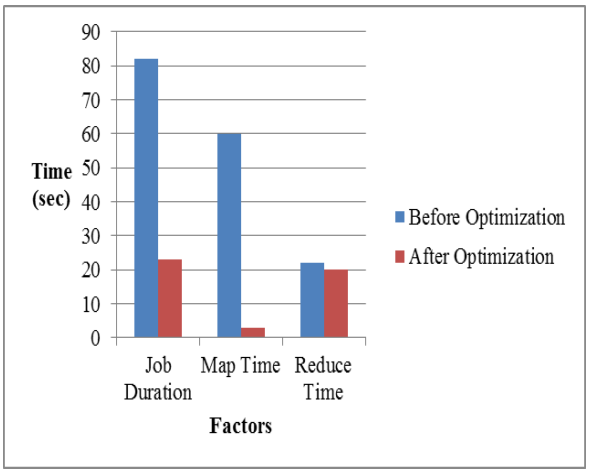


Figure.10 Performance Evaluation

7. Conclusion

In this article, an improved hadoop framework is presented. It includes analysing the internal working of the Hadoop MapReduce job, so that changes are done to the job in progress, task in progress and task tracker files. In H2Hadoop [19] the work was concentrated on the DNA dataset only and in the improved hadoop [22] they have considered the text data and the performance time was not reasonably improved. The changes performed to optimise the system in this article are in contact with the job tracker at once, launching synchronised cleanup function, cleanup later checkup, skipping feature, resetting the heartbeat interval from the response, avoiding speculative tasks resulted in an improved hadoop framework. The performance of the improved system using optimization is better in the computation time.

8. Future Work

Performance factors

The other performance factors of File System Counters, MapReduce Framework can be modified to obtain a more optimised hadoop system.

Tool for better performance

The result can be analysed using Spark to obtain a faster performance than Hadoop, as spark runs in-memory on the cluster, and it is not tied to Hadoop MapReduce paradigm. This makes repeated access to the same data much faster. Spark can run as a standalone or on top of Hadoop YARN, where it can read data directly from HDFS.

Security issues in Hadoop can be addressed at various levels, including but not limited to file system, networks, scheduling, load balancing, concurrency control, and databases.

References

[1] MMing, M., G. Jing, and C. Jun-jie, "Blast-Parallel: The parallelizing implementation of sequence alignment algorithms based on Hadoop platform", *In: Proc. of International Conf. on Biomedical Engineering and Informatics (BMEI)*, IEEE, pp. 465-470, 2013. doi**:** 10.1109/BMEI.2013.6746988

[2] cC. Schatz, B. Langmead and S. L. Salzberg, "Cloud computing and the DNA data race", *Journal of Nature biotechnology*, Vol.*28, No.*7, p.691, 2010. doi:10.1038/nbt0710-691.

[3] E. E. Schadt, et al., "Computational solutions to large-scale data management and analysis", *Journal of Nature Reviews Genetics*, Vol.11, No.9, pp. 647-657, 2010. https://doi.org/10.1038/nrg2857.

[4] Marx, *Biology: The big challenges of big data*. Nature Publishing, 498(7453): pp. 255-260, 2013. https://doi.org/10.1038/498255a

[5] N. Tiwari, S. Sarkar, U. Bellur and M. Indrawan, "Classification framework of MapReduce scheduling algorithms", *International Journal of ACM Computing Surveys (CSUR)*, Vol.47, No.3, p.49, 2015.doi:10.1145/2693315

[6] M. Zaharia, et al., *Job scheduling for multi- user mapreduce clusters*, Tech. Rep. UCB/EECS-2009-55, EECS Department, University of California, Berkeley, 2009. doi =10.1.1.649.3097

[7] R. Gu, et al., "SHadoop: Improving MapReduce performance by optimizing job execution mechanism in Hadoop clusters", *Journal of Parallel and Distributed Computing*, Vol.74, No.3, pp. 2166-2179, 2014. https://doi.org/10/1016/j.jpdc.2013.10.003

[8] Q. Chen, C. Liu, and Z. Xiao, "Improving MapReduce performance using smart speculative execution strategy", *Journal of IEEE Transactions on Computers*, Vol.*63, No.*4, pp.954-967, 2014. doi:10.1109/TC.2013.15

[9] Y. Jinshuang, et al., "Performance Optimization for Short MapReduce Job Execution in Hadoop", In: *Proc. of International Conf. on Cloud and Green Computing (CGC)*, IEEE, pp. 688-694, 2012. doi:10.1109/CGC.2012.40

[10] Apache, Centralized Cache Management in HDFS. Update date 2014.https://hadoop.apache.org

[11] S. Zhang, et al., "Accelerating MapReduce with distributed memory cache", In: *Proc. of International Conf. On Parallel and Distributed Systems (ICPADS),* IEEE, pp.472-478, 2009. doi:10.1109/ICPADS.2009.88

[12] J. Zhang, G. Wu, X. Hu and X. Wu, "A Distributed Cache for Hadoop Distributed File System in Real-Time Cloud Services", *In: Proc. of International Conf. On Grid Computing,* IEEE Computer Society, pp.12-21, 2012. doi:10.1109/Grid.2012.17

[13] L. Longbin, et al., "ShmStreaming: A Shared Memory Approach for Improving Hadoop Streaming Performance", In: *Proc. of International Conf. On* Advanced Information Networking and Applications (AINA), IEEE, pp.137-144, 2013. doi:10.1109/AINA.2013.90

[14] B. Palanisamy, et al., "Purlieus: locality-aware resource allocation for MapReduce in a cloud.", *In: Proc. of International Conf. On High Performance Computing, Networking, Storage and Analysis*, ACM, pp.1-11, 2011. doi:10.1145/2063384.2063462

[15] H. Alshammari, J. Lee and H. Bajwa, "H2Hadoop: Improving Hadoop Performance Using the Metadata of Related Jobs," in *IEEE Transactions on Cloud Computing*, vol. 6, no. 4, pp. 1031-1040, 2016.doi:10.1109/TCC.2016.2535261

[16] Hammoud, and M. F. Sakr, "Locality-aware reduce task scheduling for MapReduce", *In: International Conf. on Cloud Computing Technology and Science (CloudCom),* IEEE, pp. 570-576, 2011. doi:10.1109/CloudCom.2011.87

[17] J. B. Buck, et al., "SciHadoop: Array-based query processing in Hadoop", *In: Proc. of International Conf. on High Performance Computing, Networking, Storage and Analysis (SC)*, 2011. doi:10.1145/2063384.2063473

[18] Holmes, *Hadoop in Practice*, Manning Publications Co., 2012.

[19] Edward, W. Dean and R. Jason, *Programming hive*, O'Reilly Publications, 2012.

[20] Y. Xiao and B. Hong, "Bi-Hadoop: Extending Hadoop to Improve Support for Binary-Input Applications", In: *Proc. of International Symposium On* Cluster, Cloud and Grid Computing (CCGrid), IEEE/ACM, pp.245-252, 2013. doi:10.1109/CCGrid.2013.56

[21] https://hadoop.apache.org/docs/r2.7.1/api/org/apache/hadoop/mapreduce/JobCounter.html

[22] Thanekar, Sachin & Bagwan, A. & Subrahmanyam, K..,Improving Hadoop performance by enhancing name node capabilities.Fronteiras. 6. 1-18.,2016. doi:10.21664/2238-8869.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 328-346 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Overview on 5G Radio Frequency Energy Harvesting

Sanae El Hassani*[1], Hind El Hassani[2], Noureddine Boutammachte[3]

[1] *LTI Laboratory, ENSA El Jadida, Chouaib Doukkali University, 24 000, Morocco*

[2] *LISA Laboratory, ENSA Fès, Sidi Mohamed Ben Abdellah University, 30 000, Morocco*

[3] *3ER Team, ENSAM Meknès, Moulay Ismail University, 50 000, Morocco*

ARTICLEINFO

Article history:
Received: 21 May, 2019
Accepted: 02 August, 2019
Online: 19 August, 2019

Keywords:
Energy Harvesting
5G
Green technologies
SWIPT
Rectenna

ABSTRACT

The current industrial landscape is becoming increasingly aware of the need to optimize energy use and management for all domains, including telecommunications. Among others, RF Energy Harvesting is a promising technique for 5G systems as an alternative to traditional energy supply sources. This paper presents 5G landscape and ecosystem. It positions RF-Energy Harvesting in 5G context and reviews its techniques. It also considers constraints and research trends for different aspects needed to make RF energy harvesting ready for deployment along with 5G enabling technologies.

1 Introduction

This paper is an extension of the overview originally presented in 2017 about *RF Energy Harvesting For 5G* at the International Renewable and Sustainable Energy Conference (IRSEC) [1]. In fact, due to the rapid evolution of new technologies, the world faces new challenges concerning Energy Efficiency (EE). The transformation in users habits, driven by the installation of new services and the creation of new needs, especially in telecommunication, resulted in a significant rising in volume demand and requirements diversity. These transformations are leading the development of the fifth Generation (5G) of mobile systems, expected by 2020 [2], and research is being conducted to issue this new generation. In addition to the growing number of subscribers, new applications as for smart cities, smart grids, Internet of Things (IoT) and Machine-2-Machine (M2M) communications imply an increasing number of sensors, and thus the need of reviewing the system design. The sophisticated power hungry future handsets, the very large number of simpler sensors, and the dense network architecture necessary to respond to the growing needs are all factors to make energy management an urgent issue for future systems. EE appears thus as a major development key of 5G systems, as pointed in many 5G visions by telecommunication industry actors [3, 4, 5]. In order to enhance the energy consumption and management, techniques like Radio-Frequency Energy Harvesting (RF-EH), virtualized network functionalities, cloud technologies, etc. are proposed.

Energy aspects in 5G systems have attracted a lot of attention in the last years, and many studies, reviews and tutorials have considered these aspects. For example, in [6], authors provide a survey on energy-efficient wireless communications, with a discussion on the most relevant challenges at that time. General energy efficient technologies are discussed iwith no specific interest in EH solutions. Authors in [7] give an overview of EH techniques, in particular on EH power management policies optimizations, with specific stress on Wireless Sensor Networks (WSNs) and on ambient RF harvesting. In their survey [8], authors discuss 5G green network EE and promising technologies and architecture for its optimizations. Among these techniques, EH is discussed as well as other key enabling techniques, but specific EH integration is not detailed. The authors in [9] also consider EH from a circuit design point of view, particularly for embedded systems, and authors in [10] research RF-EH for IoT from an implementation perspective, including the design of antennas, rectifiers, and matching networks.

This paper aims to give a general review of EH in 5G systems, which includes both ambient and dedicated EH. In addition, considering the imminent commercial launching of 5G, and the ongoing standardization of 5G New Radio (5GNR) [11], an updated overview of RF-EH techniques is needed taking into account the standardization landscape. To this end, 5G requirements and use cases, as well as 5G ecosystem and some of the key 5G enabling technologies, are

*Corresponding Author: Route d'Azzemour Nationale N1 ELHAOUZIA BP 1166 El Jadida 24002 Morocco, (+212) 5 23 34 48 22 & elhassani.s@ucd.ac.ma

presented in the first section. In addition, a special subsection is dedicated to IoT, as one of the most RF-EH appealing 5G enabled use cases. The second section discusses EE in 5G networks from both network and terminal sides, while RF-EH implementation challenges and candidate solutions are reviewed in the the third section. The fourth section presents RF-EH in 5G ecosystem, giving a literature review of combinations between RF-EH and some of the most important 5G key enabling technologies. A conclusion of the paper is then given.

2 5G general technological landscape

2.1 5G Requirements

Compared to current mobile generation, 5G mobile data rate volume per area is expected to be multiplied by 1000, and user data rate as well as the number of connected devices are expected to be multiplied by 10 to 100. Battery life should also be 10 times longer, and latency 5 times shorter. 5G requirements are summarized in Figure 1, and the International Telecommunication Union (ITU), which has established an R&D programme worldwide on International Mobile Telecommunications for 2020 and beyond values(IMT-2020), has established a 5G framework as illustrated in Table 1 [12].

Table 1: 5G requirements

Requirements	Desired Value
Data Rate	1 to 10 Gbps
Data Volume	9 GB/h (busy period) and 500GB/month/user
Latency	Less than 5 ms
Battery Life	One decade
Connected devices	300000 devices/AP
Reliability	99.999%

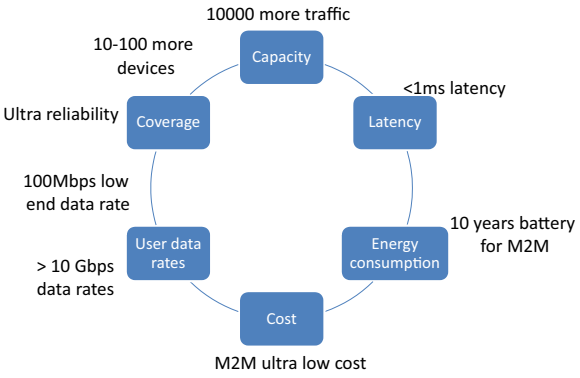


Figure 1: 5G key requirements summary (Source [13])

5G systems are meant to manage at the same time very low cost devices, such as sensors, and advanced ones, such as smartphones and tablets. For example, simple devices may choose not to use large bands. Thus, in the same system, and in some frequency bands, smart devices will use a band of 100 MHz while sensors only 10 MHz. Note that 5G will have to manage much denser networks than today, around 1 million connections/km^2, and the more predominant design parameter will thus be the cost/area.

2.2 5G use cases

Traditional as well as new use cases are addressed in 5G, and this implies new challenges. The corresponding requirements are specified by several organizations and research actions.

Each use case has its own requirements, such as throughput, delay, reliability, etc. and consequently uses different techniques. Hence, a wide range of devices, such as smartphone, wearable, MTC, etc., are being all expected to use 5G systems in a diverse heterogeneous environment. A categorization example of use cases has been developed by the Next Generation Mobile Networks (NGMN) Alliance through representative examples, which leads to 8 families of use cases [14], illustrated in Figure 2.

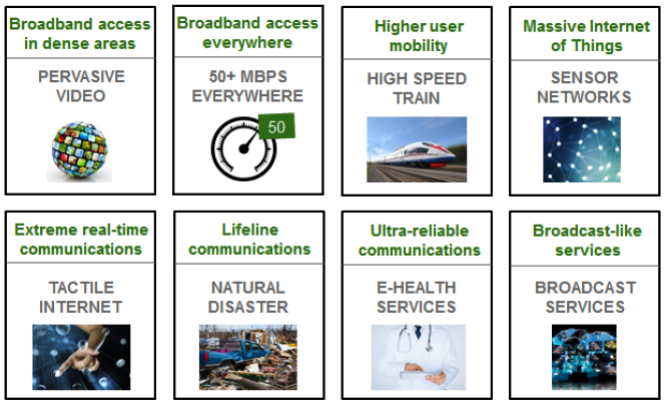


Figure 2: Families of use cases identified by NGMN (Source[14])

For example, smart cities applications including metering, city or building lights management, environment monitoring, and vehicle traffic control can be deployed as WSNs. These services, when aggregated, form a high density devices environment with heterogeneous technologies that need to be managed in a common communication and interworking framework, where very low cost devices with very long battery life may exist. 5G may provide platform for such deployments. Another example of 5G use cases is the mobile Video Surveillance, which is expected to be deployed almost everywhere. This means an automated analysis of the video footage with constraints, though relatively reduced for battery life, but high for reliability and security of the network as well as performance and interactivity with remote systems.

Note that the 3rd Generation Partnership Project (3GPP) issued and approved four Technical Reports on 2016, which outlined new Services and Markets Technology Enablers (SMARTER) [15], and contained more than 70 different use cases categorized into different groups :

- Massive Internet of Things, particularly relevant to the new vertical services, such as smart home and city, smart utilities, e-Health, and smart wearables.
- Critical Communications, for which there is more need for latency, reliability, and availability. Use case

examples are industrial control applications and tactile Internet.

- Enhanced Mobile Broadband, including use case families related to higher data rates, higher density, deployment and coverage, higher user mobility, devices with highly variable user data rates, fixed mobile convergence, and small-cell deployment.

- Network Operation, addressing functional system requirements, including for example flexible functions and capabilities, new value creation, migration and interworking, optimizations and enhancements, and security.

Example of projects exploring EE and EH in 5G systems is Sustainable Cellular Networks Harvesting Ambient Energy (SCAVENGE) [16], which is a project funded from the European Union's Horizon 2020. It defines among others an architecture framework for EH networks for Core Network (CN) and Radio Access Network (RAN).

2.3 5G ecosystem

5G systems cover a large span of use cases, with different data rates, latencies, reliabilities, densities, required powers, etc. This implies many tradeoffs, and a highly scalable architecture. As described in [17], the new 5G ecosystem is complex and involves many participants, including manufacturers for components, chipsets, network equipments and wireless devices, in addition to mobile networks operators, application developers and even users themselves, all overlaid by different standardization entities. As stated in [18], the first mover of 5G is the wireless part as it establishes foundations for the feasibility of final products and services. Many connectivity options are considered in 3GPP, which allows several deployment alternatives. However, as highlighted in [18] the focus for initial deployments is on non-standalone NR and standalone NR, as the deployment of several options may make the whole 5G ecosystem too complex. The simplification of 5G through reduction of deployment options is also adopted by other industry actors, such as Ericsson [19].

2.4 Examples of 5G key enabling technologies

5G systems require an important change in the cellular architecture as well as key wireless technologies. The requirement for massive network capacity along with the increased user experience suggest research in numerous dimensions. In particular, the RAN appears as one of the biggest challenges. For example, the increase in bandwidth demand implies the optimization of current systems spectrum use, unlicensed spectrum use and the allocation of new bands, starting with the move to millimeter wave (mmWave) bands. The optimization of spectrum use also means that, among others, new waveforms and advanced Multiple Input Multiple-Output (MIMO) will be used. Extreme densification and offloading are also key elements to enhance capacity, which implies an urgent need for virtualization and softwarization of the network. As presented in [4], the Figure 3 illustrates these needs in three distinct dimensions.

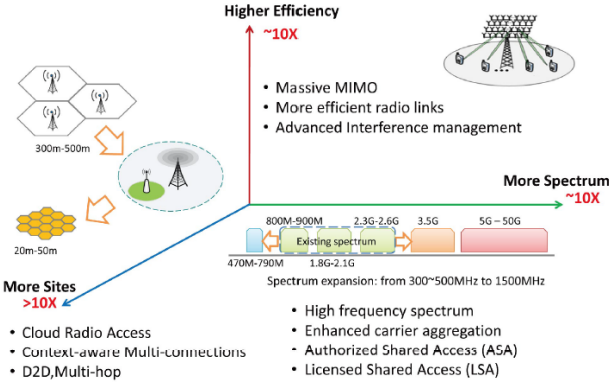


Figure 3: 5G technological research in three different dimensions (Source: [4])

Examples of key 5G enabling technologies are given in the remainder of this subsection.

2.4.1 Spectrum allocation

5G is a convergence of previous and new systems, and its spectrum is expected to be a combination of allocated and new bands. Some of the spectrum below 6GHz is thus being re-purposed for use with newer technologies; it is the primary 5G band, as decided in WRC 2015. Complementary 5G bands above 6GHz, are expected to be allocated in the 2019 World Radiocommunication Conference (WRC 2019) [20] (October). Illustration of these bands is given in Figure 4. On the The 3rd Generation Partnership Project (3GPP) Release-15, the New Radio-Unlicensed spectrum (NR-U) work item supports both the existing 5GHz unlicensed band and the new "greenfield" 6GHz unlicensed band. Other unlicensed and shared spectrum bands, including mmWave, can be expected in coming releases [11], even though it is not as mature as current frequency bands. In fact, mmWave has rather hostile propagation characteristics, as well as a current comparatively high equipment cost, which has limited its use so far. However, 5G mmWave mobile technology have been already demonstrated, as first by Samsung in May 2013 [21], and later by others, as for Huawei [22]. A more accomplished 5G pilot system was demonstrated in 2018 Winter Olympics [23].

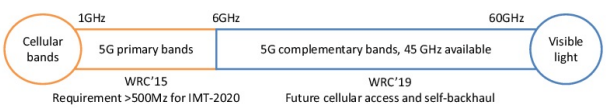


Figure 4: WRC-15 and WRC-19 5G spectrum allocation

Integration of many bands for the same transmission can be made through Cognitive Radio (CR), motivated by the underutilization of many frequency bands. In fact, CR allows dynamic allocation of underutilized resources, after their sensing, in an opportunistic way without altering the other users, and thus to expand the spectrum of a cellular networks on demand and at a relatively low cost [24, 25].

2.4.2 Carrier Aggregation (CA)

As 5G is a convergence of many systems, the 5G radio access uses simultaneously many technologies, with possibly different frequency bands. This may enable a more effective use of the crowded and fragmented spectrum, a simultaneous use of different Radio Access Technologies (RATs) and more flexibility of resource allocation. However, this implies also the need to aggregate different carriers. CA is already used in Long Term Evolution (LTE) since 3GPP Release 10 for inter and intra LTE-LTE contiguous carriers aggregation. Release 11 presented Intra-band non-contiguous CA, and Release 12 presented CA for Frequency Division Duplex (FDD) and Time Division Duplex (TDD). Release 13 presented CA for 5 GHz WiFi bands, called Licensed Assisted Access, and Release 14, which starts 5G standardization, defines more cases following the 5G adopted bands [26]. Some issues should be considered for Internet Protocol (IP) configuration, interference coordination and time synchronization. Moreover, resource scheduling over the aggregated bands should be optimized and may use advanced tools, such as game theory [27].

2.4.3 Massive MIMO

Many techniques used to enhance spectrum efficiency are surviving in 5G, while other new techniques are proposed. MIMO antennas techniques is already used for LTE and LTE-Advanced (LTE-A), and 5G proposes massive MIMO as one of its most important enabling techniques. In fact, thanks to multiple antennas at the Base Station (BS), the energy in downlink can be focused to form beams towards the User Equipment (UE). In 5G, a large number of antennas, i.e. massive MIMO, can enable the serving of many users in an accurate way. Moreover, when combined with other 5G techniques, like small cells and mmWave, it provides significant capacity, as required for 5G systems [28]. Note that massive MIMO can also be used for backhauling [29].

2.4.4 Channel modeling

In 5G systems, the expected bands to be used, especially above 6GHz, are not addressed by current models, and accurate models are needed to enable many of the physical layer techniques. Thus, research is already conducted for mmWave characterization in several environments [30]. As indicated in [31], the new channel model extends the existing 3GPP 3D channel model, and handles frequencies up to 100 GHz in several mobility scenarios. In this optic, 3GPP RAN Meeting for example was held in Busan Korea in June 2016 and approved the first standard for the mobile broadband 5G high-frequency ($6 - 100$ GHz) channel model.

2.4.5 Signal waveforms

Multicarrier modulation, particularly Orthogonal Frequency Division Multiplexing (OFDM), is adopted in LTE and enables good spectrum efficiency. However, it still presents some drawbacks, as it inserts a Cyclic Prefix (CP) that reduces the efficiency, in addition to its large sidelobes, and thus considerable Out-Of-Band (OOB) emission, resulting from the rectangular pulse shaping,as well as its high Peak to Average Power Ratio (PAPR). Most popular candidates waveform for 5G, in addition to OFDM, include :

- Filtered OFDM, which removes OOB emissions by filtering the signal [32].
- Filter Bank based MultiCarrier-Offest Quadrature Amplitude Modulation (FBMC-OQAM), which is a modified OFDM requiring no CP and thus enhancing spectrum efficiency [33]. Sub-carriers are filtered at both transmitter and receiver sides, and this reduces OOB and makes the signal more robust against Doppler effect, Inter-Symbol Interference (ISI), synchronization imperfections, spectrum fragmentation, etc. This also results in self-equalization, which can reduce the number of unused subcarriers, and blind channel tracking, reducing pilot contamination and thus enhancing massive MIMO functioning.
- Universal Filtered Multi Carrier (UFMC), which filters subbands rather than subcarriers [34]. This gives more flexibility to the design. The use of a prefix in this case depends on filters design.
- Generalized Frequency Division Multiplexing (GFDM), which also provides a flexible structure via adjustable pulse shaping and subcarrier filtering. In addition, data symbols are grouped both in time and frequency domains. However, filtering flexibility implies self-created interference, which is dealt with using iterative interference cancellation at the receiver [35].

2.4.6 Multiple access

The multiple access technique used in 4G is OFDMA, which may survive for 5G. However, massive MIMO in a dense network context, multi-user MIMO (MU-MIMO) can be enabled using Spatial Division Multiple Access (SDMA). For higher capacity, Non-Orthogonal Multiple Access (NOMA) schemes are also proposed for 5G, as it allows more than one user to share the same subband without coding or spreading redundancy [36, 37, 38]. For this scheme, users are superposed in time and frequency, and have different powers following their channel conditions. Differentiation of the users is then performed using Successive Interference Cancellation (SIC), which uses joint processing combined with subcarrier and power allocation algorithm [39, 40]. An illustration of NOMA mechanism is given in Figure. 5. So far, 3GPP NR has specified flexible technology framework that can be tuned to enable a wide range of 5G scenarios, and more specifications are expected for Release 16 [41].

2.4.7 Ultra Dense Network (UDN)

5G brings a substantial network architecture redesign, particularly through densification, in order to support the expected heavy traffic. Networks densification has been used since the second generation (2G) for capacity scaling, but the ultra densification for future networks can provide an extreme and user-centric reuse of system bandwidth on the spatial domain, and improve propagation conditions by reducing the distance between the end user and the BS. It implies

an increased spatial reuse of system resources, a large node density and irregular deployments. Challenging issues about UDN include among others interference management, heterogeneous multi-RAT optimization, the possibly high cost of a dense infrastructure, the cell-edge effect, and the backhauling.

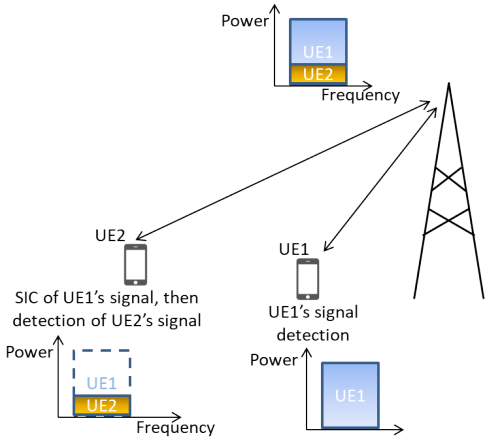


Figure 5: Downlink NOMA in a simple scenario with one BS and two UEs

2.4.8 Cloud-Based Radio Access Network (C-RAN)

As network densification is a key 5G networks enabler, it is expected that BSs will have to handle a heavy radio processing. A suggested solution in this case is the Cloud-based RAN (C-RAN). In fact C-RAN dissociates the Base-Band processing Unit (BBU) from the Remote Radio Heads (RRHs); For resource optimization, the BBU part is moved to a pool, or to the cloud, which is a virtualized cluster connected to the core network through the backhaul, and the transmission of the radio signal is performed by RRHs based on computed signals from the cloud. On the otehr hand, RRH is connected with the cloud BBU pool via a fronthaul.

RRHs are cheaper and simpler if cloud based computations are used. This allows thus more scalability and densification opportunities. It can also reduce the delay for inter-cell coordination and permits more joint optimization of processing, e.g. for cooperative interference management, scheduling, Coordinated Multi-Point, etc. [42].

C-RANs can also be self-organizing; intelligent smart cells are deployed in the area covered by a macro BS, and can enable advanced clustering and coordination schemes to enhance mobility and handovers. In addition, C-RAN in Heterogenous context lead to H-CRAN, where different types of nodes, including low power and high power nodes, use the same computing pools. The H-CRAN is illustrated in Fig. 6.

2.4.9 Network architecture

5G cellular networks are expected to be multi-RAT and multi-layer. Considering emerging 5G use cases, current network architecture is relatively incompetent to address 5G requirements. Thus, one of the central objectives in 5G architecture is network flexibility. Logical core and RAN have to be split to allow their evolution and to ease their deployment. To this end, the European Telecommunications Standards Institute (ETSI) adopted the separation of User Plane (UP) functions and Control Plane (CP) functions, and modularization of the function design, e.g. to enable flexible and efficient network slicing [44]. Similarly, a 5G architecture made of three layers and a management entity is presented by NGMN [45]. It is based on a central cloud and typically comprises multiple data centres which may be very distant, and connected to each other or to the central cloud or edge clouds. Figure 7 illustrates this topology as viewed by the project 5G NOvel Radio Multiservice Adaptative network architecture (5G NORMA), which is one of the 5G Public Private Partnership (5G PPP) projects under the Horizon 2020 framework [45].

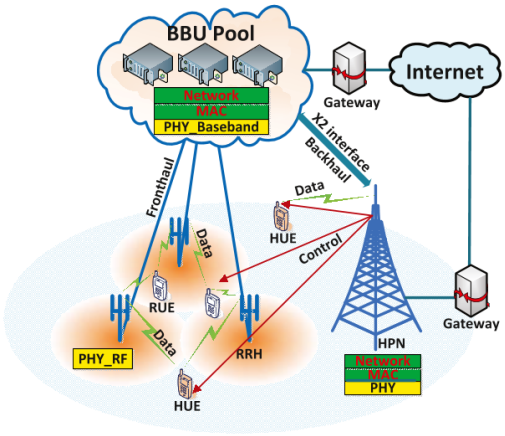


Figure 6: H-CRAN : Heterogeneous Cloud Radio Access Networks (Source [43])

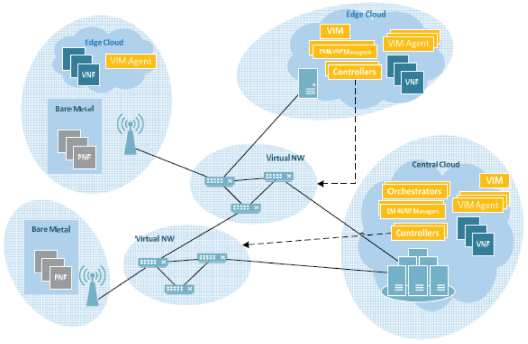


Figure 7: 5G NORMA topology view on architecture

As suggested in [46], in order to avoid redundancy, low layer functionalities are insured by radio logical network while a network cloud provides all higher layer functionalities. In addition, functions can be dynamically deployed in the network cloud through Software Defined Networks/Network Functions Virtualization (SDN/NFV).

In addition to layers, 5G uses network slicing to deliver services with different performance characteristics. It is a network technique, initially for the core network but extended to an End-to-End (E2E) concept including also RAN part. For example, massive IoT service, connecting fixed sensors that measure humidity, temperature, etc. to mobile networks, does not require advanced mobility features, critical in serving mobile phones. In contrary, a mission-critical IoT service, like autonomous driving or remote controlled

robots, requires a very low E2E latency. To avoid building a dedicated network for each service, multiple logical networks over a single physical network can be insured through network slicing. The physical network is sliced up into different virtual E2E slices, with specific network functions and RAT parameters, following the requirements of the use case or business model. Note that network slicing also limits error propagation between slices.

2.4.10 NFV/SDN

An important trend for 5G networks is virtualization, as it enables resource sharing. In addition, hardware and software have to be dissociated, and network functions to be moved towards software. Assuming separation between control and data, NFV is an important 5G architecture driver. It provides network adaptability and scalability [44], boosted by automated mechanisms; simple and fast operations will be needed to deploy new network features, and resources can be easily shared through the abstraction of physical resources to a number of virtual resources. An illustration of wireless network virtualization is given Figure 8.

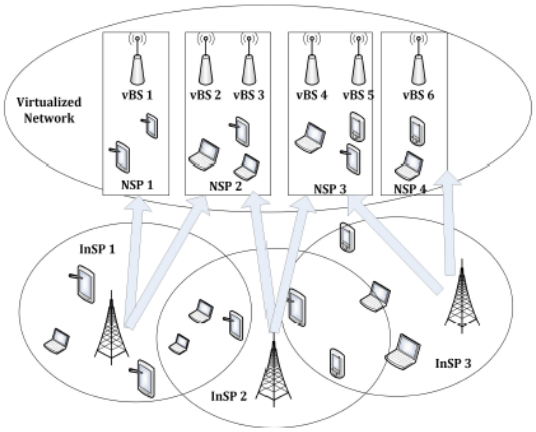


Figure 8: An example of wireless network virtualization (Source [47])

NFV uses a virtualized commercial server to set up network function software in virtual machines. RAN works thus as edge cloud while core works as core cloud. Connectivity among virtual machines located in edge and core clouds is insured by SDN. This makes thus the network architecture ultra-flexible; NFV processes network functions in the cloud, and SDN creates and exposes private virtual networks, linking thus the cloud back to the network. SDN enables the orchestration of the NFV physical and virtual infrastructure resources, through supporting provisioning and configuration of network connectivity and bandwidth.

2.4.11 Mobile Edge Computing

The diversity of 5G applications and the density of the network may lead to congestion, which may be solved by Mobile Edge Computing (MEC), mainly standardized by ETSI, 3GPP and ITU-T. MEC provides Information Technologies (IT) and cloud computing functions in the RAN, in close proximity to mobile subscribers. This enhances responsiveness to content, services and application requests, and enables thus ultra-low latency, high bandwidth radio network information.

MEC introduces new processing capabilities in the BS for new applications, with a new functions separation and a new interface between BBU and Remote Radio Unit (RRU). This is linked to C-RAN as C-RAN implements baseband processing capabilities in distributed locations. A demonstration for a 5G MEC proof-of-concept was conducted by Ericsson and Vodafone [48].

Note that MEC is currently deployed in LTE networks as an add-on to offer services in the edge. However, 5G system specifications provide new functionalities for edge computing [44].

2.4.12 Device to Device communications (D2D)

High-density expected for 5G networks can also be solved by D2D communication. In fact, unlike voice-centric systems, where no spatial proximity is assumed between communication parties, some data communication cases may imply collocated devices which try to wirelessly share content. In this case, it can be more efficient to use proximity-based D2D communication, which may also enhance capacity and coverage when devices act as transmission relays. Note that D2D is already used in 4G, and presents some challenges for 5G. Figure 9 presents typical downlink D2D scenario.

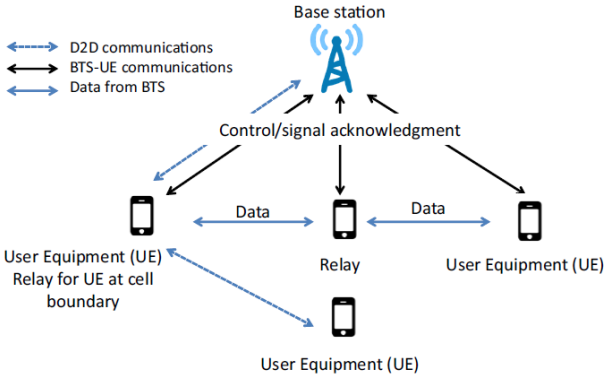


Figure 9: D2D typical scenario (Source [49])

2.5 IoT applications landscape

2.5.1 IoT characteristics

In IoT applications, a large variety of devices is implied, from small devices to big data centres. The devices can be self-empowered, battery-empowered or connected to the grid. For IoT specific applications, most devices, in number, are battery-empowered or self empowered, i.e. using EH.

IoT is distinct from WSN as it has specific characteristics. In particular, IoT energy sources have the following characteristics [50]:

- Scalability, as IoT often imply large deployments in possibly variable locations with different accessibilities. In this case, batteries are not ideal.
- Maintenance-free, as the deployments often imply a large number of devices. Wired empowering would be thus impossible and battery-based one would be costly.

- Mobility support, as for many applications, the devices are mobile.
- Long life-time, as in some applications, and because of difficult maintenance, devices are supposed to work autonomously up to several decades.
- Flexibility, as capacity and size of energy sources may vary following the IoT application.
- Low-cost, as IoT are usually deployed in a large scale, which means that a small difference in the unit cost may limit affordability of the application.
- Sustainability, as large scale deployments may imply significant power consumption, requiring thus sustainable power solutions.
- Environment-friendly, as IoT popularity may imply serious environmental impact, especialy if disposable batteries are used for devices powering.

Currently, solutions meeting these characteristics with very low power consumption are standardized, but having energy autonomy is even more attractive. EH-enabled nodes techniques, as well as batteries storing harvested energy and optimized low power circuits, have to be thus combined with power saving mechanisms and network resource management. Note that for traditional wireless networks, many factors are important in the design. For example, the acceptable latency of the transmission can lay from some milliseconds to some days. Transmission rate is also generally low, mostly some kilobytes per second. Similarly, the distance between nodes is critical, as it contributes in the definition of the transmission power. Depending on the application, the distance can range from centimeters to kilometers, and it can vary over time when mobility of the nodes is considered.

2.5.2 IoT standards and solutions

RF-EH globally provides little energy compared to other energy sources. Thus, it is more suitable for low power IoT applications. These are hence one of the most interesting motivations for RF-EH technologies development. It is also one of the vectors in international standardization bodies. Even though solutions exist since many years for low-power applications, including Wi-Fi, Bluetooth, and ZigBee, these are unsuitable for long-range cellular and M2M, which are essential for IoT. Low-Power Wireless Area Networks (LPWANs) aim to respond to such requirements. In fact, in order to achieve a low-power long-range communication, high receiver sensitivity is needed, as fine as to −130 dBm against −90 to −110 dBm in traditional technologies. This means that a higher energy per bit is needed, and thus that the modulation rate should be slower.

One of the major issues for M2M communications is the low-power long-range communication. LPWAN technologies, in order to achieve a long range capability, need to use high receiver sensitivities, up to −130 dBm compared with the −90 to −110 dBm in many traditional wireless technologies. This implies a higher energy per bit and thus a slower modulation rate. The two most popular solution for LPWAN are Sigfox and LoRaWAN. The first one is a global IoT network, aiming a low deployment cost of long-range, small-message device communication. It is a narrowband technology using extremely slow Binary Phase Shift Keying (BPSK) modulation. The second one is an LPWAN specification for long-range low-datarate wireless battery operated devices communication, promoted by LoRa Alliance, which coverage is comparable to a cellular network.

3GPP addressed categories and new standards for IoT requirements [51]. New users categories (cat1-6) are introduced starting from LTE Release 8. Release 10 defines 6-8 categories, and Release 12 defines LTE Cat-0. This category suits IoT applications and general M2M applications needs with low data rates in short bursts. This results in a reduced complexity and power. NarrowBand-IoT (NB-IoT) and LTE-M (or eMTC for enhanced Machine-Type Communications), are two complementary technoogies for IoT applications. NB-IoT is a direct candidate of Sigfox and LoRa, as it is less energy consuming and can address a large amount of connected devices. It can be deployed as a standalone solution or inside an LTE carrier, and will have more in-band deployment flexibility in 5G NR. LTE-M addresses another connected devices category which is more constrained in latency, mobility and density. Application examples are e-health, security, object tracking, etc. LTE-M however is more energy consuming. An illustration of LTE IoT evolution is given in Figure 10.

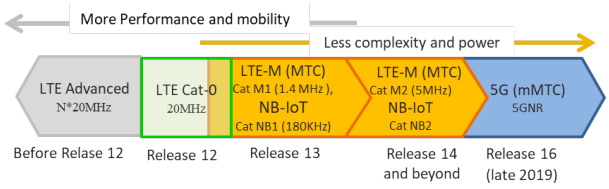


Figure 10: 3GPP IoT solutions standards

Note that many 5G technologies are specified to support IoT deployment, such as include D2D, SDN, NFV and MEC. For example, C-RAN reduces the cost of ultra-dense networks, expected especially with IoT deployments, through the simplification of BSs. This enables flexible and adaptable deployment.

3 5G Energy considerations

Today, energy is a major issue in all industries, including telecommunications. For example, the Mobile and wireless communications Enablers for Twenty-twenty Information Society (METIS) project considers EE as an important 5G topic [52]. The objectives of the project lead to technical goals with in particular a much denser network, containing low-power devices requiring higher battery life, up to 10 times longer than current batteries.

3.1 Energy efficiency on network side

5G will densify the network in order to increase the coverage and the capacity of the system. However, as power consumption is expected to increase at the same time, a challenging issue for 5G is the EE. For example, considering

data centres in the network, the author in [53] examines the variety of clusters and clouds that implies the need for power-saving control mechanisms. He presents network and cluster resources control algorithms, as well as an integrated and hierarchical control of the resources. Similarly, as treated by the project EARTH [54], improving EE at BS level need a variety of technologies. For example, it is proposed in [55] to use Cellular Partition Zooming (CPZ) to reduce energy consumption; following the present users in the area, the BS zooms in for coverage or out for sleep mode, which is energy saving.

Optimizations, categorized on green energy researches, aim BS energy use reduction [56]. In fact, the power consumption in the Operation Expenditure (OPEX) in today's mobile systems is dominated by networks, up to 90%, and only 10% is for mobile user. The network electricity supply is thus a critical issue in mobile industry, especially for regions with difficult grid power connection. Alternative energy sources, light, wind, vibrations, etc. are considered. However, alternative energies, especially renewable ones, present some challenges regarding availability and volumes. In particular, combining renewable energy sources at BS is critical. For example, the japanese operator NTT DoCoMo launched in July 2004 the *DoCoMo Eco Tower* [57], which is an experimental an self-powered 3G BS. Thid BS uses solar and wind power at once. Similarly, Alcatel proposed *Eco-sustainable Wireless Networks* to replace current generators [58]; Alcatel has an Alternative Energy Program combining alternative energy sources and energy-efficient equipments and networks.

Note that RF-EH appears as a candidate technology to further optimize energy management in 5G. In particular, EH from RF signals is actively studied and mathematical models are made in order to make it possible in the near future, as proposed in [59]. In the same scope, researches are interested in several scenarios, as cooperative multicast communication in [60]; EH relays harvest energy from the BS transmissions following two power splitting strategies at the relays for EH, and analytical expressions of outage probability are given for these strategies.

3.2 Energy efficiency on terminals side

Battery life is a major constraint for equipments' design. In fact, it can be lengthened through enhancement of its capacity, use and recharge process. The overall EE of the network would be consequently improved. 5GPPP establishes requirements for future 5G systems considering ultra-efficient 5G hardware, especially in terms of energy consumption, flexibility and compatibility with heterogeneous environments [61]. The interface between network and terminal can use massive MIMO technique to enhance EE, in addition to new waveforms and to network radio resources optimization. The air interface, procedures, and signaling can also be enhanced in terms of energy consumption through sleep modes, simplified procedures access schemes, and more efficient modulation, spreading and coding that allow low transmission powers. Energy consumption can be reduced using EH from environmental energy sources, such as light, heat and vibration [62]. However, as they vary along location, time, weather conditions, etc., it can be difficult to guarantee a given Quality of Service (QoS) that is required in some wireless applications. A possible solution for this problem is ambient radio signals harvesting, i.e. RF-powered energy-harvesting network (RF-EHN) [63, 64].

Note that circuit design has also to be coherent with EE optimization. Many technologies are discussed to prepare 5G-ready handsets, as in [65]. The research areas for that purpose include RF design, mobile cooperation, context awareness and seamless roaming among multiple technologies. New techniques for power amplification, antenna and filter design, and RF architecture are being studied for future handsets design. For example, M2M and D2D communications may use wireless power transfer technologies and optimization of sleep mode switching in order to have a battery-less sensor operation.

4 RF Energy harvesting implementation

4.1 Energy harvesting sources

EH has been already widely deployed using several energy sources. It consists on absorbing ambient energy in order to utilize it instantly or later. The most popular harvested energy sources are solar energy, wind energy, tidal waves energy, mechanical energy, electromagnetic energy, thermal energy, etc. EH sources and their corresponding sensors are illustrated in Figure. 11 [10].

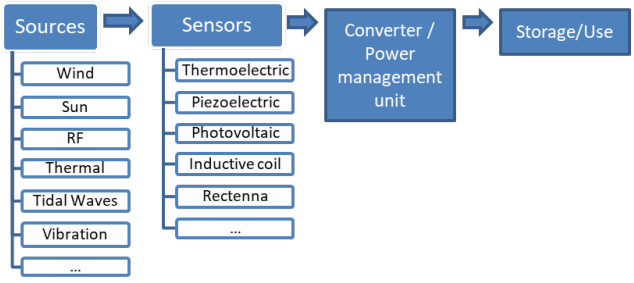


Figure 11: Energy harvesting sources and sensors

- RF sources have a power density ranging from 0.1 μW/cm^2 for ambient RF to 1000 μW/cm^2 for dedicated RF [10]. Electricity in that case is produced through magnetic inductive coupling: a loop with time varying current induces a receive loop with an open circuit voltage, and the resulting voltage can thus be directly used or stored in a battery for later use.
- Harvesting energy from heat is based on thermal energy generator, which uses the difference of temperatures to generate electricity. However, it has a low efficiency(5-10%) [50] and needs a temperature difference of 30°C. The power density in this case is about 135 mW/cm^2 at 10°C [66].
- Sun and light EH, i.e. solar and photovoltaic energy, is very popular nowadays due to its power density, efficiency, and flexibility. The light is collected by a semiconductor material, usually silicon, that produces DC power by the movement of electrons in the

semiconductor. The electricity is then stored in batteries or supercapacitors. The power density is about 100 mW/cm^3, but it requires direct exposure to light, which limits its use indoor [66].

- Mechanical energy is by the other hand harvested from vibration, pressure, etc. Among others, piezoelectric material, when exposed to a mechanical force, makes a polarization of ions in the crystal. This has a high energy density, which is about 50 μJ/N [66]. An example of piezoelectric EH for IoT applications can be found in [50].

4.2 Energy Harvesting constraints

EH presents an interesting solution to current energy consumption issues. However, it also presents some design constraints, as ambient RF waves figure among the most challenging energy sources. It suffers more particularly from :

- Causality, so that when harvested in the present, the energy is only available in the future, .
- Difficulty of concurrent EH and information transmission, at least for current designs.
- Efficiency loss when converting RF harvested signal to Direct Current (DC).
- Finite nature of energy storage.

For some applications, difficult environment access, connection cost and mobility makes it difficult to charge sensors or, at least, with a certain periodicity. In addition, if EH is used, for some EH sources, harvested energy depends on time and/or location and may not be enough for terminal's needs. Examples of time dependance are stillness time for vibration energy and night time for solar energy. Electromagnetic Radiation appears thus as an interesting stable and predictable energy source.

A cellular system can integrate EH using a standalone technique, hybrid technique or Simultaneous Wireless Information and Power Transfer (SWIPT) technique [67]. For the first technique, it allows to equip only some entities with EH. Hybrid EH empowers the system by both grid and EH, which may solve occasional unavailability of EH. SWIPT empowers the devices with harvested radio wave frequencies through which data is being received, as developed in 4.4.1.

Most of current mobile systems broadcast RF signals, carrying energy that presents a true opportunity as a source. At the same time, considering this possibility, the RF energy can supply devices. This implies the use of Wireless Power Transfer (WPT), and promising techniques are studied in order to implement such systems. The most significant examples are :

- Magnetic resonance coupling, which uses near-field signals and covers an extent ranging from some centimeters to some meters. This makes this technique adapted for charging mobile phones and hybrid electric vehicles, and its efficiency is 30% to 90% for a Tx-Rx distance ranging from 0.75m to 2.25m [68].
- Resonant inductive coupling, which exploits near-field signals and can also be used suitable for mobile phones charging, passive RF IDentification (RFID) cards and contactless cards. Frequency bands for this technique range from 10 KHz and 30 MHz and have an efficiency from 6% to 90% respectively [69].
- RF energy transfer, which uses far-field waves. Its coverage ranges from few meters to several kilometers and is suitable for wireless body and WSNs. Its efficiency is 0.4% to over 50% for −40 and −5 dBm input power, respectively [70]. This technique is being increasingly researched these last years and is gaining growing maturity, which promises new networks concepts, such as Wireless Powered Communication Networks (WPCN) [71].

The circuit suggested in [72] for ambient RF energy battery charge, as may be used in IoT-like applications, is made of a rectifier circuit and an impedance matching network. It converts RF waves to Direct Current (DC). Unlike EH circuit, information decoding uses down-conversion and sampling to process the baseband signal. For EH, a dedicated circuit directly connected to a receiving antenna, called a rectenna (rectifier antennas), captures energy which is converted then into functional DC voltage. This operation is called RF-to-DC conversion.

A rectenna has five main components: an antenna, the network matching the output impedance of the antenna to the rest of the circuit, a low pass filter that eliminates higher-order harmonics, the rectifier which is a diode, and DC filter load (peak detector). Such rectennas are designed in [73], operating at 27.5GHz, and achieve rectification with coupling between the input and output microstrip lines in the design. Note that EH-Oriented transceivers, especially rectennas, can be adopted in most cases for both indoor and outdoor applications and provide an efficient RF-to-DC conversion, reaching 80% in some configurations [74].

Figures 12 and 13 illustrate separately respectively energy and information receivers, which have to be combined for an EH scenario.

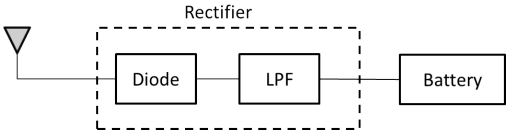


Figure 12: Simplified energy receiver scheme

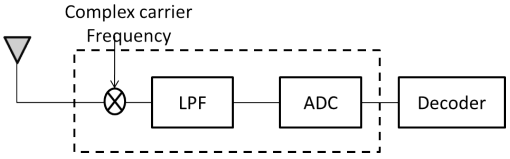


Figure 13: Simplified information receiver scheme

In the case of a receiver combining both EH and information decoding, the received power is split between the two either statically or dynamically. The splitting depends on the variations of RF energy and information, and also in the

needs of the receiver. The combined receiver can separate information and energy circuits [75], as illustrated in Figure 14, or it can integrate both in the same circuit, as illustrated in Figure 15.

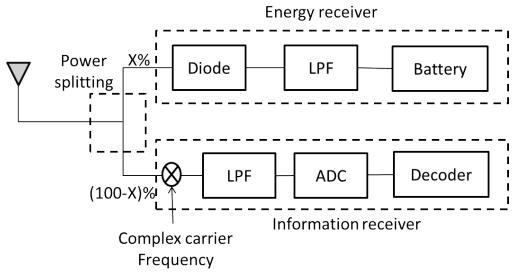


Figure 14: Separated information and energy receiver

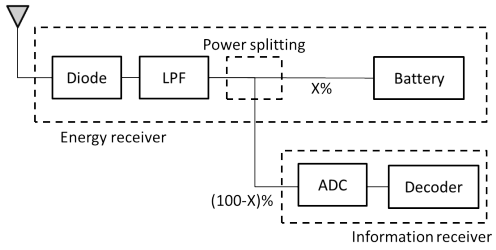


Figure 15: Integrated information and energy receiver

Authors in [75] show that there is a tradeoff exists between harvested energy and achievable rate, so that either integrated or separate receivers are selected depending on the harvested energy.

In teh remainder of this paper, RF-EH is considered. Note that RF energy by nature implies a relative time variability, as well as a sporadic aspect. On the other hand, this technology at mobile networks' scale is much younger than other energy sources harvesting, and accumulates thus less technology mastering and experience.

4.3 RF Energy Harvesting hardware

Several 5G use cases announce the need for self-sufficient and self-sustaining equipments. This energy chase covers the network as a whole including terminal devices, network elements and data centres, and EH can bring a solution for such requirements [72]. However, implementation of RF-EH presents many challenges, including, in some scenarios, the conception of rectennas to collect RF energy and the WPT [73]. Two categories can be cited for WPT: generation of electric power within few meters, i.e. near-field, and generation in an extent that can reach few kilometers, i.e. far-field. The efficiency of energy transfer in near-field case can reach up to 80% [76]. The WPT in far-field appears as RF or microwave signal and uses antennas for transmission and reception then converts it to power by rectifier circuits.

From a hardware point of view, antenna designs already exist for EH from interference signals issued from GSM, WLAN, UMTS and related applications, where antennas are designed with the potential to be used in EH systems. Several categories are also proposed in the literature for different frequency bands such as 900MHZ, 940MHz, 1.95GHz and 2.44GHz, 0,8 to 12GHz [77, 78, 79, 80, 81]. The designed antennas can be integrated in an antenna array to increase the gain and preserve a very high frequency bandwidth.

Though the harvested energy remains suitable for low power electronics, as it is rather low compared to other EH sources. However, a major challenge for RF-EH is still the low power sensitivity and legal restrictions of maximum authorized RF power radiation.

n the other hand, following [82], the global EH market is expected to reach 974.4 Million USD in 2022 when it was only 268.6 Million USD in 2015. This growth is driven by electronic sensors with ultra low-power and by embedded systems, which involve an energy issue. EH can be used to solve this issue, as suggested in [83], and EH relays can be considered, especially when the devices use WPT through RF signals. These signals carry both information and energy, and suggest for relays to charge themselves by extracting energy from the signals they overhear.

Note that RF-EH chips are already being developed as indicated in [84], where a startup has raised 30 million USD in funding to develop chips that can be powered by ambient radio frequencies. More generally, EH from other sources is also coming to reality, as the market for EH chips will approach 3.4 billion USD by 2022 according to the report in [85]. The mentioned report cites some companies that are in the race for EH chip market, and include for example Apple, National Instruments, Analog Devices, STMicroelectronics, Powercast, Renesas, Microchip Technology, Texas Instruments, Mentor Graphics, etc. An example of optimizations for RF Energy in a WSN using Powercast can be found in [86].

4.4 RF Energy harvesting approaches

In order to supply power to the nodes through WPT, it is possible to use either BSs already present in the network, or dedicated Power Beacons (PBs) specifically deployed, as illustrated in Figures 16 and 17. The latter has the advantage of being easy to deploy as it doesn't need backhauling. In both cases, energy beamforming can improve the EH efficiency, and distributed cooperative beamforming can be used to provide data rate fairness [87].

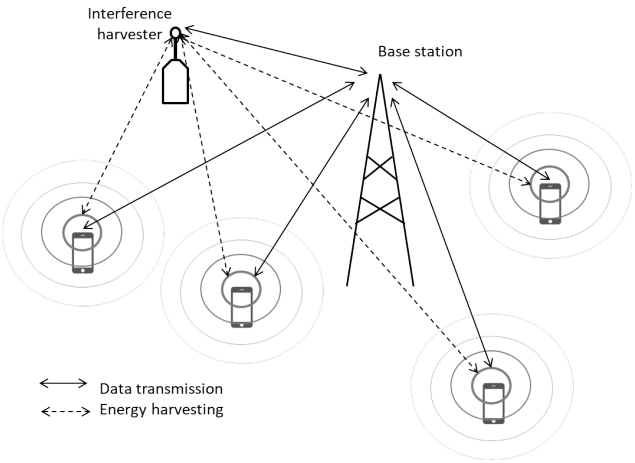


Figure 16: Illustration of the dedicated energy sources to supply RF energy

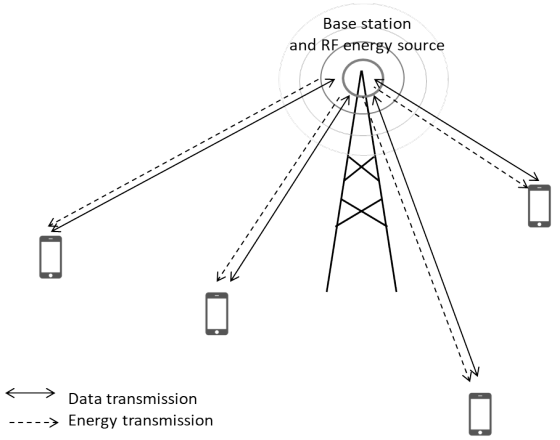


Figure 17: Illustration of the use of existing BSs to supply RF energy

4.4.1 Simultaneous Wireless Information and Power Transfer

Many suggestions for EH receiver architectures are inspired from the use of WPT in wireless networks. Actually, the same electromagnetic field can be used both for information and energy transfer to the end users and/or devices. This case is called Simultaneous Wireless Information and Power Transfer (SWIPT).

Two different tasks have to be completed by an EH receiver: decoding information and harvesting energy. The ideal case would integrate in the same receiver both functions with no power loss. However, realistic case dedicate a circuit for each task, so that EH and information decoding are performed separately. Authors in [75] suggest an intermediate solution that partially combines EH and information decoding circuits in a manner that the rectifier performs a part of the information decoding.

In a SWIPT configuration, a time slot, or a received signal power portion, are used for EH. The rest of time, or of power, are intended for decoding information. Nevertheless, time sharing scheme is more interesting as energy receivers and information receivers, which are usually distinct, have different power sensitivities. Note that there is a tradeoff between energy transfer rate and information transmission rate following the channel. For example, in an Additive White Gaussian Noise (AWGN) channel, maximum information rate target and maximum power transfer efficiency target are compatible.

There are technological issues that still need to be addressed in order to enable SWIPT use for wireless network deployment. For example, if we consider the nature of the transfers in SWIPT, the variation detection in RF signal by the receiver defines the information amount it can decode. The relationship between the two can be described by the Signal to Interference plus Noise Ratio (SINR), i.e. ratio between the power of the signal power and the disturbance power, including interference, noise and eventually any other disturbance affecting the decoding. In practice, even though a signal power can be low, its information rate could be very high. This is however not true for energy amount carried by an RF signal, which depends on its magnitude only. Hence, in a SWIPT, there is a trade-off between transferred energy amount and information rate, as it is not possible to maximize the "rate" of the two transfers simultaneously. Consequently, there is a need to adapt te network design for SWIPT implementation, leading to research issues on the design, analysis and optimization of SWIPT-based networks architectures and protocols. For example, in [88], the wireless powering of terminals in an UDN considers game theoretical approach of power allocation in a network of Full-Duplex (FD) BSs which serve an area highly dense of half-duplex (HF) user equipments. Considering EE and coverage optimization problem for such network as a mean field game gives an optimal strategy to adapt BSs for this type of network settings. SWIPT can also be considered in cooperative communication networks, as proposed in [89], where relays harvest powers to be able to amplify and forward data. An optimization, such as maximum ratio combining, can be then operated considering direct and relayed signals.

Note that there are many variants of SWIPT schemes following the integration in the model of other 5G techniques or the consideration of specific use case. An example is given in [87], where a cooperative communication network is considered with a distributed energy beamforming and simultaneous communication with several users. Moreover, when used with NOMA, efficient EH can be performed, and EE can be enhanced with the use of EH pilots for channel estimation.

4.4.2 Dedicated RF sources powering

Many tools are proposed in case of powering through dedicated RF energy source, as in [90] where optimal power allocation is suggested for frequency division multiplexing, meaning that data transmission and EH can be performed simultaneously, and time division multiplexing, meaning that data transmission and EH are implemented in different time slots but in the same frequency band. Optimal EH and information transmission strategy is then given and long-term throughput can be optimized. Note that authors in [91] considered the possible optimizations in the case of dedicated energy sources for WPT as in IoT applications. Authors showed that power control is critical for system achievable rate improvement in the case of a finite blocklength, and when the Tx power is asymptotically optimum, performance is nearly optimum.

PBs provided by the network, can be integrated in a relay network. For example, [92] suggests that relay nodes harvest transmitted power from the PB station to forward signals to destination. The information can be relayed but also power, following adequate protocols to optimize power splitting, especially in the presence of multi-antennas. Another example using dedicated RF charging in the case of smart wearables, where the energy consumption is very low, is discussed in [93]. This implies a mix of operator-deployed and user-deployed systems that will need to cooperate. Authors in fact highlight through system-level evaluation the preference for omnidirectional energy transmission in this case.

4.4.3 Interference harvesting

An increasing number of users are actively using mobile systems and demanding continuously improved quality of

experience. In addition to the diversity of end devices and of applications, each with its own constraint, 5G uses a heterogeneous UDN to respond to the current landscape, which already implies the coexistence of many RF signals on different bands. In fact, as illustrated in Figure 18, today's current deployments imply the presence of many signals, which can be harvested to power devices. In particular, GSM bands present a higher RF power density, while other sources may also be exploited for energy harvesting, such as Amplitude Modulation (AM) radio stations and Ultra High Frequency (UHF) RFID.

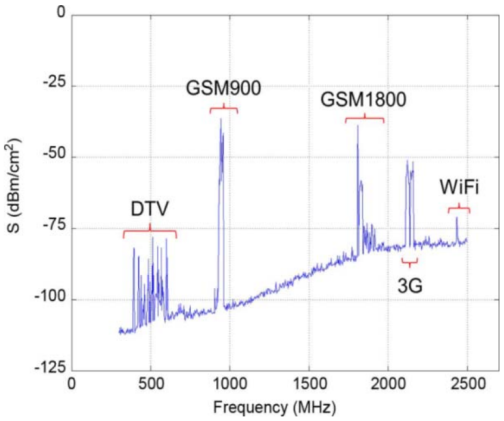


Figure 18: Example of measured RF power densities of main RF energy sources in an urban outdoor location (Source [94])

A given device using a given system is exposed to signals intended for other users in the same network, signal from other networks, or signals on other bands. These signals are usually seen as interference and many techniques and resources are used to mitigate it. It has to be carefully dealt with in an EH context for different types of channels, including broadcast channels, multicast channels, multiple access channels, and multi-user interference channels, as reviewed in [95]. However, they can also be seen as an energy source to harvest.

Figure 19 illustrates the case where signal from the user can be harvested by an interference harvester, which is a node capable of communicating itself with the BS. The harvesting and transmitting are usually sequential.

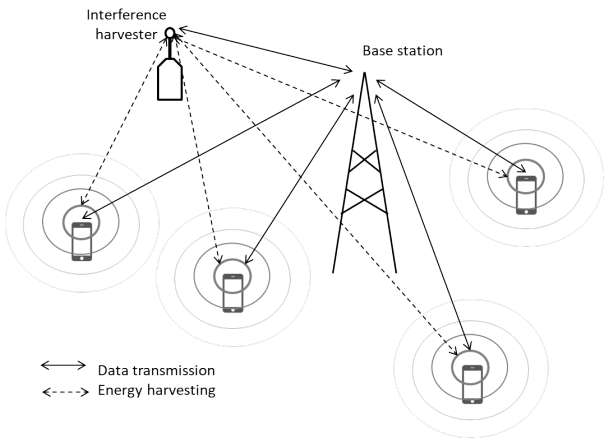


Figure 19: Example of an interference Harvesting scenario

The authors in [96] present opportunistic communication for devices that can reconfigure according to the the band they use, which makes them able to separate useful signal from possibly harvestable OOB interference. Nevertheless, interference harvesting from the same band requires additional processing, and strategies to realize this harvesting are currently being researched, and promising to be a key enabling factor for self-sustainable transmissions in 5G networks [97]. Similarly, authors in [98] present a cooperative wirelessly powered communication network protocol, where distant users overhear uplink signals of users and downlink access point broadcast signals, and time allocation for uplink and downlink are jointly optimized, which increases users rate and improves user fairness and transmission delay. Actually, the analysis of RF-EH performance has been considered in works such as [99], where the harvested energy mean, transmission outage probability and power outage probability are studied.

The energy can be also harvested from parts of the signal of the same user. An example with multicarrier modulations, as in [74], considers the fact that the Cyclic Prefix (CP) is used when still analog and thus, its energy can be harvested. The CP would remove inter-symbol interference, and at the same time contribute in SWIPT implementation, especially for WPT in downlink.

A backscatter communication architecture is suggested in [100], which combines ambient EH and reflection of this same energy modulated, with all its related signal processing aspects and multiple access techniques.

4.5 RF Energy harvesting scenarios

Globally, as enumerated in [95], three EH and transmission protocols can be identified:

- Harvest-Use (HU), for which the device is powered by EH directly, without using any buffer. Data transmission is thus possible only when harvested energy can cover the processing energy cost.

- Harvest-Store-Use (HSU), for which a storage device, typically a battery, is used to gather harvested energy. This energy is then available for use only in the next time slot. This case has to consider carefully possible storage inefficiency during the charging and leakage during the storage. For example, only about 70% of the energy is stored in Ni-MH rechargeable batteries while 95% can be reached with a supercapacitor. However, a battery implies less leakage than a supercapacitor.

- Harvest-Use-Store (HUS), which uses two energy storage devices : a supercapacitor storing the harvested energy for immediate use, and a buffer storing the remaining energy for later use.

The energy buffer is critical when using one of these protocols. Other schemes in a different optimization perspective can consider storage constraints, QoS requirements or capacity achieving targets. An example is the introduction of a sleep-and-awake mechanism to enhance the achievable rate [95].

4.6 RF Energy harvesting scheduling

In order to have an efficient EH system, and depending on the used EH protocol used, the energy consumption has to be adapted to the EH. This is to prevent energy outage or overflow. Thus, the design of energy use scheduling is critical. It can be offline or online, following the case where channel state information is available at the beginning of the transmission or only on the fly.

Energy scheduling depends on many parameters, such as data throughput, transmission competition time, mean delay, outage probability, quality of coverage, message importance, grid power consumption when there is a hybrid power supply, etc. For instance, the scheduling of RF harvested energy optimizes the time duration of information processing and EH. A tradeoff between the minimum information outage probability and the maximum average harvested energy defines the optimal mode switching between the two functions at the receiver. A review of scheduling techniques can be found in [95].

The signaling in wirelessly empowered communication networks has also to be carefully designed. In fact, if wireless powered communication networks control is centralized, signaling overhead, as well as optimizations in case of large number of devices can be prohibitively complex. Thus, a distributed EH random access protocol can be adopted, as in [101], where a slotted ALOHA is used a harvest-until-access protocol. The proposed algorithm optimizes the average channel throughput without knowledge of full channel state information, especially when the number of devices is large, which reduces the complexity and overhead of a centralized management.

Integrated spectrum and energy control mechanisms are presented for a typical heterogeneous 5G network in [102]. In this work, the authors consider the network where the devices have different EH capabilities and perform a cooperative resource sensing for CR implementation. In this case, both energy and spectrum are *"harvested"*. This leads to an architecture and model with specific spectrum and energy control mechanisms in many 5G scenarios, and consequently a Medium Access Control (MAC) protocol for this context. Another energy and spectrum efficiency optimized MAC protocol for 5G EH can be found in [103], where SWIPT is considered for cooperative and non-cooperative schemes with EH capabilities in a CR aided 5G networks.

5 RF Energy harvesting in 5G ecosystem

5G ecosystem integrates a wide range of use cases, using established and new technologies. EH is researched to be integrated with key 5G enablers, and examples of these researches are given in the remainder of this section.

5.1 UDN RF-EH-enabled

UDN are among main 5G enablers for many use cases, including IoT. Considering ambient RF-EH in dense networks, many implementation and operation challenges arise including energy availability, EH mode selection, cooperation between BSs and devices. In particular, there are tradeoffs between EE, SINR and outage probability, as highlighted in [104], where it is found that co-channel interference EH can bring efficiency improvement, especially if combined with an optimization of BSs parameters.

When the capacity is limited, the H2020 project SCAVENGE [105] suggests that cells can be split so that some are master Access Nodes (AN) while the others are EH AN. In case of heavy traffic, all ANs would be enabled to provide required capacity, while for light traffic, a part of EH AN can be switched off. This can be possible as UDN have high probability of Line-of-Signt, making the access easier for devices without the need of EH ANs. Backhauling the cells, if using wireless connection, makes the deployment cheaper and more flexible. In particular, EH ANs backhauling can use mmWave or be self-backhauling. In the second case, the access and the backhauling use the same link, shared in time and/or frequency. In the first case, mmWave can provide EH capabilities for rapid AN installation. In particular, for some deployment scenarios, micro cells with low power can be used, for which an EH powering and a wireless backhaul are suitable. The use of EH in this case, as indicated in [105], can be cost-effective. It can also provide an easy coverage and deployment in rural areas, easy cell planning, reduced environmental impact and cost saving in long-term.

Note that WPT can be used between BSs in addition to terminals. This will imply a new cell planning based for example on energy balancing connection and QoS-aware BS planning.

5.2 RF-EH in C-RAN

C-RAN and H-CRAN aim to provide a high data rate, energy efficient network. However, because of its heterogeneous nature, H-CRAN is more energy consuming as it has a larger number of RRHs than C-RAN. When the network is dense, the energy requirements become critical and several adaptations and optimizations have to be made. EH can bring thus a solution as in [106], where BSs are suggested to be powered with EH, and can also power devices. A throughput optimization in that case can be performed under receiver charged energy constraint. The grid energy consumption in the case H-CRAN downlink with one grid-powered BS and many RRHs EH-powered is studied in [67] and a model and optimization of user association, power allocation and EH constraints are given. In this scenario, the EE optimization is modeled as a fractional mixed integer nonlinear programming problem, which can be solved in few iterations. This improves the number of users, who are mostly associated to EH-RRHs, and thus decreasing the grid energy consumption.

In a UDN deployment, the implementation cost can be high due to infrastructure and small cells backhauling, and C-RAN can en be used to reduce this cost, especially for low data rate traffic as in IoT. The fronthauling then can use EH for an energy efficient deployment [105]. Similarly, the energy cost in an RF-EH combined with an energy management and scheduling in H-CRAN context is considered in [107]. RRHs in that case are EH-powered, and a hierarchical framework based on economical model and using energy sharing and energy trading between 5G actors, and more specifically internet service providers, can provide a

framework for the system design.

5.3 RF-EH enabled MEC

In MEC, computing resources are available at the edge of the mobile network, i.e. BSs, in order to ensure low latency and to reduce core network traffic. Globally, two main scenarios are considered for EH and MEC combination. Either EH capability is deployed at the MEC BS (or resource server) to prevent possible grid energy absence or unavailability in some scenarios, like natural disasters or remote areas, or it is implemented on device side, which is more suitable in case, for example, of IoT. Study of the state of the art for these two scenarios can be found in [108]. For the first scenario, energy forecast and EH enabled BSs load balancing have to be controlled. An example considers BSs equipped with EH capabilities and may reduce thus by the same occasion energy consumption. The authors in [109] suggest to optimize the BSs and virtual MEC resources use following traffic load and forecasted harvested energy parameters. The heuristic optimization employs sleep mode for BS and soft-scaling for Virtual machines. The authors announce up to 69% energy saving in single BS case, that can be enhanced by up to 16% in BSs cluster. Similarly, an illustration of the second scenario can be found in [110] and present a low-complexity dynamic computing offloading strategy for MEC with EH. This strategy decides computation offloading power and device transmit power, without the need of distributed information, and is easily implementable, besides being asymptotically optimum.

5.4 D2D with EH

D2D communication suffers from a major constraint, which is energy. EH can be thus integrated to devices in order to solve this issue, and many studies consider this case and provide theoretical performance derivation for different scenario. For example, in [111], the authors consider the scenario where UEs in a traditional cellular network assisted with relays use decode-and-forward cognitive D2D communication and EH. They provide analytical expression for main performance parameters, which helps to optimize link robustness. Another scenario can consider D2D in large-scale cognitive cellular networks as by authors in [112], where D2D transmitters harvest energy from dedicated multi-antenna PBs. A power transfer model with three possible power transfer policies can be proposed, in addition to an information signal model with two considered receiver selection cases. In addition to outage probability, secrecy is also considered in the analysis conducted by the authors and complete the framework for secure D2D communication.

5.5 RF-EH with key 5G RAN techniques

Spectral efficiency is critical in current mobile systems, particularly in RAN. Thus, spectrum optimizations have to be considered, in combination with RF-EH techniques. Reviewed techniques in the rest of this section include cooperative communications and cognitive radios, in addition to massive MIMO, NOMA and D2D. The combinations and optimizations usually consider many of these techniques at a time, in order to prepare its integration on practical systems.

5.5.1 Cooperative communications

In an EH network, nodes may have different channel and energy conditions, which may enhance performance using a cooperative approach. Transmission schemes for cooperative networks, multi-user networks, cognitive radio networks and cellular networks are thus to be considered.

Multi-user environment is an example of scenarios with EH, where interference management mechanisms have to be effective. Moreover, this environment presents spatial diversity gains that offer an opportunity to enhance system performances through collaboration between network nodes, which form a virtual MIMO system. Among studied aspects for cooperative networks, the power allocation is to be optimized under EH models, as well as throughput in delay constrained environment, and joint power allocation and time scheduling with EH.

Another example of EH integration issues is the joint optimizations of EE and frequency bands in the design of wireless-empowered transceivers, as presented in [74] where in-band interference harvesting is designed so to realize self-sustainable communication architecture.

5.5.2 NOMA

The adaptation of EH with 5G coming techniques, such as NOMA or CR networks, is an important issue to address for RF-EH deployment. In fact, there is a need for more comprehensive design for both new technologies, NOMA and RF-EH. Such combination may if established provide an efficient use of the energy and radio spectrum in energy efficient and sufficient networks, as in [113]. Other examples of issues are modeling and analysis of large scale EH networks and relay-based EH networks, as well as the design of energy cooperative strategies between harvesting devices. Note that in addition to energy and power transfers, it is fundamental to consider energy storage at end nodes, which ideally would consume harvested energy immediately within a transmission cycle.

Analysis tools for EH NOMA are intensively investigated, as in [114] where users outage probability under imperfect channel state information and imperfect SIC in a cooperative EH NOMA network is derived. Another example of mathematical modeling for user grouping, power and time allocation can be found in [115], which adopts a mesh adaptive direct search algorithm to optimize data rate and transmit power for each user with reasonable complexity. Similarly, outage probability is calculated for relay systems with NOMA and EH in [116], where the multiuser downlink is considered and BSs use multiple antennas to beamform information and energy.

NOMA use in specific wireless powered IoT relay systems with RF-EH capabilities can be found in [117], where RF-EH and information transmission system with time switching and power splitting relaying are considered. Due to the constrained nature of IoT nodes, the authors propose for IoT nodes to first harvest energy from the source node, either with time switching relaying or power splitting relaying

protocol. After that, the information of the source node is transmitted by the relay node as well as its own information using NOMA. The description of this system is mathematically derived by the authors, allowing several optimizations of system parameters.

5.5.3 Massive MIMO

Massive MIMO is a key 5G enabler, and its integration with EH schemes has been studied by many researchers. In [118], the authors explore the use of massive MIMO in heterogeneous networks. The deployment considers macrocells where users harvest energy to maximize their uplink power by minimizing uplink pathloss. It is found that user association should be optimized to maximize both harvested energy and information transfer. In this case, the downlink received power and the uplink received power should be both considered in the optimization. The stochastic model for the harvested energy and its link to uplink achievable performance demonstrates the added value of massive MIMO on that point, though the gain may deteriorate with network density. The proposed model enables however the study of massive MIMO and RF-EH integration with other technologies.

Another example of 5G techniques integration can be found in [119], where the downlink of a C-RAN enabled scenario with distributed large-scale MIMO is considered. Each RRH is powered EH and the grid in order to secure the RRHs from possible EH unavailability. The problem problem of BS energy consumption minimization under QoS constraints in this scenario can be solved using linear programming or an online energy management algorithm.

SWIPT in massive MIMO systems is studied in [120], where an analytical study of hybrid and full-dimensional processing is performed. It considers hybrid precoding and beamforming BS and derives expressions for harvested energy, achievable rates and energy-rate tradeoff for growing number of BS antennas and network nodes. A hybrid precoding and combining is proposed which optimizes harvested energy with little energy-rate tradeoff degradation.

5.5.4 mmWave

mmWave is attractive for EH as it is intended to be integrated for large antenna arrays and a dense network, presenting thus interesting harvesting opportunities. However, it propagation characteristics, i.e. poor penetration and diffraction, may limit its use to some scenarios. Many studies though consider RF-EH for mmWave communications. For example, authors in [121] consider a scenario where mmWave signal provides energy and/or information to low power devices and derive key information and EH performance parameters as well as overall coverage performance. This establishes an analytical framework to design network and device parameters, for example for antenna geometry optimization.

A study on combination of NOMA with massive MIMO using mmWave can be found in [122], where SWIPT is used to enhance EE. Hybrid precoding is used for mmWave massive MIMO, and information and energy can both be *"harvested"* by users thanks to a power splitting receiver. Algorithms are proposed to select a user per beam, and an analog precoding is then performed to all beams, then users grouping is performed and a digital precoding is applied. An iterative joint optimization of power allocation and power splitting reaches finally a maximized achievable sum-rate.

Note that antennas design for mmWave specific applications are also a challenge, as investigated in [123], where the authors consider ambient RF-EH antenna on textile for wearable in mmWave, and provides up to 40% on-body and 60% off-body radiation efficiency.

5.5.5 Cognitive Radio

Cognitive radio environment needs to be adapted with EH. As already mentionned, CR is often considered with other 5G enabling technologies. In fact, the spectrum usage is optimized as Secondary Users (SUs) use the spectrum of Primary users (PUs), following different priorities and cooperation strategies. These strategies may incorporate information transmission and spectrum sharing among nodes, but also EH strategies. In this case, both PU and SU get benefits as the PU improves his throughput through SU as relay node, and SU gets more spectrum usage opportunities from PU's spectrum. In addition, PU's bands access can be supported by SU's harvested energy. Conventional cognitive radio networks, e.g., spectrum allocation and management, spectrum sensing and handover, spectrum sharing, are reconsidered to enhance the network reliability in an EH context where, in addition, harvesting, spending or conserving energy have to be optimized.

An example of transmission CR schemes in RF-EH network can be found in [124], where an analytical study characterizes dedicated PBs. Information transfer and EH operate asynchronously in different bands. Another example considers a two-way relay assisted CR NOMA network [125]. In this example, a joint optimization of power allocation, transmit power and power splitting ratio is proposed. Similar throughput maximizations under QoS constraints can be found in [126] for a CR context.

6 Conclusion

This paper is a reflection of the increasing interest on new techniques to make energy use more efficient in 5G systems. It aims to provide a comprehensive review on RF-EH-enabled 5G networks. To this end, 5G standardization, requirements, use cases, ecosystem and key enabling technologies were reviewed, and particularly their energy aspects. RF-EH was then highlighted as a promising technique to solve energy supply and/or optimization. Different aspects of EH were exposed, including implementation and network protocols. The paper then presents literature review of RF-EH integration with 5G enabling technologies, as well as some research trends and challenges following the integrated technologies.

Conflict of Interest The authors declare no conflict of interest.

References

[1] S. El Hassani, H. El Hassani, and N. Boutammachte, "Rf energy harvesting for 5g: An overview," in *2017 International Renewable and Sustainable Energy Conference (IRSEC)*, Dec 2017, pp. 1–6.

[2] Cisco, "Cisco, visual networking index: Global mobile data traffic forecast ,update 2013-2018," Tech. Rep., 2014.

[3] "5g energy performance, ericsson white paper," Ericsson, Tech. Rep., apr 2015.

[4] "Driving the convergence of the physical and digital world - white paper on next generation mobile technology," ZTE Corp, Tech. Rep., 2014.

[5] "Sktelecom 5g white paper (white paper,version 1.0)," SKTelecom, Tech. Rep., oct 2014.

[6] S. Buzzi, C. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5g networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, April 2016.

[7] Y. He, X. Cheng, W. Peng, and G. Stuber, "A survey of energy harvesting communications: models and offline optimal policies," *Communications Magazine, IEEE*, vol. 53, no. 6, pp. 79–85, June 2015.

[8] I. B. Sofi and A. Gupta, "A survey on energy efficient 5g green network with a planned multi-tier architecture," *Journal of Network and Computer Applications*, vol. 118, pp. 1 – 28, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804518302029

[9] T. Soyata, L. Copeland, and W. Heinzelman, "Rf energy harvesting for embedded systems: A survey of tradeoffs and methodology," *IEEE Circuits and Systems Magazine*, vol. 16, no. 1, pp. 22–57, Firstquarter 2016.

[10] S. Divakaran, D. D. Krishna, and N. Nasimuddin, "Rf energy harvesting systems: An overview and design issues," *International Journal of RF and Microwave Computer-Aided Engineering*, p. e21633, 12 2018.

[11] "3gpp commits to 5g nr in unlicensed spectrum in its next release," https://www.qualcomm.com/news/onq/2018/12/13/3gpp-commits-5g-nr-unlicensed-spectrum-its-next-release, Press News, Tech. Rep., Dec 2018, accessed: 2019-07-26.

[12] "FP7 ICT project : Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society 5G," https://www.metis2020.com/, Tech. Rep., 2012, accessed: 2019-05-11.

[13] "White paper - 5g use cases and requirements future works," Nokia Networks, Tech. Rep., 2014.

[14] "5G White Paper," NGMN Alliance, Tech. Rep., 02 2015.

[15] "5g and energy," https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White_Paper-on-Energy-Vertical-Sector.pdf, 5G PPP, Tech. Rep., 2015, accessed: 2017-12-16.

[16] M. S. C. A. H2020-MSCA-ITN-ETN, "Sustainable cellular network harvesting ambient energy," http://www.scavenge.eu/, Tech. Rep., accessed: 2019-07-28.

[17] "Follow the leaders on technology, tools, and techniques : Understanding and testing the 5g ecosystem," http://literature.cdn.keysight.com/litweb/pdf/5992-3144EN.pdf, Keysight TECHNOLOGIES, Tech. Rep., jul 2018, accessed: 2019-07-30.

[18] M. Medin and G. Louie, "The 5g ecosystem: Risks & opportunities for dod," https://media.defense.gov/2019/Apr/03/2002109302/-1/-1/0/DIB_5G_STUDY_04.03.19.PDF, Defense Innovation Board, Tech. Rep., apr 2019, accessed: 2019-07-30.

[19] M. CAGENIUS, A. RYDE, J. VIKBERG, and P. WILLARS, "Simplifying the 5g ecosystem by reducing architecture options," https://www.ericsson.com/48f8e1/assets/local/publications/ericsson-technology-review/docs/2018/etr-5g-core-radio-migration.pdf, Ericsson Technology review, Tech. Rep., nov 2018, accessed: 2019-07-30.

[20] J. A. Manner, "Wrc-19 — an opportunity to bridge the digital divide for 5g: Opinion," https://news.itu.int/wrc-19-opportunity-to-bridge-digital-divide-5g/, ITU News, Tech. Rep., May 2019.

[21] "Samsung announces world's first 5g mmwave mobile technology," https://news.samsung.com/global/samsung-announces-worlds-first-5g-mmwave-mobile-technology, Samsung Corp., Tech. Rep., May 2013, accessed: 2019-07-26.

[22] "Huawei to bring 73ghz mmwave mu-mimo live demo to deutsche telekom," http://www.huawei.com/en/news/2016/2/73GHzmm-Wave-Mu-MIM-live-demo, Huawei press, Tech. Rep., Feb 2016, accessed: 2019-07-26.

[23] A. Bleicher, "First intercontinental 5g trial begins at winter olympics," http://www.huawei.com/en/news/2016/2/73GHzmm-Wave-Mu-MIM-live-demo, IEEE Spectrum, Tech. Rep., Feb 2018, accessed: 2019-07-26.

[24] X. Hong, J. Wang, C. Wang, and J. Shi, "Cognitive radio in 5g: a perspective on energy-spectral efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 46–53, July 2014.

[25] E. Paulson, M. KAMALUDIN, S. Kamilah, and U. Dauda, "Cognitive radio in 5g - a smart city perspective-," in *The 2017 Malaysia-Japan Workshop on Radio Technology*, 01 2017.

[26] R. Alkhansa, H. Artail, and D. M. Gutierrez-Estevez, "Lte-wifi carrier aggregation for future 5g systems: A feasibility study and research challenges," *Procedia Computer Science*, vol. 34, no. 0, pp. 133 – 140, 2014, the 9th International Conference on Future Networks and Communications (FNC'14)/The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC'14)/Affiliated Workshops. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050914009235

[27] M. Alam and M. Ma, "Resource matching in carrier aggregation enabling 5g networks," *Wireless Personal Communications*, 10 2016.

[28] H. PAPADOPOULOS, C. Wang, O. BURSALIOGLU, X. Hou, and Y. Kishiyama, "Massive mimo technologies and challenges towards 5g," *IEICE Transactions on Communications*, vol. E99.B, pp. 602–621, 03 2016.

[29] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-mimo-based wireless backhaul for the 5g ultra-dense network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13–21, October 2015.

[30] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *Access, IEEE*, vol. 1, pp. 335–349, 2013.

[31] "5g channel model for bands up to100 ghz," *IEEEGC*, dec 2015.

[32] X. Zhang, M. Jia, L. Chen, J. Ma, and J. Qiu, "Filtered-ofdm - enabler for flexible waveform in the 5th generation cellular networks," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[33] A. Farhang, N. Marchetti, F. Figueiredo, and J. P. Miranda, "Massive mimo and waveform design for 5th generation wireless communication systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 70–75.

[34] F. Schaich, T. Wild, and Y. Chen, "Waveform contenders for 5g - suitability for short packet and low latency transmissions," in *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*, May 2014, pp. 1–5.

[35] N. Michailow, M. Matthé, I. Gaspar, A. Navarro Caldevilla, L. Mendes, A. Festag, and G. Fettweis, "Generalized frequency division multiplexing for 5th generation cellular networks (invited paper)," *IEEE Transactions on Communications*, vol. PP, pp. 1–18, 08 2014.

[36] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (noma): Concept, performance evaluation and experimental trials," in *2015 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct 2015, pp. 1–6.

[37] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5g systems," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1647–1651, Oct 2015.

[38] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink non-orthogonal multiple access for 5g wireless networks," in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, Aug 2014, pp. 781–785.

[39] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer International Publishing, 2018.

[40] S. R. Islam, M. Zeng, and O. Dobre, "Noma in 5g systems: Exciting possibilities for enhancing spectral efficiency," 06 2017.

[41] M. Aldababsa, M. Toka, S. Gokceli, G. K. Kurt, and O. Kucur, "A Tutorial on Nonorthogonal Multiple Access for 5G and Beyond," *arXiv e-prints*, p. arXiv:1902.08992, Feb 2019.

[42] C. Mobile, "C-ran the road towards green ran (white paper,version 3.0)," Tech. Rep., dec 2013.

[43] J. Li, M. Peng, Y. Yu, and Z. Ding, "Energy-efficient joint congestion control and resource optimization in heterogeneous cloud radio access networks," *CoRR*, vol. abs/1602.05351, 2016. [Online]. Available: http://arxiv.org/abs/1602.05351

[44] G. T. . version 15.3.0 Release 15) Technical specification, "System architecture for the 5g system," https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.03.00_60/ts_123501v150300p.pdf, Tech. Rep., sep 2018, accessed: 2019-07-28.

[45] D. d3.1 : Functional network architecture and security requirements, "5g novel radio multiservice adaptive network architecture (5g norma)," https://5gnorma.5g-ppp.eu/wp-content/uploads/2016/11/5g_norma_d3-1.pdf, Tech. Rep., dec 2015, accessed: 2019-07-28.

[46] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5g network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov 2014.

[47] Z. Chang, Z. Zhou, S. Zhou, T. Chen, and T. Ristaniemi, "Towards service-oriented 5g: Virtualizing the networks for everything-as-a-service," *IEEE Access*, vol. 6, pp. 1480–1489, 2018.

[48] R. Wireless, "Vodafone, ericsson connect german labs in 5g mobile edge poc," http://www.rcrwireless.com/20160701/europe/germany-5g-mobile-edge-poc-tag17, Tech. Rep., jul 2016, accessed: 2019-07-28.

[49] V. K. Singh, H. Chawla, and V. A. Bohara, "A proof-of-concept device-to-device communication testbed," *CoRR*, vol. abs/1601.01398, 2016. [Online]. Available: http://arxiv.org/abs/1601.01398

[50] M. Shirvanimoghaddam, K. Shirvanimoghaddam, M. M. Abolhasani, M. Farhangi, V. Z. Barsari, H. Liu, M. Dohler, and M. Naebe, "Towards a Green and Self-Powered Internet of Things Using Piezoelectric Energy Harvesting," *arXiv e-prints*, p. arXiv:1712.02277, Dec 2017.

[51] 3GPP, "Standards for the iot," https://www.3gpp.org/news-events/3gpp-news/1805-iot_r14, Tech. Rep., dec 2016, accessed: 2019-07-28.

[52] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann4, and O. Bulakci, "Metis research advances towards the 5g mobile and wireless system definition," *EURASIP Journal on Wireless Communications and Networking*, 2015.

[53] P. Arabas, "Energy aware data centers and networks: a survey," *Journal of Telecommunications and Information Technology*, no. 4, pp. 26–36, 2018.

[54] "Final report on green radio technologies, earth project delivrable d4.3," https://www.ict-earth.eu/publications/deliverables/deliverables.html, accessed: 2017-12-16.

[55] S. Zhang, X. Xu, Y. Wu, L. Lu, and Y. Chen, "A survey on 5g new waveform: From energy efficiency aspects," pp. 1939–1943, Nov 2014.

[56] L. C. Wang and S. Rangapillai, "A survey on green 5g cellular networks," in *2012 International Conference on Signal Processing and Communications (SPCOM)*, July 2012, pp. 1–5.

[57] "Ntt docomo launches eco-friendly cell phone station," http://www.japanfs.org/en/news/archives/news_id025703.html, accessed: 2017-06-28.

[58] R. Garafola, "Eco-sustainable wireless networks: Ready for prime time," *Alcatel-Lucent*, 2009.

[59] F. Mukhlif, K. A. B. Nooridin, Y. A. AL-Gumaei, and A. S. AL-Rassas, "Energy harvesting for efficient 5g networks," in *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, July 2018, pp. 1–5.

[60] M. Ashraf, J. W. Jang, J.-K. Han, and K.-G. Lee, "Video multicast cooperative communication in 5g systems with radio frequency energy harvesting," *Computer Communications*, vol. 118, pp. 60 – 68, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014036641730508X

[61] "5g ippp (2014) 5g infrastructure ppp: The next generation of communication networks will be 'made in eu'," http://ec.europa.eu/information_society/newsroom/cf/dae/itemdetail.cfm?item_id=14424, accessed: 2017-12-16.

[62] J. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *Pervasive Computing, IEEE*, vol. 4, no. 1, pp. 18–27, Jan 2005.

[63] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with rf energy harvesting: A contemporary survey," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 2, pp. 757–789, Secondquarter 2015.

[64] S. Devasenapathy, R. Venkatesha Prasad, V. Rao, and I. Niemegeers, "Impact of antenna directionality and energy harvesting rate on neighbor discovery in eh-iots," in *Consumer Communications and Networking Conference (CCNC), 2013 IEEE*, Jan 2013, pp. 302–307. [Online]. Available: http://www.es.ewi.tudelft.nl/papers/2013-Devasenapathy-EHND.pdf

[65] A. Radwan and J. Rodriguez, *Energy Efficient Smart Phones for 5G Networks*, ser. Signals and Communication Technology. Springer International Publishing, 2016. [Online]. Available: https://books.google.co.ma/books?id=jCkTvgAACAAJ

[66] Z. Kausar, A. Reza, M. U. Saleh, H. Saleh, and H. Ramiah, "Energizing wireless sensor networks by energy harvesting systems: Scopes, challenges and approaches," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 973–989, 07 2014.

[67] N. Chughtai, "Energy harvesting in 5g networks," Ph.D. dissertation, 08 2018.

[68] A. Kurs, K. A.S., R. Moffatt, J. Joannopoulos, P. Fisher, and M. Soljai, "Wireless power transfer via strongly coupled magnetic resonances," *Science*, Jul 2007.

[69] R. Zhang and C. K. Ho, "Mimo broadcasting for simultaneous wireless information and power transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[70] C. Mikeka and H. Arai, "Sustainable energy harvesting technologies—past, present and future, hap. design issues in radio frequency energy harvesting system," *InTech*, 2011.

[71] H. Ju and R. Zhang, "Throughput maximization in wireless powered communication networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 4086–4091.

[72] T. Kaźmierski and S. Beeby, *Energy Harvesting Systems: Principles, Modeling and Applications*, ser. SpringerLink : Bücher. Springer New York, 2010. [Online]. Available: https://books.google.co.ma/books?id=e4QQc6AEG1IC

[73] P. EFTHYMAKIS, "A rectenna for 5g energy harvesting," 2014.

[74] M. Maso, *Energy Harvesting Oriented Transceiver Design for 5G Networks*. Cham: Springer International Publishing, 2016, pp. 67–95.

[75] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, November 2013.

[76] S. Kim, R. Vyas, J. Bito, K. Niotaki, A. Collado, A. Georgiadis, , and M. M. Tentzeris, "Ambient rf energy-harvesting technologies for selfsustainable standalone wireless sensor platforms," in *Proc. IEEE*, November 2014, pp. 1649–1666.

[77] H. Kanaya, S. Tsukamaoto, T. Hirabaru, D. Kanemoto, R. K. Pokharel, and K. Yoshida, "Energy harvesting circuit on a one-sided directional flexible antenna," *IEEE Microwave and Wireless Components Letters*, vol. 23, no. 3, pp. 164–166, March 2013.

[78] B. Pham and A. Pham, "Triple bands antenna and high efficiency rectifier design for rf energy harvesting at 900, 1900 and 2400 mhz," 06 2013, pp. 1–3.

[79] L. M. Borges, N. Barroca, H. M. Saraiva, J. Tavares, P. T. Gouveia, F. J. Velez, C. Loss, R. Salvado, P. Pinho, R. Gonçalves, N. B. Carvalho, R. Chavéz-Santiago, and I. Balasingham, "Design and evaluation of multi-band rf energy harvesting circuits and antennas for wsns," in *2014 21st International Conference on Telecommunications (ICT)*, May 2014, pp. 308–312.

[80] N. A. Zainuddin, Z. Zakaria, M. N. Husain, B. M. Derus, M. Z. A. A. Aziz, M. A. Mutalib, and M. A. Othman, "Design of wideband antenna for rf energy harvesting system," in *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*, Nov 2013, pp. 162–166.

[81] J. SCHNEIDER1, M. MRNKA, J. GAMEC, M. GAMCOVA, and Z. RAIDA, "Vivaldi antenna for rf energy harvesting," *RADIOENGINEERING*, vol. 25, pp. 666–671, 12 2016.

[82] "Energy harvesting system market by technology (light, vibration, electromagnetic, thermal), application (building and home automation, consumer electronics, industrial, transportation, security), and geography - global forecast to 2022," http://www.marketsandmarkets.com/Market-Reports/energy-harvesting-market-734.html, Apr 2016, accessed: 2019-05-19.

[83] K. hao Liu and P. Lin, "Toward self-sustainable cooperative relays: state of the art and the future," *Communications Magazine, IEEE*, vol. 53, no. 6, pp. 56–62, June 2015.

[84] J. Morra, "Startup's energy harvesting chips run on radio frequencies," https://www.electronicdesign.com/analog/startups-energy-harvesting-chips-run-radio-frequencies, January 2019, accessed: 2019-05-19.

[85] R. Pell, "Boom time for energy harvesting chip market," https://www.eenewseurope.com/news/boom-time-energy-harvesting-chip-market-0, September 2018, accessed: 2019-05-19.

[86] A. M. Zungeru, L. Ang, S. R. S. Prabaharan, and K. P. Seng, "Radio frequency energy harvesting and management for wireless sensor networks," *CoRR*, vol. abs/1208.4439, 2012. [Online]. Available: http://arxiv.org/abs/1208.4439

[87] A. Rajaram, R. Khan, S. Tharranetharan, D. N. Jayakody, R. Dinis, and S. Panic, "Novel swipt schemes for 5g wireless networks," *Sensors*, vol. 19, p. 1169, 03 2019.

[88] X. Ge, H. Jia, Y. Zhong, Y. Xiao, Y. Li, and B. Vucetic, "Energy efficient optimization of wireless-powered 5g full duplex cellular networks: A mean field game approach," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2019.

[89] S. Mahama, D. K. P. Asiedu, and K. Lee, "Simultaneous wireless information and power transfer for cooperative relay networks with battery," *IEEE Access*, vol. 5, pp. 13 171–13 178, 2017.

[90] Q. Li, J. Gao, H. Liang, L. Zhao, and X. Tang, "Optimal power allocation for wireless sensor powered by dedicated rf energy source," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2791–2801, March 2019.

[91] T. A. Khan, R. W. Heath, and P. Popovski, "Wirelessly powered communication networks with short packets," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5529–5543, Dec 2017.

[92] T. Do, "Optimal energy harvesting scheme for power beacon-assisted wireless-powered networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, pp. 802–808, 09 2017.

[93] O. Galinina, H. Tabassum, K. Mikhaylov, S. Andreev, E. Hossain, and Y. Koucheryavy, "On feasibility of 5g-grade dedicated rf charging technology for wireless-powered wearables," *IEEE Wireless Communications*, vol. 23, pp. 28–37, 2015.

[94] A. Nechibvute, A. CHAWANDA, N. Taruvinga, and P. LUHANGA, "Radio frequency energy harvesting sources," *Acta Electrotechnica et Informatica*, vol. 17, pp. 19–27, 12 2017.

[95] M. L. Ku, W. Li, Y. Chen, and K. J. R. Liu, "Advances in energy harvesting communications: Past, present, and future challenges," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1384–1412, Secondquarter 2016.

[96] N. Zhao, F. Yu, and V. Leung, "Wireless energy harvesting in interference alignment networks," *Communications Magazine, IEEE*, vol. 53, no. 6, pp. 72–78, June 2015.

[97] M. Maso, S. Lakshminarayana, T. Q. S. Quek, and H. V. Poor, "The price of self-sustainability for block transmission systems," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1549–1562, Aug 2015.

[98] W. Shin, M. Vaezi, J. Lee, and H. V. Poor, "Cooperative wireless powered communication networks with interference harvesting," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3701–3705, April 2018.

[99] H. Kong, I. Flint, P. Wang, D. Niyato, and N. Privault, "Exact performance analysis of ambient rf energy harvesting wireless sensor networks with ginibre point process," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3769–3784, Dec 2016.

[100] M. L. Memon, N. Saxena, A. Roy, and D. R. Shin, "Backscatter communications: Inception of the battery-free era—a comprehensive survey," *Electronics*, vol. 8, no. 2, 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/2/129

[101] H. Choi and W. Shin, "Harvest-until-access protocol based on slotted aloha for wireless powered dense networks," in *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, Jan 2019, pp. 1–6.

[102] Y. Liu, Y. Zhang, R. Yu, and S. Xie, "Integrated energy and spectrum harvesting for 5g wireless communications," *Network, IEEE*, vol. 29, pp. 75–81, 06 2015.

[103] J. Ghosh and Z. Jakó, "Mac protocol operation in energy harvesting for 5g networks," *ICST Transactions on Mobile Communications and Applications*, vol. 3, p. 153554, 01 2018.

[104] A. Ghazanfari, H. Tabassum, and E. Hossain, "Ambient RF energy harvesting in ultra-dense small cell networks: Performance and trade-offs," *CoRR*, vol. abs/1512.03122, 2015. [Online]. Available: http://arxiv.org/abs/1512.03122

[105] M. S. C. A. H2020-MSCA-ITN-ETN, "Sustainable cellular network harvesting ambient energy, deliverable d3.1 : Wp3 - sustainable networking, d3.1 - intermediate report," http://www.scavenge.eu/wp-content/uploads/2018/02/D3.1.pdf, Tech. Rep., jan 2018, accessed: 2019-07-28.

[106] Z. Chen, Z. Chen, L. X. Cai, and Y. Cheng, "Energy-throughput tradeoff in sustainable cloud-ran with energy harvesting," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[107] G. Qiao, S. Leng, Y. Zhang, M. Zeng, and L. Xu, "Multiple timescale energy scheduling with energy harvesting aided heterogeneous cloud radio access networks," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, July 2016, pp. 1–6.

[108] Q. Pham, F. Fang, H. Vu, M. Le, Z. Ding, L. B. Le, and W. Hwang, "A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art," *CoRR*, vol. abs/1906.08452, 2019. [Online]. Available: http://arxiv.org/abs/1906.08452

[109] T. Dlamini, Ángel Fernández Gambín, D. Munaretto, and M. Rossi, "Online supervisory control and resource management for energy harvesting bs sites empowered with computation capabilities," *Wireless Communications and Mobile Computing*, p. 17, 2019.

[110] Y. Mao, J. Zhang, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, Dec 2016.

[111] H.-S. Nguyen, T.-S. Nguyen, and M. Voznak, "Wireless powered d2d communications underlying cellular networks: design and performance of the extended coverage," *Automatika*, vol. 58, no. 4, pp. 391–399, 2017. [Online]. Available: https://doi.org/10.1080/00051144.2018.1455016

[112] Y. Liu, L. Wang, S. A. R. Zaidi, M. Elkashlan, and T. Duong, "Secure d2d communication in large-scale cognitive cellular networks: A wireless power transfer model," *IEEE Transactions on Communications*, vol. 64, 11 2015.

[113] L. S. Mohjazi, M. Dianati, G. K. Karagiannidis, S. Muhaidat, and M. Al-Qutayri, "Rf-powered cognitive radio networks: technical challenges and limitations," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 94–100, 2015.

[114] D. Do, M. Vaezi, and T. Nguyen, "Wireless powered cooperative relaying using NOMA with imperfect CSI," *CoRR*, vol. abs/1810.00276, 2018. [Online]. Available: http://arxiv.org/abs/1810.00276

[115] M. Basharat, "Joint user grouping and time allocation for noma with wireless power transfer," 2017.

[116] P. T. Hiep and T. M. Hoang, "Non-orthogonal multiple access and beamforming for relay network with rf energy harvesting," *ICT Express*, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S240595951930150X

[117] A. Rauniyar, P. E. Engelstad, and O. N. Østerbø, "Rf energy harvesting and information transmission based on noma for wireless powered iot relay systems," in *Sensors*, 2018.

[118] Y. Zhu, L. Wang, K. Wong, S. Jin, and Z. Zheng, "Wireless power transfer in massive mimo-aided hetnets with user association," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4181–4195, Oct 2016.

[119] R. Hamdi, E. Driouch, and W. Ajib, "Energy management in large-scale mimo systems with per-antenna energy harvesting," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[120] R. Shrestha and G. Amarasuriya, "Wireless energy harvesting in massive mimo with low-dimensional digital precoding," in *Proc. IEEE Global Communication Conference (GLOBECOM)*, 12 2018, pp. 1–7.

[121] T. A. Khan, A. Alkhateeb, and R. W. H. Jr., "Millimeter wave energy harvesting," *CoRR*, vol. abs/1509.01653, 2015. [Online]. Available: http://arxiv.org/abs/1509.01653

[122] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive mimo-noma with simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 131–141, Jan 2019.

[123] M. Wagih, A. S. Weddell, and S. Beeby, "Millimeter-wave textile antenna for on-body rf energy harvesting in future 5g networks," in *Wireless Power Week 2019, IEEE Wireless Power Transfer Conference. (21/06/19)*, June 2019, pp. 1–4. [Online]. Available: https://eprints.soton.ac.uk/431955/

[124] N. I. Miridakis and T. A. Tsiftsis, "A new interweave cognitive radio scheme for out-band energy harvesting systems," *IEEE Access*, vol. 6, pp. 72 225–72 232, 2018.

[125] W. Zhao, R. She, and H. Bao, "Energy efficiency maximization for two-way relay assisted cr-noma system based on swipt," *IEEE Access*, vol. 7, pp. 72 062–72 071, 2019.

[126] B. Alzahrani and W. Ejaz, "Resource management for cognitive iot systems with rf energy harvesting in smart cities," *IEEE Access*, vol. 6, pp. 62 717–62 727, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 334-340 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Modular Multipurpose, Parameter Centered Electronic Health Record Architecture

Norbert Maggi[1,2], Luca Douglas Magnoni[2], Carmelina Ruggiero[1, 2], Roberta Gazzarata[2], Mauro Giacomini[1, 2,*]

[1]*Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genova, Genova, Italy*

[2]*Healthropy S.r.l., Savona, Italy*

A R T I C L E I N F O

Article history:
Received: 15 June, 2019
Accepted: 06 September, 2019
Online: 22 October, 2019

Keywords:
EHR
HL7
Interoperability

A B S T R A C T

Health Information Technology is playing a key role in healthcare. Specifically, the use of electronic health records has been found to bring about most significant improvements in healthcare quality, mainly as relates to patient management, healthcare delivery and research support. Health record systems adoption has been promoted in many countries to support efficient, high quality integrated healthcare. The objective of this work is the implementation of an Electronic Health Record system based on a relational database. The system architecture is modular and based on the concentration of specific pathology related parameters in one module, therefore the system can be easily applied to different pathologies. Several examples of its application are described. It is intended to extend the system integrating genomic data.

1. Introduction

This paper is an extension of work originally presented in IEEE 4th International Forum on Research and Technology for Society and Industry [1]

The collection of health information for patient data management including patient characteristics, treatments and outcomes has been over the years a most important aspect of health care [2, 3]. EHRs, which replace paper medical records to collect people's health information, integrate different and heterogeneous data and are one of the most widely used Health Information Technology (HIT) tools. HIT contributes to the improvement of quality of care, to patient's safety and to the reduction of health care costs by making available accurate documentation and rapid information retrieval and management. The adoption of EHRs has been promoted in many countries, both at government level an in the private sector, even though their adoption is still rather limited, mostly because of standardization, certification, security and privacy aspects and concerns. The adoption of standard based EHR systems has been found to be able to reduce spending for healthcare to a great extent [4]. EHRs can improve patient care and safe practice, however potential risks related to medical error, systems failure, and security and legal responsibility aspects should be considered, identifying barriers for EHRs adoption and ways to address them [5-9]. Moreover, organizational and human factors can promote or hamper the adoption of a EHR [10]. Further relevant aspects of EHRs are interfaces with other information systems, that should be based on standard formats, such as Health Level 7 Clinical Document Architecture (HL7 CDA), HL7 Version 3, openEHR and on controlled vocabularies such as Logical Observation Identifiers Names and Codes (LOINC), "International Classification of Diseases (ICD), Systematized Nomenclature of Medicine (SNOMED) and others [11]. The need for generic and interoperable EHRs has been recognized, specially as relates to interoperability, for the preservation of clinical information across heterogeneous systems [12]. Standards are needed to implement interoperability between systems.

An EHR system contains quantitative data (clinical data), qualitative data (text documents), and medication records. The related amount of data is rapidly expanding, and big data technology is going to play a key role in this respect, specially as relates to natural language processing, clinical decision support, integration of genomics and system biology with EHR data, knowledge dissemination and interaction with patients [13].

The set of requirements that must be met by EHR systems is defined by ISO 18308:2011. According to this standard the scope of the EHR is regarded as broader than documentation, prevention and treatment of illnesses. ISO 18308 relates to shared EHR information and to the EHR system aspects that may be used for the support and coordination of patient-centred continuity of care. An EHR system includes data repositories, directory services,

*Corresponding Author: Mauro Giacomini. Via all'Opera Pia, 13 – 16145 Genoa (Italy) email: mauro.giacomini@dibris.unige.it

knowledge services containing terminological system, care pathways and workflows, user applications and other modules and services. An EHR architecture can be classified according to a data approach, a concepts approach and a process/service approach. These approaches are present in existing systems and in specific domain languages and modelling languages [14]. A medical record system has been developed, adopting the three approaches above. The medical record system that has been developed is based on Health Level 7 (HL7) standard and can be easily tailored to the needs of different clinical wards. The use of this standard allows the system to support interoperability between the main components of the Hospital Information System (HIS), such as the Laboratory Information System (LIS) and the Pharmaceutical Information System (PIS), and can be easily extended for example to the radiology information system (RIS) and to the picture archiving and communication system (PACS). A Service Oriented Architecture (SOA) has been adopted, that can be defined as an open extensible, composable architecture based on autonomous, interoperable and reusable services, implemented as a Web services [15]. SOA has been employed in healthcare after adoption in other sectors, such as banking, transport and industrial automation [16]. SOA is widely used for real time implementation in large distributed systems, mostly because it supports the easy integration of new software within the existing one, reducing the impact on service users and cost [17-20]. Moreover, the SOA paradigm has been successfully implemented in the design of several distributed healthcare systems [21-23].

2. Materials and Methods

The system has been implemented taking into account the specific needs of different departments of the San Martino hospital in Genoa (Italy).

The workflow relating to a visit includes the identification of the patient through the information already present in the HIS obtained by a pseudo-anonym code according to the specifications of UE General Data Protection Regulation, (GDPR) [24]. The information on laboratory analyses is obtained by query of the hospital LIS and the information on therapies is obtained by query of the PIS of the hospital. The data are stored on different platforms which do not interact with each other, therefore its direct use is cumbersome and with possible misunderstandings. In order to address this problem, a system based on a relational database developed with Microsoft SQL Server 2017 has been set up, which has made it possible to create a consistent and logical representation of the information. It is worth noting that, in order to achieve GDPR compliance, the data in the database are stored by pseudo-anonymization techniques. A pseudo-anonymous code is used to query the database, in a secure fashion. The information relating to the identity of the patient is present only at the physician interface level, after authentication [23].

The database has been designed developing a conceptual model of the data, abstracting relevant aspects of the activities and processes relating to patient management. A conceptual model is a set of rules and conventions that allows to move from a reality of interest to a conceptual scheme of the data. The model has been divided into three levels that represent respectively the protagonists of the project (level 0), the clinical event (level 1), and the format of the parameters and results (level 2).

In level 0 the main actors and features of the project have been considered, defining the fundamental entities and the relationships among them. The Entity-Relation (E-R) diagram shows the general logistical aspects of the platform (Figure 1). Patients are followed up and subsequently entered into the database by the operators. Both patients and operators are linked to the centres involved in each clinical study. Each operator may be authorised to access more than one clinical study, and each patient can be involved in more than one study. Moreover, the clinical events relating to each patient may take place in different centres (the latter aspect being of fundamental importance, since different centres may use different units of measurement and ranges of normality). Each study may consider more than one clinical event, some of which may be common to those of other studies, while others may be specific to certain studies only.

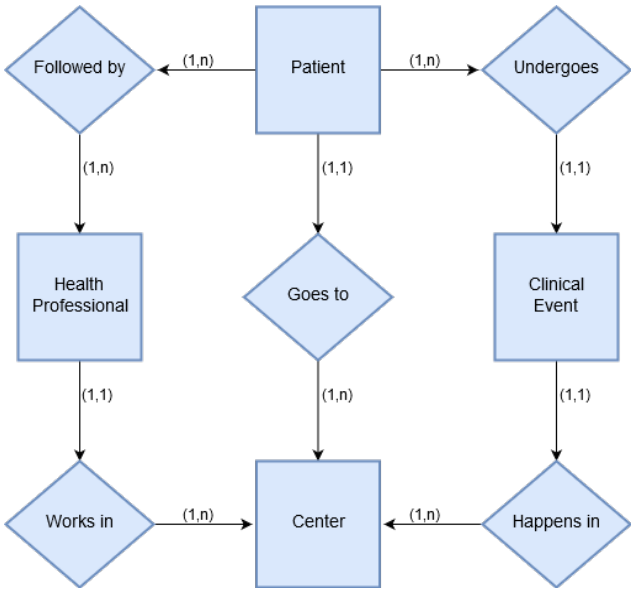


Figure 1. Entity relationship diagram for level 0

Level 1 of the E-R diagram (Figure 2) shows the meta-descriptive approach that has been adopted to manage clinical events and make the database structure as simple and general as possible. The main element is the definition of different Event Types. Within the conceptual definition of systems implemented with E-R diagrams [25], attributes are the elements that describe in detail the characteristics of entities and relationships. By transforming E-R diagrams with relational logic schemes, these attributes become columns of tables in databases. This approach, if applied to our case, binds the set of elements considered in a medical record (which happens in many third-party applications), blocking it in the data base (DB) construction.

When developing the system, it has been chosen to store the names of the elements that are regarded as rows of the table (called "parameters"). Another table (called "categories") has been categorized, and related with the clinical reality through the links in the table that derive from the relation shown in in figure 2 (diamond "considered aspect divided by"). This choice allows to use the same DB and the same software for several tables and to easily characterize the individual parameters in a strong semantic way (using international codes such as - LOINC - ICD - ...). This greatly helps interoperability.

The most important aspects have been divided into different "Categories". These are "containers" of parameters, homogeneous

from the conceptual point of view, but not necessarily homogeneous from a data format point of view (for example, a clinical event of the "Therapy" type has a "Therapy information" category, including parameters such as "Adherence" and "Therapy start date"). Each clinical event has Results, which are the values of the parameters of that particular event.

Events are grouped into types. Parameters are also grouped into types, which are translated and mapped in LOINC using standardized dictionaries, in order to support data reuse.

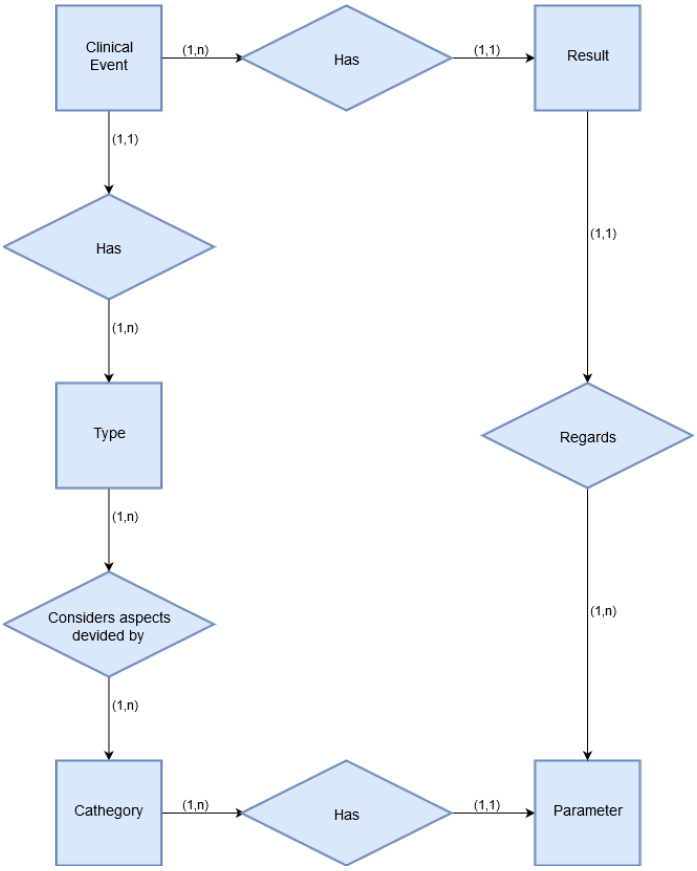


Figure 2. Entity relationship diagram for level 1

Level 2 shows the entities and relationships which are taken into account to manage the Parameters and Results of Clinical Events. Each Parameter has a specific format (typically real integer, string or Boolean), so the results (i.e. the records corresponding to the values of the parameters of an event) can have different formats. The E-R diagram (Figure 3) takes into account the fact that all parameters stored in the database have their own units of measurement and range of normality, but these units and ranges are different for each centre that may be involved in the network. This is of fundamental importance for the work carried out by the physician. In the transition from the conceptual scheme to the logical scheme, great attention has been paid to this aspect, in order to allow each user to work with the tools that are most familiar to him/her. Moreover, the differences between the units of measurement have been taken into account in the data extraction algorithm. All records relating to a single parameter are converted into a standard unit of measurement, defined by the users as the ones which are most widely used.

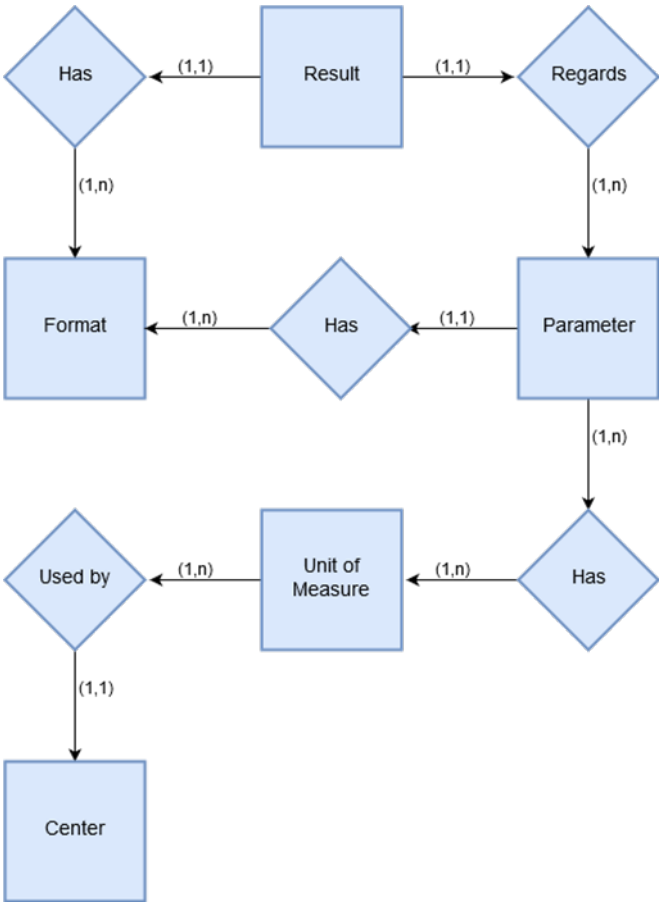


Figure 3. Entity relationship diagram for level 2

The implementation, including the creation of a web client and a web service, has been carried out in Visual Basic .NET using Microsoft Visual Studio 2015 software.

The system architecture has been designed with special reference to interoperability, according to HL7 in the HSSP (Healthcare Services Specification Project). The HSSP, started in 2005 by HL7, focuses on standard software services for healthcare, enabling the use of SOA, which has been recognised as extremely useful for interoperability, code reuse and new needs response [26]. Standards service interfaces for infrastructure reuse capability and specific health facility were considered providing a methodology for the implementation of relevant services [27].

Specifically, the HL7 Clinical Document Architecture (CDA) has been used for the encapsulation of clinical data that can be easily exchanged between applications, retaining their semantic significance [28].

Clinical parameters are associated differently with specific names and internal codes in different hospital and/or health centres. In order to allow the communication among different centres, the codes used by different hospitals/centres have been translated into LOINC language with regard to clinical trials [29], international Anatomical Therapeutic Chemical Classification System (ATC), national Marketing Authorisation (AIC) for drugs, International Classification of Diseases 9th version (ICD9) for diagnosis and ISO Country 3166-1 for the classification of the country of origin of patients.

Table 1. Clinical events for the Maculopathies, HIV, and Tuberculosis

Maculopathies	**HIV**	**Tuberculosis**
Personal data	Blood sample taken	First visit
Anamnestic	Admission history and physical note	Visit
Calendar	Medical history unknown	Microbiological examination
Eye examinations	Therapy	Laboratory examination
Maculopathies	Discharge summarization note	Resistance development
Laser Injection Therapies	Specialist visit	Instrumental examination
Primary Diagnosis	Submission questionnaire	Therapeutic treatment
Secondary Diagnosis	Non-Haart Therapy	Alcohol test
Clinical Folder	Adverse event	
Two Eyes	Qualitative therapy	
Pharmacovigilance Card	Repeated medical history	
Ozurdex Monitoring Card	Vaccination	
Visual Functioning Questionnaire	Microbiological culture	
Retina Plus - Personal data	Failure	
Retina Plus - Access	Instrumental examination	
Retina Plus - Questionnaire	Visit Control	
Telescreening	Pregnancy	
Laboratory Exams		
Plaquenil Protocol		
Certificates		
Allergies		

The “San Martino” hospital in Genoa is equipped with a network infrastructure made of two Local Area Networks (LANs) that are not connected to each other (the hospital LAN and the university LAN). Specifically, the parts of the HIS are placed in the hospital LAN, which is not accessible from external networks for security reasons, because of the hospital firewall. The access to data and to the required information is granted by a virtual private network (VPN), capable to securely connect the two LANs, allowing access for clients belonging to the university network.

The original data are sent through the VPN and, by tunnelling protocols and advanced encryption, are secured by a protective barrier before they are shipped over a potentially dangerous infrastructure such as the Internet. Tunnelling is a multi-protocol (IPsec) data encapsulation process within another IP packet that is shipped over the network. In order to transmit the data to authorized end recipients only, the data are packaged twice by a specific authentication system in which access levels depend on user role.

The strong modularity of the system allows to use the general structure described above for different diseases. The system has so far been used for the implementation of the EHR for the Ligurian HIV network [30-32], for ophthalmology [33, 34] and for infectious diseases [35]. For each of these cases, different clinical events, parameter categories and parameters have been implemented in order to provide a tool which is very suited to the specific needs of physicians and caregivers. Table 1 and table 2 show, respectively, the clinical events and the categories of parameters for the three scenarios.

3. Results

The system that has being developed is hosted in Genoa university network. The user interface of the system is divided into several parts, each concerning a specific section of the medical visit. The EHR system is managed through a web interface that, for security reason, can only be accessed within the Genoese hospital and university LANs. The system interaction with the physician or caregiver takes place as follows. When the user logs in he/she is redirected to the page relating to the specific patient selection and the examination starts. After obtaining access permission, the caregiver can either manually enter the data of a new patient or see the data for a patient who is already present in the data base.

For a new entry, the related information is retrieved from the hospital databases and, after translation into LOINC, is stored in the EHR system database. The clinical parameters are retrieved from the LIS and the information on the laboratory tests is stored by an array of CDA which is sent to the web service which stores it in the system database.

A section of the clinical diary allows physicians to inspect the exams that are available for each patient. The physician can see all specific visits, their dates and parameters. The information for each clinical event is retrieved and updated by algorithms that taking into account the ethical committee permissions.

The creation of the web pages for user interface is also guided by the structure of the DB, i.e. a page is formed for each event, within which all parameters related to that event are presented. The type of value for each parameter (integer, real, list or free text, date) automatically inserts the selection of the most suitable input tool for that type (textbox, dropdown list, calendar, ...). The measurement units for the numerical parameters and the normal ranges also (possibly linked to the patient's personal data, age, sex, height, weight) are also automatically inserted.

Table 2. Categories for Maculopathies, HIV, and Tuberculosis

Maculopathies	**HIV**	**Tuberculosis**
Personal data	Phenotype and viremia	Symptomatology
Anamnestic	Carbohydrate metabolism	Inflammation indices
Calendar	Cardiovascular profile	Diagnosis of tubercular infection
Visuals and symptoms	Liver functionalities	Sputum microbiological examination
Front and rear segment	Renal function	Microbiological examination of bronchoaspirate
Laser Injection Therapies	Oxidative stress	Microbiological examination of urine
Primary diagnosis	NK phenotype	Microbiological examination of faeces
Secondary diagnosis	Anamnestic	Microbiological examination of other material
Topical therapies in progress	Non-Haart Therapies	Resistance
Expert	Inflammation parameters	Instrumental examinations
Folder	Leukocyte formula hemochrome	Localization
Exclusion criteria	Monocyte populations	Inpatient
Prohibited treatments	Mocyte activation markers	Comorbidity
Interruption of treatment	Inflammation mediators	Treatment
Pharmacovigilance Card	Lymphocyte Proliferation	Side effects
Ozurdex Monitoring Card	Therapy information	First-line oral agents
Visual Functioning Questionnaire	Drugs therapy	B: Injectable agents II line
Appendix of Optional Additional Questions	Information about hospitalization	A: Fluoroquinolones
Retina Plus Personal data	Diagnosis of hospitalization	C: Other second line agents
Retina Plus Access	Diagnostic and therapeutic procedures	D1: Deputy agents
Retina Plus Questionnaire	Discharge information	Laboratory examination
Telescreening	Comorbidity	CBC
Grape harvesting	Discharge diagnosis	D2: Deputy agents
Laboratory Exams	Non-Haart Drugs	D3: Deputy agents
Instrumental examinations	Objective examination	
Microbiology tests	Chemistry	
Plaquenil start therapy	Specialist visit information	
Plaquenil	HIV information	
Blood pressure	HBV information	
Certified	HCV information	
Allergies	Hepatitis	
	T lymphocyte activation markers	
	Side effects therapy	
	Submission questionnaire	
	Non-Haart Drugs	
	Information on drugs therapy	
	Details of side effects therapy	
	T reg	
	Concomitant therapy	
	Adverse event therapy	
	HCV adverse event	
	Qualitative HCV therapy	
	Qualitative HIV Therapy	
	Vaccination	
	Microbiology	
	Non-hepatitis serologies	
	Instrumental examination information	
	Control visit information	
	Resistance	
	Pregnancy	
	Barthel ADL (Activities of Daily Living)	
	Barthel MOB	
	IADL	

Table 3. Use of EHRs for pathologies

	Initial year	Patients	Parameters	Events	Values	Sessions
HIV	2013	4540	802	200424	6481673	9033
Maculopathies	2014	5554	801	52335	955753	2166
Tuberculosis	2016	324	125	6460	103691	2324

4. Discussion and Conclusions

An EHR system has been set up in order to meet the needs of medical and clinical personnel. The main aspects that have been considered relate to the management of visits and to the setting up of a clinical diary according to the standards currently in use.

The objectives of the study have been met as shown in table 3, in which the system's application to maculopathies, HIV, and tuberculosis are described. Specifically, for the three pathologies the table shows the initial year, the numbers of patients which have been recorded, the numbers of related parameters, the number of events and the number of values that have been stored. The last column, named "session", shows the number of days in which an operator queries an EHR.

The system aims to simplify physician's, nurses' and caregiver's work for administration and management of treatment and follow-up. Moreover, it aims to support multicentric tracking for patients who are treated in several hospital facilities and the conduction of clinical trials. The system was set up focusing on parameters and parameter types, grouping them in one specific module. This facilitates the use of the system for different pathologies, because the pathology related data are in one module per pathology. The use of one single section for the management of different pathologies allows for easy use by different departments of one hospital and by different hospitals. This design choice also facilitates data reuse.

The main problems and limitation that have been encountered relate to data availability, because of data anonymity requirements for security and privacy. In the first place, this problem has been circumvented requiring the hospital to make the data pseudo-anonym, while keeping the link between the pseudo anonym code and the patient identity separate from the data and managed by the hospital system only, on physically separated platform. Moreover, the data which are copied onto the EHR system platform are encrypted by up to date techniques which protect them from attacks. Therefore, even if the data were red this would not provide any useful information

It is intended to evaluate the system performance by questionnaires addressing usability by physician, nurses and care giver.

Further developments mostly relate to the integration of genomic data in the system [36, 37]. This aspect is becoming increasingly important and the parameter focused architecture of the system is suitable in this respect. The integration of genotype and phenotype information in the EHR and the insertion of genomic information in clinical workflows would facilitate the planning, administration and evaluation of genetically enabled care.

References

[1] N. Maggi, C. Ruggiero, M. Giacomini, "Towards an Electronic Medical Record for Precision Oncology: Information and Knowledge Management," in 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI), 2018. https://doi.org/10.1109/rtsi.2018.8548422

[2] B. Blobel, *Analysis, design and implementation of secure and interoperable distributed health information systems*. Amsterdam ; Oxford: IOS Press, 2002.

[3] P. Coorevits, M. Sundgren, G. O. Klein, A. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, A. Schmidt, P. Singleton *et al.*, "Electronic health records: new opportunities for clinical research," *Journal of Internal Medicine,* 274(6), 547-560, 2013. https://doi.org/10.1111/joim.12119

[4] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, R. Taylor, "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs," *Health Affairs,* 24(5), 1103-1117, 2005. https://doi.org/10.1377/hlthaff.24.5.1103

[5] V. Palabindala, A. Pamarthy, N. R. Jonnalagadda, "Adoption of electronic health records and barriers," *Journal of Community Hospital Internal Medicine Perspectives,* 6(5), 2016. https://doi.org/10.3402/jchimp.v6.32643

[6] J. M. Gesulga, A. Berjame, K. S. Moquiala, A. Galido, "Barriers to Electronic Health Record System Implementation and Information Systems Resources: A Structured Review," *Procedia Computer Science,* 124(544-551, 2017. https://doi.org/10.1016/j.procs.2017.12.188

[7] S. Ajami, T. BagheriTadi, "Barriers for Adopting Electronic Health Records (EHRs) by Physicians," *Acta Informatica Medica,* 21(2), 2013. https://doi.org/10.5455/aim.2013.21.129-134

[8] A. Boonstra, M. Broekhuis, "Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions," *BMC Health Services Research,* 10(1), 2010. https://doi.org/10.1186/1472-6963-10-231

[9] C. M. DesRoches, D. Charles, M. F. Furukawa, M. S. Joshi, P. Kralovec, F. Mostashari, C. Worzala, A. K. Jha, "Adoption Of Electronic Health Records Grows Rapidly, But Fewer Than Half Of US Hospitals Had At Least A Basic System In 2012," *Health Affairs,* 32(8), 1478-1485, 2013. https://doi.org/10.1377/hlthaff.2013.0308

[10] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, D. Z. Sands, "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption," *Journal of the American Medical Informatics Association,* 13(2), 121-126, 2006. https://doi.org/10.1197/jamia.M2025

[11] C. J. McDonald, "The barriers to electronic medical record systems and how to overcome them," *J Am Med Inform Assoc,* 4(3), 213-21, May-Jun 1997.

[12] D. Kalra, "Electronic Health Record Standards," *Yearbook of Medical Informatics,* 15(01), 136-144, 2018. https://doi.org/10.1055/s-0038-1638463

[13] T. B. Murdoch, A. S. Detsky, "The Inevitable Application of Big Data to Health Care," *Jama,* 309(13), 2013. https://doi.org/10.1001/jama.2013.393

[14] B. Blobel, "Interoperable EHR Systems–Challenges, Standards and Solutions," *European Journal for Biomedical Informatics,* 14(2), 10-19, 2018.

[15] T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*. Pearson Education, Limited, 2016.

[16] S. R. Loya, K. Kawamoto, C. Chatwin, V. Huser, "Service Oriented Architecture for Clinical Decision Support: A Systematic Review and Future Directions," *Journal of Medical Systems,* 38(12), 2014. https://doi.org/10.1007/s10916-014-0140-z

[17] R. Gazzarata, F. Vergari, T. S. Cinotti, M. Giacomini, "A standardized SOA for clinical data interchange in a cardiac telemonitoring environment," *IEEE J Biomed Health Inform,* 18(6), 1764-74, Nov 2014. https://doi.org/10.1109/JBHI.2014.2334372

[18] E. Vasilescu, S. K. Mun, "Service Oriented Architecture (SOA) implications for large scale distributed health care enterprises," in 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, Conference Proceedings, 2006. https://doi.org/10.1109/Ddhh.2006.1624805

[19] R. Gazzarata, B. Giannini, M. Giacomini, "A SOA-Based Platform to Support Clinical Data Sharing," *Journal of Healthcare Engineering,* 2017(Article ID 2190679), 1-24, 2017. https://doi.org/10.1155/2017/2190679

[20] R. Gazzarata, M. Giacomini, "A Standardized SOA for Clinical Data Sharing to Support Acute Care, Telemedicine and Clinical Trials," *European Journal for Biomedical Informatics,* 12(01), 2016. https://doi.org/10.24105/ejbi.2016.12.1.9

[21] H. Kondylakis, B. Claerhout, M. Keyur, L. Koumakis, J. van Leeuwen, K. Marias, D. Perez-Rey, K. De Schepper, M. Tsiknakis, A. Bucur, "The INTEGRATE project: Delivering solutions for efficient multi-centric clinical research and trials," *J Biomed Inform,* 62(Aug 2016), 32-47, 2016. https://doi.org/10.1016/j.jbi.2016.05.006

[22] R. Alonso-Calvo, S. Paraiso-Medina, D. Perez-Rey, E. Alonso-Oset, R. van Stiphout, S. Yu, M. Taylor, F. Buffa, C. Fernandez-Lozano, A. Pazos *et al.*, "A semantic interoperability approach to support integration of gene expression and clinical data in breast cancer," *Comput Biol Med,* 87(Aug 2017), 179-186, 2017. https://doi.org/10.1016/j.compbiomed.2017.06.005

[23] G. Gazzarata, R. Gazzarata, M. Giacomini, "A standardized SOA based solution to guarantee the secure access to EHR," *Conference on Enterprise Information Systems/International Conference on Project Management/Conference on Health and Social Care Information Systems and Technologies, Centeris/Projman / Hcist 2015,* 64(1124-1129, 2015. https://doi.org/10.1016/j.procs.2015.08.582

[24] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, 2016.

[25] A. Silberschatz, H. F. Korth, S. Sudarshan, *Database system concepts*, 6th ed. ed. New York: McGraw-Hill Higher Education ; London : McGraw-Hill [distributor], 2011.

[26] M. W. Bridges, "SOA in healthcare, Sharing system resources while enhancing interoperability within and between healthcare organizations with service-oriented architecture," *Health Management Technology,* 28(6), 6, 8, 10-6, 8, 10, 2007.

[27] K. Kawamoto, A. Honey, K. Rubin, "The HL7-OMG Healthcare Services Specification Project: motivation, methodology, and deliverables for enabling a semantically interoperable service-oriented architecture for healthcare," *J Am Med Inform Assoc,* 16(6), 874-81, Nov-Dec 2009. https://doi.org/10.1197/jamia.M3123

[28] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. Shabo Shvo, "HL7 Clinical Document Architecture, Release 2," *J Am Med Inform Assoc,* 13(1), 30-9, Jan-Feb 2006. https://doi.org/10.1197/jamia.M1888

[29] *Logical Observation Identifiers Names and Codes (LOINC)*. Available: http://loinc.org

[30] P. Fraccaro, V. Pupella, R. Gazzarata, C. Dentone, G. Cenderello, P. De Leo, F. Bozzano, G. Casalino Finocchio, A. De Maria, D. Fenoglio *et al.*, "The Ligurian Human Immunodeficiency Virus Clinical Network: A Web Tool to Manage Patients With Human Immunodeficiency Virus in Primary Care and Multicenter Clinical Trials," *Medicine 2.0,* 2(2), 2013. https://doi.org/10.2196/med20.2712

[31] B. Giannini, N. Riccardi, G. Cenderello, A. Di Biagio, C. Dentone, M. Giacomini, "From Liguria HIV Web to Liguria Infectious Diseases Network: How a Digital Platform Improved Doctors' Work and Patients' Care," *AIDS Research and Human Retroviruses,* 34(3), 239-240, 2018. https://doi.org/10.1089/aid.2017.0064

[32] M. Bonetto, S. Mora, N. Maggi, C. Ruggiero, M. Giacomini, "Standards for the Reuse of Clinical Data in Research: the Connection of the Liguria HIV Network and the CISAI Cohort," in 2019 Global Medical Engineering Physics Exchanges/ Pan American Health Care Exchanges (GMEPE/PAHCE), 2019. https://doi.org/10.1109/gmepe-pahce.2019.8717316

[33] P. Fraccaro, M. Nicolo, M. Bonetto, M. Giacomini, P. Weller, C. E. Traverso, M. Prosperi, D. Osullivan, "Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach," *BMC Ophthalmology,* 15(1), 2015. https://doi.org/10.1186/1471-2415-15-10

[34] M. Bonetto, M. Nicolo, R. Gazzarata, P. Fraccaro, R. Rosa, D. Musetti, M. Musolino, C. E. Traverso, M. Giacomini, "I-Maculaweb: A Tool to Support Data Reuse in Ophthalmology," *IEEE Journal of Translational Engineering in Health and Medicine,* 4(1-10, 2016. https://doi.org/10.1109/jtehm.2015.2513043

[35] N. Riccardi, B. Giannini, M. L. Borghesi, L. Taramasso, E. Cattaneo, G. Cenderello, F. Toscanini, M. Giacomini, E. Pontali, G. Cassola *et al.*, "Time to change the single-centre approach to management of patients with tuberculosis: a novel network platform with automatic data import and data sharing," *ERJ Open Research,* 4(1), 2018. https://doi.org/10.1183/23120541.00108-2017

[36] N. Maggi, R. Gazzarata, C. Ruggiero, C. Lombardo, M. Giacomini, "Cancer precision medicine today: Towards omic information in healthcare systems," *Tumori Journal,* 105(1), 38-46, 2018. https://doi.org/10.1177/0300891618792473

[37] N. Maggi, L. M. Douglas, C. Ruggiero, R. Gazzarata, M. Giacomini, "Information Technology System Including Patient Generated Health Data for Cancer Clinical Care and Research," *Studies in health technology and informatics,* 261(289-293, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 340-350 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Hybrid Technique for Enhancing Underwater Image in blurry conditions

Chrispin Jiji[*,1], Nagaraj Ramrao[2]

[1] *Kalasalingam Academy of Research and Education, Krishnankoil-626126, Tamilnadu, India*

[2] *Department of ECE, The Oxford College of Engineering, Bengaluru-560068, Karnataka, India*

ARTICLE INFO

Article history:
Received:03 March, 2019
Accepted:10 April, 2019
Online: 17 April, 2019

Keywords:
Image Restoration
Image Enhancement
Color Adjustment
Hybrid Technique

ABSTRACT

Enhancing underwater visualization using hybrid technique is generally employed into oceanic production. Through growing oceanic learning, undersea processing has drawn extra importance owing towards necessary task of picture towards attaining data. Although, suitable to reality of dust-like constituent and beam reduction, undersea descriptions continually experience small contrast and color alteration. In this paper, we estimate submerged beam transmission progression also intend an effectual means to defeat the backscatter trouble. Our scheme generally includes three steps; first, we reconstruct the picture using adaptive regularization. Second, we separate the reconstructed picture with weighted decomposition; third, we exploit color adjustment along with dehazed process via gradient guided filter towards holding dual mechanisms independently; at last, re-establish fine effect, we use hybrid technique for enhancing the picture. The tentative outcome illustrates that our proposed process extensively get better quality of unclear submerged descriptions. In common, our proposed method verified as well-performed and effectual than existing technique.

1. Introduction

In the modern years, scientists and researchers have exposed their main concern in capturing the aquatic life beneath the water. Although the statement that 70% of the globe encloses water includes enormous energy, abundant mineral assets and biological resources. The living beneath the water is still not as much look at. Conversely, challenges related by capturing descriptions beneath the water have been complex to defeat, owing to carry-over haze and color cast. The muddy nature of water suitable to process element such as mineral deposits, sand and plankton, showed a huge difficulty in undersea look at region. These elements set up lack of clarity in undersea descriptions which spoils the clearness and ocular analysis of the picture and in turn, inappropriate for any advance use. Researchers have finished that the spreading and absorption property of beam underneath are the chief basis for this deformation [1]. While water is roughly 800 times denser than air, so when beam penetrate from air to water, it come across reflection fact at the surface owing to which only unfair measure of beam go through water. Afterwards the beam undergoes spreading cause when it strikes and element and minerals suspend in water. Scattering diverges the beam in dissimilar directions which diminish the measure of beam falling on the object captured underneath. The end effect of spreading outcome is that the undersea captured descriptions are dim in look [2]. A further problem is causal to lack of clarity of picture that the water molecules take up certain measure of beam when beam hit on them. Because the mass of water is larger than air, it employs dissimilar absorption possessions for diverse beam wavelengths, so the color of undersea descriptions is typically deformed. Generally, red part vanishes first as of its highest wavelength, where blue beam, by small wavelength, convey the greatest path beneath water. Therefore, undersea descriptions are often dominated via blue color. In common, color cast and beam scattering reason deviate color along with contrast degradation into descriptions attained undersea as shown in Figure 1.

In contrast, our system uses dual phases specifically Adaptive Regularization for restoration and Hybrid technique intended for improving its quality. In this work we suggested for merging the thought mainly dictionary knowledge by structural cluster along with picture restoring functions.

Our advance establishes novel regularization to beat a superior stability among local and non-local strength. Generally, our algorithm interprets learning through local and non-local similarity give inference of each picture elements. The restored picture decomposes into dissimilar elements for signifying their

[*]Chrispin Jiji, 8951627124 & chrispinjij@gmail.com

basic depiction, later uses diverse development for every part individually along with hybrid technique introduced for rebuilding enhanced picture. The projected process is close towards or even slightly superior to the state-of-the-arts. Tentative results exposed that projected process extensively outperforms many important process in existing arts.

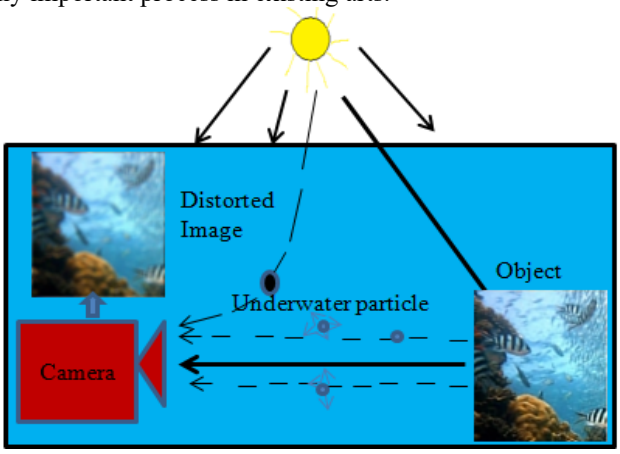


Figure 1: Underwater Imaging System

The summary of this work described as follow. Section 2 portray existing art. Section 3 projects hybrid technique using gradient guided filter for enhancing undersea descriptions. Section 4 illustrates numerical experiments. At last, Section 5 concludes this work.

2. Existing art

Enhancement system for undersea degradation typically explore the undersea optical form and balance bad possessions caused by water along with elements, or simply via undersea processing process for enhancing unclear imagery. State-of-arts are generally considered meant for single image enhancement, since multi-images may limit performance choice, slow down the processing speed. Instinctively, handling this concern well, undersea optical form must consider and investigated at first. Among different undersea imaging methods, the representation method [3], give a complete undersea optical form, within which beam transmission path, beam reduction relation of dissimilar color channel, vertical as well as horizontal deepness of undersea surroundings, non-natural beam possessions and etc are in use. This process is entirely with real form and meet outstanding for a few undersea problems, conversely, its evaluation procedure is difficult and time overwhelming, which is hard to meet, but it gives a full-scale study of undersea surroundings. Fusion-based method, which combines processing outcome by color correction system and histogram equalization method, it also makes a superior outcome for several submerged environments, though, it merely uses scheme of image processing as well as ignoring submarine features, hence ending results undergo more or less fog, data defeat as well as over-enhancement difficulty. To explain these problems well, both processes combined and projected a new method. Semi-inverse approach in [4] employed towards perceiving foggy region and then expected atmospheric rate in plain areas. At last, they meet enhanced picture using a simple weighted fusion.

For few picture, particularly scenes underneath, mist exposure using semi-inverse system will be unsuccessful suitable to the small dissimilarity among true and the inverse picture. This means if hazy region exposure not succeeds and obtains a bad performance. To solve this problem, a simple fusion-based dehazed method using a dissimilar descriptions also resultant from true picture [5], [6]. The primary picture estimate white balance process, and the next input picture through a simple linear change for enhancing contrast of hazy picture. At last, the pixel-fusion scheme uses three weight maps for combining that imagery for enhancing visibility of foggy picture also extended fusion-based scheme [7]. Fusion-based dehazed procedure is easy, speedy as well as meets alike result as the DCP scheme in [8] and [9]. But it is not succeeding when the picture includes inhomogeneous haze. The cause is that this process did not take into account the deepness data of a hazy picture. The dual descriptions attained by gamma correction by various scales for sandstorm enhancement [10]. The enhanced fusion-based approach using various descriptions uses single picture dehazed scheme in [11]. Fusion approach is new and efficient intended for single picture defogging. The primary picture is for enhancing dissimilarity; in addition, later picture mainly for compensating color deformation as well as decrease halo effect plus noise. While above techniques are fast as well as simple, they cannot make excellent picture. Amusingly, the multi-scale progression can be approximated with well-organized and visually satisfying single-scale process [12]. HR fusion in [13] has to be measured as a possible balance, useful to our process when high resolution is required.

The algorithm [14] are effective methods with low computational complexity also great performance, which stretch histograms of picture for improving the dynamic range compress the display area towards enhancing dissimilarity, but these methods often cause over-enhancement, amplify the noise and show poor effect on color correction. Owing towards the drawbacks of single image processing, some relatively complicated methods are proposed. Meanwhile, [15] proposes a retinex-based approach to deal with undersea issues, however, performance of this system is not distinct, color cast as well as contrast degradation still exist. The system may guess a uniform [16, 17] or an extra reasonable non-uniform [18] background-radiance. Enhancement technique eliminates the colour cast [19, 20] with overall white balancing.
This work is extensive to the effort primarily obtained through ICCIC [21]. Moreover, the projected work is for Enhancing Underwater Image in blurry conditions using hybrid technique. Revised outcome much improves lighting plus color cast trouble during submerged descriptions. Excellence process not essential to resolve substantial type of corrupted representation, other than fairly enhances dissimilarity also gets superior picture excellence as individual illustration view. Experimental results show good performance meant for underneath representation.

2.1. Motivation

Underwater distorted descriptions typically to explore optical model also compensating bad effects caused by underwater element uses processing scheme to restore performance of distorted images. The state-of-arts are mostly designed for single-image enhancement, since multi-images may limit performance range and slow down processing speed. Towards handling this problem well, initially we considered Image re-establishment process. The physical system of ocular imaging, reverse

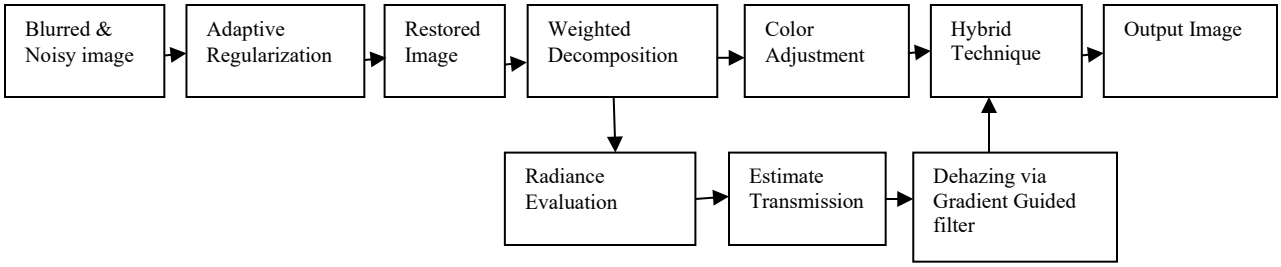


Figure 2: Block diagram of Proposed Method

degradation process and balance distortion to get clear imagery but it has small visibility and color cast problem. Image enhancement process do not get the source of picture degradation into account, but mainly use targeted version to get better dissimilarity, detail, and ocular possessions. It provides complete undersea ocular form, where the illuminated transmission path, beam reduction of dissimilar color path taken into concern totally based on physical model and achieves excellent for some underwater backscatter issues. However, its evaluation development is difficult and time-consuming, but it gives full-scale analysis of undersea situation. To enhance dissimilarity, visibility, color spread of restored imagery, fusion-based model which boost picture by fusing several input imageries have received much attention. The processing results by color constancy and dehazed algorithm, generates good results. However, it only uses technique of image processing and so its results suffer from more or less haze, information failure as well as over-enhancement problem.

In order to solve this issue well, we proposed a Hybrid Technique for Enhancing Underwater Image in blurry conditions by joining dehazed technique via gradient guided filter with color balance of the restored imagery.

3. Proposed Method

As depicted in Figure 2 our new scheme primarily to Enhance Undersea picture in blurry conditions using hybrid technique. Initially, we recreate the picture using adaptive regularization. Secondly, we split reconstructed picture using weighted decomposition. Thirdly, improve contrast and texture information using Dehazing via gradient guided filter. Here we estimate global radiance by obtaining average rate vector as brightest rate, transmission map evaluation mainly for enhancing texture feature. Fourth, a color adjustment towards enhancing color. At last, we used hybrid technique for obtaining better picture by computing the mechanisms of dual weights in addition to efficient.

3.1. Adaptive Regularization for Restoration

For an image $d \in P^{HXW}$, we believe patch based depiction $d_m = Pd \in P^N$, $m = 1,2,.....N$ where P indicates a parameter to extract patch d_m as of d . Suppose a known dictionary φ_m meant for d_m , we get $d_m = \varphi_m \alpha_m$ entails α_m as sparse depiction of d . In particular, restoration describes

$$D = \left(\sum_{i=1}^{N} P_i^T P_i \right)^{-1} \left(\sum_i P_i^T d_i \right) \tag{1}$$

The iterative shrinkage key towards l1-estimation is widely used in previous art. Since the evaluation of $\hat{d}$ get simplified by convergence, trained by a novel dictionary. At present let us include adaptive regularization term [22, 23]. For every patch d_m k-nearest-neighbors patch signify $d_m^{(1)},.....d_m^{(k)}$. The weighted guess is representing

$$\hat{d_i} = \sum_{t=1}^{k} w_{m,t} D_i^{(t)}, w_{m,t} = \exp\left(-\left\| D_m - D_m^{(t)} \right\| / h \right) \tag{2}$$

where h stands for constant to manage the window. Since (2) local form by picture roughly considered with

$$D_L(d) = \sum_{m=1}^{M} \left\| d_m - a_{k_m}^T \chi_m \right\|_2^2 \tag{3}$$

In matrix type (3) represents

$$D_L(d) = \sum_{m=1}^{M} \left\| (I - V) d_m \right\|_2^2 = \left\| (I - V) \varphi \alpha \right\|_2^2 \tag{4}$$

where I specify identity matrix and

$$V(m,n) = \begin{cases} a_i, if\ d_n \text{ is an element of } \chi_m, a_m \in a_{k_m} \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

It follows from (2) that the nonlocal similarity in a picture roughly considered through

$$D_{NL}(d) = \sum_{m=1}^{M} \left\| d_m - \sum_{t=1}^{k} w_{m,t} d_m^{(t)} \right\|_2^2 \tag{6}$$

which can be written more efficiently in a matrix type as

$$D_{NL}(d) = \sum_{m=1}^{M} \left\| (I - W) d_m \right\|_2^2 = \left\| (I - W) \varphi \alpha \right\|_2^2 \tag{7}$$

Where I stands for the identity and

$$W(m,l) = \begin{cases} b_i^l, if\ d_m^l \text{ is an element of } \beta_m, b_m^l \in b_m \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

Joining (4) and (7), we have the next novel regularization as

$$\hat{\alpha} = \arg\min \left\{ \left\| d - H\varphi\alpha \right\|_2^2 + \sum_{m=1}^{M} \sum_{n=1}^{N} \lambda_{mn} \left| \alpha_{m,n} \right| \right\} + \gamma \left\| (I - V) \varphi \alpha \right\|_2^2 + \eta \left\| (I - W) \varphi \alpha \right\|_2^2 \tag{9}$$

where a novel regularization factor η manage stability among local variation with nonlocal strength. The nonlocal regularization term not only imposes nonlocal similarity however as well not directly assist the evaluation of signal variance essential with λ_{mn} as $\lambda_{mn} = \frac{2\sqrt{2}\sigma_w^2}{\left|\hat{\sigma}_{m,n}\right| + \varepsilon}$ where $\sigma_{m,n}$ and σ_w signifies the variance of signal and noise correspondingly.

3.2. Weighted Decomposition

Weighted decomposition used to separate imagery into various parts and then perform several schemes according to their characteristics. Techniques like wavelet transform and weighted decomposition are most well-liked process. Former scheme is time consuming and cannot take out the characteristics. Later on initiate a weighted decomposition means, which divide true picture into dual part, one conveys the illuminant part as another states the reflectance part of picture as:

$$I_c(d) = I_c^R(d) + I_c^I(d) \tag{10}$$

where $c \in \{r,g,b\}$ signifies every color part of picture, $I_c^R(d)$ denotes reflectance part and $I_c^I(d)$ represents illumination part.

$$I_c^R(d) = \Upsilon . I_c(d) \tag{11}$$

$$I_c^I(d) = (1-\Upsilon) . I_c(d) \tag{12}$$

where Υ denotes weighted factor which improve the dissimilarity of reflectance part for removing backscatter as

$$\Upsilon = \xi . \frac{I_c(d)}{I_c^{\max}} \tag{13}$$

Where $I_c^{\max}$ denotes maximal pel rate of color, ξ signify a direct cause by resolving reflectance weight, $\xi \in (0,1)$, if $\xi = 0$, the entire picture treated as illuminate part, while $\xi = 1$, the entire picture is denoted by reflectance part.
The backscatter outcome simply exists with illuminate part, whereas reflectance part simply experiences as of color deformation.

3.3. Radiance Evaluation

To approximate background beam, a supreme approach is to select a picture element, while color deformation and dissimilarity degradation are distance dependent. Though, in this system, scenes are vivid than surrounding beams, may perhaps an unwanted collection outcome. To get precise outcome, we exploit a hierarchical searching means using quad-tree subdivision [24]. Initially, the picture broken up into four identical rectangular areas, after that used for all area by estimating mean rate subtracted from standard deviation as

$$Score_l = \frac{1}{3N} \sum_{c\in\{r,g,b\}} \sum_{d=1}^{N} I_l^c(d) - \frac{1}{3} \sum_{c\in\{r,g,b\}} \sqrt[2]{\frac{\sum_{d=1}^{N} \left(I_l^c(d) - \bar{I}_l^c \right)^2}{N}} \tag{14}$$

Where $l = 1,2,3,4$ stand for four image areas, N denotes the picture element number contained by the area, $I_l^c(d)$ signifies picture element of d rate c section in l area, $\bar{I}_l^c$ represents mean picture element rate of c part in l area. Later, we choose the area by least variance, as well as split into quadrant areas. This procedure continues until the size is smaller than threshold. The average rate attained as surrounding beam and considered as brightest rate with the complete representation.

3.4. Estimate Transmission

The transmission rate cannot be selected randomly; dissimilarity improvement with data defeat lessening must chose keen on concern. First, dissimilarity improvement cost task, E_c, with data loss cost task E_i reduce dual function concurrently, while dissimilarity improvement cost as negative summation of C_{MSE} of every color channels, furthermore data defeat cost as summation of square of reduced picture elements as

$$E_c = -\sum_{c\in\{r,g,b\}} \sum_{d=1}^{N} \frac{\left(J_c^T(d) - \bar{J}_c^T \right)^2}{N} = -\sum_{c\in\{r,g,b\}} \sum_{d=1}^{N} \frac{\left(I_c^T(d) - \bar{I}_c^T \right)^2}{N.t(d)^2} \tag{15}$$

$$E_i = -\sum_{c\in\{r,g,b\}} \sum_{d=1}^{N} \left\{ \left(\min\left\{0, J_c^T(d)\right\} \right)^2 + \left(\max\left\{0, J_c^T(d) - 255\right\} \right)^2 \right\} \tag{16}$$

At last, for every local region the best transmission rate $t(d)$ is approximated by reducing the later task as,

$$E = E_c + \gamma . E_i \tag{17}$$

where γ denotes weighted factor to organize the control of data loss cost.

3.5. Dehazing via Gradient Guided Filter

The Guided filter in [25] refine transmission map but cannot signify fine next to a few boundaries. While there might be a few halos in the descriptions. The halos decrease the visual excellence of the resultant descriptions. Hence Gradient guided filtering is an effectual and process transmission map, use input picture as guidance can have executed by linear system as

$$(L + \lambda . U) . t = \lambda . \tilde{t} \tag{18}$$

where L represents matting Laplacian matrix, U denotes identity matrix, and λ signify regularization factor. The Laplacian matrix represents

$$L_{ij} = |\omega|\left(\delta_{ij} - W_{ij}\right) \tag{19}$$

where $|\omega|$ stands for number of picture elements in a block, δ_{ij} denotes Kronecker delta, and W_{ij} represents gradient guided filter [26] kernel weight as:

$$W_{ij} = \frac{\left(\chi\left(D^{'}\right)+\varepsilon\right)}{mn}\sum_{p}^{m}\sum_{p}^{n}\frac{1}{\chi\left(D^{'}\right)+\varepsilon} \tag{20}$$

where $\chi\left(D^{'}\right) = \sigma_{G,1}\left(D^{'}\right)\sigma_{G,\xi_1}\left(D^{'}\right)$, ξ_1 denotes dimension of filter and D signify guidance picture.

To reach a precise transmission map as well as fine points, the advanced diffusion map split up into smooth and detailed part as

$$t = t_{smooth} + t_{\det ail} \tag{21}$$

where t_{smooth} signify smooth part of transmission map while $t_{\det ail}$ denotes detailed part. For smooth part, we obtain Gaussian low-pass filter as effective smoothing filter, and denoted as:

$$G\left(d,b\right) = \frac{1}{2\pi\sigma^2}.e^{\frac{\left(d^2+b^2\right)}{2\sigma^2}} \tag{22}$$

where σ denotes scale factor of Gaussian blur. Therefore, the smooth and detailed part can estimate as

$$t_{smooth} = t * G \tag{23}$$

$$t_{\det ail} = t - t_{smooth} \tag{24}$$

Following, the improved transmission map evaluated as:

$$t_{enhanced} = t_{smooth} + \alpha.t_{\det ail} \tag{25}$$

where α represents the enhancement factor to manage the improved measure of detailed part.

3.6. Color Adjustment

The reflectance part returns the texture and fine points of submerged picture, and only undergoes color deformation caused with energy absorption underneath. Simplest color adjustment [27] is an outstanding means for confirming vast stability among adjustment concert in addition to data loss. The energy absorption in undersea causes color distortion. To overcome this problem, simplest color balance scheme for only extend the picture element of color means whereas conserve data as much as probable with physically set dissimilar truncation relation of color channel, so their histograms are capable towards maximum range [0,255]. For quick stretch, apply an affine transform function on each color channel for mapping picture element from 0 to 255. Though, only some unusual pels previously map most and least standards, for improving color act with small picture element by peak and least rates prior to affine transform task. In fact, this procedure source extra or less white and black area may seem to be not natural. In general, this algorithm not focuses for proper color allotment, but give color adjustment with contrast to some degree.

3.7. Hybrid Technique

A hybrid technique mainly combines dual imagery to generate clear representation. The vital task is mainly for computing weight maps of descriptions furthermore joins them mutually by luminance, saliency along with exposedness weight map for deriving normalized weighted maps [28].

Luminance weight map [29] represents luminance factor of every picture part, created with computing STD among color channels along with luminant rate, where luminant rate l represents:

$$l = \alpha.r + \beta.g + \gamma.b \tag{26}$$

where $\alpha + \beta + \gamma = 1$, each denotes weight constraint of color part.

It is capable for exactly reflect luminant measure of picture and illustrate better improvement; it confirms negative possessions on contrast and colourfulness. Headed for avoiding these, introduces successive weight maps.

Saliency weight-map [30], reflects salient things and points, intends to highlight this discerning submerged sight. Main drawback is for overestimating important area, as a result exposedness weight-map exploits to assure precision as well as defend middle color in picture.

Exposedness weight-map [30], assess significance of exposed picture elements. It protects local dissimilarity be non-exaggerated or non-understated. In general, picture element rates close to average rate are likely to have superior look. Gaussian-modelled remoteness to average rate represents

$$W_E = e^{-\frac{\left(I(d)-\bar{I}\right)}{2\sigma^2}} \tag{27}$$

where σ denotes standard deviation, $I(d)$ signify pel rate set at location d and $\bar{I}$ represent mean rate. As a result, the entire weight maps proficient for generating fine look for output imagery. The normalized weight map is simply computes:

$$W_{norm} = \sum_{k=1}^{K} W_k \tag{28}$$

where W_k specify the kth weight map.

The last free backscatter and color deformation picture be derived as:

$$R_c = \sum_{n=1}^{N} W_{norm}^{n}.J_c^n \tag{29}$$

where R_c denotes color part of fused image, J_c^n signifies color part of nth input. To get enhanced output, we basically exploit a multi-scale pyramid decomposition system for achieving fusion technique. The joint development represents

$$R_c^l = \sum_{n=1}^{N} G^l\left\{W_{norm}^n\right\}.L^l\left\{J_c^n\right\} \tag{30}$$

where $L^l\left\{J_c^n\right\}$ denotes Laplacian pyramid of color part of n^{th} input representation, $G^l\left\{W_{norm}^n\right\}$ is n^{th} stabilized weight map and l represents pyramid stages. While this scheme achieves quite rapid with the enhanced output image can make excellent result.

4. Experimental Results and Analysis

The projected technique evaluates the state-of-art processes; by imagery composed as of representational undersea test imagery and dissimilar metrics utilized to compare enhancement performances of these processes.

4.1. Qualitative Evaluations

To verify the effectiveness adaptive regularizations, we primarily use reconstruction outcome taking place in picture *seer* with projected technique in Figure 2. Adaptive regularization shows improved results to remove ringing artifacts near boundaries. The projected technique in Figure 3 composed of adaptive regularization to reconstruct the imagery, weighted decomposition process to separate illuminant and reflecting part. The illuminant part is to improve contrast and texture details using dehazing via gradient guided filter. The reflecting part is to enhance color using color adjustment. Finally, Hybrid technique used to enhance the result. As of Figure 4 weights of dual mechanisms applies a proficient technique for attaining excellence picture. In order to represent different features, we used numerous weight map process to compute the descriptions of true imagery and then join to obtain normalized weight maps. Evaluation of a projected scheme with existing technique shown in Figure 5. The scheme in [21] better reconstructs the picture but fails to enhance its visibility. The scheme in [14] fails to correct and enhance visibility of degraded picture, amplify noise as well as show poor effect on color adjustment. The process in [15] over-enhance distorted picture, which causes severe, color distortion along with noise amplification, thus integral performance of enhanced picture is worse. The scheme [30] gets good result, not only correcting the color casting but also enhancing the contrast for removing haze. Though, dissimilarity is overstretched in [24], a few picture elements reduced using over-flow or under-flow. The over-enhancement problem in [25] for test images and some information loss exist as well as in boundary effect due to guided filter introduced. Our projected technique is close towards or even somewhat improved than state of the arts. The RGB color space of dissimilar patch dimension in Figure 6 attains fewer pel map result. Moreover, RGB color space in Figure 7 of original imagery with improved results that all the pels of original picture maps into RGB color space gather in left corner, where is large green rate and small blue and red rate, and other method make more or less stretch of the pixel mapping results. The mapping outcome of our scheme is most excellent and it maps the largest area in the RGB color space, which means it has the best integral dynamic compression range. However, many mapping pixels of our method mapped the boundary, i.e., dark region and brightest region, it indicates that there is information loss of our method.

4.2. Quantitative Evaluations

The submerged quality metrics for evaluating and analyzing excellence of underneath imagery, including gray mean, standard deviation, mean gradient etc. The gray mean rate of an image reflects integral intensity, and higher the gray mean is, higher the intensity is, while standard deviation of a picture reflects high frequency part of a picture, which relates picture contrast, higher the standard deviation is, higher the contrast is as well as greater color data. Meanwhile, [31] show that a picture shows good integral quality when gray mean among 100 and 200 and its standard deviation among 35 and 80 after they analyzed and statistic a large measure of imagery. Mean gradient reflects speed of the changes of minor details in picture; it can represent the features of texture transform and the degree of clearness well. Table 1 review metric outcome for reconstruction used for different patch dimension also patch by small dimension offer better result. Following, Table 2 achieves best results in [15], μ_{RG} and $\boldsymbol{\mu_{YB}}$ which are close to zero. The largest $\boldsymbol{\sigma_{RG}}$ in [25], largest $\boldsymbol{\sigma_{YB}}$ in original imagery. Overall, our method obtains the highest score after computing the PSNR, SSIM, UICM, UIConM and UIQM value, lowest score in SSIM. Consequently, projected system is better than existing system giving enhanced outcome.

$$Mean = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}E(u,v) \tag{31}$$

$$\sigma^2 = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\left\{E(u,v)-Mean\right\}^2 \tag{32}$$

$$AG = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\sqrt{\frac{\left[E(u,v)-E(u+1,v)\right]^2+\left[E(u,v)-E(u,v+1)\right]^2}{2}} \tag{33}$$

Where σ^2 denotes standard deviation, *Mean* represents mean picture element, *AG* represents mean gradient and *UV* signify amount of picture elements. Efficient metric UIQM composed of three independent measurements [32-34] as described below.

a) Undersea Image Colorfulness Measurement (UICM)

The performance of color correction used for evaluating performance of undersea picture improvement progression, with red-green (rg) along with yellow-blue (yb) color parts are used:

$$rg = r - g \tag{34}$$

$$yb = \frac{r+g}{2} - b \tag{35}$$

Considering that submerged picture often suffers heavy noise, the asymmetric alpha-trimmed statistical standards [35] used for measuring submerged picture colorfulness,

$$\mu_{rg} = \frac{1}{N - T_L - T_r}\sum_{x=T_L}^{N-T_r} Intensity_{rg}(x) \tag{36}$$

where N signify total number of pels in the rg part and all pels sorted as $x_1 < x_2 < ..... < x_N$, $T_L = \alpha_L N$ and $T_r = \alpha_r N$ are the amount of least and most picture element to be reduced as of

sorted order $x_1 < x_2 < ..... < x_N$. The average rate μ_{rg} corresponds to chrominance gray, and average rate close towards nil in rg-yb color part imply superior for balancing color. Further, the second-order statistic variance is defined by:

$$\sigma_{rg}^2 = \frac{1}{N}\sum_{x=1}^{N}\left(Intensity_{rg}(x) - \mu_{rg}\right)^2 \quad (37)$$

σ_{rg}^2 represents the pixel activity. The first and second order statistic information μ_{yb} and σ_{yb}^2 of the yellow-blue part be computing in similar way. The whole, colorfulness coefficient parameter is expressed as

$$UICM = -0.0268.\sqrt[2]{\mu_{rg}^2 + \mu_{yb}^2} + 0.1586.\sqrt[2]{\sigma_{rg}^2 + \sigma_{yb}^2} \quad (38)$$

b) Undersea Image Sharpness Measurement (UISM)
Sharpness reflects details and edges of an image, and fine captured images are likely to show better sharpness. For more efficient, enhancement measure estimation (EME) as

$$EME = \frac{2}{s.t}\sum_{K=1}^{s}\sum_{L=1}^{t}\log\left(\frac{I_{\max,K,L}}{I_{\min,K,L}}\right) \quad (39)$$

where picture divided by $s.t$ blocks, and obtain the maximal and minimal pel values in each block, $\frac{I_{\max,K.L}}{I_{\min,K,L}}$ indicates comparative dissimilarity relation in every block. Then the undersea image sharpness measure (UISM) can be written as:

$$UISM = \sum_{c=1}^{3}\lambda_c.EME\left(grayscale\ \text{edge}_c\right) \quad (40)$$

where λ_c stands for weight coefficient of each color part.

c) Undersea Image Contrast Measurement (UICM)
Intended for submerged descriptions, dissimilarity Distortion is usually happened with backscattering. The contrast performance be measured by:

$$\log AMEE = \frac{1}{s.t}\sum_{K=1}^{s}\sum_{L=1}^{t}\frac{I_{\max,K,L} - I_{\min,K,L}}{I_{\max,K,L} + I_{\min,K,L}}.\log\left(\frac{I_{\max,K,L} - I_{\min,K,L}}{I_{\max,K,L} + I_{\min,K,L}}\right) \quad (41)$$

and submerged picture contrast measure be written as:

$$UIConM = \log AMEE\left(\text{I}ntensity\right) \quad (42)$$

d) Undersea Image Quality Measurement (UIQM)
Undersea descriptions modelled as linear superposition of absorbed as well as spreaded part. Meanwhile, absorption and scattering by dusk-like particles causes color casting, sharpness attenuation with dissimilarity distortion. Consequently, logical use of linear form to create entire UIQM as:

$$UIQM = \alpha.UICM + \beta.UISM + \gamma.UIConM \quad (43)$$

where α , β , and γ signifies weight coefficients to organize each measure as well as balance their values.



a) True Picture b) Blurred c) Restored Picture



a) Reconstructed Image b) Reflectance c) Color Balance f) Final Output



d) Illuminance e) Dehazing viagradient guided filter

Figure 3: Results on the process of Proposed Method



a) Luminance weight 1 b) Saliency weight 1 c) Exposedness weight 1 d) Normalized weight 1



e) Luminance weight 2 f) Saliency weight 2 g) Exposedness weight 2 h) Normalized weight 2



a) [21] b) [14] c) [30] d) [24]



e) [15] f) [25] g) Ours

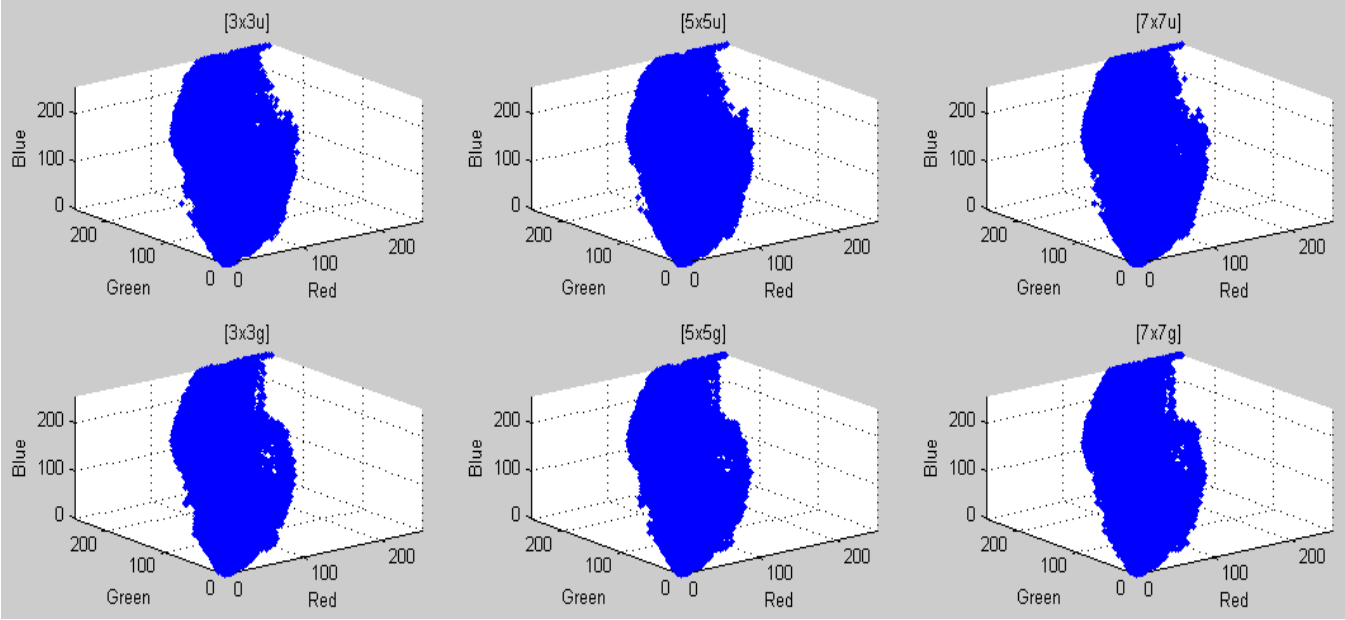


Figure 6: RGB color space mapping reconstruction results on seer image

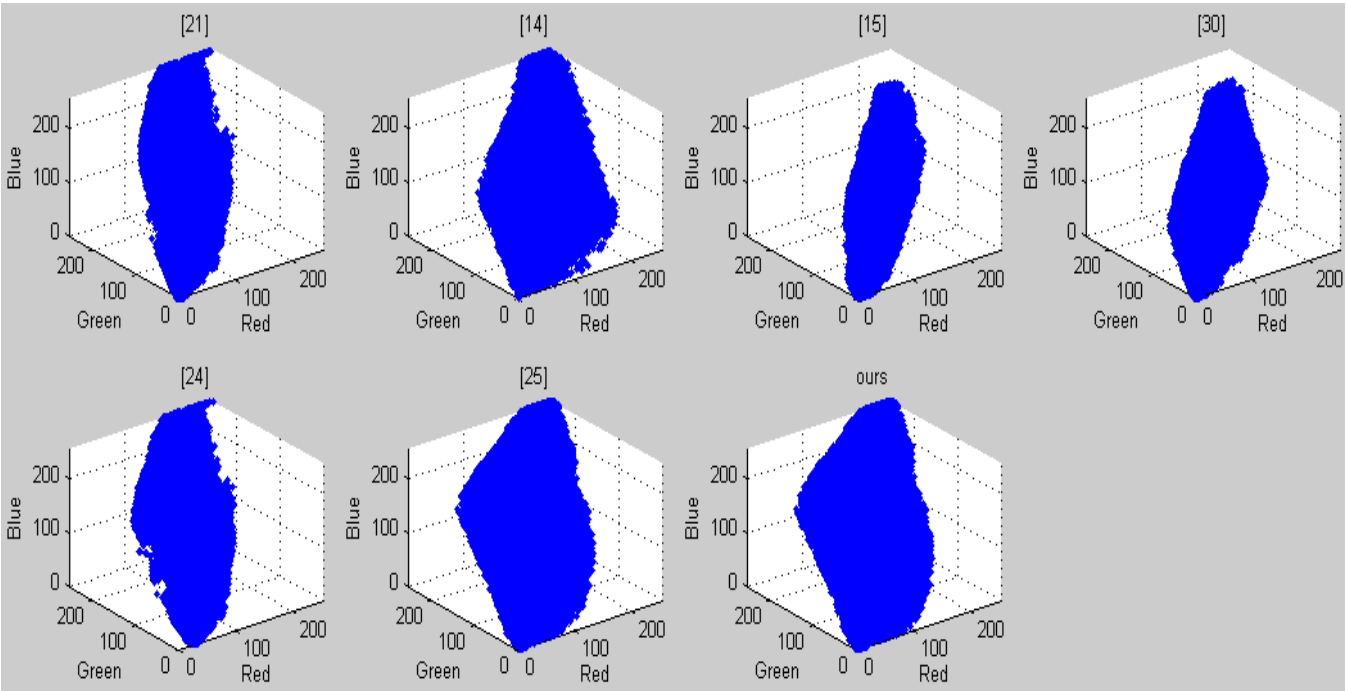


Figure 7: RGB color space mapping Enhancement results on seer image

Table 1: Reconstruction metric results of seer image

	Uniform blur			Gaussian blur		
	7x7	**5x5**	**3x3**	**7x7**	**5x5**	**3x3**
Mean	127.33	127.35	127.36	127.35	127.37	127.37
SD	71.266	71.184	71.136	70.148	70.065	70.027
AG	9.767	8.910	8.476	6.721	6.262	6.129
PSNR	26.16	26.28	26.35	23.96	24.04	24.05
RMSE	17.827	17.469	17.270	13.689	13.320	13.169
SSIM	0.741	0.754	0.756	0.618	0.623	0.622
μ_{RG}	-45.531	-45.543	-45.558	-45.535	-45.538	-45.537
μ_{YB}	27.93	27.92	27.906	27.734	27.73	27.729
σ_{RG}	-41.214	-41.311	-41.352	-41.272	-41.326	-41.34
σ_{YB}	27.004	27.038	27.052	26.892	26.91	26.918
UICM	-45.531	-45.543	-45.558	-45.535	-45.538	-45.537
UIConM	0.745	0.737	0.736	0.716	0.704	0.704
UISM	7.059	6.890	6.906	6.512	6.276	6.244
UIQM	3.467	3.388	3.389	3.199	3.0891	3.080

Table 2: Enhancement metric results of seer image

	Original	**[21]**	**[14]**	**[30]**	**[24]**	**[15]**	**[25]**	**Ours**
Mean	127.39	127.34	129.26	104.05	128.99	103.41	136.21	**136.21**
SD	66.272	71.236	68.047	56.586	70.056	64.509	82.112	**83.494**
AG	3.790	9.226	**15.819**	11.89	9.326	8.660	12.212	12.517
PSNR	21.53	24.10	27.670	28.483	16.557	28.073	30.004	**30.013**
RMSE	21.369	13.717	10.544	9.602	37.900	10.022	8.059	**8.052**
SSIM	0.466	0.619	0.724	0.786	0.839	0.765	0.841	**0.842**
μ_{RG}	-45.685	-45.533	-4.937	-24.524	-33.344	**-0.960**	-37.795	-37.801
μ_{YB}	-41.664	-41.224	-13.078	-25.712	-36.852	**-1.732**	-35.456	-35.472
σ_{RG}	27.789	27.93	22.343	23.581	33.732	11.964	**29.896**	29.888
σ_{YB}	**27.155**	27.016	20.332	22.625	30.378	9.647	26.214	26.226
UICM	4.505	4.516	-4.937	4.223	-33.344	-0.960	4.917	**4.917**
UIConM	0.621	0.741	0.836	0.827	0.790	0.872	0.480	**1.0318**
UISM	5.771	7.005	**7.132**	7.055	7.003	6.911	7.013	6.997
UIQM	2.637	3.434	4.957	4.3512	3.952	5.132	2.723	**4.689**

5. Conclusion and Future Work

This work is mainly for Enhancing undersea representation in blurry conditions using Hybrid technique. Our scheme composed with Adaptive regularization for recreation, weighted decomposition procedure to separate recreated imagery into color change along with dehazed schemes separately, then multi-scale fusion technique towards enhancing picture. The strong with effective Hybrid technique which incorporate the description of all fused imagery into resultant outcome constitute backbone of projected scheme. In fact, the projected scheme generally beats other methods in terms of qualitative as well as quantitative evaluations. Our technique confirms excellent act for undersea imagery, but border effect cause data loss. Meantime, our technique does not outperform some of the state-of-art system and even direct over improvement for precise cases. The challenging work for future would be to find out a well-organized means to about vertical deepness and optimize the transmission deepness of undersea imagery more precise, at last build up absolute, effective and well-organized system.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank the reviewers for their thorough and helpful remarks.

References

[1]. Ruo W, Yaf W, Jun Z, Xianp F. "Review on Underwater Image Restoration and Enhancement Algorithms", 7th Intern Confrnc on Intrnt Multimedia Comput and Serv ACM, 2015.

[2]. Schetni R, Cors S. "Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods", EURASIP Jrn Advn sgnl Pr, Jan 2010.
[3]. John C, Yin C.C. Underwater Image Enhancement by Wavelength Compensation and Dehazing, IEEE Transc Img Pr-2012.
[4]. C. Ancut, Hermns, and Beka, "A Fast Semi Inverse Approach to Detect and Remove the Haze from a Single Image',' *Comp Vis*: Springer, Nov2011
[5]. O. Ancu, C. Anc, and Bekae," Effective Single Image Dehazing by Fusion," *Procd IEEE Intr. Confr. Img Proc*, Sep-2010.
[6]. Ancut and Ancut C, ``Single Image Dehazing by Multi-Scale Fusion," *IEEE Transct Img Proc*, Aug-2013.
[7]. C. O. Ancu, T. Habe, and Bekae, "Enhancing Underwater Images and Videos by Fusion," *Proce. IEEE Confrn. Comp. Visn. Pattrn Recogn,* Jun. 2012.
[8]. K. Hem, Jia S, and Xiao T, "Single Image Haze Removal using Dark Channel Prior,", *IEEE Transc. Patrn Analy. Macnn*, Dec. 2011.
[9]. Raan F, "Single Image Dehazing", *ACM -Transc Graphic*, Aug- 2008.
[10]. Xue F, Yu H, Del Z, Xia P and Xing D, "A Fusion-based Enhancing Approach for Single Sandstorm Image", *IEEE Intrn. Wrkshp Mult S/gl Proc*, Sept-2014.
[11]. Fan G, Jin T and Zix C, "Fusion Strategy for Single Image Dehazing", *Intrn. Jrn. Dig. Contt Techl. App*, Jan. 2013.
[12]. C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Single-scale fusion: An effective approach to merging images," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 65–78, Jan. 2017.
[13]. H. Lu, Y. Li, S. Nakashima, H. Kim, and S. Serikawa, "Underwater image super-resolution by descattering and fusion," *IEEE Access*, vol. 5,pp. 670–679, 2017.
[14]. Zuiderv K., "Contrast limited adaptive histogram equalization", pub book Graphcs gem IV, 1994.
[15]. Fu X, Zhug P, Huang Y, Yin L, Xia P and Xing D, "A Retinex-based Enhancing Approach for Single Underwater Image", Img Procd IEEE Intrl Confr, 2014.
[16]. Emb S., Chi L., Cava A, "Underwater image and video dehazing with pure haze region segmentation", Comp Vis and Im Unde, 2018.
[17]. Peng Y, Cos P,"Underwater image restoration based on image blurriness and light absorption",IEEE Tran on Im Pro, 2017.
[18]. Li C, Cava, "Background light estimation for depth-dependent underwater image restoration",IEEE Intl Conf on Ima Processg, 2018.
[19]. AncuC., Ancut., D Vlee, Bekt, P, "Color balance and fusion for underwater image enhancement" IEEE Transa Imge Pro, 2018.
[20]. Ber D., Trei T., Avid S, "Diving into haze-lines: Color restoration of underwater images", Brith Mach Visn Confer, 2017
[21]. Chrispin jiji and Nagaraj Ramro "Deblurring Underwater Image Degradations based on Adaptive Regularization" IEEE Inter Conf on Computat Intelli and Comput Rese, 2017
[22]. Weis D, Guang S, Xi L, Lei Z and Xiaol, "Image Reconstruction with Locally Adaptive Sparsity and Nonlocal Robust Regularization", S/gl Procg: Imag Comn, Nov-2012.
[23]. Dong, Zhan, Gua S and Xiaoli, "Image Deblurring and Super Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization", IEEE Transa Ime Proces, July 2011.
[24]. Kwa K, Wo J , You S "Optimized Contrast Enhancement for Real Time Image and Video Dehazing", Jrn Visu Commn Img Repr, Feb-2013.
[25]. Zhang, "Removing backscatter to enhance the visibility of underwater objects",msc dissertn, Nanyang Univ, 2016
[26]. F Ko, Weih C, Changy W, Zhengg L, "Gradient Domain Guided Image Filtering", *IEEE Transc Img Proceg*, 2015.
[27]. Cod. Ancut , Cosm Chri , Philip, "Color Balance and Fusion for Underwater Image Enhancement", IEEE Tra Img Procg, Jan 2018
[28]. C.Anc, Ancu Cod, T Hab, Phil B,"Fusion-based Underwater Image Enhancement by Weight Map Techniques",IC3-2018,Chap *in* Adv in Intell Sys and Comp, Jan 2019.
[29]. Ancti, Habr T and Phil B. Fusion-based Restoration of the Underwater Images", Intrn Confrn Img Proc, Sep-2011.
[30]. Ancti C, Ancu O, Haer "Enhancing Underwater Images and Videos by Fusion", Compr Visn Patn Recogn , Jun-2012.
[31]. Jobson D, Rahman Z, Woodel "The Sstatistics of Visual Representation", Procd SPE- 2002.
[32]. Panet K, Ga C and Agai S,"Human-Visual-System-Inspired Underwater Image Quality Measures", Jr of Ocen E-2016.
[33]. Cha Y, Ric M, Andr C, "An Online Platform for Underwater Image Quality Evaluation", Intl Confr on Patn, CPR 2018
[34]. Priya S and Sim, "Quality measure technique for Underwater images using Artificial neural network" Jon of Eng and Appl Sc, 2018
[35]. Bedr and Watt, "Alpha-Trimmed means and their relationship to Median Filters," IEEE Transc Acost. Spch Sigl Proc, 1984.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 347-358 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Using Privacy Enhancing and Fine-Grained Access Controlled eKYC to implement Privacy Aware eSign

Puneet Bakshi*, Sukumar Nandi

Department of Computer Science and Engineering, Indian Institute of Technology, Guwahati, 781039, India

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 23 July, 2019
Online: 19 August, 2019

Keywords:
eSign
eKYC
Electronic Signature
Privacy
Access Control
Authentication
Token
BAN logic

ABSTRACT

eSign is an online electronic signature service which is recently gaining more prominence in India. eSign is based on two online services from UIDAI, viz. a viz., Aadhaar based authentication and retrieval of resident's eKYC information after taking his/her consent. With increased adoption of Aadhaar based services, privacy of user data has become more and more important. Present method of taking boolean consent from resident through non-UIDAI entity may not be acceptable for two main reasons, first is that the consent does not include in itself a proof from resident that the consent is indeed taken from him/her and second is that the resident may wish to have better privacy and fine grained access control rules to access his/her eKYC data. Bakshi et.el have introduced a mechanism to improve amortized performance of eSign using a digital access token. In this work, the digital access token is enhanced to include Privacy Enhancing and Fine-Grained Access Control (PEaFGAC) Statements for facilitating Privacy Aware eSign. These tokens can be used by other entities to access eKYC data of the resident with better access controls enforced by the resident. This paper briefly describes the present model of eSign, the earlier proposed model of eSign followed by the proposed model of Privacy Aware eSign. The proposed model of Privacy Aware eSign is also analyzed using BAN logic assuming Dolev-Yao security environment.

1 Introduction

eSign is an online electronic signature service in India which is being promoted by Government of India as part of its Digital India Initiative. As opposed to traditional dongle based electronic signature, *eSign* provides benefits such as less cost, no manual authentication, no requirement of special hardware device and no requirement for the end user to keep any key secret. With the passage of *Information Technology Act (ITA-2000)*, an electronically signed digital document is considered equivalent to a handwritten signed paper document. In India, *eSign* is being regulated by *Controller of Certifying Authority (CCA)* and is being operated by certain designated empaneled agencies known as *eSign Service Providers (ESP)*. *ESP* provides its services to application specific agencies known as *Application Service Providers (ASP)*. *ASP* provides *eSign* service to the end users. *eSign* is governed by *Public Key Infrastructure (PKI)* which is further governed in legal matters by the national legislature of the country.

To avail *eSign* service, a resident needs to enroll with *Unique Identification Authority of India (UIDAI)* and receive a 12-digit identity number called *Aadhaar* [1] [2]. As part of the enrollment process, resident needs to provide information about his/her identity and address to *UIDAI* such as *Name*, *Date of Birth*, *Address*, *Phone number*, *Email-id*, *Biometric (fingerprint-scan, iris-scan)* etc. The process of obtaining this information from the end user is known as *Know Your Customer (KYC)* and is initially introduced by *Reserve Bank of India (RBI)* for financial banks [3]. Traditionally, this process involves submission of a self-attested physical form along with necessary physical documents, followed by verification and approval by receiving organization. eKYC is an online service which facilitates completion of KYC process electronically. eKYC has some significant benefits over traditional KYC, eKYC eliminates submission of physical documents by customer, is faster and is less error prone. *UIDAI*'s eKYC service facilitates a third entity to retrieve resident's identity, address and other details after taking explicit consent and authorization from the resident.

With increased adoption of *Aadhaar* based identification,

*Corresponding Author: Puneet Bakshi, IIT Guwahati, b.puneet@iitg.ac.in

many online services are now using *Aadhaar* based services and with its such wide adoption, privacy of user data has become even more important. Although *Aadhaar* based eKYC service provides access to eKYC data only after taking an explicit consent from the resident, this way of taking consent from resident has two shortcomings. First is that the consent is taken by *non-UIDAI* entity and does not encode in itself a proof from resident that the consent is indeed given by the resident. Second is that providing a boolean consent is too broad, either an unconditional access is given to the whole eKYC information or no access is given at all. A resident may wish to have a better privacy enhancing fine-grained access control to his/her eKYC data. Resident may wish to define a privacy and access control policy dictating the *scope* of information which can be provided, the *purpose* for which the information can be provided and *recipients* to whom the information can be provided. For example, a resident may wish to disclose only his/her name and address, only for electronic signature purpose and only to a specific *eSign* Service Provider.

In [4], the author explained two limitations of present *eSign* model. The first limitation is that the eKYC data access reflects a restrictive *self-only*, *full-resource* and *unlimited* access control. Author pointed out that a resident may wish to have a better access control mechanism which allows third entities to access part of a resource which is to be used for a specific purpose and for a limited time period. The second limitation is that for each *eSign* request, resident has to authenticate itself each time and to include the authentication proof in each such request. Author pointed out that if a resident needs to *eSign* multiple times, time taken by initial authentication phase will be a major bottleneck. The author proposed that amortized performance of eSign can be improved using *digital access token* which encodes in itself the authentication proof and other information such as how many *eSign* requests can be made using this token and the expiry time of the token.

This paper is an extension of [4] and the digital access token is enhanced to implement *Privacy Aware eSign*. Our main contribution in this paper is to introduce a method to implement *Privacy Aware eSign* using *Privacy Enhancing and Fine-Grained Access Control (PEaFGAC) Statements*. A resident can encode these statements in digital access token for better access to his/her eKYC data. This token can be provided to third entities so that they can present this token for claiming protected resource from *UIDAI*. This paper also presents security analysis of the proposed scheme using *Burrows-Abadi-Needham (BAN)* logic. The analysis shows that in the proposed scheme, even if the network is unreliable, the exchanged information is reliable and is secured against eavesdropping.

The remainder of this paper is organized as follows. Section 2 presents related work. Notations used in the paper are reported in figure 1. Section 3 presents Aadhaar based eKYC service. Section 4 presents *eSign* version 2.0 model. Section 5 presents *eSign* model proposed in [4] to improve amortized performance of *eSign*. Section 6 presents proposed *Privacy Aware eSign* model using privacy enhancing and fine-grained access controlled eKYC. Section 7 presents formal security analysis of the proposed model using *BAN logic* and finally section 8 concludes the paper.

$\{X\}_Y$	X is signed by key Y
SK_Y	Symmetric key of entity Y
SK_{Y_Z}	Symmetric key shared by entities Y and Z.
PR_Y	Private asymetric key of entity Y
PB_Y	Public asymetric key of entity Y
R_i	Resident
ASP_i	Application Service Provider
ESP_i	eSign Service Provider
$UIDAI$	Unique Identification Authority of India
ID_{R_i}, ID_{ASP_i} ID_{ESP_i}	Identities of R_i, ASP_i and ESP_i
TID_{ESP_i}, TID_{ASP_i}	Unique transaction identifiers generated by ESP_i and ASP_i
PW_{R_i}	Password of R_i for login to ASP_i portal
$AadhaarNo_{R_i}$	Aadhaar No of R_i
C_{R_i}	Cookie associated with R_i's logged-in session, assigned by ASP
PR_{B_i}, PR_{ASP_i} PR_{ESP_i}, PR_{UIDAI}	Private keys of R_i browser, ASP_i, ESP_i and $UIDAI$
PB_{B_i}, PB_{ASP_i} PB_{ESP_i}, PB_{UIDAI}	Public keys of R_i browser, ASP_i, ESP_i and $UIDAI$
$n*_!$	nonces such as $n1_{ASP_i}$, where $*$ is any integer and ! can be R_i, ASP_i, ESP_i or $UIDAI$
$DataR_i(DataA_i,$ $DataE_i, DataU_i)$	Intermediate data in plaintext to be send by R_i (ASP_i, ESP_i, $UIDAI$)
$SignR_i(SignA_i,$ $SignE_i, SignU_i)$	$\{H(DataR_i)\}_{PR_{Ri}}$
$consent_{use_ekyc}$	Consent from resident whether his/her eKYC can be used
$consent_{genuse_at}$	Consent from resident whether a digital access token can be generated for later use
$License_{ASP_i}$ $License_{ESP_i}$	License for ASP_i (ESP_i) to use services from ESP (UIDAI)
M_i	Message (in plaintext) to be eSign
DSC_{Ri_Mi}	Digital Signature Certificate generated for message M_i for resident R_i
$\{M\}_{eSign_R_i_ESP_i}$	eSigned message (by R_i) through ESP_i
$H()$	One way cryptographically secure hash fn
$\|\|$	Concatenation operator
$\oplus$	XOR operator

Figure 1: Notations used in this paper

2 Related Work

Digital tokens are increasingly being used in many cryptography related applications to achieve varied objectives.

U-Prove [5] is an identity management solution based on blind signatures [6] which uses digital tokens to achieve objectives of privacy and anonymity. *U-Prove* consists of two protocols, viz., issuance protocol and presentation protocol. In issuance protocol, identity provider issues digital token to the subscriber which (s)he can later present to the verifier in presentation protocol so that the service provider can grant resource access to the subscriber. A *U-Prove* token consists of a unique token identifier, a public key of the token which aggregates information in the token, a token information field which encodes token specific information, a prover information field which is opaque to the issuer, issuer signature on all the other token contents and a boolean value which indicates whether the token is protected by a device. *U-Prove* uses digital tokens effectively by encoding necessary information in it in cryptographically secure way to achieve objectives such as privacy and anonymity.

OAuth2 [7] is an authorization framework which allows delegation of access to protected resources to a third party by using digital tokens referred to as access tokens. Access tokens are issued to Clients by *Authorization Server* after taking permission from *Resource Owner*. An access token can be of two types, viz., a bearer token and a MAC token. A bearer token is an opaque string which can be used to claim access to a resource by any entity who presents the

token. A MAC token is essentially a shared symmetric key which is used to sign a challenge by the client to prove its possession of the token to authorization server. *OAuth2* uses digital tokens effectively for access delegation and is used by many organizations for data sharing.

Bitcoin [8] is a decentralized digital currency which can be transacted over peer-to-peer bitcoin network. A bitcoin network is composed of cryptographically secure linear chains of blocks with each block containing a header and a collection of transactions. A transaction is essentially a digital token that changes ownership of bitcoins from one entity to another. Each transaction in bitcoin network is broadly composed of three parts, viz., input, output and amount. Input refers to the previous owner of the bitcoins, output refers to the new owner of the bitcoins and amount refers to the amount of bitcoin that is transacted. Bitcoins uses cryptographically secure digital information containers (similar to digital tokens) effectively for the realization of digital currency.

Although Attribute Based Encryption (ABE) is also evolved to protect privacy of user data, it is based on Identity Based Encryption (IBE). An agency may not shift from PKI to IBE framework for a number of reasons.

3 Aadhaar based e-KYC service (v2.1)

Aadhaar based eKYC service is available to general citizens only through certain empaneled agencies such as *eSign Service Provider (ESP)* and the infrastructure network is secured by certain designated agencies known as *Authentication Service Agency (ASA)* and *KYC User Agency (KUA)*. eKYC service is hosted as a stateless REST based web service over HTTPS and the details are sent as input data encoded in XML. Figure 2 depicts Aadhaar's eKYC webservice as specified in eKYC v2.1 specification [9]. The specification provides following information about element *rc* which represents the resident consent.

"rc – (mandatory) Represents resident's explicit consent for accessing the resident's identity and address data from Aadhaar system. Only valid value is "Y". Without explicit consent of the Aadhaar holder application should not call this API [9]."

As can be seen from the specification, *rc* is a boolean consent and assumes that it has been transferred from resident to *UIDAI* unaltered. Although intermediate communication channels between various entities from resident to *UIDAI* are well secured and access to eKYC data is provided only after receiving explicit consent from resident, this way of taking consent from resident has two shortcomings. First is that the consent is taken by a *non-UIDAI* entity and does not encode in itself a proof from resident that it is (s)he who provided the consent. This is because residents do not have any registered tamper proof crypto device which can be used to encrypt user consent using resident specific PIN or password. Second is that providing a boolean consent is too broad, either an unconditional access is given to the whole eKYC information or no access is given at all. A resident may wish to have a better privacy enhancing fine-grained access control to his/her eKYC data indicating details on *scope*, *purpose* and *recipient*.

```
URL:
  https://<host>/kyc/<ver>/<ac>/<uid[0]>/<uid[1]>
        /<asalk>

Input Data:
  <Kyc ver="" ra="" rc="" lr="" de="" pfr="">
       <Rad>base64 encoded fully valid Auth XML for
           resident
       </Rad>
  </Kyc>

Response Data:
  <Resp status="" ko="" ret="" code="" txn="" ts=""
        err=""> encrypted and base64 encoded KycRes
        element
  </Resp>

  <KycRes ret="" code="" txn="" ts="" ttl="" actn=""
      err="">
      <Rar>base64 encoded fully valid Auth response
          XML for resident
      </Rar>
      <UidData uid="">
          <Poi name="" dob="" gender="" />
          <Poa co="" house="" street="" lm="" loc=""
               vtc="" subdist="" dist="" state=""
               country="" pc="" po=""/>
          <LData lang="" name="" co="" house=""
               street="" lm="" loc="" vtc=""
               subdist="" dist="" state=""
               country="" pc="" po=""/>
          <Pht> base64 encoded JPEG photo of the
               resident
          </Pht>
          <Prn type="pdf"> base64 encoded signed
               Aadhaar letter for printing
          </Prn>
      </UidData>
      <Signature/>
  </KycRes>
```

Figure 2: Aadhaar's eKYC 2.1 API

4 Present model of eSign in India

In *eSign* version 2.0 [10], a resident first registers itself with a front end application specific agency viz. a viz., *Application Service Provider (ASP)*. A resident can use either OTP based authentication or biometric based authentication. In case of OTP based authentication, OTP generation request is forwarded to *UIDAI* via *ASP* and *ESP*. *UIDAI* generates an OTP and sends it to resident's registered mobile number. In case of biometric based authentication, resident gets his fingerprint/iris scanned through a registered device. After authentication phase, resident now initiates an eSign request through *ASP* by providing it the consent to use his/her eKYC, the document to be signed and his/her *Aadhaar* number. *ASP* calculates cryptographic hash of the document and sends it along with the resident's consent and resident's *Aadhaar* number to *ESP*. *ESP* takes authentication proof from resident, creates a random symmetric key SK_{ESP_UIDAI} and a

Personal Identity Data Object (*PID*). *PID* encodes in itself the resident's authentication proof and the cryptographic hash of the *PID* object ($SHA256(PID)$). *ESP* first encrypts *PID* with SK_{ESP_UIDAI}, second encrypts cryptographic hash of *PID* ($SHA256(PID)$) with SK_{ESP_UIDAI} and third encrypts SK_{ESP_UIDAI} with public key of *UIDAI* (PB_{UIDAI}). *ESP* now wraps them in a new object called "Auth" and sends it to *UIDAI* in eKYC request. *UIDAI* provides eKYC information to *ESP*. Using received eKYC information, *ESP* generates a Digital Signature Certificate (DSC) and provides it to *ASP*. *ASP* provides the digitally signed document to the resident.

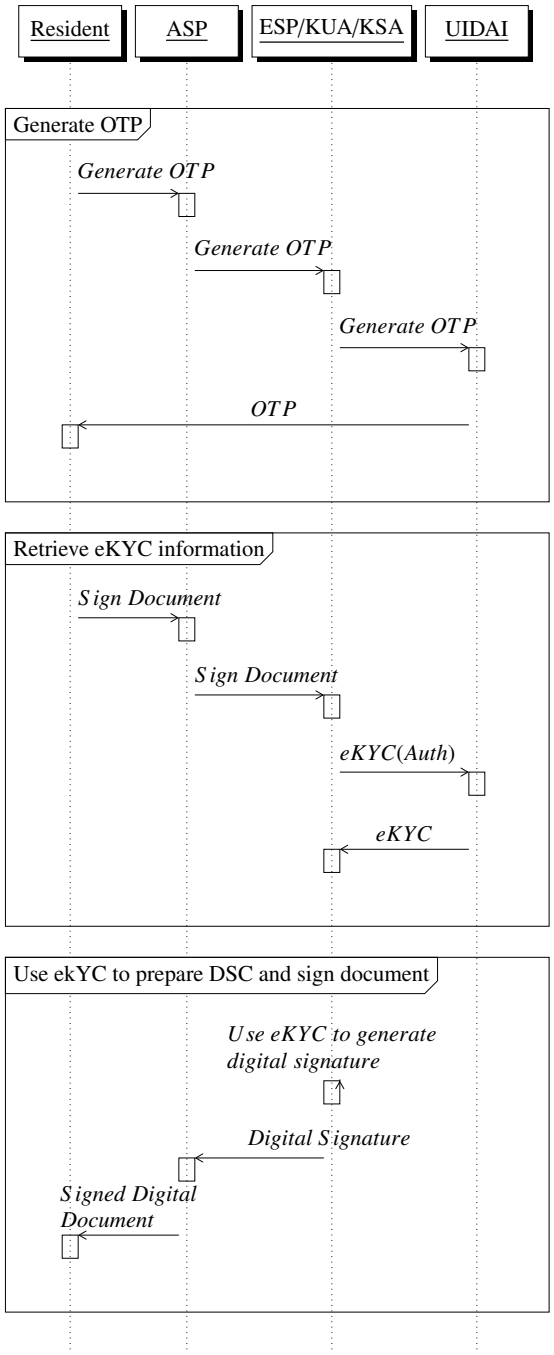


Figure 3: Sequence diagram of eSign 2.0

In practice, the initial authentication phase in *eSign* request is most time consuming since it involves either the manual text input (in case of OTP based authentication) or the physical scan of the fingerprint/iris (in case of biometric based authentication). Other than that, in some use cases such as *Create Birth Certificate*, *Create Death Certificate*, *Student Enrollment*, etc., the application is most heavily used during a certain time period (nearing the end of a deadline) which puts a sudden nationwide load on *UIDAI* services.

5 eSign model as proposed in [4]

In [4], author explained two limitations of present *eSign* model and proposed a mechanism to address the same. The first limitation is that in present model of *eSign*, eKYC data access reflects a restrictive *self-only*, *full-resource* and *unlimited* access control. Author pointed out that the resident may wish to have a better access control mechanism reflecting *third-entity-also*, *partial resource*, *use-limited* and *time-limited*.

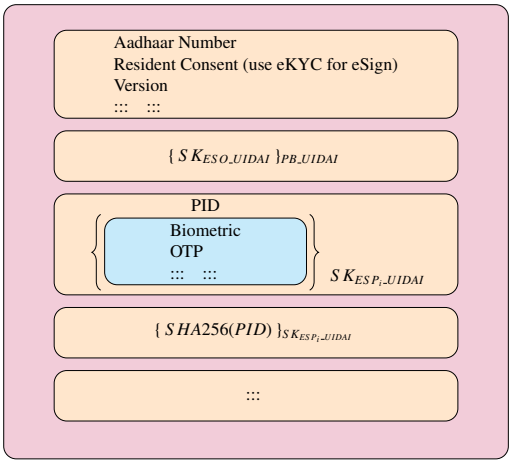


Figure 4: Auth Object (eSign 2.0)

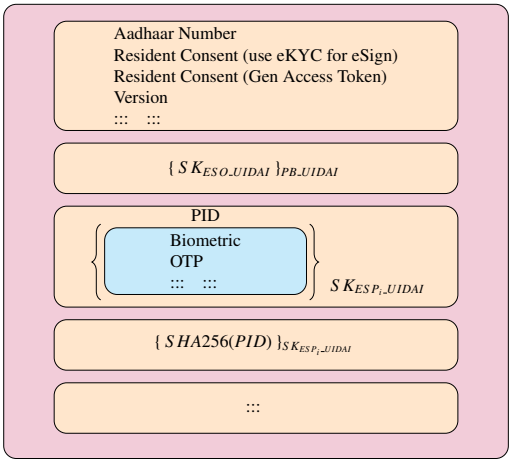


Figure 5: Auth Object as proposed in [4]

The second limitation is that a resident has to authenticate himself/herself for each *eSign* request and include the corresponding authentication proof in each *eSign* request.

If a resident wishes to *eSign* a large number of documents, the initial authentication phase will comprise most of the overall *eSign* time. After taking necessary consent from the resident, his/her authentication proof be stored with *ESP* in first request and is reused in rest of the requests.

A digital access token [figure 6] can be used to include claims from participating entities (*ESP* and *UIDAI*). A new service named *GenerateAccessToken* is proposed to be introduced by *UIDAI*.

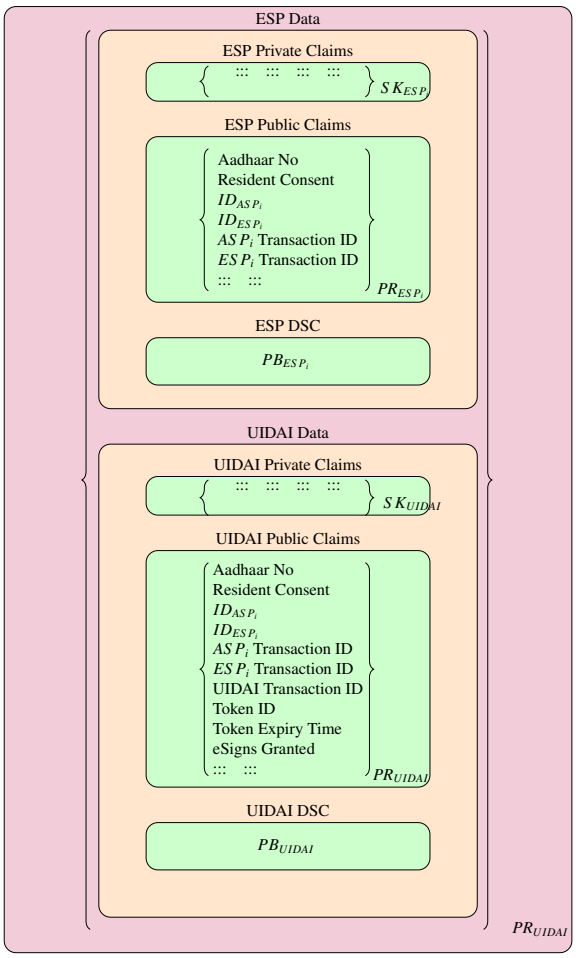


Figure 6: Access Token Structure as proposed in [4]

In this proposed model of eSign [figures 7, 8], resident first authenticates himself/herself using OTP or biometric based authentication and sends *eSign* request to *ASP*. *ASP* forwards this *eSign* request to *ESP*. *ESP* takes OTP and permission to generate access token from resident and creates an "Auth" object. This "Auth" object is created as before but additionally including *ESP* claims as well. *ESP* sends *GenerateAccessToken* request to *UIDAI* including "Auth" object. After receiving this request, *UIDAI* creates an access token including its own claims as well as claims received from *ESP*. *UIDAI* sends this access token back to the *ESP*. Now, *ESP* sends eKYC request to *UIDAI* including this access token instead of the "Auth" object. After receiving eKYC information from *UIDAI*, *ESP* generates *Digital Signature Certificate (DSC)* from it and provides the same to *ASP*. *ASP* embeds DSC in the document and sends the digitally signed document to the resident. For all rest of the *eSign* requests, *ESP* can reuse the same access token in eKYC requests and avoid the initial authentication phase.

The paper also presented two usability scenarios, based on whether the eKYC information can be cached by *ESP* or not. If *ESP* is permitted to reliably and securely store eKYC information of the resident, even the repeated eKYC requests from *ESP* to *UIDAI* can be avoided.

The paper also presented performance comparison analysis of proposed model with present model and found substantial improvement in amortized performance of *eSign*.

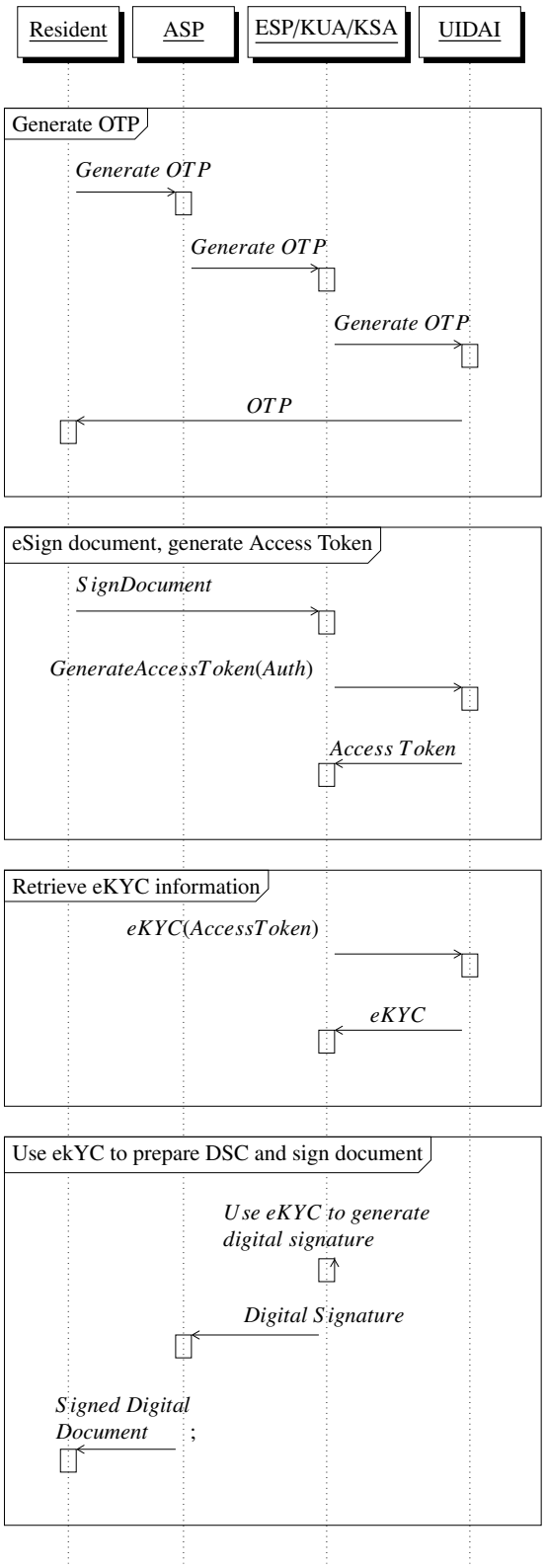


Figure 7: First call to eSign in eSign model proposed in [4]

6 Privacy Aware eSign model

In earlier proposed model [section 5] digital access token is used to increase amortized performance of *eSign* by storing necessary claims from *UIDAI* and *ESP*. The same token can be enhanced to include claims from resident as well. A resident can encode claims related to privacy and fine grained access control of his/her eKYC data. A stricter *PEaFGAC* statements may be enforced centrally at *UIDAI* level and an

overriding less strict rule can be supplied with each eKYC request to grant access to the requesting entity.

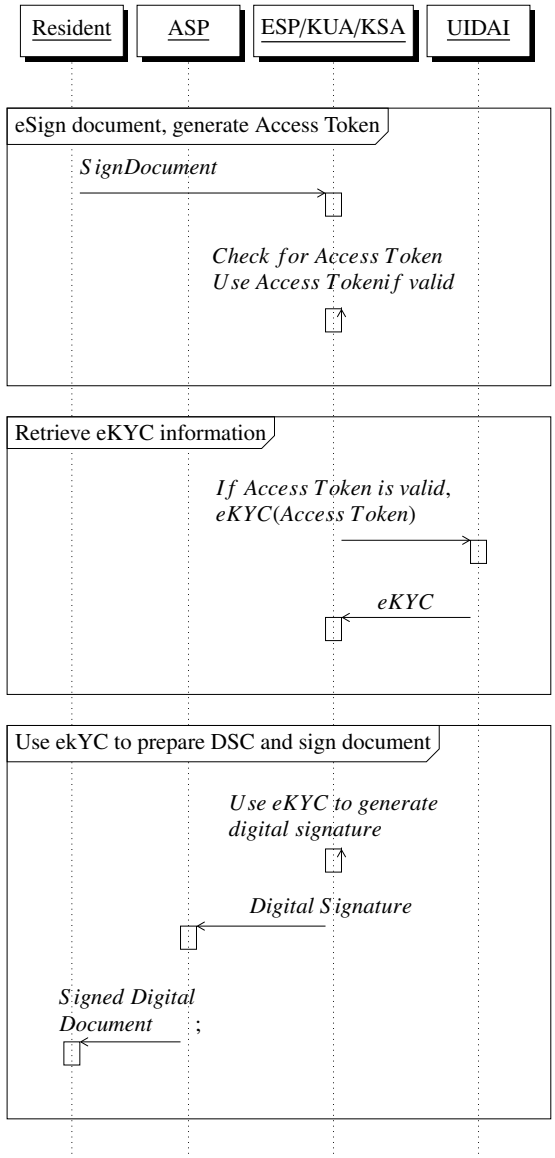


Figure 8: Second call to eSign in case eKYC needs to be fetched again

6.1 Privacy and Fine-Grained Access Control (PEaFGAC) Statements for eKYC data

A *PEaFGAC* statement encodes in itself the *scope* of information which can be provided, the *purpose* for which the information can be provided and attributes of *recipients* to whom the information can be provided. These statements are comprised of small sub-statements which are combined using relational operators. Each statement is identified by a numeric *id* and an alphanumeric *tag*.

An example of a *PEaFGAC* statement is presented in figure 10. This statement encodes in it that the purpose for seeking eKYC information should be *eSign*, seeking entity must either have the email in domain *finance.iitg.ac.in*, or must be working in *finance* department of *Indian Institute of Technology, Guwahati (IITG)*, or must have a designation of *director* or above. The statement is uniquely identified by a statement identifier (*ID*) and has a small alphanumeric representational string (*TAG*). Other than these, the staement also encodes in it the purpose for which eKYC can be accessed (*Purpose*), required (*eKYC*) attributes of information seeker (*AP*) and eKYC information which can be provided to the requester (*scope*). If required, a user can have multiple privacy statements for his/her eKYC data represented by different tags.

```
AadhaarNo
POI:
  name, dob
POA:
  co, house, street, lm, loc, vtc, subdist, dist,
  state, country, pc, po
LData:
  lang, name, co, house, street, lm, loc, vtc,
  subdist, dist, state, country, pc, po
Pht:
  <Base64 encoded JPEG photo of resident
Org:
  dep, desig
Other:
  email
```

Figure 9: Least information assumed to be available from eKYC for this paper

```
PS.ID:
  5
PS.Tag:
  eSignDoc
PS.Purpose:
  eSign
PS.AP:
  (email = *@finance.iitg.ac.in)       OR
  (org = IITG AND org.dep = finance)   OR
  (desig >= director)
PS.Scope:
  poi.name, poi.dob, poa.country, Ldata.lang
```

Figure 10: Example of a PEaFGAC statement

It is assumed that all entities which request eKYC data also have their eKYC information available with *UIDAI*. This include not just the users but the organizations such as *ESP*s as well. To better know an entity (both users and organizations), it is proposed that eKYC fields are expanded to include more details such as entity type (indicating whether the subject is a human or an organization), resident's organization, resident's department, resident's designation, etc. When an entity attempts to access eKYC data of a resident, entity's eKYC data and purpose for which the eKYC data is sought are verified against *PEaFGAC* statement protecting eKYC data to decide whether the requisite access can be granted or not. Only if the access can be granted, will the eKYC data be provided to the requesting entity. The eKYC data provided to the entity is limited in scope by *PEaFGAC* statement. For rest of this paper, eKYC data is assumed to consists of at least the information presented in figure 9.

6.2 PEaFGAC Token

The token structure introduced in [4] can be enhanced to include resident claims including *PEaFGAC* statement [fig-

ure 11]. Before sending an *eSign* request, resident creates a *PEaFGAC* token by sending a token generation request to *UIDAI* through *ASP* and *ESP*. During token generation process, resident is redirected to *UIDAI* web page where (s)he provides OTP value for authentication and *PEaFGAC* statement for privacy and fine-grained access to his/her eKYC data. Subsequently *UIDAI* verifies the OTP value and signs cryptographic hash of the statement with its private key and stores the signed hash in resident's private claims and stores the statement in plaintext in resident's public claims section of *PEaFGAC* digital access token.

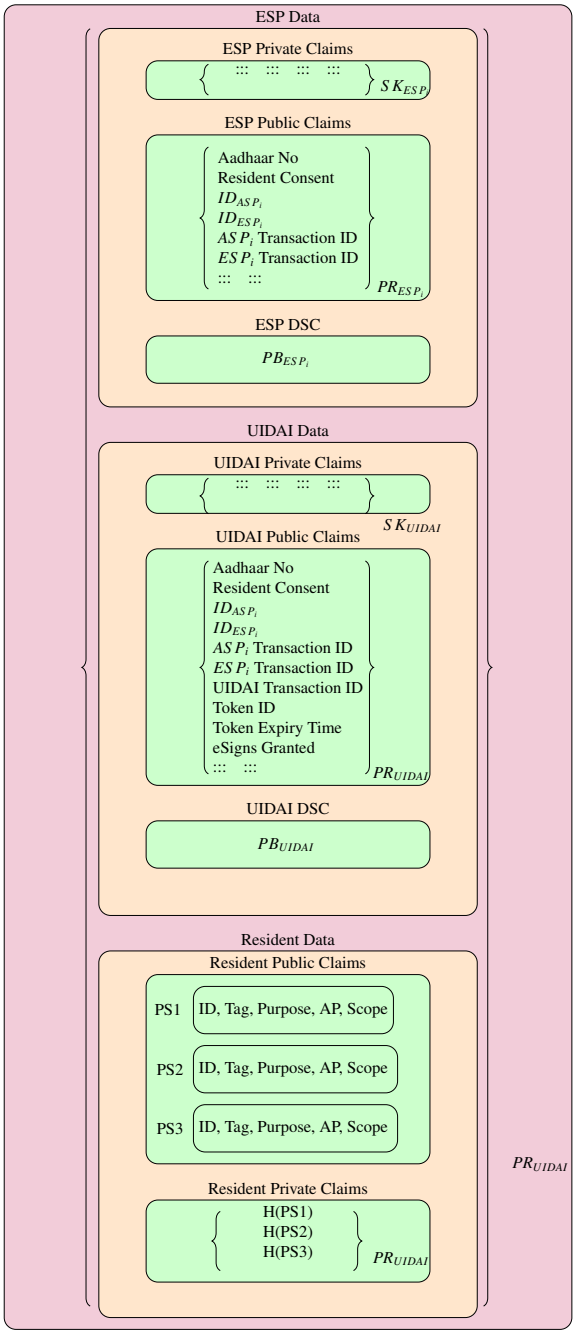


Figure 11: Proposed PEaFGAC Token Structure

Figure 12 depicts the sequence and details of communication messages among participating entities for generation of a token. First column indicates the message identifier, second column indicates the participating entities and the direction of communication and third column indicates what message is sent and how it is constructed.

Login to ASP		
$TGM1$	$R \rightarrow ASP$	$Generate\ nonce\ n1_{R_i}$ $DataR_1 = ID_{R_i} \| PW_{R_i} \| PB_{B_i} \| n1_{R_i}$ $SignR_1 = \{H(DataR_1)\}_{PR_{B_i}}$ $\{\ loginReq(DataR_1, SignR_1)\ \}_{PB_{ASP_i}}$
$TGM2$	$R \leftarrow ASP$	$DataA_1 = C_{R_i} \oplus (n1_{R_i} + 1)$ $SignA_1 = \{H(DataA_1)\}_{PR_{ASP}}$ $\{\ loginRes(DataA_1, SignA_1)\}_{PB_{B_i}}$
Generate OTP		
$TGM3$	$R \rightarrow ASP$	$Generate\ nonce\ n2_{R_i}$ $DataR_2 = AadhaarNo_{R_i} \| C_{R_i} \| n2_{R_i}$ $SignR_2 = \{H(DataR_2)\}_{PR_{B_i}}$ $\{\ getotpASPReq(DataR_2, SignR_2)\ \}_{PB_{ASP_i}}$
$TGM4$	$ASP \rightarrow ESP$	$Generate\ nonce\ n1_{ASP_i}$ $DataA_1 = AadhaarNo_{R_i} \| ID_{ASP_i} \|$ $License_{ASP_i} \| TID_{ASP_i} \| n1_{ASP_i}$ $SignA_1 = \{H(DataA_1)\}_{PR_{ASP_i}}$ $\{\ getotpESPReq(DataA_1, SignA_1)\ \}_{PB_{ESP_i}}$
$TGM5$	$ESP \leftarrow UIDAI$	$Generate\ nonce\ n1_{ESP_i}$ $DataE_1 = AadhaarNo_{R_i} \| ID_{ESP_i} \|$ $License_{ESP_i} \| TID_{ESP_i} \| n1_{ESP_i}$ $SignE_1 = \{H(DataE_1)\}_{PR_{ESP_i}}$ $\{\ getotpReq(DataE_1, SignE_1)\ \}_{PB_{UIDAI}}$
$TGM6$	$R \leftarrow UIDAI$	$\{\ OTP\ \}_{SecureCellularNetwork}$
$TGM7$	$ESP \leftarrow UIDAI$	$DataU_1 = returnStatus \| TID_{ESP_i} \|$ $MaskedMobileNo \| (n1_{ESP_i} + 1)$ $SignU_1 = \{H(DataU_1)\}_{PR_{UIDAI}}$ $\{\ getotpRes(DataU_1, SignU_1)\}_{PB_{ESP_i}}$
$TGM8$	$ESP \leftarrow ASP$	$DataE_2 = returnStatus \| TID_{ASP_i} \|$ $MaskedMobileNo \| (n1_{ASP_i} + 1)$ $SignE_2 = \{H(DataE_2)\}_{PR_{ESP_i}}$ $\{\ getotpESPRes(DataE_2, SignE_2)\}_{PB_{ASP_i}}$
$TGM9$	$ESP \leftarrow R$	$DataA_2 = returnStatus \|$ $MaskedMobileNo \| (n2_{R_i} + 1)$ $SignA_2 = \{H(DataA_2)\}_{PR_{ASP_i}}$ $\{\ getotpASPRes(DataA_2, SignA_2)\}_{PB_{B_i}}$
Generate Token		
$TGM10$	$R \rightarrow ASP$	$Generate\ nonce\ n3_{R_i}$ $DataR_3 = C_{R_i} \| n3_{R_i}$ $SignR_3 = \{H(DataR_3)\}_{PR_{B_i}}$ $\{\ gentokASPReq(DataR_3, SignR_3)\ \}_{PB_{ASP_i}}$
$TGM11$	$ASP \rightarrow ESP$	$Generate\ nonce\ n2_{ASP_i}$ $DataA_3 = AadhaarNo_{R_i} \| ID_{ASP_i} \| License_{ASP_i} \|$ $TID_{ASP_i} \| n2_{ASP_i}$ $SignA_3 = \{H(DataA_3)\}_{PR_{ASP_i}}$ $\{\ gentokESPReq(DataA_3, SignA_3)\ \}_{PB_{ESP_i}}$
$TGM12$	$ESP \rightarrow UIDAI$	$Generate\ nonce\ n3_{ESP_i}$ $DataE_3 = AadhaarNo_{R_i} \| ID_{ESP_i} \|$ $License_{ESP_i} \| TID_{ESP_i} \|$ $n3_{ESP_i}$ $SignE_3 = \{H(DataE_3)\}_{PR_{ESP_i}}$ $\{\ gentokUIDAIReq(DataE_3, SignE_3)\ \}_{PB_{UIDAI}}$
$TGM13$	$UIDAI \leftarrow ESP$	$Generate\ nonce\ n1_{UIDAI}$ $DataU_1 = UIDAIRedirectURL$ $(ForTakingPrivacyStatements) \|\|$ $PB_{UIDAI} \| TID_{ESP_i} \| n1_{UIDAI}$ $SignU_1 = \{H(DataU_1)\}_{PR_{UIDAI}}$ $\{\ genpsUIDAIReq(DataU_1, SignU_1)\ \}_{PB_{ESP_i}}$
$TGM14$	$ESP \leftarrow ASP$	$Generate\ nonce\ n4_{ESP_i}$ $DataE_4 = UIDAIRedirectURL$ $(ForTakingPrivacyStatements)$ $\| PB_{UIDAI} \| TID_{ASP_i} \| n4_{ESP_i}$ $SignE_4 = \{H(DataE_3)\}_{PR_{ESP_i}}$ $\{\ genpsESPReq(DataE_4, SignE_4)\ \}_{PB_{ASP_i}}$
$TGM15$	$ASP \leftarrow R$	$Generate\ nonce\ n4_{R_i}$ $DataA_4 = \{\ Present\ UIDAIRedirectURL\ to$ $Resident\ which\ requests\ him$ $to\ provide\ OTP\ Value\ and\ Privacy$ $statements \| PB_{UIDAI} \| n4_{R_i}\}$ $SignA_4 = \{H(DataA_4)\}_{PR_{ASP_i}}$ $\{\ genpsASPReq(DataA_4, SignA_4)\ \}_{PB_{B_i}}$
$TGM16$	$R \rightarrow UIDAI$	$\{\ PEaFGACPrivacyStatements\ \}_{PB_{UIDAI}}$
PEaFGAC Token Generation Request		
$TGM17$	$R \rightarrow ASP$	$DataR_4 = C_{R_i} \| (n4_{R_i} + 1)$ $SignR_4 = \{H(DataA_4)\}_{PR_{B_i}}$ $\{\ genpsASPRes(DataA_4, SignA_4)\ \}_{PB_{B_i}}$
$TGM18$	$ASP \leftarrow ESP$	$DataA_5 = (n4_{ESP_i} + 1)$ $SignA_5 = \{H(DataA_4)\}_{PR_{ASP_i}}$ $\{\ genpsESPRes(DataA_4, SignA_4)\ \}_{PB_{ESP_i}}$
$TGM19$	$ESP \leftarrow UIDAI$	$DataE_5 = (n1_{UIDAI} + 1)$ $SignE_5 = \{H(DataE_5)\}_{PR_{ESP_i}}$ $\{\ genpsUIDAIRes(DataE_5, SignE_5)\ \}_{PB_{UIDAI}}$
$TGM20$	$ESP \leftarrow UIDAI$	$Create\ PEaFGACTokenAT_{Ri}$ $DataU_2 = AT_{Ri} \| (n3_{ESP_i} + 1)$ $SignU_2 = \{H(DataE_5)\}_{PR_{ESP_i}}$ $\{\ gentokUIDAIRes(DataE_5, SignE_5)\ \}_{PB_{UIDAI}}$
$TGM21$	$ASP \leftarrow ESP$	$DataE_6 = (n2_{ASP_i} + 1)$ $SignE_6 = \{H(DataE_5)\}_{PR_{ESP_i}}$ $\{\ gentokESPRes(DataE_6, SignE_6)\ \}_{PB_{ASP_i}}$
$TGM22$	$R \leftarrow ASP$	$DataA_5 = (n3_{R_i} + 1)$ $SignA_5 = \{H(DataE_5)\}_{PR_{ASP_i}}$ $\{\ gentokESPRes(DataA_5, SignA_5)\ \}_{PB_{B_i}}$

Figure 12: Proposed PEaFGAC Token Generation protocol

6.3 Proposed Privacy Aware eSign model using PEaFGAC statement for eKYC data and PEaFGAC token

This section presents how *PEaFGC* statement for eKYC data and *PEaFGAC* token can be used to implement *Privacy Aware eSign*. It is assumed that *PEaFGAC* token has already been generated as explained in section 6.2. It is also assumed that the communication channel between resident and *ASP* is secured using SSL/TLS, between *ASP* and *ESP* is secured using SSL/TLS and between *ESP* and *UIDAI* is secured using dedicated secure leased lines.

Figure 13 depicts the sequence and details of communication messages transferred in *eSign* request in case eKYC needs to be fetched again.

Login to ASP		
$M1$	$R \rightarrow ASP$	$Generate\ nonce\ n1_{R_i}$ $DataR_1 = ID_{R_i} \| PW_{R_i} \| PB_{B_i} \| n1_{R_i}$ $SignR_1 = \{H(DataR_1)\}_{PR_{B_i}}$ $\{\ loginReq(DataR_1, SignR_1)\ \}_{PB_{ASP_i}}$
$M2$	$R \leftarrow ASP$	$DataA_1 = C_{R_i} \oplus (n1_{R_i} + 1)$ $SignA_1 = \{H(DataA_1)\}_{PR_{ASP}}$ $\{\ loginRes(DataA_1, SignA_1)\}_{PB_{B_i}}$
Initiate eSign request		
$M3$	$R \rightarrow ASP$	$Generate\ nonce\ n2_{R_i}$ $DataR_2 = M_i \| consent_{use_ekyc} \|$ $consent_{genuse_at} \| C_{R_i} \| n2_{R_i}$ $SignR_2 = \{H(DataR_2)\}_{PR_{B_i}}$ $\{\ signdocASPReq(DataR_2, SignR_2)\ \}_{PB_{ASP_i}}$
$M4$	$ASP \leftarrow ESP$	$Generate\ nonce\ n1_{ASP_i}$ $DataA_3 = H(M_i) \| AadhaarNo_{R_i} \| ID_{ASP_i} \|$ $License_{ASP_i} \| TID_{ASP_i} \|$ $consent_{use_ekyc} \| consent_{genuse_at} \|$ $n1_{ASP_i}$ $SignA_3 = \{H(DataA_3)\}_{PR_{ASP_i}}$ $\{\ signdocESPReq(DataA_3, SignA_3)\ \}_{PB_{ESP_i}}$
Retrieve eKYC (reusing access token) and sign document		
$M5$	$ESP \rightarrow UIDAI$	$Generate\ nonce\ n1_{ESP_i}$ $If\ AT_{R_i}\ is\ valid\ use\ it$ $DataE_5 = AT_{R_i} \| H(M_i) \| n1_{ESP_i}$ $SignE_5 = \{H(DataE_5)\}_{PR_{ESP_i}}$ $\{\ kycESPReq(DataE_5, SignE_5)\ \}_{PB_{UIDAI}}$
$M6$	$ESP \leftarrow UIDAI$	$Retrieve\ eKYC_{ESP_i}$ $Retrieve\ AT_{R_i} \rightarrow UC \rightarrow AP$ $Verify\ whether\ access\ can\ be\ granted$ $based\ on\ above\ two\ parameters.$ $eKYC_{R_i} = eKYC\ of\ resident\ scoped$ $by\ AT_{R_i} \rightarrow UC \rightarrow scope$ $DataU_3 = eKYC_{R_i} \| (n1_{ESP_i} + 1)$ $SignU_3 = \{H(DataU_3)\}_{PR_{UIDAI}}$ $\{\ kycESPRes(DataU_3, SignU_3)\ \}_{PB_{ESP_i}}$
$M7$	$ASP \leftarrow ESP$	$Generate\ key\ pair\ PB_{R_i}\ PR_{R_i}\ using\ eKYC_{R_i}$ $SignChain = \{\ PB_{R_i}\ \}_{PR_{ESP_i}} \|$ $\{\ PB_{ESP_i}\ \}_{PR_{CCA}}$ $DSC_{R_i_M_i} = \{\ eKYC_{R_i} \| H(M_i)\ \}_{PR_{R_i}} \|$ $PB_{R_i} \| SignChain$ $Delete\ PR_{R_i}$ $DataU_2 = DSC_{R_i_M_i} \| TID_{ASP_i} \| (n1_{ASP_i} + 1)$ $SignU_2 = \{H(DataU_2)\}_{PR_{ESP_i}}$ $\{\ signdocESPRes(DataU_2, SignU_2)\ \}_{PB_{ASP_i}}$
$M8$	$R \leftarrow ASP$	$\{M_i\}_{eSign_R_i_ESP_i} = M_i \| DSC_{R_i_M_i}$ $DataA_6 = \{M_i\}_{eSign_R_i_ESP_i} \| (n2_{R_i} + 1)$ $SignA_6 = \{H(DataA_6)\}_{PR_{ASP_i}}$ $\{\ signdocASPRes(DataA_6, SignA_6)\ \}_{PB_{B_i}}$
$M9$	$R \leftrightarrow R$	$Verify\ correctness\ of\ eKYC,\ H(M)\ and$ $SignChain\ in\ \{M_i\}_{eSign}$

Figure 13: Proposed Privacy Aware eSign model

7 Formal Security Analysis of the proposed model using BAN Logic

This section presents formal security analysis of the proposed scheme using *Burrows-Abadi-Needham (BAN) logic* [11]. Because of space limitation, it is assumed that PEaFGAC token has already been generated securely.Analysis of the token generation request can be done similarly.*BAN logic* is a well-known model used to find beliefs of participants in a cryptographic protocol.

Operator Usage	Description
$P \models X$	P believes statement X
$P \lhd X$	P sees statement X
$P \mapsto X$	P controls X
$\#(X)$	Message X is fresh
$P \overset{K}{\leftrightarrow} Q$	P and Q share key K
$\overset{K}{\mapsto} P$	P has K as its public key
$P \overset{X}{\rightleftharpoons} Q$	Formula X is a secret known only to P and Q
$\{X\}_K$	Formula X is encrypted using K
$\langle X \rangle_Y$	Formula X is combined with formula Y

Figure 14: Fundamental BAN operators

Operator Usage	Description
$M_{eSign_R_ESP_CCA}$	eSign of message M is done by Resident R through ESP approved by CCA $M_{eSign_R_ESP_CCA}$ $= M \| DSC_{R_M}$ $= M \| \{eKYC_R \| H(M)\}_{PR_R} \|$ $PB_R \| SignChain$ $= M \| \{eKYC_R \| H(M)\}_{PR_R} \|$ $PB_R \| \{PB_R\}_{PR_{ESP}} \|$ $\{PB_{ESP}\}_{PR_{CCA}}$
$P \models E_i \overset{secure}{\longrightarrow} E_j$	P believes that communication from entity E_i to E_j is secure
$P \models E_i \overset{secure}{\longleftarrow} E_j$	P believes that communication from entity E_j to E_i is secure
$P \models E_i \overset{secure}{\longleftrightarrow} E_j$	P believes that communication between entities E_i and E_j is secure in both directions
$P \models E_i \overset{ACTPerm}{\longleftrightarrow} E_j$	P believes that entity E_i has given permission for action ACT to entity E_j
$P \models C_R \rightsquigarrow E_i$	P believes that cookie C_R is associated with logged-in entity E_i
$E_R \models E_R \overset{C_{E_R} \rightarrow ID_{E_R}}{\longrightarrow} E_R$	E_R believes that it has securely communicated its identity ID_{E_R} to entity E_R through cookie C_{E_R}

Figure 15: Extended BAN operators

The security environment is assumed to be based on *Delev-Yao model* in which all messages are communicated over public channels and an attacker can see, modify, compose and replay messages but cannot break cryptographic principles. The security environment also assumes that an attacker can decipher messages if he has a valid decryption key. Some of the fundamental operators used in *BAN logic* are defined in figure 14. An extension to *BAN* logic, defined in figure 15 is required to analyse the proposed model.

Rules of Inference

[R1:] *Message meaning rules* concern the interpretation of messages. They all derive beliefs about the origin of messages.

For shared secrets, the inference rule is

$$\frac{P \models Q \overset{Y}{\rightleftharpoons} P,\ P \lhd \langle X \rangle_Y}{P \models Q \rhd X}$$

That is, if P believes that the secret Y is shared with Q and sees $\langle X \rangle_Y$, then P believes that Q once said X.

[R2:] The *nonce-verification* rule expresses the check that a message is recent, and hence, that the sender still believes in it:

$$\frac{P \models \#(X),\ P \models Q \rhd X}{P \models Q \models X}$$

That is, if P believes that X could have been uttered only recently and that Q once said X, then P believes that Q believes X.

[R3:] The *jurisdiction* rule states that if P believes that Q has jurisdiction over X, then P trusts Q on the truth of X:

$$\frac{P \models Q \Rightarrow X,\ P \models Q \models X}{P \models X}$$

[R4:] The *seeing* rule states that if a principal sees a formula, then he also sees its components, provided he knows the necessary keys:

$$\frac{P \lhd (X,Y)}{P \lhd X},\quad \frac{P \lhd \langle X \rangle_Y}{P \lhd X},\quad \frac{P \models Q \overset{K}{\leftrightarrow} P(.) P \lhd \{X\}_K}{P \lhd X},$$

$$\frac{P \models \overset{K}{\mapsto} P,\ P \lhd \{X\}_K}{P \lhd X},\quad \frac{P \models \overset{K}{\mapsto} P,\ P \lhd \{X\}_{K^{-1}}}{P \lhd X}.$$

Note that if P sees X and P sees Y it does NOT follow that P sees (X, Y) since that means that X and Y were uttered at the same time.

[R5:] The *fresh* rule states that if one part of the formula is fresh, then the entire formula must be fresh.

$$\frac{P \models \#(X)}{P \models \#(X, Y)}.$$

[R6:] The *belief* rule states that if P believes one part of the formula, then it also believe part of the formula.

$$\frac{P \models (X, Y)}{P \models (X)}.$$

Extended Rules of Inference

[R7:] If receiver entity E_R believes that C_R is a cookie associated with a unique session from resident R, PB_B is public key with browser used by resident R, PB_{E_R} is public key of receiver entity E_R, n_R is a fresh nonce generated by R, E_R receives message of the form $\{CommMsgReq(X||C_R||n_R, \{H(X||C_R||n_R)\}_{PR_{B_i}})\ \}_{PB_{E_R}}$, then E_R believes that X is sent by entity R and communication channel from R to E_R is secure and no message is observed, modified or replayed by an intruder.

$$\begin{array}{lll} E_R & \models & \overset{PB_{E_R}}{\longmapsto} E_R, \\ E_R & \models & C_R \rightsquigarrow S, \\ E_R & \models & \#n_R, \\ E_R & \models & \{\{Y\}_{PR_R}\}_{PB_R} = Y \\ E_R & \lhd & \{\ CommMsgReq\ (\\ & & \quad X||C_R||n_R, \\ & & \quad \{H(X||C_R||n_R)\}_{PR_{R_i}})\ \}_{PB_{E_R}} \\ \hline E_R & \models & R \xrightarrow{Secure} E_R, \\ E_R & \models & R \rhd X \end{array}$$

[R8:] If receiver entity E_R believes that PB_{E_R} is public key of receiver entity E_R, n_{E_R} is a fresh nonce generated by E_R, E_R receives message of the form $\{\ CommMsgReq\ (X||n_{E_R}, \{H(X||n_{E_R})\}_{PR_{E_R}})\ \}_{PB_{E_R}}$, then E_R believes that X is sent by entity E_R and communication channel from E_R to E_R is secure and no message is observed, modified or replayed by an intruder.

$$\begin{array}{lll} E_R & \models & \overset{PB_{E_R}}{\longmapsto} E_R, \\ E_R & \models & \{\{Y\}_{PR_{E_R}}\}_{PB_{E_R}} = Y, \\ E_R & \models & \#n_{E_R} \\ E_R & \lhd & \{\ CommMsgReq\ (\\ & & \quad X||n_{E_R}, \\ & & \quad \{H(X||n_{E_R})\}_{PR_{E_R}})\ \}_{PB_{E_R}} \\ \hline E_R & \models & E_R \xrightarrow{Secure} E_R \\ E_R & \models & E_R \rhd X \end{array}$$

[R9:] If receiver entity E_R believes that communication from all possible sender entities E_{R_i} to E_R $(\forall i = 1...n)$ is secure, then E_R believes that communication channel to E_R is secure and no message is observed, modified or replayed by an intruder.

[R10:] If resident R believes that C_R is a cookie associated with a unique session from resident R, PB_{E_R} is public key of entity E_R, n_R was a fresh nonce generated by R and used in a previous request call from R to E_R, R receives a message of the form $\{CommMsgRes(X||(n_R+1), \{H(X||(n_R+1)\}_{PR_{E_R}})\}_{PB_{E_R}}$, then E_R believes that X is sent by entity R and communication channel from R to E_R is secure and no message is observed, modified or replayed by an intruder.

[R11:] If sender entity E_R believes that PR_{E_R} is private key of sender entity E_R, PB_{E_R} is public key of receiver entity E_R, n_{E_R} was a fresh nonce generated by R and used in a previous request call from E_R to E_R, E_R receives message of the form $\{CommMsgRes\ (X||(n_{E_R}+1), \{H(X||(n_{E_R}+1)\}_{PR_{E_R}})\}_{PB_{E_R}}$, then E_R believes that X is sent by entity E_R and communication channel from E_R to E_R is secure and no message is observed, modified or replayed by an intruder.

[R12:] If sender entity E_R believes that communication from all possible receiver entities E_{R_i} $(\forall i = 1...n)$ is secure, then E_R believes that communication channel to E_R is secure and no message is observed, modified or replayed by an intruder.

[R13:] An electronic signature (M_{ieSign}) is a valid signature only when resident verifies that three main parts in signature, viz., $eKYC$, $H(M)$ and $SignChain$ are as expected.

Assumptions

The protocol makes several assumptions. The assumptions relevant for the discussion of this paper are listed below.

[A1:] It is assumed that all sessions from all residents R_i keeps their cookie C_{R_i} secret.

[A2-A6:] The scheme makes several assumptions about public keys. For example, R_i believes that PB_{ASP_i} is public key of ASP_i. Similar to this, other entities also make similar assumptions. These assumptions are listed below.

$$\begin{array}{llll} R_i & \models & \overset{PB_{ASP_i}}{\longmapsto} & ASP_i \quad ...(A2) \\ ESP_i & \models & \overset{PB_{ASP_i}}{\longmapsto} & ASP_i \quad ...(A3) \\ ASP_i & \models & \overset{PB_{ESP_i}}{\longmapsto} & ESP_i \quad ...(A4) \\ UIDAI & \models & \overset{PB_{ESP_i}}{\longmapsto} & ESP_i \quad ...(A5) \\ ESP_i & \models & \overset{PB_{UIDAI}}{\longmapsto} & UIDAI \quad ...(A6) \end{array}$$

[A7:] ASP_i assumes that all valid cookies C_{R_i} are associated with a valid ongoing session from a unique valid user R_i already logged in to ASP_i portal.

$$ASP_i \models C_{R_i} \rightsquigarrow ID_{R_i} \quad \forall i = 1..n$$

[A8-A15:] R_i and ASP_i assumes that all nonce $n*_{R_i}$ (where $*$ is any integer used in the scheme) are fresh. Similar to this, other entities also make similar assumptions. These assumptions are listed below.

$$\begin{array}{lll} R_i & \models & \#n*_{R_i} \quad ...(A8) \\ ASP_i & \models & \#n*_{R_i} \quad ...(A9) \\ ASP_i & \models & \#n*_{ASP_i} \quad ...(A10) \\ ESP_i & \models & \#n*_{ASP_i} \quad ...(A11) \\ ESP_i & \models & \#n*_{ESP_i} \quad ...(A12) \\ UIDAi & \models & \#n*_{ESP_i} \quad ...(A13) \\ UIDAI & \models & \#n*_{UIDAI} \quad ...(A14) \\ ESP_i & \models & \#n*_{UIDAI} \quad ...(A15) \end{array}$$

[A16:] It is assumed that when ASP_i receives message of the form $CommMsg(DataA_j, SignA_j)_{PB_{ASP_i}}$ from ESP_i, it has verified the validity of data, i.e., $\{SignA_j\}_{PB_{ESP_j}} = H(DataA_j)$. The same assumption is made for all entities receiving messages of this form.

Goals to be achieved.

Following are the goals which are envisaged to be achieved by the proposed model.

[G1-G6:] Sender entity must be sure that the data received by receiver entity is same as what was sent by it and is not modified, observed or replayed by an intruder after it was sent by the sender entity. Similarly, receiver entity must be sure that the data received by it is same as what was sent by sender entity and is not modified, observed or replayed by an intruder after it was sent by the sender entity.

$$\begin{array}{llll} ASP_i & \models & R_i & \overset{secure}{\longleftrightarrow} \quad ASP_i \\ R_i & \models & R_i & \overset{secure}{\longleftrightarrow} \quad ASP_i \\ ASP_i & \models & ASP_i & \overset{secure}{\longleftrightarrow} \quad ESP_i \\ ESP_i & \models & ASP_i & \overset{secure}{\longleftrightarrow} \quad ESP_i \\ ESP_i & \models & ESP_i & \overset{secure}{\longleftrightarrow} \quad UIDAI \\ UIDAI & \models & ESP_i & \overset{secure}{\longleftrightarrow} \quad UIDAI \end{array}$$

[G7:] Resident R_i must be sure that at the end what he receives is indeed a digital signature on message M_i, signed by resident's private key and generated by the genuine ESP_i. $R_i \models M_{ieSign} = M_{eSign_R_i_ESP_i_CCA}$

Idealization

BAN idealization of communication messages in communication phase is shown in table 1

Table 1: BAN Idealization of Proposed Protocol (Part I: M1-3) and (Part II: M4-8)

$M1$	ASP_i	$\lhd$	$\{login\ ($ $ID_{R_i} \| PW_{R_i} \| PB_{B_i} \| n1_{R_i}$ $\{H(ID_{R_i} \| PW_{R_i} \| PB_{B_i} \| n1_{R_i})\}_{PR_{B_i}}$ $)\}_{PB_{ASP_i}}$
$M2$	R_i	$\lhd$	$\{loginRes\ ($ $C_{R_i} \oplus (n1_{R_i}+1)$ $\{H(C_{R_i} \oplus (n1_{R_i}+1))\}_{PR_{ASP_i}}$ $)\}_{PB_{B_i}}$
$M3$	ASP_i	$\lhd$	$\{signdocASPReq\ ($ $M_i \| consent_{use_ekyc} \| consent_{genuse_at} \| C_{R_i} \| n2_{R_i})$, $\{H(M_i \| consent_{use_ekyc} \| consent_{genuse_at} \| C_{R_i} \| n2_{R_i})\}_{PR_{B_i}}$ $)\}_{PB_{ASP_i}}$
$M4$	ESP_i	$\lhd$	$\{signdocESPReq\ ($ $H(M_i) \| AadhaarNo_{R_i} \| ID_{ASP_i} \| License_{ASP_i} \| TID_{ASP_i} \| consent_{use_ekyc} \| consent_{genuse_at} \| n1_{ASP_i}$, $\{H(H(M_i) \| AadhaarNo_{R_i} \| ID_{ASP_i} \| License_{ASP_i} \| TID_{ASP_i} \| consent_{use_ekyc} \| consent_{genuse_at} \| n1_{ASP_i})\}_{PR_{ASP_i}}$ $)\}_{PB_{ESP_i}}$
$M5$	$UIDAI$	$\lhd$	$\{kycESPReq\ ($ $AT_{R_i} \| H(M_i) \| n1_{ESP_i}$, $\{H(AT_{R_i} \| H(M_i) \| n1_{ESP_i})\}_{PR_{ESP_i}}$ $)\}_{PB_{UIDAI}}$

$M6$	ESP_i	$\triangleleft$	$\{kycESPRes\ ($ $eKYC_{R_i} \| (n1_{ESP_i} + 1)$ $\{H(eKYC_{R_i} \|$ $(n1_{ESP_i} + 1))\}_{PR_{UIDAI}}$ $\}_{PB_{ESP_i}}$
$M7$	ASP_i	$\triangleleft$	$\{signdocESPRes\ ($ $\{eKYC_{R_i} \| H(M_i)\}_{PR_{R_i}} \|$ $PB_{R_i} \| \{PB_{R_i}\}_{PR_{ESP_i}} \|$ $\{PB_{ESP_i}\}_{PR_{CCA}} \|$ $TID_{ASP_i} \| (n1_{ASP_i} + 1),$ $\{H(eKYC_{R_i} \| H(M_i)\}_{PR_{R_i}} \|$ $PB_{R_i} \| \{PB_{R_i}\}_{PR_{ESP_i}} \|$ $\{PB_{ESP_i}\}_{PR_{CCA}} \|$ $TID_{ASP_i} \|$ $(n1_{ASP_i} + 1))\}_{PR_{ESP_i}}$ $)$ $\}_{PB_{ASP_i}}$
$M8$	R_i	$\triangleleft$	$\{signdocASPRes\ ($ $M_i \| \{eKYC_{R_i} \| H(M_i)\}_{PR_{R_i}}$ $\| PB_{R_i} \| \{PB_{R_i}\}_{PR_{ESP_i}} \|$ $\{PB_{ESP_i}\}_{PR_{CCA}} \| (n2_{R_i} + 1)$ $\{H(M_i \| \{eKYC_{R_i} \| H(M_i)\}_{PR_{R_i}}$ $\| PB_{R_i} \| \{PB_{R_i}\}_{PR_{ESP_i}} \|$ $\{PB_{ESP_i}\}_{PR_{CCA}} \|$ $(n2_{R_i} + 1))\}_{ASP_i}$ $)$ $\}_{PB_{B_i}}$

Analysis

[P1-P6:] Using messages M1, M3 and rule R7, it can be deduced that ASP_i believes that communication from R_i to ASP_i is secure. Using messages M2, M8 and rule R11, it can be deduced that ASP_i believes that communication from ASP_i to R_i is secure. From these two deductions, it can further be deduced that ASP_i believes that communication between R_i and ASP_i is secure in both directions.

Using M1, M3, R7, R8,
$ASP_i \models R_i \xrightarrow{secure} ASP_i$ ($I1$)

Using M2, M8, R11,
$ASP_i \models R_i \xleftarrow{secure} ASP_i$ ($I2$)

Using I1 and I2,
$ASP_i \models R_i \xleftrightarrow{secure} ASP_i$ ($G1$: *Proved*)

Using M1, M3, R10, R11
$ASP_i \models R_i \xrightarrow{secure} ASP_i$ ($I3$)

Using M2, M8, R8,
$ASP_i \models R_i \xleftarrow{secure} ASP_i$ ($I4$)

Using I3 and I4,
$ASP_i \models R_i \xleftrightarrow{secure} ASP_i$ ($G2$: *Proved*)

Using M4, R7,
$ESP_i \models ASP_i \xrightarrow{secure} ESP_i$ ($I5$)

Using M7, R11,
$ESP_i \models ASP_i \xleftarrow{secure} ESP_i$ ($I6$)

Using I5 and I6,
$ESP_i \models ASP_i \xleftrightarrow{secure} ESP_i$ ($G3$: *Proved*)

Using M4, R11,
$ASP_i \models ASP_i \xrightarrow{secure} ESP_i$ ($I7$)

Using M7, R8,
$ASP_i \models ASP_i \xleftarrow{secure} ESP_i$ ($I8$)

Using I5 and I6,
$ASP_i \models ASP_i \xleftrightarrow{secure} ESP_i$ ($G4$: *Proved*)

Using M5, R7,
$UIDAI \models ESP_i \xrightarrow{secure} UIDAI$ ($I7$)

Using M6, R11,
$UIDAI \models ESP_i \xleftarrow{secure} UIDAI$ ($I8$)

Using I7 and I8,
$UIDAI \models ESP_i \xleftrightarrow{secure} UIDAI$
($G5$: *Proved*)

Using M5, R11,
$ESP_i \models ESP_i \xrightarrow{secure} UIDAI$ ($I9$)

Using M6, R8,
$ESP_i \models ESP_i \xleftarrow{secure} UIDAI$ ($I0$)

Using I9 and I10,
$ESP_i \models ESP_i \xleftrightarrow{secure} UIDAI$ ($G6$: *Proved*)

[P7:] Using message M9 and rule R13, it can be deduced that R_i believes that $\{M_i\}_{eSign}$ is a valid electronic signature.

Using M9 and R13,
$R_i \models \{M_i\}_{eSign} = \{M_i\}_{eSign_R_i_ESP_i_CCA}$
($G7$: *Proved*)

8 Conclusion

This work is an extension of the work [4] on enhancing amortized performance of *eSign* by using digital access tokens including claims from *ESP* and *UIDAI*. In this work, the digital access token introduced in [4] is extended to include privacy and fine-grained access control statements for access to resident's eKYC data. This enhanced token can be used by third entities to access the protected eKYC data with better privacy and fine-grained access control rules enforced by the resident. A formal security analysis of the proposed model using *BAN logic* is also presented.

References

[1] UIDAI, "Unique Identification Authority of India", [Online]. Available: https://uidai.gov.in (2017).

[2] Wikipedia, "Aadhaar", [Online]. Available: https://en.wikipedia.org/wiki/Aadhaar (2017).

[3] Reserve Bank of India, "Master Direction - Know Your Customer (KYC) Direction, 2016", [Online]. Available: https://rbidocs.rbi.org.in/rdocs/notification/PDF s/18MDKYCD8E68EB1 3629A4A82BE8E06E606C57E57.PDF, 2018.

[4] Bakshi, Puneet, Neelakantan Subramanian, and Sukumar Nandi. "Using digital tokens to improve amortized performance of eSign." 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2018.

[5] Paquin, Christian, and Greg Zaverucha. "U-prove cryptographic specification v1. 1." Technical Report, Microsoft Corporation (2011).

[6] Chaum, David. "Blind signatures for untraceable payments." Advances in cryptology. Springer, Boston, MA, 1983.

[7] Hardt, Dick. The OAuth 2.0 authorization framework. No. RFC 6749. 2012.

[8] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).

[9] UIDAI, "Aadhaar e-KYC API Specification - Version 2.1", [Online]. Available: http://www.cca.gov.in/cca/sites/default/files/files/eSign-APIv2.1.pdf, 2017

[10] CCA, "eSign Service", [Online]. Available: http://cca.gov.in/cca/?q=eSign.html (2017).

[11] Burrows, Michael, Martin Abadi, and Roger Michael Needham. "A logic of authentication." Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 426.1871 (1989): 233-271.

[12] Rivest, Ronald L., Adi Shamir, and Leonard Adleman. "A method for obtaining digital signatures and public-key cryptosystems." Communications of the ACM 21.2 (1978): 120-126.

[13] Merkle, Ralph C. "A digital signature based on a conventional encryption function." Conference on the theory and application of cryptographic techniques. Springer, Berlin, Heidelberg, 1987.

[14] Paquin, Christian, and Greg Zaverucha. "U-prove cryptographic specification v1. 1." Technical Report, Microsoft Corporation (2011).

[15] Chaum, David. "Blind signatures for untraceable payments." Advances in cryptology. Springer, Boston, MA, 1983.

[16] Hardt, Dick. The OAuth 2.0 authorization framework. No. RFC 6749. 2012.

[17] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).

[18] PKIA2017, "Development of Smart Authentication and Identification in Asia", [Online]. Available: http://pkiindia.in/pkia/#preceding, 2017.

[19] CCA, "Pki framework in India", [Online]. Available: http://www.cca.gov.in/cca/?q=pki˙frame work.html (2015).

[20] CCA, "Public key certificate classes", [Online]. Available: http://www.cca.gov.in/cca/?q=node/45 (2017).

[21] CCA, "Empanelled eSign Service Providers", [Online]. Available: http://www.cca.gov.in/cca/?q=service-providers.html (2017).

[22] Jones, Michael, John Bradley, and Nat Sakimura. Json web token (jwt). No. RFC 7519. 2015.

[23] Jones, Mike, et al. Cbor web token (cwt). No. RFC 8392. 2018.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 351-356 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Location Prediction based on Variable-order Markov Model with Time Feature and User's Spatio-temporal Rule

Ying Xia*[1], Yu Gong[1], Xu Zhang[1], Hae-young Bae[2]

[1]School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, 400065, China

[2]Department of Computer Engineering, Inha University, 22212, South Korea

ARTICLE INFO

Article history:
Received: 28 January, 2019
Accepted: 05 April, 2019
Online: 24 April, 2019

Keywords:
Location prediction
Active grid
Variable-order Markov model
Time feature
Spatio-temporal rule

ABSTRACT

Location-based service has been widely used in modern life. It brings a lot of convenience to our lives. Improving the accuracy of location prediction can provide better location-based service. We propose a location prediction method based on the variable-order Markov model with time feature and user's spatio-temporal rule. First, the user's trajectory data needs to be abstracted, and then the useful stay points in the user's trajectory are extracted. The location prediction is performed by scoring each candidate area, and the score is composed of scores in time and space dimensions. Finally, for the possible zero frequency problem, it is solved by mining the spatio-temporal rule of the user. Experiments using the actual data set GeoLife show that the proposed method improves the prediction accuracy.

1. Introduction

With the popularity of mobile location devices, the acquisition of location information is becoming more and more easier. Location-based services have become common services in our lives. It is necessary for us to mine location data and use the acquired knowledge to extend location services or improve service quality. Location prediction is an important research point in the field of location data mining, which can provide us with better recommendation services, social networking services and so on.

This paper is an extension of work originally presented in conference name [1].

Markov chain is a set of discrete random variables with Markov and the probability of each event depends only on the previous event. The Markov process is named after the Russian mathematician Andrey Markov. In this process, the future state is only related to the current state and has nothing to do with the past. Markov chain is divided into continuous Markov chain and discrete Markov chain. In discrete state-space, discrete time random process is called discrete Markov chain, and continuous time random process is called continuous Markov chain. [2]

Markov chain is a common statistical model, which is widely used in various kinds of prediction. It has the advantages of simplicity and efficiency, and is a good representation of time series data. Therefore, it can also be used for location prediction.

In literature [3], a location prediction method based on adaptive higher-order Markov model was proposed, which significantly improves the accuracy of location prediction. The difference between a higher-order Markov model and a Markov model is that the future state depends on the current state of the model and the past k-1 state of the model, not just the current state. However, it does not solve the problem of not finding current state in the history data, which is called the zero frequency problem. Literature [4] used variable-order Markov model for location prediction. It is based on high-order Markov model and makes the best use of history data. The zero frequency problem is also solved by an approach called escaped rule. Moreover, we found that many people have a regular life activity. By analyzing these people' history data, certain rules can be mined in time and space. These rules can be used in location predication. In the literature [5], the author not only considered the influence of user space factors on location prediction, but also considered the time factor, which further improved the accuracy of location prediction. This lets us notice the importance of time factor in location predication. Literature [6] provided a method for location prediction by using space-time rules, which is not only simple, but also can obtain good prediction results. These studies gave us a lot of inspiration.

*Corresponding Author: Ying Xia, Email: xiaying@cqupt.edu.cn

With lots of studies, a location predication method based on variable-order Markov model with time feature and user's spatio-temporal rule was proposed. It included the preprocessing of data, the construction of variable-order Markov model with time feature and the analysis of the time and spatial rules of users. First, data preprocessing makes the location data more suitable for location prediction. Since time and spatial factors both have impact on location predication, a variable-order Markov model with time feature is constructed for location predication. Finally, for the case that the variable-order Markov model cannot be used to perform the prediction, the prediction is made by analyzing users' time and spatial rules.

2. Data preprocessing

Before predicting the location, location data needs to be preprocessed, which includes three steps, that is, data cleansing, location information abstraction and stay point extraction.

2.1. Data cleaning

The purpose of data cleansing is to make the data more real, regular, and easy to handle. First, you need to modify the irregular data to make it more regular. Second, data not in research range needs to be deleted. In this study, Beijing was chosen as the research object so the range of location data control is between east longitude 115.25 ˚ to 117.3 ˚ and north latitude 39.26˚ to 41.03˚. If not, it is regarded as data that is not in this range. Finally, incomplete or unreal data are cleaned up.

2.2. Location information abstraction

Location data is usually obtained by mobile device. And it is mainly composed of coordinate location data. The data are very detailed and it is difficult for us to analyze them. Location data needs to be abstracted into area data to perform conventional mining. The abstraction of location data also helps decrease computational complexity. In literature [7], the idea of building active grid was proposed. We use this method to abstract the location data:

- First, the user's activity range is set in a large rectangular area and the edge is parallel to the longitude and latitude lines.
- Then, the rectangular area is divided into a square grid. And the user's location point is mapped to the grid.
- Finally, active grid is connected based on time passage. This is user's active grid sequence.

The raw location data is composed of points. This is original trajectory data. $\{p_1,p_2...p_i...p_n\}$. P_i is the location point in the trajectory data. After the abstraction, the original user location sequence $\{p_1,p_2...p_n\}$ is converted to active grid sequence $\{z_1,z_2...z_i...z_n\}$. z_i is the grid in the abstract trajectory data.

2.3. Stay points extraction.

The stay points of moving objects are important points or areas such as shopping centers, workplaces, places of residence and so on. These points should be emphasized in location-based services. Therefore, the stay points in the trajectory need to be extracted. And there are many ways to extract stay points. The method of considering time and space factors is not only simple but also effective. It is based on the principle that a place is a stop point when you stay longer than the set time. The extraction method includes the following steps:

- Set the time value *t*, which is the standard of residence time.
- Search the area with residence time exceeding t in the trajectory data. The residence time is calculated by the difference between the maximum residence time and the minimum residence time.
- The stay points by time chronological order constitutes the new track.

The trajectory data composed of residence points are extracted by the above method. Algorithm 1 is the pseudocode used for the residence point extraction.

Algorithm 1 Stay points extraction

Input: User's historical data d_data and stay time threshold t

Output: User's stay points s_area

The maximum time value to stay in the area time_max

The Minimum time value to stay in the area time_min

Sort the data of d_data by time

for each area of d_data **do**

  stay time = time_max -time_min

  if stay time > t

    The area add to the stay area s_ area

end for

return s_ area

3. Location Prediction

The method of location prediction includes constructing variable-order Markov model with time feature and analysis of user's time and spatial rules. It's named location prediction based on variable-order Markov model with time feature and user's spatio-temporal rule (VMTST).

3.1. Variable-order Markov Model with Time Feature construction and prediction methods

Variable-order Markov model is an improved model based on high-order Markov model. It makes use of the excellent prediction ability of the high-order Markov model and further realizes the automatic selection of high-order Markov model order. Time feature are also important for location prediction. Thus, time feature and variable-order Markov model are combined, and the scoring method is used to achieve the location predication. The method is as follows:

- Extract the historical track data composed of user's stay points.

- Search history tracks based on the user's current track. If the matching historical trajectory information is found, the transition probability of the next grid in these trajectories is calculated. It is called *P*.
- If the matching track information is not searched, some active grids must be removed and then new current track is created. The earliest active grids need to be removed because the earlier an active grid arrives, the less impact it will have on its future location. This process requires multiple executions until matching trajectory data is found.

  For example, Figure 1 shows the user's historical trajectory. If the user's current trajectory (Z_2-Z_3), all historical trajectory data containing (Z_2-Z_3) are searched. Z_4 after calculation is the highest region with transition probability. Region Z_4 is the predicated result. But if the user's current trajectory is (Z_4-Z_5-Z_6). The current trajectory is not found in history data. Z_4 is the earliest region and has the least influence on location predication, so Z_4 region should be deleted and search it again from the historical track.

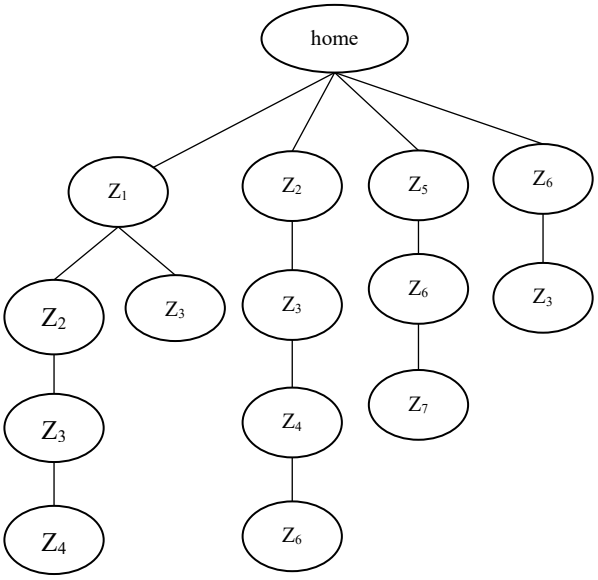


Figure 1: User's historical trajectory

- For each candidate region, their scores are composed of spatial accessibility scores and time feature scores. The scoring formula is as follows:

$$Score = W_1 * P + W_2 * S \quad (1)$$

  W_1 and W_2 are the weights which are possible to reach spatial scores and time characteristic scores in the total scores.

  Time information includes two parts: date and time. Therefore, the score of time feature *S* is composed of two parts: the periodic coefficient of the date and the similarity score of time.

$$S = C * Si \quad (2)$$

  Si is the similarity score of time. Clustering and normalization idea is used here. It is obtained by the following formula.

$$Si = 1 - t/12 \quad (3)$$

$$t = th - tn \quad (4)$$

  th is the hourly portion of historian data record.

  tn is the hour component of current status.

  C is the periodic coefficient of date. It is divided into workdays and holidays. It is calculated by the following rules:

  If record date and forecast date are workdays, it means the location data is regular. Thus, the value of *C* is greater than 1.

  If historical date and forecast date are holidays, it means the location data is irregular. Thus, the value of *C* is less than 1.

  In other cases, *C* is 1 .

- The zone with the highest score is used as the predicting outcomes.

The location prediction algorithm of variable-order Markov model with time feature is as follows:

Algorithm 2 Construction of the variable-order Markov model with time feature

Input: users historical data *h_data*,users current trajectary *c_try*

Output: users next area *n_zone*

The weight of the space in the total score W_1

The weight of the time in the total score W_2

while Search from historical data (*h_data*) with user's current trajectory(*c_try)* return false

c_try = remove the earliest zone from the user's current trajectory

if the number of areas in user's current trajectory ==1

break

Calculate the transition probability *P* of each next areas in the matched trajectory

for each area of next areas **do**

Calculate the time similarity score *Si* and date coefficient *C* for this area

Time feature score for this area *S* = *C*Si*

Score for this area *score* = W_1*P+W_2*S

end for

Search for the highest scored area *n_zone* in the next area of historical data

return *n_zone*

3.2. Analyze Spatio-temporal Rule of Moving Objects

When users have less historical data, it is difficult to search matching historical track information. Location predication is done by analyzing the user's time and spatial rules. Spatio-temporal rule mainly refers to the regularity of users' time and spatial positions at different times. The following methods are used to analyze the spatio-temporal rule of users:

- The time interval is set to *t*.
- In the user's historical data, count the number of occurrence of user in each interval and each active grid. The region with the maximum number is the hot spot region during this time period.
- Set the numbered ID for each time period in time order.
- The result of spatio-temporal rule is composed of ID, hot region.

During this period, the more time periods are divided, the better the prediction effect is, but the more time is needed for calculation. It is easy to understand that the more active grids are present, the more likely they are to become a mobile destination for users. Therefore, hot spot active grids with the maximum number of occurrences are the prediction area based on spatio-temporal rule. Algorithm 3 describes this process.

In the location prediction, the user's current activity time matches the ID in the spatio-temporal rule. Next ID hotspot region is searched in the spatio-temporal rule, which is considered a place where users may arrive in the future. This process is algorithm 4.

Algorithm 3 Analysis of user's spatio-temporal rule

Input: users historical data *h_data*, time period *t*
Output: users spatio-temporal rule result *s_t_result*
Dividing the time period according to the value of *t*
for each time period **do**
  Set a number for each time period *t_ID*
  Calculate the area with the most user arrivals during this time period *max_area*
  t_ID and *max_area* added to the spatio-temporal rule result *s_t_result*
end for
return *s_t_result*

Algorithm 4 Location prediction using user's spatio-temporal rule

Input:users spatio-temporal rule result *s_t_result*
recording time of the user's current trajectory *r_time*
Output:location prediction area obtained by users spatio-temporal rule *next_area*
Get the number of the *r_time* named *n_ID*
for each number in the time period **do**
  if *n_ID* = number in the time period
    get the hot area of the next time period *hot_area*
    next_area=hot_area
return *next_area*

Prediction accuracy and recall rate were used as evaluation indexes in the experiment. The formula is as follows:

4. Experiment

4.1. Experimental design

In order to verify the effectiveness of the proposed prediction method, experiments were performed using the GeoLife [8] trajectory data set. This data set includes trajectory data of 182 users over five years. It has about 18,670 tracks, including longitude, latitude, time and other information. And it records track data like users' work, entertainment and sports as well as other outdoor activities.

After data cleansing, extract the user's ID, longitude, latitude, date, time and other attributes. The experiment data set is shown in table 1. 10% data is randomly selected for test data and the rest are training data. Experimental research area is east longitude 115.25˚ to 117.3˚ and latitude 39.26˚ to 41.03˚. The size of grid unit is 100m*100m, the limit time of residence time point t is 10 min and the time interval selected for the analysis of users' spatio-temporal rule is 2 hours.

Contrast model:

MM: High order Markov location predication model.

VOMM: Location prediction model proposed in literature[4] uses variable-order Markov model for location prediction and escaped rules to solve the zero frequency problem.

VMSTPM: The location prediction model is proposed in literature [1]. This method uses the variable-order Markov model for location prediction and also uses spatio-temporal rule to solve the zero frequency problem.

Table 1: Sample of Data

ID	latitude	longitude	date	time
10	39.921695	116.472345	2007-08-04	03:30:34
10	39.961675	116.472315	2007-08-04	03:30:36
10	39.924583	116.472290	2007-08-04	03:30:38
10	39.921572	116.472290	2007-08-04	03:30:39
10	39.921565	116.475588	2007-08-04	03:30:40
10	39.921577	116.472210	2007-08-04	03:30:41
10	39.921570	116.472300	2007-08-04	03:30:42
10	39.921576	116.475588	2007-08-04	03:30:44
10	39.921580	116.472290	2007-08-04	03:30:46
10	39.921576	116.472321	2007-08-04	03:30:50

4.2. Evaluation Factors

$$Precision = \frac{accurate}{accurate+wrong} \quad (5)$$

$$Recall = \frac{accurate+wrong}{total} \quad (6)$$

Table 2: The value of W_1 and W_2 in different situation

	case 1	case 2	case 3	case 4	case 5	case 6	case 7	case 8	case 9
W_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
W_2	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

Table 3: The value of C_1 and C_2 in different situations

	case 1	case 2	case 3	case 4	case 5	case 6	case 7	case 8	case 9
C1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
C2	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

In the above equations, accurate indicate the number of correct predictions, wrong indicate the number of incorrect predictions total indicate the number of all testing trajectories.

4.3. Analysis of Experimental Results

First, the coefficients in the candidate region scoring formula were tested by experiments and prediction accuracy was used as the evaluation index. The experiment selected nine different values, as shown in table 2. From table 2, we can see that the prediction accuracy is the highest when W_1 and W_2 are 0.3 and 0.7. It proves the importance of time feature to location prediction. Thus, when the weight of time feature is great, the prediction accuracy is higher.

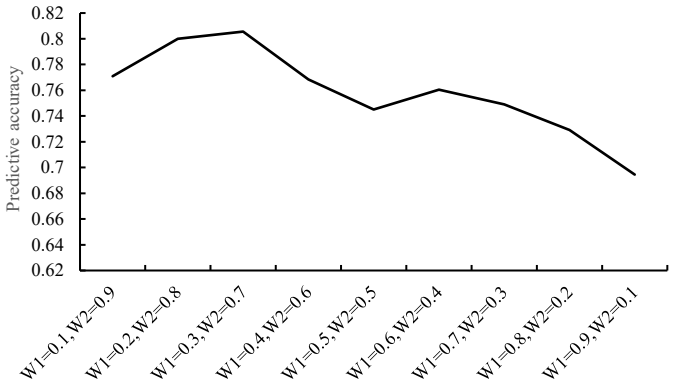


Figure 2: Prediction precision of W_1 and W_2 in different situations

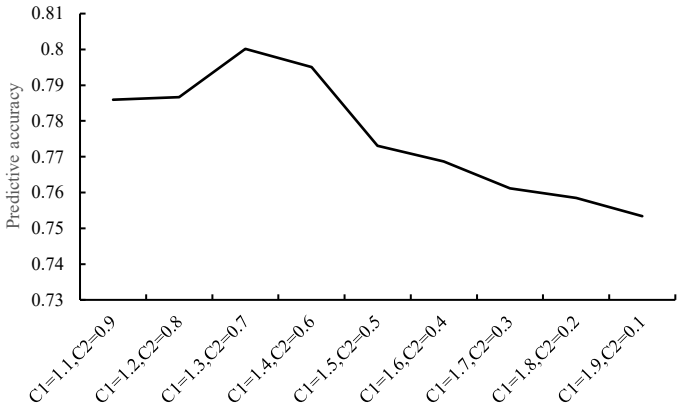


Figure 3: Prediction precision of C_1 and C_2 in different situations

Besides, the period coefficient value of the date obtained the best value by experiments. The value of C is C_1 when record date and prediction date are both workdays. And the value of C is C_2 when both historical date and prediction date are holidays. C_1 and C_2 take the values in table 3. It can be seen from figure 3 that the prediction outcome is best when C_1 is 1.3 and C_2 is 0.7. Therefore, other experiments were performed with these values.

Then, lots of prediction experiments were performed to verify the prediction outcome of the model. The prediction outcome is shown in figure 4. And VMTST at different orders showed more accurate prediction result. The best prediction result is obtained in the second order, which is about 30% higher than the original high-order Markov model and about 20% higher than the VOMM model. In figure 5, we can see that VMTST model is more stable in higher-order predictions. Therefore, it is better to use the current state and a history state for location prediction. This will get the optimum result. These results show that VMTST model has good predictive performance.



Figure 4: Prediction precision with different orders

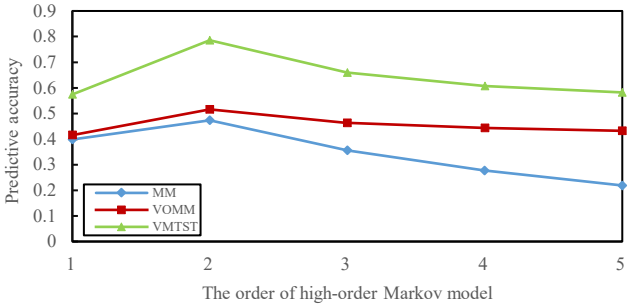


Figure 5: Prediction precision with different models

Experiments were conducted to verify the importance of time in prediction. The prediction accuracy is still used as the evaluation index in the experiment. And the results are shown in figure 6. It can be seen that the prediction accuracy of VMTST is higher than that of VMSTPM with different orders, which proves that time feature is a very important factor in location prediction.

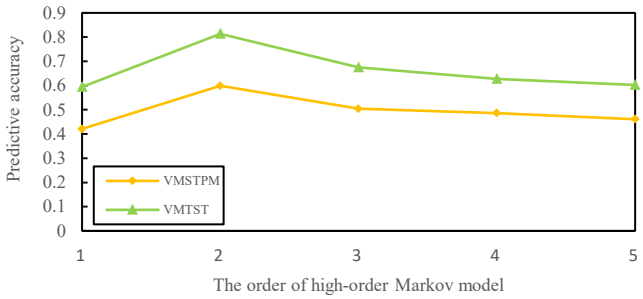


Figure 6: Prediction precision of VMSTPM and VMTST

The figure 7 is the comparison of recall rates of three models. As you can see, the recall rate of VMTST is 1.0, which is higher than high order Markov model. This means that it is effective to solve the zero frequency problem by using the user's spatio-temporal rule. However, the recall rate of VOMM model is also 1.0, which means VOMM model also solves the zero frequency problem. More experiments are required to verify the performance of VMTST.

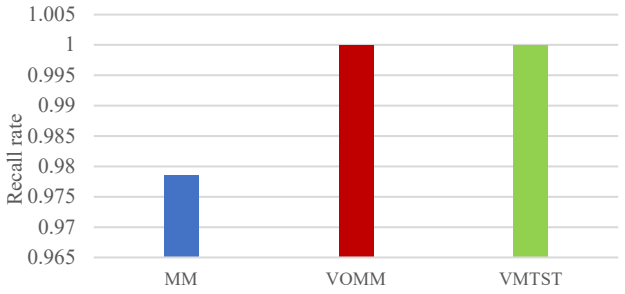


Figure 7: Recall rate of different models

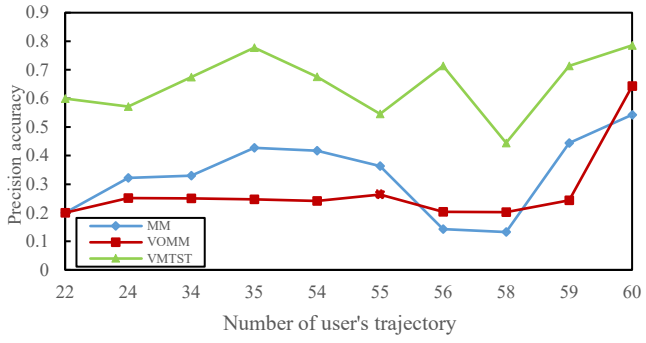


Figure 8: Prediction precision of users with fewer trajectory

The above experiments show that the spatio-temporal rule can solve the zero frequency problem. But, it is necessary to verify the performance of spatial and temporal laws in predicting accuracy through experiments. The experiment selected 10 users with less historical data. Therefore, it is easier to generate zero frequency problem. And it can be seen from the experimental results in figure 8 that when the trajectory quantity is small, the prediction accuracy of VMTST model is still higher than other models. Therefore, the spatio-temporal rule of users is more suitable to solve the problem of zero frequency.

5. Conclusion

A location prediction method was proposed in this paper, including preprocessing of location data, construction of variable-order Markov model with time feature and analysis of user's spatio-temporal rule. The preprocessing of position data makes the research scope fixed and the track data easier to analyze and process. Variable-order Markov model is an improved model of high-order Markov and it has higher location prediction accuracy. The time features are added to the variable-order Markov model to further increase the prediction accuracy. The problem of zero frequency can be solved by analyzing the spatio-temporal rule of users. Finally, lots of experiments were carried out using the actual trajectory data set.

By summing up the experiments, we can conclude that VMTST model improves the prediction accuracy. Time features are very important for location prediction. And the spatio-temporal rule can solve the problem of zero frequency, and be also more suitable.

In the future, we will look for new methods to mine the spatio-temporal rule of users so as to get the higher prediction accuracy. Also we will study the application of location prediction method in practice.

References

[1] Y. Xia, Y. Gong, X. Zhang and H. Y. Bae, "Location Prediction Based on Variable-order Markov Model and User's Spatio-temporal Rule" in International Conference on Information and Communication Technology Convergence, Korea, 2018. https://doi.org/10.1109/ICTC.2018.8539593

[2] P. A. Gagniuc, Markov Chains: From Theory to Implementation and Experimentation, John Wiley & Sons, 2017.

[3] M. Q. Lü, L. Chen, G. C. Chen, "Position Prediction Based on Adaptive Multi-Order Markov Model" Journal of Computer Research and Development, 47(10), 1764-1770, 2010

[4] J. Yang, J. Xu, M. Xu, N. Zheng, Y. Chen, "Predicting next location using a variable order Markov model." ACM Sigspatial International Workshop on Geostreaming, USA, 2014. https://doi.org/10.1145/2676552.2676557

[5] X. Li, C. Yu, L. Ju, J. Qin, Y. Zhang, L. Dou, S. Y. Qing, "Position Prediction System Based on Spatio-Temporal Regularity of Object Mobility" Information Systems, 75, 43-55, 2018. https://doi.org/10.1016/j.is.2018.02.004

[6] H. M. Wong, S. V. Tseng, C. C. J. Tseng. "Long-Term User Location Prediction Using Deep Learning and Periodic Pattern Mining." in International Conference on Advanced Data Mining and Applications, 2017. https://doi.org/10.1007/978-3-319-69179-4_41

[7] F. Giannotti, M. Nanni, F. Pinelli, "Trajectory pattern mining." in Acm Sigkdd International Conference on Knowledge Discovery & Data Mining ACM, 2007. https://doi.org/10.1145/1281192.1281230

[8] Y Zheng, X Xie, W. Y. Ma, "GeoLife: A Collaborative Social Networking Service among User, location and trajectory" IEEE Data Eng, 33(2), 32-39, 2010

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 357-369 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Computer Security as an Engineering Practice: A System Engineering Discussion

Robert M. Beswick*

Mission Design and Navigation Section, NASA, Jet Propulsion Laboratory, California Institute of Technology 91109 USA

ARTICLE INFO

Article history:
Received: 29 June 2018
Accepted: 03 April 2019
Online: 24 April 2019

Keywords:
Computer Security Engineering
Security Fault Tolerance
Ground Data System
Defense in Depth
Defense in Breadth
Least Privilege
Vulnerability Removal
Absolute Security
Sterility in Implementation
Time Based Security
Communications Security
Operations Security
Complementary Intersection

ABSTRACT

We examine design principles from more than 20 years of experience in the implementation and protection of mission critical flight systems used by the Mission Design and Navigation Section at NASA's Jet Propulsion Laboratory. Spacecraft navigation has rigorous requirements for completeness and accuracy, often under critical and uncompromising time pressure. Fault tolerant and robust design in the ground data system is crucial for the numerous space missions we support, from the Cassini *orbital tour of Saturn to the Mars rover* Curiosity. *This begins with the examination of principles learned from fault tolerant design to protect against random failures, and continues to the consideration of computer security engineering as a derivative effort to protect against the promotion of malicious failures. Examples for best practice of reliable system design from aviation and computer industries are considered and security fault tolerance principles are derived from such efforts. Computer security design approaches are examined, both as abstract postulates (starting from cornerstone principles with the concepts of Confidentiality, Integrity, and Availability) and from implementation. Strategic design principles including* defense in depth, defense in breadth, least privilege, *and* vulnerability removal *are target points for the design. Additionally, we consider trust in the system over time from its sterile implementation, viewed against the backdrop of Time Based Security. The system design is assessed from external access data flows, through internal host security mechanisms, and finally to user access controls. Throughout this process we evaluate a* complementary intersection *– a balance between protecting the system and its ease of use by engineers. Finally, future improvements to secure system architecture are considered.*

1. Introduction

This paper examines approaches used in the process of securing the computer systems employed by the Navigation Ground Data System. This text is meant to serve as a discussion about best practices in computer security engineering. It is based on twenty years of the author's practical, "in the trenches", field experience on systems involved in this effort. These systems comprise a multi-mission network that encompasses the navigation elements of more than forty current and previous interplanetary flight missions here at the Jet Propulsion Laboratory. This paper does not seek to be a prescriptive document (do this one thing, buy this product, etc.), but instead seeks to examine a process of how secure systems are designed – i.e., what general security principles we have found valuable [1].

Few systems require as much resiliency or have as much risk of causing negative (and final) outcomes as the computer systems used in support of Flight Operations. Scant resources are often available for the maintenance of complex hardware and software architectures, and these high-availability/high-reliability systems are often expected to function without (and moreover cannot *tolerate*) the regular sorts of software updates expected in other computational environments.

As an example, one of these missions, Cassini, from its launch in October of 1997, would spend seven years crossing the solar system, arriving at and entering into Saturn orbit in June, 2004. It would then orbit Saturn nearly three hundred times over thirteen years, conducting hundreds of targeted flybys of Saturn's largest moons during its mission lifespan, finally coming to a fiery end in Saturn's atmosphere in September of 2017. This operational effort would be conducted on a network of computer systems having a requirement for no more than two minutes of unplanned downtime

*Robert M. Beswick, 4800 Oak Grove Dr., Pasadena, CA 91109 USA, 1-818-393-0539, Robert.M.Beswick@jpl.nasa.gov

a year (99.9995% availability) [2]. For a network that could not be upgraded (except in minor incremental steps) during the length of the entire 13-year Saturn orbital tour, maintaining and keeping such a computer network secure, as well as supporting backwards compatibility (and obsolescent hardware) would be a challenging problem. In fact, the systems used over the launch and interplanetary voyage across the Solar System contained less processing power, storage, and *total* memory than the author's current iPhone – and these systems would have to be kept functional and secure in case there was a need for backwards comparison.

While some organizations may have little need for a computational environment larger than a client-server configuration of a few systems, others may require a more complex environment. The Ground Data System for Navigation is constructed in the manner of a classic Development and Operations model with development and production networks of several hundred servers and workstations. This network is used by teams of engineers working, often under critical time pressures, with rigorous requirements for accuracy [2].

For this environment, it should be clear that a particular cryptographic protocol, a set of software, or even an appliance will not solve these security challenges. This is not an application problem but instead a systems problem. As in other engineering disciplines, in such an environment the design must be based on good *principles*, because, like with the pouring of a foundation of a building, you only get one such opportunity.

Part of the effort in this discussion stems from the author's own struggles to find a good systemic set of definitions and guidance while endeavoring to build a more secure network. What you see here is a reflection on the insights I was trying to find for my own efforts. Furthermore, many current efforts in computer security research are based not on building a secure system but tearing a system down. Indeed, a great deal of present investigation in computer security is based around the twin ideas of "if it ain't broke, don't fix it", and "try and see if can be broken". Accordingly, extensive research across the spectrum of computer security is being conducted in the field of penetration testing, where the primary methodology is to break into an existing setup and then fix the discovered problems. While this can be helpful in discovering specific flaws, it does not provide much help in trying to understand a more general architecture model of security.

An example of this is perhaps best exemplified by the software tool known as "Chaos Monkey" – part of an open source suite of tools called "The Simian Army" that was originally designed for the cloud infrastructure of the streaming media service Netflix [3]. This software set comprises a series of tools that help the design of resiliency in a set of virtual machine instances. It does so by randomly shutting down members of the set of virtual machines. By forcing developers and systems engineers to prepare for unexpected failure, a more fault tolerant network design will emerge (it is hoped). This has been further expanded and generalized to even more powerful tools that comprise "Chaos Gorilla", which randomly simulates the shutdown of an entire Amazon Availability Zone, or the even more devastating "Chaos Kong", which simulates the shutdown of an entire Amazon Region [4]. This design is an example of "survival of the fittest", incorporating a genetic algorithm-like approach to the design of secure systems.

Critical questions that should be raised in conjunction with this software are, "How does a developer or systems engineer build a more stable and secure system? What principles should be used? What methods should be avoided?"

Continuing with the metaphors given above, consider an analogy to the above (genetic) algorithm: a team of stone-age architects trying to build a bridge across a river, first by using a captive monkey, then a captive gorilla, and then a giant mythical beast to try throwing stones and batter a pile of rocks into a working bridge. While such a method will produce some results (albeit very slowly), what about other engineering approaches to design? What of the arch and the use of suspension? What about the consideration of tension and compression and the use of different materials in the construction of a bridge?

Clearly other engineering principles can be useful here in order to produce an initial design and improve upon it before bringing someone to attempt to tear it down. In like manner, this paper is a study on security architecture design, and hopes to add to such efforts by discussing principles on "how to build a (better) bridge".

A word about the expected audience of this paper: the design principles discussed in Sections 2 and 3 are aimed at top level design of secure systems, and may be of greatest use for project management and systems engineers (it could also be titled "how to avoid buying crap"), while Section 4 covers an example implementation targeted more for computer systems architects and systems administrators. The paradigms covered in Section 2 and 3 are observed derivations from fault tolerance and have wide applicability in systems engineering, while Section 4 applies this methodology to securing a specific computer system.

2. System design approaches: Fault Tolerance and Security

Fault Tolerance

2.1. Fault Tolerance

There are several valuable definitions for the concept of fault tolerance. Fault Tolerance, according to Carl Carson, comprises "...a design that enables a system to continue its intended operation, possibly at a reduced level, rather than failing completely, when some part of the system fails" [5, p. 167].

This has been categorized by Barry Johnson in the approaches taken for fault tolerance in microprocessor design as [6]:

- Minimize the number of points where a single fault will cause the whole system to fail.
- Graceful degradation – known also as "Fail-Gently", a system's ability to continue operating in the event of a failure, having a decrease in operating capacity no worse than necessary for the severity of the failure.
- Redundancy in components – both in space, having multiple parts that can be utilized, and in time, with the repeatability of an operation.

Jim Gray and Daniel Siewiorek help characterize such steps to provide high availability in computer systems by examining single points of failure to promote [7]:

- Independent failure – where each module functions so that if it fails, it does not impact other modules in the system.
- Design diversity – using hardware and software from differing organizations to promote independent failure modes (e.g. differing *types* of failure).

There are abundant examples of fault tolerant design. This is a well understood area of systems engineering. From the 135-year old Brooklyn Bridge in New York [8], to the current Eastern Interconnection power grid of the United States [9], approaches such as these are needed where one must trust the behavior of a system to work, and to work in predictable ways.

As an aspirational example, consider one such computer design further at the highest end of the reliability spectrum. The primary flight control system of the Boeing 777 achieves ultra-high reliability metrics. It is a highly redundant, highly available system comprising the fly-by-wire avionics controls and is a rare example of a true Triple Modular Redundant (TMR) system (with little exception the highest level of redundant design, having three redundant components for each single point of failure), in both computer nodes, software, and hardware. It has service metrics requiring a maximum rate of failure of the flight computers of 1.0×10^{-11} failure/hours (i.e. a failure of the flight computers less than 1 in 100 billion flight hours) [10]. This is an example of what can be done with sufficient effort and due diligence – a computer system millions of passengers a year put their trust in.

2.2. Security Fault Tolerance

In like manner, this paper considers computational systems that one can trust – as is done with the physical systems described above. From fault tolerance, we can derive similar approaches to deal with the actions of intelligent actors[1] rather than random chance or stress failure modes. Such an approach was described affectionately by Ross Anderson and Roger Needham as "Programming Satan's Computer" [11]. While the application of these principles we apply here to the flight computer and network security of our systems, the principles are applicable across the board to security design. Such design promotes:

- Secure systems should be resilient from random chance and predicted stress modes of failure.
- Secure systems should also be resistant to direct action.
- Simplicity of design, known as the popularized KISS principle, is a golden virtue in secure systems [12].

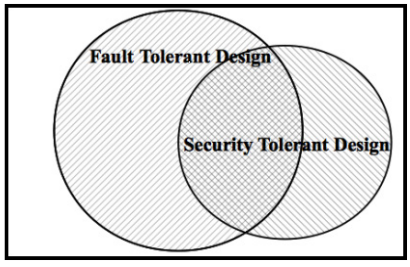


Figure 1: Relation of Fault Tolerant and Security Tolerant Design

These principles can be considered in four approaches:

2.2.1. For the most critical systems, use a machine that has only one function.

This approach has the benefit that in the event of a failure, you have lost only that one function. There are numerous examples in security design. Indeed, "appliance" IT systems, such as web servers, email servers, firewalls, and similar systems use this approach. As a good analogy: in (most) kitchens, refrigerators and ovens are not a part of the same appliance – even though they have the same task of changing and maintaining food temperature.

2.2.2. For redundancy in a security context ideally differing configurations should be used.

This technique considers that the systems in a given setup should not all have the same potential vulnerability (and therefore not truly redundant against a threat). It is an example of the previously discussed design diversity. Examples of this include the use of multiple arrays of web servers – running different OS/web server software, or application servers that use differing configurations and password sets, or multiple (different) backup systems – e.g. tape, disk, and cloud service providers. As an analogy, consider the commute from a major urban center: it is better to have several differing options for transport, be it freeway, commuter train, bus, or even surface streets. In the event one method is impacted, other options are available to return home.

2.2.3. Single points of failure of a system should be few, and truly independent.

As observed above, in a security context, systems that are single points of failure should be truly *independent* of other points of failure. Such single points of failure should be examined closely to ensure that they are actually single, independent, points of failure.

This goes hand in hand with the first point about single use. One example of this can be seen in network file server "appliances" that perform the function of serving files – other application sets such as virus checkers and configuration management tools may be run on the files themselves, but they are seldom loaded on the file servers. This design approach is often misunderstood in poor implementations, especially where security may be seen as a software package or an add on feature. As an analogy consider tires that serve as critical single points of failure for a modern car – in a similar manner, it would be absurd if a failure of the GPS system caused the tires to fail!

2.2.4. Systems should Fail-Safe and/or Fail-Gently.

In case of failure, a security system should degrade safely (known as Fail-Safe), and/or compromise only a limited part of the overall system or otherwise take an "acceptable" amount of time to fail ("Fail-Gently"). Examples of this abound, many of them utilizing cryptography, such as disk and file encryption, password login for most computer systems, firewalls, network segmentation, and network Intrusion Prevention Systems (IPS). An analogy can be found in the mechanical realm in that high grade safes are rated

[1] To be clear: this term is meant here to describe human attackers, not intelligent software agents. They may also be described as malicious actors, or threat actors depending on context. The colloquial and overlapping terms of hackers, crackers, or security hackers are also used in the general news media.

for time against tools, cutting/welding torches, and explosives. Some safes may have mechanisms that break or cause the safe to be unopenable if drilled or otherwise tampered with.

3. Computer Security Design

3.1. Abstract Principles

This computer security design incorporates the above techniques through increasing grades of refinement. Abstract concepts of Confidentiality, Integrity, and Availability (the traditional trinity of security) help to determine the "who, what, when, and where" of the security needs for navigation computation. These principles provide definition for the key concerns for securing a system, not in terms of technique, or subsystem protected, but rather in terms of what features in the computational environment must be protected.

3.1.1. Confidentiality

Confidentiality is referred to as "the concealment of information or resources" [13, p. 4]. This can be of high priority for some financial systems, where customer data is not only a crucial part of business operations, but also where strong legal regulations may come into play for control of customer information. It may be critical in military computer systems (for some cases it may be more desirable to destroy the system than allow the unauthorized release of information). This is significantly less crucial in the field of navigation computation. Such information for navigation comprises mechanisms used to authenticate access to the systems, network and system configuration information (possibly of use to subvert security), and restricted navigation software. An example of a baseline concern for Confidentiality can be seen even in the naming of individual systems, as such host names may reveal a great deal of information about the underlying network design [14]. Consider a hypothetical, badly named example from JPL: the host name "cas-web-serv3", which immediately gives information the system in question is running a web server daemon, for the Cassini project, and that most likely there are (at the minimum), two similar servers.

3.1.2. Integrity

Integrity is "the trustworthiness of data or resources ... usually phrased in terms of preventing improper or unauthorized change" [13, p. 5]. Integrity is particularly critical for navigation computation, as improper or unauthorized modification of the environment could cause very serious problems. Corrupted data sets, results or software could cause terminal errors in spacecraft control. With regard to the accuracy required for navigation of spacecraft missions, the sensitivity placed on the accuracy and integrity of the data, software, and corresponding results cannot be overstated. One such example of the critical nature of Integrity is seen in the previous case of the Cassini project. Integrity was critical on such a mission, which consisted of one of the largest teams of navigation engineers ever assembled, flying on the navigation computer system one of the most complex orbital and interplanetary trajectories ever designed [15].

3.1.3. Availability

Availability is "the ability to use information or resources" [13, p. 6]. Conceptually, it is the most direct derivation of principles coming from fault tolerant design. Availability in the context of navigation computation, has an additional concern that someone could deliberately try to deny access to a system service or data sets. The 24-hour-a-day, 365-day-a-year activity calendar of spacecraft navigation indeed makes this a crucial concern [15]. An event which causes the system to be unavailable, which in other industries might be cause for concern, could lead to the premature end of a multimillion dollar space mission.

3.2. Implementation

From these abstract principles, there are a number of paradigms that are employed for the implementation of better protection. Accordingly, strategic archetypes are used in the work toward good security design.

3.2.1. Defense in Depth

Defense in depth is a military security term with a long historical pedigree. Instead of a single defensive stronghold or chokepoint, this strategic principle is based around a series of overlapping defenses, forming a type of fault tolerant redundancy [16, pp. 40-44] in that a failure of one part of the system does not lead to the failure of the whole system. Considering a networked computer system, this can be seen on a variety of levels in overlapping defenses: starting at the network perimeter firewall, the network segmentation rules used on the network switches, the host based firewall, the protections applied to the processes that connect to the network, and then the internal host security measures that must be bypassed by a potential attacker. This is analogous to a defensive perimeter made up of a series of overlapping walls and fortifications, much like that seen in that of a medieval castle, or World War I trench warfare, presenting so many obstacles to entry that an adversary, having a choice, will pick a less well defended target [17, p. 259].

3.2.2. Defense in Breadth

Defense in breadth is a term serving as the converse for *defense in depth*. It involves the concept that a secure system design must consider the whole system in its implementation, or an attacker will simply ignore a well fortified part of the system and instead chose an easier route into the system. All of the previous paradigms of overlapping defenses, minimization, elimination, and security over the system life, must be considered for *every* part of the system. This is a difficult task and it does lay bare the challenge of the security engineering process. By definition this is an effort to examine the totality of all of the possible attacks against the system. Indeed, this defensive principle is referred to by Sami Saydjari as *Ensure Attack-Space Coverage*. He acknowledges that achieving this comprehensive coverage is a large and challenging problem, and one that is usually not well understood, but he underlines its necessity as a security engineering principle when he remarks "Depth without breadth is useless; breadth without depth weak. " [18, pp. 133-135] Consider an analogy from a failure of this principle: putting a well-locked door with a security system on a Hollywood film set facade; while it is possible to break through the front door, it is much easier to simply walk around the side or back of such a facade because most such sets do not even have a back wall. History is rife with excellent examples of the failure to consider the whole scope of ways an adversary could carry out an

attack. One such example can be seen in the WWII failure of the Maginot line on the French-German border. [2]

3.2.3. Least Privilege

The principle of *least privilege* is that subjects (processes and users by extrapolation) should be given only the rights and accesses for the tasks the subjects need to perform, and no more [13, p. 457]. This avoids many possibilities for the manipulation of a process (or manipulation by a user) to get them to do things they were not intended to do. There are many similarities to ideas in finance and accounting where financial mismanagement and corruption in corporations are prevented by not allowing any one person to take charge of all the corporate finances – also known as the "two man" rule. As an example, consider a hypothetical printing utility that can write to a central spooling directory as a privileged user. It is discovered that by manipulating the input files sent to the printer, it is possible to write to other directories on the system. With some effort and cleverness it may become possible to overwrite critical system files or even get remote access to the system. By changing the permissions of the spooling directory and removing the ability to run as a privileged user, the printing utility can continue to function, without enhanced access, thereby removing its capability to be misused. A similar analogy is a classic safe deposit box repository in a bank, where two people (each with a different key) are required for access to a given safe deposit box.

3.2.4. Vulnerability Removal

When we remove potential vulnerabilities, we can elude the many problems that come with those vulnerabilities, and sometimes avoid other issues as well. Matt Bishop refers to this as *The Principle of Economy of Mechanism* [13, pp. 459-460]. An example of *vulnerability removal*: a self-replicating worm that attacks email server processes can do little damage if you do not have an email server on your system. Likewise, if you do not have a web server running on your system, there is no need to monitor for such and keep up with attendant patches. Making sure that (along with the software and its own attendant maintenance) possible vulnerabilities in buggy software are not installed, can save tremendous resources and time over the system lifetime. As an analogy, this is akin to not just locking a side door into a building, but instead never constructing the door into the building in the first place.

3.2.5. Sterility and Absolute Security

We now consider the idea of *sterility.* This is a point along a spectrum of levels of trust in a given system. Understanding how much one can trust such a system may help determine how it should be used and where it should be placed on the network. A few (very rare) systems we consider to be sterile and have *absolute security*, the top of this spectrum, or "security beyond any reasonable doubt". [3] To be clear: this is a theoretical initial state, a starting point for definition, and *no* connected or exposed system can be said to be sterile and in this state. This is a valuable concept, especially as we consider the next paradigm. It is comparable to a definition of an initial starting point (a *t0*) that is useful when considering how that system might evolve over time.

Let us further consider, that a system implemented on a closed network, built from a secure image is secure, or *sterile* to the limits of confidence one has in the hardware (and the hardware vendors), network, and the software media (DVD or secure image and/or patch servers, etc.) used to construct the system. On such a spectrum of trust this system would be seen as approaching *absolute security* considering these limits of confidence. This useful idea can be seen as extrapolated (as systems in miniature) from credentials, passwords, or certificates that also have not been exposed. We can liken this to issues seen in medical care, or food preparation and handling, where much effort for hand washing, cleanliness, or cooking is undertaken to produce sterile areas, or clean food for consumption.

3.2.6. Security over time or robustness

The previous paradigms examine the implementation of a static system. Considering the evolution of a system state over time, one should also consider how the system is secure during its implementation and over its whole lifecycle. How robust is this system? How much can it be trusted and how does that trust change over time? This is similar to principles of Communications Security (COMSEC) and Operations Security (OPSEC). These are military security terms that examine the communications and background operations processes of military organizations, working to prevent compromise of communications, and the processes of such organizations. As in a military organization, our trust in a system will change over time. This is strongly dependent on the environment and networks that connect to this system. It is very important to examine these connections and method of contact to characterize the *security over time* of a system.

This concept is akin to what Winn Schwartau called "Time Based Security" [18, p. 34]. In Time Based Security, a system would be considered "secure" if the time to break through the system protection (Pt) was greater than the time required to detect (Dt) and to respond (Rt) to the intrusion. This can be expressed as:

$$Pt > Dt + Rt \quad (1)$$

Time Based Security suffers in the great difficulty of quantifying these values in (1), for the system protection, detection, and reaction time. Nonetheless Time Based Security does help illustrate well that security over time is a derivation from the fault tolerant concept of Fail-Gently, as discussed above.

Coming from the previous point about sterility, trust in a sterile system must change when it is attached to outside networks. Our confidence in such a system is then based on the security configuration of the external network, internal host, and user

[2] The Maginot line was designed to protect France from a German incursion in the event of a second world war. It was an excellent example of the principles discussed here, providing significant fault tolerant, defense in depth, with carefully considered access and both perimeter and internal security. A marvel of military engineering, its primary construction protected the French-German border. Most elements of the line were operational and fully prepared to continue fighting *after* the fall of France, and had to be ordered to surrender with the French capitulation. The designers famously did not consider (mostly due to the staggering costs to expand the fortification) an assault across the French-Belgian border.

[3] This is a functional term that in our model is akin to, but not the same as, the mathematical term *perfect security* – such as the unbreakable one-time pad [36].

controls of the system. Log mechanisms can help give indications of change to the system. If this trust is misplaced, the system could be compromised, and the system could become untrustworthy and itself a source of security concerns. As in the previous section we consider that this is akin to handling a compromised password or certificate, and in most cases the system must be rebuilt.

There is considerable kinship between these ideas and the study of infection and food-borne illness. The frequent admonition about hand washing before meals comes to mind. What things did one touch or pick up? Do you have a thermometer to see if the food is cooked? A brisket set out too long at a picnic may "go bad" and become a host to viruses, worms and other sources of contagion. It may be possible to sterilize it, or at least to clean the dish containing the item. Much care is required to safely clean dishes under food handling regulations, but having gone bad, food, no matter how savory, almost all of the time must be thrown out.[4]

4. System Evaluation

Keeping these principles of secure system design in mind, we consider an example case: a single computational node and the techniques and mechanisms put to use to create a secure system. Our model system is a server running Red Hat Enterprise Linux 7. Particular care is taken in its construction as it serves as a fundamental "building block" in our system plan. This security design takes on several design principles found in architecture in that the strength of a system can often be improved, not by what one adds, but by what one takes away. Also, this serves as an effective approach in considering the overall security of our environment. Moreover, many systems are actually allocated in a "one engineer, one workstation" ratio. Similar efforts can just as easily be a network segment, implemented in a network group.

It is useful to reiterate here: this is *example*. These security architecture guidelines can be applied in many other approaches and types of system setups. This is an outline that other system architects and system engineers can follow to create their own secure systems. It is hoped that this demonstration will help to make these key concepts and their application in design more clear. This system-design model is a simple one. It is to serve as a means to explain principles and their use in an architecture. As we extrapolate to a network, or a network-of-networks we get a clearer understanding of these principles and their use in a model. Complexity may be introduced in efforts to move towards a particular architecture model.

For example, one such effort called Zero Trust computing, moves towards a very atomic model of trust, that of so called "de-perimeterisation" where the outer defenses of a network system are assumed to be ineffective. Zero Trust uses the same design goals, but with an aggressive application of the principle of *least privilege* applied to every element of a network [20, pp. 2-3, 12-18]. It involves the hardening of every computational node, and every network connection and data access, and moves away from the traditional specialization of gateways, firewalls, and trusted subnets or zones. The model we discuss is agnostic to such efforts, as approaches we take here are equally valid for such designs. A more "classic" network security model is, simpler to explain.

Our fundamental design goal is to create a system that can be trusted and relied upon to perform the difficult task of spacecraft navigation. This design effort sought to find an optimal tradeoff between usability and security, the point of both maximal utility (for the users) and security: a point where users could work without being fettered by obtrusive security measures, while confident that their work is safe. Indeed, it is a mistaken, albeit popular, notion that security must interfere with usability, and that secure systems must necessarily be hard to use. Much of the time this misconception is due to a poor implementation of a secure feature, or a security control that was added on as an afterthought to a software system. Failing to consider these requirements and needs of the users also can lead to the perceived adversarial role of security in an organization. The idea of examining the security controls in consideration with the user environment is referred to by Matt Bishop as the *Principle of Psychological Acceptability* [13, pp. 464-466]. The strategy we consider here strives to find a complementary intersection between usability and security, an optimal point where these two groups of competing requirements reach their maximum effectiveness [21]. This is similar to the concept known as the Center of Percussion (known also as the so-called "sweet spot") as found in weapons, aircraft, sound engineering, and sports equipment design – where multiple factors combine into an optimum response from a given amount of effort [22]. Analogously, these competing needs are similar to the economic concept of the law of supply and demand.

Finding this complementary intersection requires the evaluation of both security and user requirements. For an example of this optimal intersection, consider the often maligned password change requirements of most institutions.[5] Usually eight or more alpha-numeric, and special characters are required, often with other "randomness" requirements. These passwords often have to be changed every 180 to 90 days (or less). As a matter of cryptographic security, the longer the encrypted password hash, the more difficult it becomes to conduct a brute force attack to obtain the unencrypted password. Hand in hand with this, changing the password more frequently helps ensure that even if the password is broken or obtained by some other means, the window of exposure is short before a new password is created. Evaluating these security requirements, without user input, leads to more "random" and longer passwords, approaching something looking akin to line noise, with changes occurring in shorter and shorter time intervals. Taken alone, this might be considered to be a positive trend.

However, it has been widely known for some time that the user response to such policies tends to decrease the security of the system [23]. Users, unable to remember their ever-changing passwords, proceed to create easily guessable password combinations, or worse still, write passwords down in easy-to-find locations such as next to computer monitors or under keyboards. Evaluating both security and user utility requirements can instead lead to an optimal solution for these concerns. This could include

[4] There are limitations to this analogy. One of my associates who is a professional chef cited several counter examples – however he was quick to note that this holds true for 99% of food handling concerns.

[5] A memorable and humorous example of this can be found at http://xkcd.com/936/

a longer password rotation with a longer password, or possibly a smart card and a password (for two- or three-factor authentication).

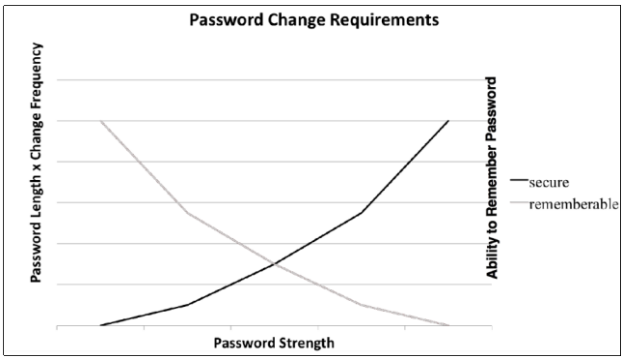


Figure 2: Complementary intersection of password controls.

But how can this optimal intersection point of security and usability for this system be determined? For this system design an initial setup is undertaken stemming from security design requirements and user needs. Numerous manuals for securing systems can be used *a priori* to determine a baseline configuration. We have used recommendations from the National Institutes for Standards and Technology (NIST) [24], the Information Assurance Directorate (previously named the System and Network Attack Center) of the National Security Agency (NSA) [25], or the Center for Internet Security (CIS) [26]. Using the security principles we discuss, these (and other) security approaches are evaluated, determining what user needs are crucial while seeking as secure a configuration as practicable. Areas of conflict (and overlap) are considered and from this iterative changes can be made.

Changes are then evaluated on the privilege, area and criticality of the alteration. Iteration towards an optimal state is a manual process of testing, examination, and consultation with users. For our model, there are three areas of concern: external network access, internal host security, and user level access controls. These areas have differing complementary intersections: the optimum trade-space differs between these areas of concern. Once this optimal state is determined, we then consider the evolution of this security state over time.

4.1. External Network Access

Under the greatest amount of scrutiny, the area of greatest care for the system design is the configuration of external network access. This is an exercise in optimization and extreme minimization. While users will be surveyed for input on desired services, security needs predominate in this area over almost all user concerns. Every access point (open network port or service) is of the highest concern as network access control is the first and best line of defense against intrusion. As we note above this approach can be generalized: this principle applies to both individual computer nodes, as well as networks of computer systems. For an analogy, if the network is compared to a street, an IP address would be the street address of the building and the ports would be the doors of that building. A firewall could be a wall surrounding that building (having its own doors/gates), or around several buildings in a compound. This analogy can be further extended much like that of a design of a military base (or a medieval fortress) with layered defenses. Again, this system design model is a simple one, and other design choices may factor into such an extension. In the area of external network access, the interplay between security controls and user needs will be very heavily tilted toward security because this represents the first, best chance to block an attacker (at the gate).

As mentioned previously this security design could be used in other system approaches. When we examine network access points to the model system, we should also consider other access points, such as physical connections to the system like the ubiquitous USB or Thunderbolt connectors. One can also consider similar methods by extrapolation to evaluate a database system, or a file storage system, by examining the controls on the data, and the access points to that data.

The approach utilized involves cutting down to the absolute minimum the number of open network ports, and then over those ports minimizing the content and amount of data transferred. Among the network testing tools we use, two include the Nmap port scanner (providing a comprehensive scan of all open ports on a machine) [27], and Nessus, a widely available network security scanner that helps probe for security vulnerabilities on the remaining open ports [28]. Indeed, although the user community may with regularity request the installation of numerous software applications that will have their own unique web server on an additional open port, such services will usually be disallowed. This is necessarily an iterative process. Sometimes if taken too far and too many ports are closed off and services shut down, the machine will not be able to communicate over the network. There may be times as well when certain user applications really do need to have a port opened for them. These should be examined most carefully. Those few services that are still available will be expected to be extremely secure against external subversion, such as the SSH remote login/file transfer/port forwarding service. In this approach:

- First test the system absent the host based firewall. This is to ensure the underlying system is secure without it. The firewall should not be the only point where outside connections are minimized and controlled. *Almost all network services can be shut down on a secure system with no real difficulties to the underlying function of the operating system!*
- Next test the system with the host based firewall engaged. In many cases stateful packet filtering can be used to trace (and allow) replies to network calls, which will eliminate the necessity for most holes in the firewall.

Table 1 depicts the results of this process for our example. From more than twenty open network ports on our initial image (which is a great improvement over previous Red Hat Linux versions) we reduced the number of open network services to a bare minimum of essential network services:

Table 1: Number of open network services with and without Firewall – Final Minimum Services

Firewall State	**TCP**	**UDP**	**ICMP**
No Firewall	3	6	2
Firewall	2	3	2

This is an excellent example of the point made earlier that the strength of a system can many times be improved by what is taken away from it. Contrasting these results with the hundreds or thousands of network services on a general home Mac or Windows computer can be enlightening as to the current default state of computer security.

4.2. Internal Host Security

A highly secure system that is impossible to use for its primary duties is not very *useful*. In evaluating our system's internal host security, the interplay between security controls and user requirements must be nearly equal due to the competing concerns to both keep it working as efficiently, and as secure, as possible for the users of the system.

By analogy, inside the building, the use of the building by its inhabitants is an important design consideration. Is it a home? A bank? A police station? Safe construction approaches are dictated for all buildings by fire codes, but how a building is used will determine much of a building's interior design.

For our own computation environment, this process of system hardening considered functions and services the user community required:

- The completion of operational duties such as supplying critical data sets to mission partners and utilization of mission critical flight software.
- The supply of required capability for a task or job role (e.g. visualization and formatting tools, viewers, editors, functional compilers, and third party software).
- The provision of *useful* utility functions and software (e.g. BR/DVD/CD archiving, external export of interesting graphical output, the capacity to safely load removable file systems, and network printing).

From these characterizations, a definition evolved of what behaviors were expected of a user and conduct they should *never* be expected to engage in. With such a set of definitions we removed from the system those privileges and programs a user was never expected to use, and added further constraints so users would have difficulty in performing anomalous activities.

Similar to the approaches used to secure the external network access of the machine, a software tool was used to evaluate the internal security settings of the machine and produce benchmarks for enhancement. Created by the Center for Internet Security (CIS), the CISscan package was used to evaluate the internal host security for the system [26]. It provides a set of metrics for assessing security configurations for a number of package distributions, as well as operating system platforms, allowing comparisons between variant software sets, alternate configurations, and operating systems. Representing an effort to create a "best practices" industry standard guideline to system security, the CISscan package is based around a set of consistent standards, that can be used to evaluate a given system against such metrics. With this tool, it was possible to create a security baseline, make sure that configuration changes and patching continued to meet that baseline, and iterate to a desired system configuration.

Categorical changes to the internal security settings included:

- Removal of extraneous executable permissions. The ability to change users in a program is an especially serious concern. SetGID and SetUID executables have the ability to change the group and user of the process that runs them. With this ability such programs present an extreme security risk. Some programs however, such as the ones that are required for users to log into the system need such a capability. Like our evaluation of open network services, almost all operating systems have a needlessly high amount of these types of programs by default. Even the latest versions of Red Hat Enterprise 7 contain numerous examples of programs with too many privileges. In our system, with effort we reduced the number of these very sensitive programs from 45 SetUID executables in the initial install to a total of 15 in the final secured configuration (see Table 2 for full discussion).
- Limit the installation of new software. Every new software installer has the potential to open up new vulnerabilities. It is far better to avoid such problems by not installing the software at all. As noted, this may minimize many maintenance issues as well, including keeping the software secure and patched. A related question for discussion is, "does your software really *need* its own web server?"
- Limit or otherwise shut down all unnecessary system services. For some minimal operating system configurations, the number of system services running on startup can be counted on one hand. This is a far cry from the average computer, which can have more than a hundred processes running at any given moment before anyone has logged in. Such extraneous processes can represent significant vulnerabilities on a system, especially if they have a network port open.
- Modify key network and kernel data structures. By changing the settings of key network and kernel configuration values, it is possible to significantly increase the protection of the system against denial of service attacks and other security failures with the occasional cost of small increases in memory use. Indeed, lowering user limits (number of users) and per-process (number of processes) on the system, significantly lower the odds an application can harm system function, blocking a denial of service attack by preventing such software from consuming system resources. Fortunately this area is one of the least likely to result in negative feedback and consequent user pushback.
- Restrict file system permissions on untrusted and/or critical file systems, secure application sets, and use caution for removable file systems. Crucial file systems are mounted read-only on most machines in our environment (especially the ones containing critical software). Also, file systems that are removable cannot execute programs with SetGID or SetUID privilege as it is trivial to create such programs that can give the user elevated privileges (thus bypass system security) on a different system and load them from a removable device.

These categorical changes are summarized both from the metrics of the CISscan tool and with direct metrics (where possible). Results of these efforts are shown in Table 2 and Table 3, taken from our model system image at four different revision times.

Table 2 examines direct metrics that can be compared across several differing system images. This includes the total number of SetGID and SetUID programs in the image, the number of "classic" System V and "new" Systemd services, as well as kernel settings (variables) changed for hardening the internal system and network from attack. For the first four rows of metrics, lower is better, for the last two rows, higher is better:

Table 2: Navigation OS Direct Metrics

Metric	Initial Install [6]	June 2016	April 2017	May 2018 [7]
(Lower numbers are better)				
Total SetUIDs	45	15	15	15
Total SetGIDs	23	18	19	19
Systemd Services	136	55	58	58
Sys V Services	5	2	2	2
(Higher numbers are better)				
Kernel variable changes	3	11	14	14
Network changes	2	4	4	4

Table 3: Navigation Red Hat Enterprise 7 Security Metrics (Center for Internet Security Benchmarks)

CIS category	May 2015	June 2016	April 2017	May 2018 [6]
Updates, Add. Security	38%	45%	45%	47%
OS	12%	11%	11%	11% (est.)
Special Services	67%	67%	67%	79%
Net Configuration	91%	92%	92%	91%
Log & Audit	50%	55%	55%	57%
System Access	45%	41%	41%	39%
User Account	14%	14%	29%	43%
Tests passed/failed	71/66	87/79	88/78	110/79 [8]

Table 3 tracks more general changes in *types* of configurations, and in order to track such changes over time, the CISscan tool is used. Breaking the OS into areas of concern, these metrics measure how secure such categories are compared to a reference model. For some cases, with the addition of software sets or new features requested by users, these scores went down. While the results are broad and vary by software sets and installed feature, they give the ability to check the aggregate state of security and compare changes over time. This may be invaluable when a software installer decides to reset a host of security settings. Such changes are very apparent in these results, and this tool also proves itself useful in providing aspirational (and safe) examples for academic papers. Relevant CIS categories along with numeric percentages for several system images are given. For these values a higher percentage is better, but it can be nearly impossible, and (mentioned previously) less than useful to reach a 100% score.

4.3. User Access Controls

The controls directly placed on user action in this security model is an area where considerable compromise is necessary in the interplay between system security and user needs. The domain is one where (for the Navigation systems) the need for usability is a greater priority than security. By analogy, consider the use of badge controlled rooms in a building. While such security may be appropriate for certain high security buildings where this type of security is required, it may not be as suitable for a supermarket or a home. Two such systems are in general available for most OS platforms.

4.3.1. Discretionary Access Control

In Linux this user control is based on standard process and file permissions, allowing for access control of critical system files and granting users the ability to manage access to data sets and user files. This mechanism allows for a separation of users and data, with these management controls based on Discretionary Access Control (DAC) mechanisms, such as traditional Unix process and file management. Such controls have been standard on almost all operating systems for some time. Care should be taken to avoid overly restrictive user settings. Unlike the previous areas, it is hard to come up with a general standard for these controls. There might be significant difference in file system and data permissions even for different teams on the same system. Indeed, all that we can recommend here is due diligence in the application of these protections.

4.3.2. Mandatory Access Control

Mandatory Access Control (MAC) is an alternate security mechanism that is available under Linux, referred to as SELinux [29]. Part of a series of research projects out of the United States' National Security Agency, this series of software modules seeks to bring aspects of highly secure operating system designs to the Open Source community [30]. These features, which allow for a second, very granular control over user actions, previously were available only on highly secure and expensive computer designs. These modules are included with the Red Hat Enterprise Linux distribution and are a part of our model image. User access can be granted or revoked to a specific file or process, categorized by role. For example, on a system without SELinux, the critical system password file (/etc/shadow), cannot normally be read, however the time it was last modified and the size can be read by an unprivileged user of the system. Such information, such as when the file was last changed (i.e. the last time a password was modified) could be useful to an attacker. For a machine running SELinux with restrictive settings however that file cannot even be listed by an unprivileged user of such a system: no information about that file will be returned *at all*.

[6] The initial install was not suitable for general use and underwent significant change. It is close but not equivalent to the May 2015 release of Table 3.

[7] It should be noted in both Table 2 and Table 3 (below) that the system contained almost five times as many software packages in the May 2018 release as the initial install due to new feature requests and open source changes.

[8] CISscan was updated and several tests were added and expanded. This category had changed significantly in the new version and the value was derived from the previous test method.

However, as we note, how well the users can use the system is a crucial concern for this area. Unlike the specialized intelligence and military computer systems that utilized Mandatory Access Control initially, users of general purpose engineering machines, like those of the navigation computing environment, typically do not have the same expectations or same security concerns about how such machines should function. The use of such constraints could potentially make the user experience miserable, or simply untenable. Furthermore, familiar applications could fail in unexpected ways. Such controls are also quite difficult to configure in a correct manner to support the Navigation user community. There are few use cases where this level of security restriction for these *User Access Controls* area, has been found useful, in comparison with the application of greater restrictions on the *External Network Access* and *Internal Host Security* subsystems.

A trial run of the SELinux system was undertaken on a limited set of operations workstations to evaluate the utility of these controls. The proposed idea was to apply an iterative process to determine the areas of the system utilized normally by the user base. SELinux was run in permissive mode over the course of several weeks, and the SELinux system was examined to determine those areas where the users would be blocked by the SELinux policy. From these conflicts a model of user activity was built up over several months, which allowed for changes to be made to the SELinux policy rules, or changes to be made to user software and/or user behavior. Over time the sum of these changes helped define *expected* user actions. Areas of the system that users did not use and did not need, would be blocked off from user activity.

In principle this is a good approach. However a formal effort examining this method noted two major difficulties [31] and these difficulties have not changed significantly:

- System analysis and configuration: Getting the computer system and its processes working, and working with the correct permission sets was significantly more technically challenging than expected. Many errors were only solved by trial and error. A considerable amount of time was spent to get the system to a working and stable state.
- User analysis: This process (even on earlier versions of SELinux) was not unusually difficult. However downtime could be required during the analysis iterations, and these iterations could make it difficult to accommodate system and user activity changes (users do not like having to log out of their systems repeatedly). This may be an area that functions better for environments that are more static.

At this time methods of Mandatory Access Control such as SELinux are still being evaluated for use in Navigation. Although it offers much improved control mechanisms, and even allowing the automatic design of user access from user actions, our current examination of the intersection between system security and user utility in this area does not recommend its use at this time.

4.4. Evolution of system security over time

Unlike the previous parts of this section which deal with the issues of creating an initial, secure system state, here we consider the evolution of that state over time. As we observed in our discussion on *security over time* how this state changes is strongly subject to the environments and networks that connect to our system.

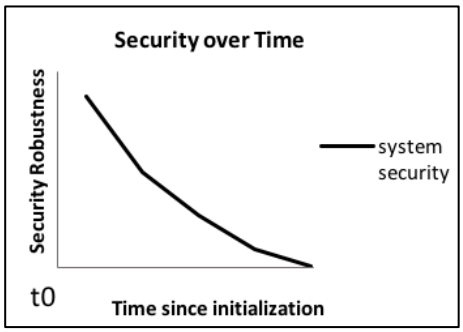


Figure 3: Security over time.

How this confidence in a system changes, and hence the *slope* of the security state in the figure above is very dependent on events in the local environment and the vulnerabilities that emerge over time. It may be nearly linear over time, or an exponential die-off due to newly discovered vulnerabilities. For example, a powered down system on a shelf (or offline storage for a VM) will not change at all. However once connected, many factors will strongly affect the slope of the curve, such as the age of the patching and software set, the security controls, and the environment to which it is connected. This is highly particular to the environmental factors acting on the system: even systems at the same site may have wide variance in this evolution. Understanding this, and remediating such changes is a large part of the maintenance lifecycle of a system. The only clear guidance that can be offered is that, as with a broken password, a system compromise will cause the curve to have a cliff (or stair-step) function in the trust of the system. This will require a system reset or reinstall to resolve.

5. Future Considerations and Conclusions

Considering the evolution of the cyber threat environment and trying to discern what new challenges will arise and how to meet these concerns is a subject of ongoing interest and research for this author. The navigation computational environment conducts critical activity on a 24/7/365 schedule, and cannot afford to be the victim of a security compromise. It is an important system that needs protection. While tips and techniques for security have dramatically changed since this author started as a system administrator, the underlying principles discussed here have changed little in two decades, or not at all. Such archetypes are examined in this paper, and while the particular technical points can change, the foundational ideas will likely remain the same.

One case in point, as noted “de-perimeterisation” approaches like Zero Trust computing may help our approach to computational and network security by encouraging rigorous security at every computational node and network connection. This is a refreshing model which brings to mind some of the most secure intelligence facilities and military fortifications of the last century. It is a vast improvement on the simple, single-defensive layer (sometimes called the hard exterior / soft interior) model which ends up being the most common approach taken in network security. Indeed, we advocate an approach in this paper of being hard on the exterior *and* hard on the interior!

While it is heartening to see such enthusiasm, and much good can come from this, it is important to remember that these efforts (still) need to consider both the complementary intersection of security and user design requirements, as well as the importance of the fault tolerant protection of *all* of the critical computational elements. As we mention above, solving the totality of this defense in breadth problem is a *hard* one. Indeed, not considering these two avenues of concern can lead to situations where the computational environment is perceived by the users as as more akin to something like a prison, or worse still, end up promoting a failure of protection akin to that seen in the previously mentioned Maginot line. Especially in the second case, aggressive efforts as Zero Trust are expensive and challenging to implement and can (from exhaustion of resources in time and effort) lead to the promotion of the failure they sought to prevent! If even a simple approach cannot be executed correctly, attempting to implement a more complex model (like Zero Trust) can be a dubious undertaking.

Another example showcasing this fundamental difference is the spate of attacks based in speculative execution on modern x86 processors (as well as methods to remediate them). These have been of tremendous concern and rapid development for the entire computer industry over the past 12 months [32]. However, the architecture problems at the source of these issues are fundamental design concerns that were identified in their implementation nearly twenty years ago [33].

In addition, a few observations may be derived from the principles espoused above. To the author's mind these should be regarded as obvious and self-evident. Unfortunately, that is often not be the case.

Of these, the most significant is *don't engineer in single points of failure*. For example, after considerable effort to construct a secure, fault tolerant network design, with a hardened firewall and a focused IDS, it is a contradictory effort to then layer a poorly secured Active Directory server for single sign-on for the network (or a poorly configured VPN to access such systems). Such a setup is self-defeating – at best. Fault tolerant and secure design should not be ignored for "just one service". In a similar manner the implementation of a poorly designed security tool designed to examine (or worse, modify) the entire enterprise environment can serve to help implement a large security hole for the entire enterprise. Indeed, as Sami Saydjari relates, there is particular concern with such enterprise security tools as such systems are the highest priority targets of attackers [19, p. 346]. Ease of management of an enterprise network is a nice goal. However, one of the questions that should come up is, "by *whom*?" Single points of failure must be closely, closely watched, as they can be doors that potentially open up the whole network to attack.

This goes with the concept that uniformity of design is not necessarily good. "Balkanization" or the promotion of multiple differing configurations (as long as each of those configurations are secure) promotes design diversity, as discussed in Section 2. While this may not be desirable from a management perspective, such diversity can lead to independent modes of failure, and hence increased security fault tolerance.

Similar to this is the desire to patch often, in the effort to be secure. These two concepts are not, as popularly believed, the same. While this may go against the grain of the currently popular DevOps and Agile software engineering paradigms, it is clear that, absent other testing schemes, it is wise to take the approach of "wait and see" when patching systems.[9] Such updates may introduce bugs, new vulnerabilities, or other unforeseen issues. This is especially true in times of stress after the announcement of major computer security bugs. Large companies can and do make mistakes, and a patch failure that in individual systems may cause pain and irritation could lead to a catastrophic failure or outage if on a critical system.[10]

Finally, one of the bigger trends in current computation lies in the increase in the use of virtual and cloud computing systems. Distributed systems such as these can offer significantly lower costs if (and this is still a big "if") their security and reliability metrics can meet the requirements of their customers. Architecture principles discussed here are (as much as they can be made to be) platform agnostic. One computer security researcher relates the major difference in cloud computing as, "Fundamentally, cloud security is a primary concern due to loss of control ... We've seen this before – [with] outsourcing..." [34, p. 24]. Does the cloud service provider follow fault tolerant and security fault tolerant design principles? Can such questions even be answered by such services?

This loss of control over the configuration for the security environment and for critical events, continues to be problematic for the Navigation use case for cloud computing, especially with concerns over Availability [34, p. 96]. With cloud computing, ownership of the systems and operational processes is by an outside organization. As a part of the operations of our computational environment, it has sometimes taken extraordinary effort to ensure the continuing functioning of our systems in times of stress. This has sometimes meant the difference between mission success and failure. In an emergency, wide latitude is given to operational staff to keep systems running and restore failed systems as fast as possible to meet the needs of flight operations. With an emergency in a remote cloud computing environment, can one even expect to get a responsible engineer on the phone? Without control over that environment, it is obvious we cannot, as Flight Director for NASA Chris Kraft said, ''...take any action necessary for mission success'' [35, p. 392].

It is hoped that the ideas in this paper will provide assistance to administrators and system engineers, and especially the astronautical community. Aspirational goals are presented here, in the hope of providing a guideline to follow for your own design

[9] A major point of defense in depth is that a single failure should not provide the opportunity for compromise on a wide scale. This should be one of the benefits of such an architecture focused approach. Many times a significant vulnerability in our systems crossed our desks with an organizational rallying cry of "patch now!", "patch now!" and we would discover that we had limited to no exposure from the vulnerability because of other defense in depth controls that were in place. Our goal is to be proactive rather than reactive. While we do have emergency patching mechanisms in place, we prefer not to use them unless truly necessary.

[10] It may be instructive to examine the release schedule (and following systems failure reports) for the firmware and software updates for the MELTDOWN/SPECTER vulnerability of 2018. Having a *running* server is in most cases preferable to one that is "currently patched", but "awaiting motherboard replacement."

efforts. These principles can be applied in Ground Data Systems design problems, as well as in other areas of systems engineering. This is particularly valuable in a community of austere budgetary realities where our bespoke systems engineering is based around missions with only a few, or (more often) only a single deliverable.

In an ideal world we would be able to trust our computer systems in the same way that when we drive a car, we trust its brakes. I believe that this optimistic idea is a badly needed one in the design of the increasingly complex and intertwined computer systems that comprise our world today. With the current state of computer security this can appear hopeless, however it is clear, as with the encouraging Boeing flight control system [10], there *are* computer systems that have been designed to be truly fault tolerant. A hard problem is not necessarily an impossible problem. There is no reason we can not also do this in computer security.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgements

To the Systems Engineers and System Administrators who have helped in support of the security architecture discussed here I would like to give my sincere thanks. Additionally, of the editors who assisted with this paper, Zachary Porcu did tremendous service and to him a considerable debt of gratitude is owed. In addition it is important to note that a great deal of appreciation is also due Hal Pomeranz, contributing author of the Center for Internet Security, and senior Fellow of the SANS Institute, who spent valuable time reviewing the ideas in this paper, and raised some insightful questions that helped refine this effort.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference to any specific commercial product, process, or service by trade name, trademark, manufacturer or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology. © 2019 California Institute of Technology. Government sponsorship acknowledged.

References

[1] This paper is an extension of work originally presented in, R. M. Beswick, "Computer Security as an Engineering Practice: A System Engineering Discussion," IEEE: 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT), 27-29 September, 2017. doi: 10.1109/SMC-IT.2017.18

[2] R. Beswick, P. Antreasian, S. Gillam, Y. H. Hahn, D. Roth and J. and Jones, "Navigation Ground Data System Engineering for the Cassini/Huygens Mission," AIAA 2008-3247, SpaceOps 2008 Conference, Heidelberg, Germany, May 12-16, 2008. doi:10.2514/6.2008-3247

[3] The Netflix Tech Blog, "The Netflix Simian Army," Medium, 19 July 2011. [Online]. URL: https://medium.com/netflix-techblog/the-netflix-simian-army-16e57fbab116. [Accessed 3 September 2018].

[4] The Netflix Tech Blog, "Chaos Engineering Upgraded," Medium, 25 September 2015. [Online]. URL: https://medium.com/netflix-techblog/chaos-engineering-upgraded-878d341f15fa. [Accessed 3 September 2018].

[5] C. Carson, Effective FMEA's – Achieving safe, reliable and economical products and processes using Failure Mode and Effects Analysis, Hoboken, NJ: Wiley & Sons, 2012.

[6] B. Johnson, "Fault-Tolerant Microprocessor-Based Systems," *IEEE Micro,* vol. 4, no. 6, IEEE Computer Society Press, Los Alamitos, CA, pp. 6-21, 1984.

[7] J. Gray and D. P. and Siewiorek, "High-Availability Computer Systems," IEEE Computer Society, Los Alamitos, CA, p. 39-48, September 1991. doi: 10.1109/2.84898

[8] D. McCullough, The Great Bridge: The Epic Story of the Building of the Brooklyn Bridge, New York: Simon & Schuster, 2012.

[9] U.S. - Canada Power System Outage Task Force, "Final Report on the August 14th, 2003 Blackout in the United States and Canada – Causes and Recommendations," April 2004. [Online]. URL: https://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/Blackout Final-Web.pdf. [Accessed 25 May 2018].

[10] Y. C. Yeh, "Safety critical avionics for the 777 primary flight controls system," IEEE - Digital Avionics Systems, Daytona Beach, FL, DASC. 20th Conference, October 14-18, 2001. doi: 10.1109/DASC.2001.963311

[11] R. Anderson and R. Needham, "Programming Satan's Computer," *Computer Science Today, Berlin, Springer, Lecture Notes in Computer Science,* vol. 1000, pp. 426-441, 1995.

[12] B. R. Rich, "Clarence Leonard (Kelly) Johnson, 1910-1990, A Biographical Memoir," in *National Academy of Sciences*, Washington, D.C., National Academies Press, p. 231, 1995.

[13] M. Bishop, Computer Security, Art and Science, 2nd. Ed., New York: Addison-Wesley, 2019.

[14] D. Libes, "RFC 1178 – Choosing a name for your computer," August 1990. [Online]. URL: http://www.faqs.org/rfcs/rfc1178.html. [Accessed 25 May 2017].

[15] I. Roundhill, "699-101: Cassini Navigation Plan, JPL D-11621," Jet Propulsion Laboratory, Pasadena, CA, 1 August 2003.

[16] S. Garfinkel, G. Spafford and A. Schwartz, Practical UNIX and Internet Security, 3rd Ed., Sebastopol, CA: O'Reilly and Associates, 2003.

[17] W. R. Cheswick, S. M. Bellovin and A. D. Rubin, Firewalls and Internet Security, Repelling the Wily Hacker, 2nd Ed., New York: Addison-Wesley, 2003.

[18] W. Schwartau, Time Based Security, Seminole, FL: Interpact Press, 1999.

[19] O. S. Saydjari, Engineering Trustworthy Systems, New York: McGraw-Hill Education, 2018.

[20] E. Gilman and D. Barth, Zero Trust Networks, Sebastopol, CA: O'Reilly Media Inc., 2017.

[21] B. Youn and P. Wang, "Complementary Intersection Method for System Reliability Analysis," *ASME Journal Mechanical Design,* vol. 134, no. 4, April 2009. doi: 10.1115/1.3086794

[22] R. Cross, "Center of percussion of hand-held implements," *American Journal of Physics,* vol. 72, no. 5, pp. 622-630, May 2004. doi:10.1119/1.1634965

[23] S. Chiasson and P. C. v. Oorschot, "Quantifying the security advantage of password expiration policies," *Designs, Codes and Cryptography,* vol. 77, Issue 2-3, pp. 401-408, December 2015.doi: 10.1007/s10623-015-0071-9

[24] National Vulnerability Database, "National Checklist Program Repository," National Instituite of Standards and Technology, [Online]. URL: https://nvd.nist.gov/ncp/repository. [Accessed 30 March 2018].

[25] Information Assurance Directorate, " Operating Systems guidance," National Security Agency, [Online]. URL: https://www.iad.gov/iad/library/ia-guidance/security-configuration/operating-systems/index.cfm. [Accessed 20 April 2017].

[26] Center for Internet Security, "CIS - Center for Internet Security," CIS, [Online]. URL: http://www.cisecurity.org. [Accessed 30 March 2018].

[27] NMAP, "Nmap," [Online]. URL: http://www.nmap.org. [Accessed 30 March 2018].

[28] Nessus, "Tenable Security," Tenable Inc, [Online]. URL: http://www.tenable.com/products. [Accessed 30 March 2018].

[29] SELinux, [Online]. URL: http://selinuxproject.org/page/Main_Page. [Accessed 25 May 2018].

[30] NSA, "SELinux," National Security Agency, [Online]. URL: https://www.nsa.gov/what-we-do/research/selinux/.[Accessed 25 May 2018].

[31] R. M. Beswick and D. C. Roth, "A Gilded Cage: Cassini/Huygens Navigation Ground Data System Engineering for Security," AIAA 2012-1267202, SpaceOps 2012 Conference, Stockholm, Sweden, June 11-15, 2012. doi:10.2514/6.2012-1267202

[32] P. Kocher, D. Genkin, D. Gruss, W. Haas, M. Hamburg, M. Lipp, S. Mangard, T. Prescher, M. Schwarz and Y. Yarom, "Spectre Attacks: Exploting Speculative Execution," *arXiv.org:1801.01203,*3 January 2018.

[33] Z. Wang and R. B. Lee, "Covert and Side Channels due to Processor Architecture," *Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06),* 11-15 Dec. 2006. doi:10.1109/ACSAC.2006.20

[34] D. Shackleford, "Security 524: Cloud Security Fundamentals - Day 1," in *[seminar], SANS National Conference*, Orlando, FL, April 7th, 2017.

[35] G. Kranz, in *Failure Is Not an Option: Mission Control From Mercury to Apollo 13 and Beyond*, New York, Simon & Schuster, 2009.

[36] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal,* vol. 28, no. 4, pp. 656-715, 1949. doi:10.1002/j.1538-7305.1949.tb00928.x

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 359-365 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Hidden Surface Removal for Interaction between Hand and Virtual Objects in Augmented Reality

Takahiro Ishizu[1], Makoto Sakamoto[*,1], Masamichi Hori[1], Takahiro Shinoda[1], Takaaki Toyota[1], Amane Takei[1], Takao Ito[2]

[1]*Department of Computer Science and System Engineering, University of Miyazaki, 889-2192, Japan*

[2]*Institute of Engineering, Hiroshima University, 739-8511, Japan*

ARTICLEINFO

Article history:
Received: 31 May, 2019
Accepted: 27 July, 2019
Online: 19 August, 2019

Keywords:
Augmented Reality (AR)
Virtual object
Occlusion problem
Canny edge detection
Hidden surface removal
Hand area

ABSTRACT

Recently, augmented reality (AR) technology has been applied to the interaction technology between human and virtual objects. In AR interaction technology, it is necessary for users to be able to manipulate virtual objects intuitively. Thus, we focus on manipulating virtual objects directly with the user's bare hands in this study. On the other hand, in AR technology, since the 3-dimensional (3D) model is superimposed on the image of the real space afterwards, it is always displayed on the front side than the user's hand (occlusion problem). Thus, it becomes an unnatural scene in some cases. In this study, this system detects hand area of the user by extracting depth information, color information and using canny edge detection in the user's hand. Thus, this system performs hidden surface removal along the area of the user's hand by considering the object-context relations between the user's hand and the virtual object. In the evaluation experiment, it is confirmed that the hidden surface removal in this study make it possible to distinguish between finger boundaries and to clarify and process finger contours. This work is an extension of the paper entitled "Hidden Surface Processing for Interaction of Hand and Virtual Objects Using Leap Motion Controller" published in 2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT).

1 Introduction

Augmented reality (AR) is used in many fields including education, medicine and entertainment, and its applications became widespread [1, 2, 3]. Also, the interaction with the virtual objects is required in those papers. Therefore, we think that more precise virtual object manipulations in AR will be performed in the future.

On the other hand, since hands are our main means of interaction with objects in real life, AR interfaces should also be able to be for free hand interaction with virtual objects.

Furthermore, it is important to be able to adapt to more detailed tasks by hand in order to handle a wide range of fields such as medical care and education.

In the conventional study [4], by putting a marker, a special glove, etc. on a user, the user's position information was acquired and interaction with a virtual object was realized. However, these methods may give the user a sense of discomfort such as weight due to wearing. Therefore, bare hand interaction is required.

However, since the 3-dimensional (3D) model displayed by the AR is superimposed on the image of the real space afterwards, the 3D model is always displayed on the front side and user's hand is hidden by virtual objects. Thus, the scene may become an unnatural scene, and the user cannot see the object-context relations of the virtual object and his/her hand, and feels that it is difficult to manipulate the virtual objects.

In the existing study [5], the system used transparent 3D models and the 3D models followed each fingers of the user based on the depth information of the user's hand. In this way, they performed hidden surface removal and tackled the occlusion problem. However, since the 3D models which follow the user's fingers are larger than the finger, a wider range than the actual finger was displayed on the front (see Figure 1). In addition, no description has been given of a case where there is a multiple virtual objects in the existing study.

In other related study [6], using the Kinect sensor in order to obtain depth information of the user, hidden surface removal was realized when an arm was inserted between multiple virtual objects. Also, it was the same even when

*Corresponding Author: Makoto Sakamoto, 1-1, Gakuen-Kibanadai-Nishi, Miyazaki City, Miyazaki, 889-2192, Japan, sakamoto@cs.miyazaki-u.ac.jp

it came to deformable virtual object. However, the Kinect sensor is a sensor that roughly recognizes human motion and depth information, and the study didn't focus on the interaction between hands and virtual objects.

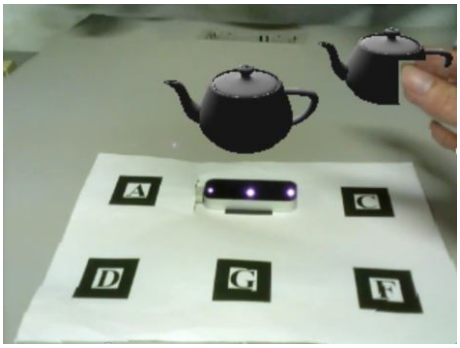


Figure 1: Hidden surface removal in existing study

In this study, we use Leap Motion Controller to acquire depth information of the user's hand. Additionally, this system acquires RGB (Red, Green and Blue) information from the web camera. Based on these pieces of information, the contour of the hand is extracted to realize hidden surface removal along the contour of the hand.

In this study, we aim to cope with more detailed virtual object manipulations by conducting hidden surface removal along the outline of the user's hand. Moreover, we conduct a manipulation of grasping a virtual object of primitive model by hand and confirm whether it works properly.

2 Proposal

In order to conduct hidden surface removal along the contour of the user's hand, this system detects hand area of the user by extracting depth information, color information and edge detection in the user's hand.

First, in order to extract hand area, we use the RGB image based on color information. Next, based on the depth information of the hand, a blue point group is created on the thumb, and a green point group is created on the other fingers. These point groups are displayed at their positions only when the hand is in front of the virtual object. In addition, the hidden surface removal along the fingers is realized by using the hand area extracted first and the result of edge detection. Also, in order to cope with multiple virtual objects and distinguish the object-context relations of the blue point group and the green point group, we use the Z-buffer method.

In this way, when the user's hand is near from camera than the virtual object, the user's hand is correctly displayed on the front side of the virtual object, and the user can grasp the anteroposterior relationship between the user's hand and the virtual object.

2.1 System Component

This system consists of a Web camera, the Leap Motion Controller, PC and AR marker. In this study, since this system needs accurate position information of user's fingers, we use the Leap Motion Controller that can acquire various data relating to user's hand with high accuracy.

The purpose of use of each of the devices shown above is as follows:

- Web camera: Acquisition of real image and recognition of marker.
- The Leap Motion Controller: Obtaining 3D coordinates of the user's fingers.
- Monitor: Video output.
- AR marker: Acquisition of position and tracking of virtual objects.

The Leap Motion Controller and the webcam both use camera features, which causes the camera features to interfere with each other. Therefore, this system addresses this problem by using network programming. This system acquires data of 3D coordinates of the user's fingers with the Leap Motion Controller on the server side. In addition, this system sends the data to the client side. On the other hand, this system acquires images with the Web camera, and conducts image processing and video output using the data which received from the server side on the client side.

This system controls the position and orientation of the virtual objects is by recognizing the AR marker. Therefore, the AR marker needs to be recognized by the web camera. However, when the user manipulates virtual objects, the AR marker may be covered by his/her own hand. Therefore, this system adopts a method of treating Marker A to Marker F as one marker (see Figure 2). Thus, even when a part of the marker is covered, the virtual objects can be displayed properly.

We place the Leap Motion Controller on the position of Marker B. This system displays the virtual objects on the position of Marker C and Marker G (see Figure 2).

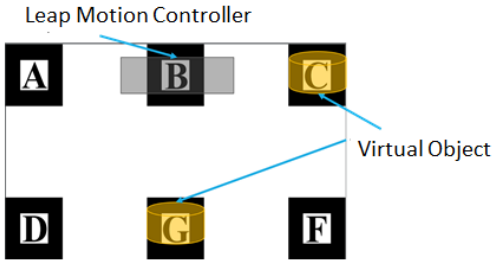


Figure 2: AR marker

2.2 Development Environment

We use ARToolKit for AR marker recognition and camera control. And, we develop this system using OpenGL for displaying virtual objects and OpenCV for various image processing. Here, the resolution of the camera is 320×240 pixels.

2.3 Flow of Image Processing

The processing flow comprises of several steps as shown in Figure 3. In Figure 3, the block of light red background indicates that it is generated image.

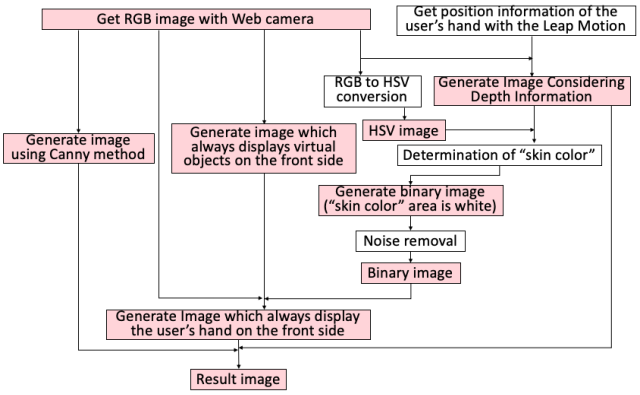


Figure 3: Flow of image processing in this system

2.4 Processing of Camera Image

This system gets RGB image with the web camera (see Figure 4). In this study, in order to use the HSV (Hue, Saturation and Value) color space often used in the processing of color images, this system converts RGB images to the HSV color space.

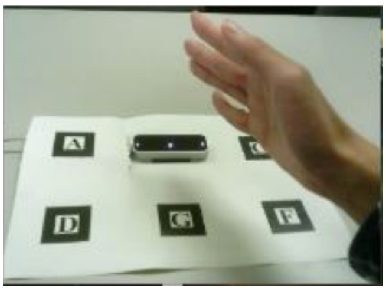


Figure 4: RGB Image

2.5 Generation of Binary Image

2.5.1 Determination of "skin color"

This system needs to extract the area of skin color for detection of the user's hand area. At first, this system puts red point on the finger which are the closest to the camera (see Figure 5). A certain range is given to the HSV value obtained from the pixel of that point, and it is defined as "skin color".

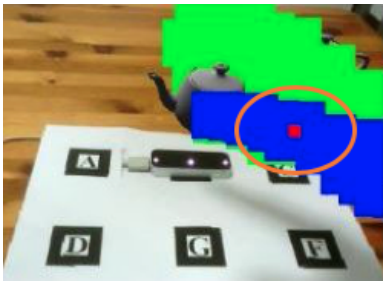


Figure 5: Red point

Next, this system detects the user's hand area as binary images.

The procedure is shown below.

1. Get the 3D coordinates of the proximal phalanx of the finger which are the closest to the camera with the Leap Motion Controller.
2. Generate images that red point is plotted around the acquired position.
3. Obtain the 2D coordinate value of pixel units in the red point.
4. In Figure 4, this system gets average values of HSV in the pixel values corresponding to the coordinates acquired in step 3.
5. Apply an appropriate range to each value of Hue, Saturation, Value and decide it as "skin color".

2.5.2 Noise removal

The binary image generated by this system contains salt and pepper noise at first. Therefore, this system conducts erosion and dilation twice in 8-neighbor in order to remove noise.

However, that alone cannot remove the noise sufficiently (see Figure 6). This is because, in this method, this system has to conduct erosion and dilation multiple times in order to remove loud noise. However, it needs complicated calculation.

Therefore, regarding to the loud noise, this system removes the noise in the area of the hand by conducting the following processing (see Figure 7).

1. Determine white area including red point (which is put on the finger which is closest to the camera) as hand area.
2. Fill the inside of the hand area with black and the outside of the hand area with white.
3. Invert white and black.

In this way, all noise in the hand area is removed (see Figure 8).

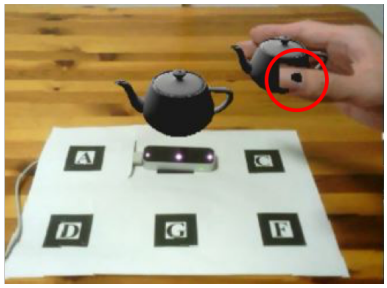


Figure 6: Result image including noise

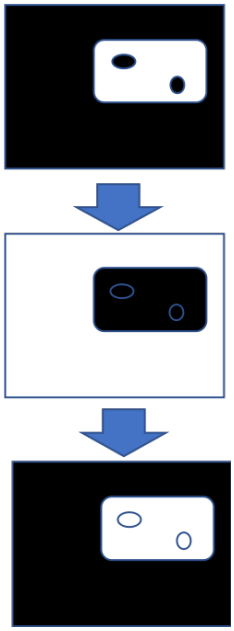


Figure 7: Noise removing process

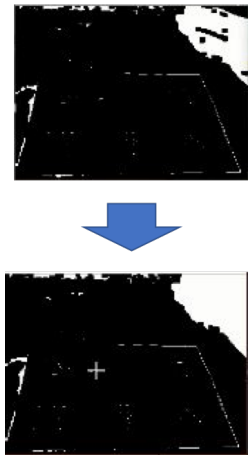


Figure 8: Noise removing

2.6 *Generate Image Considering Depth Information*

This system acquires 3D coordinates of the distal bone, middle phalanx, basal bone and metacarpal bone of the user's hand (see Figure 9) with Leap Motion Controller.

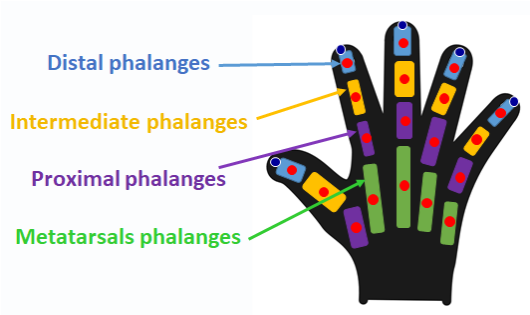


Figure 9: Hand joint

Based on the acquired position information of the fingers, this system generates images plotting blue point group (following the position of the thumb) and green point group (following the position of a finger other than the thumb).

Blue point group displays at the positions of each joint of the thumb when the user's thumb is front of the back object (see Figure 10).

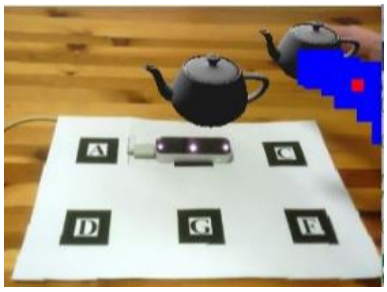


Figure 10: Blue point group following the thumb

In addition, green point group displays at the positions of each joint of the fingers except thumb when the user's fingers except thumb are front of the back object (see Figure 11).

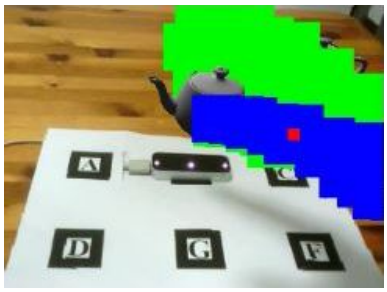


Figure 11: Blue point group and green point group following the whole hand

Furthermore, this system makes it possible to handle multiple virtual objects of different depths. The object-context relations between the back virtual object and the hand of the user is determined depending on whether or not blue point group and green point group are displayed. On the other hand, the object-context relations between the virtual object on the near side and the user's hand is determined by using the Z-buffer method. Using the Z-buffer method, when a finger is positioned behind the virtual object, blue point group and green point group following the position of the finger are hidden by the virtual object. In this way, this system can determine the object-context relations between the user's hand and the virtual objects even if there are multiple virtual objects.

2.7 *Generate Images which Always Display The Hand in Front of Virtual Objects*

This system synthesizes the original RGB image (see Figure 12-a) only in the white area in Figure 13 to the image in Figure 12-b. Thus, this system generates images that the user's hand is always displayed on the front of the virtual object (see Figure 12-c).

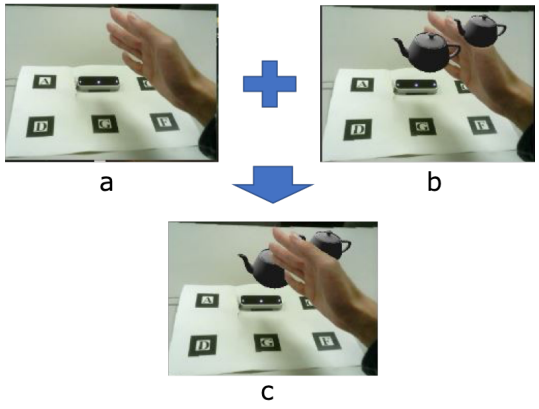


Figure 12: Generation of images which hand is displayed front

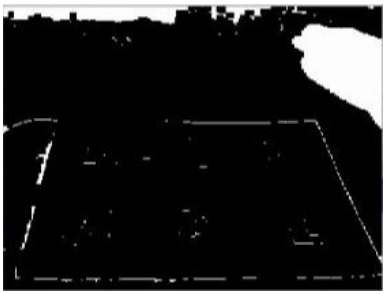


Figure 13: Binary image

2.8 Generate Result Images

The area in Figure 12-b corresponding the area of the blue point group, the green point group and the red point group (in the case where the position of the proximal phalanges of thumb is in front of the virtual object) of Figure 14-a are replaced with the image shown in Figure 14-b (Figure 12-c). Thus, based on the depth information, this system generates images that the hand is displayed in front of the virtual object (see Figure 14-c).

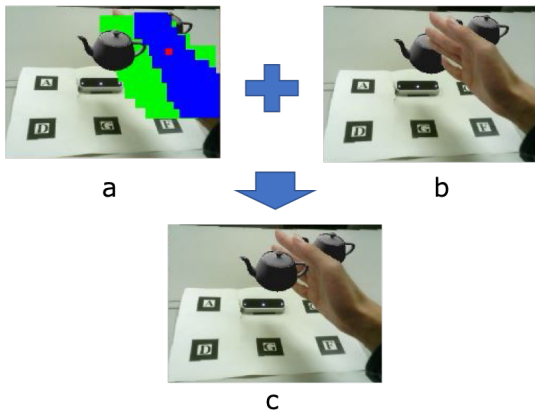


Figure 14: Generation of result images

2.9 Distinction between Fingers Using Canny Edge Detection

The white area in Figure 13 indicates the hand area. However, no distinction between fingers is made from this image. Therefore, if hidden surface removal is conducted based on this image, even if only the thumb should be subjected to hidden surface removal like Figure 15, other fingers are involved and processing is conducted.

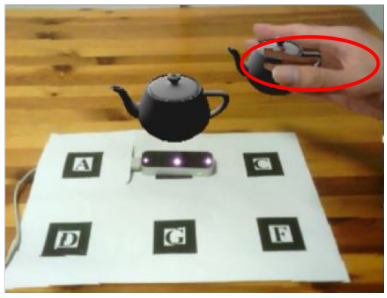


Figure 15: Involved other fingers

Therefore, we need edge detection of the user's thumb in order to solve this problem.

Edge detection is the name for a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or more formally, has discontinuities. The point at which image brightness changes sharply are typically organized into a set of curved line segments termed as edges.

We need to detect finger edge in order to distinguish between thumb and other fingers. Furthermore, at the time of the edge detection, the edge must be one clear line, since the area of the thumb extracted by edge detection is filled in the area surrounded by the edges in the later processing.

In this study, we use Canny edge detection algorithm [7] to find an edge of the user's thumb. Because it provides better results and efficiency than other available algorithm when we want to detect clear edges [8].

Canny edge detector uses two thresholds and if the intensity of that pixel is below that lower threshold than it is assigned "0" value that can be none edge and if the intensity of any pixel is above that the higher threshold than it is assigned "1" value and can be edge.

In this study, we distinguish the thumb from the other fingers. In order to detect the thumb, the thumb is surrounded by an ellipse, and the inner area surrounded by it and the boundary displayed by edge detection is extracted. The area is the thumb area (see Figure 16).

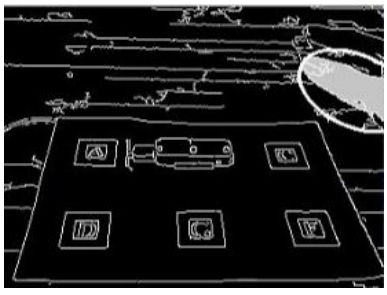


Figure 16: Extraction of the thumb area

The following process is conducted to detect the thumb.

1. Get two coordinates of upper left end and lower right end of blue point group (see Figure 17-a).

2. Make an ellipse based on a line connecting the two points (which are the coordinates acquired in step 1).

3. Overlap the ellipse surrounding the thumb on the image generated based on the Canny edge detection (see Figure 17-b).

4. Fill the area surrounded by the ellipse and the edge in Figure 17-b which corresponding to the position including the red point group in Figure 17-a with gray.

5. Regards this gray area as the thumb area.

In this way, this system generates result image (see Figure 17-c).

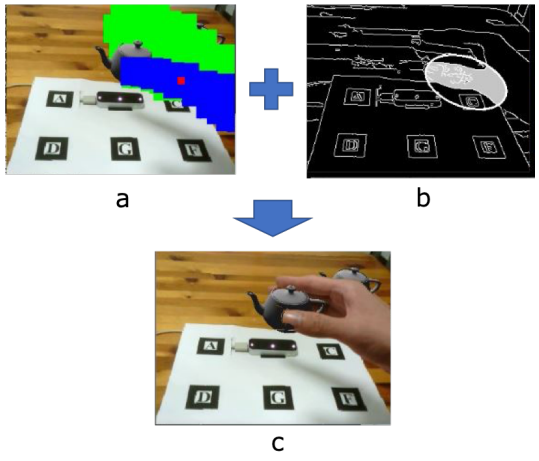


Figure 17: Distinction between the thumb and the other fingers

3 Experimental Results

We executed this system to confirmed whether hidden surface removal along the fingers is possible or not.

First, we checked whether this system can cope with multiple virtual objects. In this system, the object-context relations between the back virtual object and the hand of the user is determined depending on whether or not blue point group and green point group are displayed. In order to confirm whether this is properly processed, we placed the virtual object at position C in Figure 2 and performed manipulation that a user grabs the virtual object. On the other hand, the object-context relations between the virtual object on the near side and the user's hand is determined by using the Z-buffer method. In order to confirm whether this is properly processed, we placed the virtual object at position G in Figure 2 and performed manipulation that a user grabs the virtual object.

Furthermore, we performed grasping manipulations on virtual objects of 3D primitive figures from various directions in order to confirm whether this system is applicable to virtual objects of various shapes.

3.1 In the case of an object behind

Since only the thumb comes to front side than Marker C, the blue point group follows only the thumb (see Figure 10) and draw the thumb properly on the front of the virtual object (see Figure 18).



Figure 18: Result at the position of C in the AR marker

3.2 In the case of an object in front

This system conducted the hidden surface removal considering multiple virtual objects by the Z-buffer method. The blue point group follows only the thumb and the green point group follows the other fingers (see Figure 11). Besides, only the point group of the thumb was drawn on the front of the virtual object by the Z-buffer method. As a result, it was possible to draw only the thumb on the front (see Figure 19).

The blue point group and green point group in Figure 11 are based on the depth information of the user's fingers. Also, the anteroposterior relationship between the blue point group and the green point group is correctly displayed by the Z-buffer method. In this way, when the thumb is positioned behind the other fingers, the green point group is displayed on the front side of the blue point group.



Figure 19: Result at the position of G in the AR marker

3.3 Various trials

We tested three movement patterns for each 3D objects. Here, we use cube (see Figure 20), cone (see Figure 21) and torus (see Figure 22) as 3D primitive objects.

In this section, we display the virtual object at position B.

Each movement is as follows.

1. Grip the virtual object from the side.

2. Grip the virtual object from the top.

3. Grip the virtual object while the user's thumb is behind the other fingers.

The results are shown below. Here, "Grip the virtual object from the side", "Grip the virtual object from the side", and "Grip the virtual object from the side" in order from the left in Figure 20 - Figure 22.



Figure 20: Cube



Figure 21: Cone



Figure 22: Torus

4 Consideration

We confirmed that this system can perform correctly hidden surface removal even when a user manipulates multiple virtual objects by Z-buffer method from the experiment result. In addition, we also confirmed that this system is effective for virtual objects of various shapes.

The distinction between the virtual object and the user's hand was not clear when using the Z-buffer method when considering the depth between the virtual object and the user's hand. Therefore, if there is only one virtual object or if the depth coordinates of the virtual objects are all at the same position, it is better to have blue point group follow the thumb of the user only than to have blue point group and green point group follow the palm of the user's hand.

Regarding gripping the virtual object from the side (see Figure 20), it was confirmed that the hidden surface removal was accurately distinguished without involving a nearby finger.

Regarding gripping the virtual object from the top (see Figure 21), sometimes the hidden surface removal of the base of the thumb did not work well, and a virtual object was displayed superimposed on the thumb. This is because the area of the thumb extracted by the Canny method is smaller than the area of the real thumb.

Regarding gripping the virtual object while the user's thumb is behind the other fingers (see Figure 22), it was confirmed that the hidden surface removal was properly treated even when the fingers other than the thumb were on the near side. This is an effect of the present system because the determination of the skin color is conducted on the foremost finger. Also, at this time, since the blue point group is moved to the back from the green point group, processing such as the Canny method for the thumb is ignored.

5 Conclusion

The future of AR as interaction technology in users and virtual objects looks bright. In recent years, AR has been applied to a wide range of fields such as education, medicine, entertainment and a guide for various tasks. Among them, the user directly manipulates virtual objects with their bare hands. In the future, we expect further development of interaction manipulation, and more detailed virtual object manipulation will be required.

Thus, we proposed and implemented hidden surface removal along the fingers, aiming at being able to apply to more detailed works about interaction manipulation in AR. In this study, we paid attention to detection the contour of the hand to realize hidden surface removal along the user's fingers. In this system, processing based on hand depth information, color information and edge detection is conducted to detect hand contours.

As for edge detection, the Canny method was used to reduce noise and obtain a clear contour. By these processes, appropriate hidden surface removal was realized without involving other fingers.

In the evaluation experiment, it was shown that it can cope with multiple virtual objects. In addition, we tested various grip methods for virtual objects of various shapes, and showed that they manipulate normally.

As future works, since the generation of binary images is not stable at the time of skin color detection, it needs the solution. Moreover, since the contour of the thumb is interrupted at the time of the Canny edge detection, it needs to set more appropriate threshold in Canny edge detection.

References

[1] Kangdon Lee, "Augmented Reality in Education and Training", TechTrends, vol.56, no.2, pp.13-21, 2012.

[2] Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M, "Augmented reality technologies, systems and applications", Multimedia Tools and Applications 51:341-77, 2011.

[3] Abdelkader Bellarbi, Christophe Domingues, Samir Otmane, Samir Benbelkacem and Alain Dinis, "Underwater augmented reality game using the DOLPHYN", Proceedings of the 18th ACM symposium on Virtual reality software and technology, ACM, 2012.

[4] Buchmann V, Violich S, Billinghurst M, Cockburn A. "FingARtips – Gesture Based Direct Manipulation in Augmented Reality", 2nd international conference on computer graphics and interactive techniques, ACM Press, pp.212–221

[5] R. Katahira, M. Soga. "Development and Evaluation of a System for AR enabling Realistic Display of Gripping Motions using Leap Motion Controller", Procedia Computer Science 60:1595–1603, 2015.

[6] J. Adrián Leal-Meléndrez, Leopoldo Altamirano-Robles, and Jesus A. Gonzalez. "Occlusion Handling in Video-based Augmented Reality Using the Kinect Sensor for Indoor Registration", Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 18th Iberoamerican Congress, CIARP 2013, Proceedings, Part II, 2013.

[7] J. Canny. "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.8, no 6, pp.679-698, 1986.

[8] Abhinandan Julka and Sandeep Bhargava. "A Static Hand Gesture Recognition Based on Local Contour Sequence", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 2, 370-379 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Distribution of Bit Patterns in Binary Sequence Generated Over Sub Extension Field

Md. Arshad Ali[*,1], Yuta Kodera[1], Takuya Kusaka[1], Yasuyuki Nogami[1], Satoshi Uehara[2], Robert H. Morelos-Zaragoza[3]

[1]*Graduate School of Natural Science and Technology, Okayama University, 7008530, Japan*

[2]*Faculty of Environmental Engineering, The University of Kitakyushu, 8028577, Japan*

[3]*Department of Electrical Engineering, San Jose State University, CA 95192, United States*

ARTICLEINFO

Article history:
Received: 05 March, 2019
Accepted: 16 April, 2019
Online: 28 April, 2019

Keywords:
Pseudo-random sequence
Distribution of bit patterns
Primitive polynomial
Trace function
Legendre symbol

ABSTRACT

The distribution of bit patterns is an important measure to check the randomness of a sequence. The authors of this paper observed this crucial property in a binary sequence which generated by using a primitive polynomial, trace function, and Legendre symbol defined over the sub extension field. The authors create a new dimension in the sequence generation research area by considering the sub extension field, whereas all our previous works are focused in the prime field. In terms of distribution of bit patterns property, this research work has notable outcomes more specifically the binary sequence (defined over the sub extension field) holds much better (close to uniform) bit distribution than the previous binary sequence (defined over the prime field). Furthermore, the authors theoretically proved the distribution of bit property in this paper.

1 Introduction

In this IoT era, we communicate with each other through the internet. Therefore, secure communication is the major matter of concern. We use symmetric cryptosystems (Advanced Encryption Standard (AES) [1]) and asymmetric cryptosystems (Rivest Shamir Adleman (RSA) [2], and Elliptic Curve Cryptography (ECC) [3]) to establish a secure communication. A pseudo-random number is one of the crucial parts of these cryptosystems. More specifically, in case of cryptography, to generate the keys (public key, private, session key, and so on) a pseudo-random number generator is used. A prominent pseudo-random number generator is essential to generate pseudo-random number having randomness property (along with other good statistical properties). Consequently, the security of these cryptosystems deliberately depends upon the randomness property regarding a sequence. Thus, it is mandatory to evaluate the randomness of a sequence before utilized them in any cryptosystems. Basically, two crucial properties namely the linear complexity [4] and the distribution of bit patterns regarding a sequence are nowadays well-known to check the randomness of a pseudo-random sequence. In this work, the authors restrict the discussion on the distribution of bit patterns property to evaluate the randomness of a sequence.

Most renowned pseudo-random number generators are the Mersenne Twister (MT) [5], Blum-Blum-Shub (BBS) [6], Legendre sequence [7, 8], and M-sequence [9]. Among those the former two pseudo-random number generators (MT and BBS) are well-known considering their applications in cryptography rather than the theoretical aspect. On the other hand, the M-sequence and Legendre sequence are prominent geometric sequences regarding the theoretical aspect. As a result, the authors attracted in the pseudo-random sequence generation research area by observing the theoretical prospect on the M-sequence and Legendre sequences.

The Legendre sequence [7, 8] is generated by applying the Legendre symbol over the odd characteristic field. It has a long period, high linear complexity, and the distribution of bit patterns of the Legendre se-

[*]Md. Arshad Ali, Graduate School of Natural Science and Technology, Okayama University, 3-1-1, Tsushima-naka, Kita, Okayama, 7008530, Japan, +81-8042661986, arshad@s.okayama-u.ac.jp

https://dx.doi.org/10.25046/aj040246

quence is known to be close to uniform [10, 11]. On the other hand, M-sequence is generated by a linear recurrence relation over the finite field. It has a maximum period but minimum linear complexity. In addition, M-sequence [9] is well-known for its uniform distribution of bit patterns [12]. Our previous work on geometric sequence [13] combines the features of the Legendre sequence and M-sequence. As mentioned previously, linear complexity and distribution of bit patterns are the important measures to evaluate the randomness of a sequence. So, regarding the linear complexity, our previous sequence [13] always possess high value. Unlike the linear complexity, the distribution of bit patterns in [13] doesn't reaches up to the mark alike the Legendre sequence and M-sequence. Hence, its a scope to improve the distribution of bit patterns in our previous sequence.

The trace calculation is an important step during our sequence generation procedure. Lets focus on the important aspect regarding this calculation. In case of prime field $\mathbb{F}_p$, the trace function maps an element of the extension field $\mathbb{F}_{q^M}$ to an element of the prime field $\mathbb{F}_p$. Therefore, the number of possible trace outputs will be in the range of $\{0 \sim p-1\}$. In other words, if we calculate the trace over the prime field, then it will output p kinds of values. On the other hand, in case of the sub extension field $\mathbb{F}_q$, the trace function maps an element of the extension field $\mathbb{F}_{q^M}$ to an element of the sub extension field $\mathbb{F}_q$ and the number of possible trace outputs will be in the range of $\{0 \sim q-1\}$ which means the trace outputs q kinds of values. It should be noted that here $M = m/m'$, $q = p^{m'}$, and m' be one of the factors of m. From the theoretical perspective, more variation in the trace values contribute to the better appearance of bits (0 and 1) in a sequence. This is one of the important aspects to consider the sub extension during the sequence generation procedure to improve the distribution of bit patterns in our previous sequence [13]. After utilizing the sub extension field, the detailed improvement in distribution of bit patterns is introduced in the result and discussion section in this paper.

Recently, the authors started to consider the sub extension field during the sequence generation procedure, which is a new dimension of our research work on generation of pseudo-random sequence (whereas our previous works on binary sequence [13] and multi-value sequence [14, 15] are considered in the prime field). As a result, our recent works on binary sequence [16] and multi-value sequence [17] experimentally observed the linear complexity, autocorrelation properties, respectively. As mentioned previously, the distribution of bit patterns is an important measure to evaluate the randomness of a sequence. Thus, the authors of this paper consider the distribution of bit patterns in a binary sequence which generated over the sub extension field.

The Legendre sequence and M-sequence are the base of the sequence research area. Their properties are already proven, therefore many researchers attracted by those sequences. As mentioned previously, our sequence also generated by the idea of the Legendre and M-sequences. Consequently, the authors thought that its properties can be theoretically proven and fortunately its proven (which shown in the later section of this paper). This is one of the contributions of the authors in this paper. Moreover, they also make a comparison between the binary sequence defined over the sub extension field with their previous work on binary sequence in terms of distribution of bit patterns property. According to the comparison result, binary sequence (defined over sub extension field) holds much better (close to uniform) distribution of bit patterns than the previous binary sequence [13]. Finding this improvement by considering the sub extension field is the major contribution of this paper.

The authors of this paper observed the distribution of bit patterns in a binary sequence which generated by a primitive polynomial, trace function, and Legendre symbol over the sub extension field. In brief, the sequence generation procedure is as follows: at first, it uses a primitive polynomial over the odd characteristic field $\mathbb{F}_p$ to generate maximum length vector sequence as elements in $\mathbb{F}_{q^M}$, then the trace function maps the extension field $\mathbb{F}_{q^M}$ elements to the sub extension field $\mathbb{F}_q$ elements, and finally the Legendre symbol binarizes the sub extension field $\mathbb{F}_q$ elements to a binary sequence. The authors already observed the period, autocorrelation, cross-correlation, and linear complexity properties of the binary sequence (which generated over the sub extension field) in their previous works [16, 17]. Thus, this paper focused on the distribution of bit patterns property. In brief, the authors count the number of appearances for each n-bit patterns (where $1 \leq n \leq (m/m')$). After observing many experimental results, the authors found that the number of appearances of each bit pattern is related to the number of zeros contained in each bit pattern. Furthermore, the authors theoretically proven the distribution of bit patterns equation. Moreover, they also make a comparison with their previous work [13].

Throughout this paper, p and q denote an odd prime number and its power $q = p^{m'}$, respectively, where m be a positive integer which mainly denotes extension degree and m' be one of the factors of m. In addition, $M = m/m'$ and $\mathbb{F}_q^*$ denotes the multiplicative group of $\mathbb{F}_q$, that is $\mathbb{F}_q^* = \mathbb{F}_q - \{0\}$.

2 Preliminaries

This section briefly explains some fundamental concepts of the finite field theory such as group, field, primitive polynomial, trace function, Legendre symbol, and dual bases. Then, binary sequence is introduced along with its period and distribution of bit patterns properties.

2.1 Group

A group is a non-empty set $\mathbb{G}$ with a binary operation $\circ$ on its elements denoted as $< \mathbb{G}, \circ >$, which satisfies

the following axioms.

- **Closure** For $\forall_{a,b} \in \mathbb{G}$, the result of $a \circ b$ also exists in $\mathbb{G}$ and it is uniquely given.
- **Associativity** Elements in group $\mathbb{G}$ should follow associativity. i.e. $(a \circ b) \circ c = a \circ (b \circ c)$, where $a, b, c \in \mathbb{G}$.
- **Identity element** There exists an element $e \in \mathbb{G}$ such that $\forall_a \in \mathbb{G}$, $a \circ e = e \circ a = a$.
- **Inverse element** For $\forall_a \in \mathbb{G}$, there exists an element $b \in \mathbb{G}$ such that $a \circ b = e = b \circ a$, where b is called inverse element of a.

Commutative group A group $\mathbb{G}$ will be commutative if $a \circ b = b \circ a$ for all $a, b \in \mathbb{G}$.

Group generator For a given group $\mathbb{G}$ if there is an element $g \in \mathbb{G}$ such that for any $a \in \mathbb{G}$ there exists an unique integer i with $a = g^i$ then g will be called as a generator of $\mathbb{G}$.

Order of a group The order of a group $\mathbb{G}$ often denoted as $\#\mathbb{G}$ is the number of elements in the group $\mathbb{G}$.

Cyclic group A group $\mathbb{G}$ will be cyclic if there exists at least one generator $g \in \mathbb{G}$ and it is denoted as $\mathbb{G} =< g >$. From the definition of cyclic group, it can be visualized that each element in a cyclic group can be generated with iterative operations of generator g which shown in the following Figure 1.

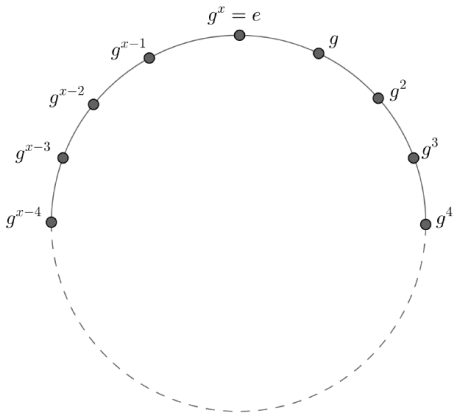


Figure 1: cyclic group.

Multiplicative group A cyclic group is called multiplicative if we tend to write its group operation in the same same way we do multiplication, that is

$$f = g \cdot x \text{ or } f = g^x.$$

2.2 Field

A field $< \mathbb{F}, +, \cdot >$ is a set that obeys two binary operations denoted by + and $\cdot$, such that

- $\mathbb{F}$ is a commutative group with respect to addition (+) having identity element 0.
- Let $\mathbb{F}^*$ is a subset of $\mathbb{F}$ having non-zero elements of $\mathbb{F}$ i.e. $\mathbb{F}^* = \mathbb{F} - \{0\}$. Then $\mathbb{F}^*$ will be called a commutative group with respect to multiplication $(\cdot)$, where every element should have multiplicative inverse in $\mathbb{F}^*$.
- For all $a, b, c \in \mathbb{F}$ the distributive law will be followed, i.e. $a \cdot (b+c) = a \cdot b + a \cdot c$ and $(b+c) \cdot a = b \cdot a + c \cdot a$.

Sub field Let $\mathbb{F}_1$ is a sub field of a field $\mathbb{F}$. Then $\mathbb{F}_1$ will be called a sub field if $\mathbb{F}_1$ obeys the laws of field with respect to the field operation inherited from $\mathbb{F}$. In addition, if $\mathbb{F}_1 \neq \mathbb{F}$, then $\mathbb{F}_1$ is a proper sub field of $\mathbb{F}$.

Prime field Let p be a prime. The ring of integers modulo p is a finite field of characteristic p having field order p denoted as $\mathbb{F}_p$ is called a prime field.

Extension field A subset $\mathbb{F}_0$ of a field $\mathbb{F}$ that is itself a field under the operations of $\mathbb{F}$ will be called a sub field of $\mathbb{F}$. In this case, $\mathbb{F}$ is called an extension field of $\mathbb{F}_0$. An extension field of a prime field $\mathbb{F}_p$ can be represented as m-dimensional vector space that has m elements in $\mathbb{F}_p$. Let the vector space be the m-th extension field be denoted as $\mathbb{F}_{q^M}$. The order of a extension field $\mathbb{F}_{q^M}$ is given as p^m (here $q = p^{m'}$ and $M = m/m'$).

In very brief, it can be said that a prime field ($\mathbb{F}_p$) is a subset of sub extension field ($\mathbb{F}_q$) and sub extension field $\mathbb{F}_q$ is also a sub set of extension field $\mathbb{F}_{q^M}$ which shown in Figure 2.

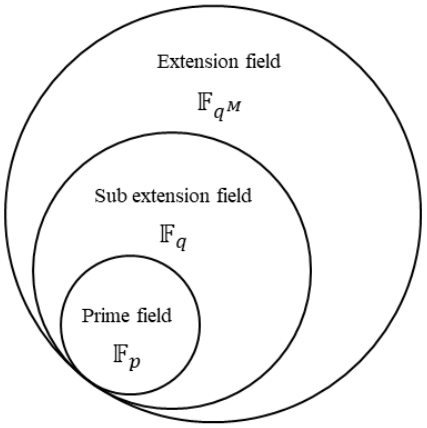


Figure 2: $\mathbb{F}_p \subset \mathbb{F}_q \subset \mathbb{F}_{q^M}$.

2.3 Primitive Polynomial

Consider a polynomial $f(x)$ of degree m over prime field $\mathbb{F}_p$. If it is not factorized into smaller degree polynomials over the prime field $\mathbb{F}_p$, it is called an irreducible polynomial. Consider the smallest number e such that $x^e - 1$ is divisible by $f(x)$ over $\mathbb{F}_p$, it is known that e becomes a factor of $q^M - 1$. Then $f(x)$ is especially called a primitive polynomial, when e is equal to $q^M - 1$. Its zero ω belongs to the extension field $\mathbb{F}_{q^M}$ and it becomes a primitive element in $\mathbb{F}_{q^M}$ that generates every non-zero element in $\mathbb{F}_{q^M}$ as its power ω^i (for $i = 0, 1, 2, \ldots, q^{M-2}$).

2.4 Trace Function

This work utilizes the trace function to map an element of the extension field $X \in \mathbb{F}_{q^M}$ to an element of the sub extension field $x \in \mathbb{F}_q$ as,

$$x = \mathrm{Tr}_{q^M|q}(X) = \sum_{i=0}^{\frac{m}{m'}-1} X^{p^{im'}}. \tag{1}$$

A crucial point, the above trace becomes an arbitrary element in $\mathbb{F}_q$ and the trace function has a linearity property over the sub extension field $\mathbb{F}_q$ as follows,

$$\mathrm{Tr}_{q^M|q}(aX + bY) = a\mathrm{Tr}_{q^M|q}(X) + b\mathrm{Tr}_{q^M|q}(Y), \tag{2}$$

where $a, b \in \mathbb{F}_q$ and $X, Y \in \mathbb{F}_{q^M}$.

2.5 Legendre Symbol

The Legendre symbol $\left(a/q\right)_2$ is used to check the quadratic residue for any arbitrary element a in $\mathbb{F}_q$. It is defined as,

$$\begin{aligned}\left(a/q\right)_2 &= a^{(q-1)/2} \\ &= \begin{cases} 0, & \text{if } a = 0, \\ 1, & \text{else if } a \text{ is a QR in } \mathbb{F}_q^*, \\ p-1, & \text{otherwise } a \text{ is a QNR in } \mathbb{F}_q^*. \end{cases}\end{aligned} \tag{3}$$

Here, QR and QNR stand for Quadratic Residue (QR) and Quadratic Non-Residue (QNR), respectively. Additionally, the non-zero element a is called the QR if it has a square root in $\mathbb{F}_q$, otherwise a is called the QNR. In this paper, the Legendre symbol is used for translating a vector sequence generated by the trace function over $\mathbb{F}_q$ to a binary sequence. Above mentioned QR and QNR in $\mathbb{F}_q$ holds the following important property.

Non-zero elements are the roots of $x^{q-1} - 1$ in $\mathbb{F}_q^*$ over $\mathbb{F}_q$ without any duplicates. Since it is factorized as follows:

$$x^{q-1} - 1 = \left(x^{(q-1)/2} - 1\right) - \left(x^{(q-1)/2} - 1\right). \tag{4}$$

It is thus found that the number of QR's and QNR's in $\mathbb{F}_q^*$ are the same and it is given by $(q-1)/2$. In addition, these numbers are important part in proving the theorem in the later section of this paper.

2.6 Dual Bases

The dual bases plays an important role in proving the theorem shown in this paper. It is defined as follows:

Let $\mathbb{F}_{q^M}$ be a finite field and $\mathbb{F}_q$ be a finite extension of $\mathbb{F}_{q^M}$. Then the two bases $\mathcal{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{m-1}\}$ and $\mathcal{B} = \{\beta_0, \beta_1, \ldots, \beta_{m-1}\}$ of $\mathbb{F}_q$ over $\mathbb{F}_{q^M}$ are said to be the dual (or complementary) bases if

$$\mathrm{Tr}_{q^M|q}\left(\alpha_i \beta_j\right) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $1 \le i, j \le m-1$.

2.7 Binary Sequence and Its Properties

This paper introduces a binary sequence along with its period and distribution of bit patterns properties as follows.

2.7.1 Generation Procedure and Period

Let ω be a primitive element in the extension field $\mathbb{F}_{q^M}$, where $M = m/m'$, m be a composite number which denotes the extension degree of the primitive polynomial, and m' be one of the factors of m. Then, by utilizing the trace function and Legendre symbol a binary sequence $\mathcal{S}$ is generated as follows:

$$\mathcal{S} = \{s_i\}, s_i = f_2\left(\left(\mathrm{Tr}_{q^M|q}\left(\omega^i\right)\Big/p\right)\right), \tag{6}$$

where $i = (0, 1, 2, \ldots, \lambda - 1, \ldots)$, $s_i \in 0, 1$ and $f_2(\cdot)$ be a mapping function, which translates the 0, 1, and $p-1$ values sequence generated by the Legendre symbol to a pseudo-random binary sequence. This mapping function is defined as follows:

$$f_2(s) = \begin{cases} 0, & \text{if } x = 0, 1 \bmod q, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

After observing many experimental results, the authors derive the equation for the period λ of the binary sequence as,

$$\lambda = \frac{2(q^M - 1)}{q - 1}. \tag{8}$$

2.7.2 Distribution of Bit Patterns

From the viewpoint of security, the distribution of bit patterns is as important as the linear complexity. If a sequence holds the uniform distribution of bit patterns, then it becomes difficult to guess the next bit after observing the previous bit patterns. For example, let's assume a binary sequence having a period of 12 as $\mathcal{S}_{12} = \{1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0\}$. If we observe the 1-bit pattern in this sequence, then we can find that it has a uniform distribution of 1 and 0. In other words, 1 and 0 appear same in number. However, when we check 2-bit patterns on $\mathcal{S}_{12}$, we find that it only has two types of patterns (10 and 01). In this case, we can easily predict the next bit patterns after observing the previous patterns. For example, let us make a sub-sequence of $\mathcal{S}_{12}$ as $\{1, 0, 1, 0, x_5, x_6\}$, we can easily guess x_5 and x_6 as $x_5 = 1$ and $x_6 = 0$. Therefore, it is also essential to evaluate the distribution of bit patterns of the sequence to confirm its randomness. In other words, the uniformity of the distribution contributes to the randomness from the viewpoint of unpredictability.

3 Distribution of Bit Patterns in Binary Sequence

In this section, we will introduce the bit distribution of binary sequence which generated over the sub extension field. In addition, bit distribution of M-sequence and Legendre sequence is also introduced here. Throughout this section $b^{(n)}$, $Z(b^{(n)})$ and $D_{\mathcal{S}_\lambda}(b^{(n)})$ denotes a bit pattern of length n, number of 0's in $b^{(n)}$, and number of appearance of $b^{(n)}$ in $\mathcal{S}_\lambda$, respectively. For example, in a binary sequence of period 15, a 3-bit pattern $b = 101$ appears 4 times. Then, these notations become $b^{(3)} = 101$, $Z(b^{(3)}) = 1$, and $D_{\mathcal{S}_{15}}(b^{(3)}) = 4$.

3.1 Bit Distribution of M-sequence

The M-sequence [9] is generated by a linear recurrence relation over the finite field. M-sequence has a maximum period and uniform distribution of bit pattern except for the case of $Z(b^{(n)}) = n$ but it has minimum linear complexity. Let, $f(x) = x^4 + x + 1$ be a primitive polynomial over $\mathbb{F}_2$, then using the linear recurrence relation a M-sequence of period 15 becomes as follows.

$$\mathcal{S}_{15} = \{1,0,0,0,1,1,1,1,0,1,0,1,1,0,0\}. \quad (9)$$

The distribution of n-bit pattern in (9) is shown in Table 1, here $1 \le n \le m$. In the case of M-sequence, except the all-zero pattern, every pattern appears same in number. For example, when $n = 3$ all patterns appear 2 times (except 000 pattern). In other words, they are uniformly distributed. Every M-sequence has such good distribution of bit pattern feature.

Table 1: Bit distribution of the M-sequence $\mathcal{S}_{15}$.

n	$b^{(n)}$	$Z(b^{(n)})$	$D_{S_{15}}(b^{(n)})$
1	0	1	7
	1	0	8
2	00	2	3
	01	1	4
	10	1	4
	11	0	4
3	000	3	1
	001	2	2
	010	2	2
	100	2	2
	011	1	2
	101	1	2
	110	1	2
	111	0	2

3.2 Bit Distribution of LegendreSequence

Legendre sequence [7, 8] is generated by applying the Legendre symbol over the odd characteristic field. Legendre sequence has a long period, high linear complexity, and the distribution of bit pattern is close to uniform. Let, $p = 23$, then the Legendre sequence of period 23 becomes as follows.

$$\mathcal{S}_{23} = \{0,0,0,0,0,1,0,1,0,0,1,1,0,0,1,1,0,0,1,0,1,0,1,1,1\}. \quad (10)$$

The distribution of n-bit pattern in (10) is shown in Table 2. In case of Legendre symbol, bit patterns appearance is close to uniform.

Table 2: Bit distribution of the Legendre sequence $\mathcal{S}_{23}$.

n	$b^{(n)}$	$Z(b^{(n)})$	$D_{S_{23}}(b^{(n)})$
1	0	1	12
	1	0	11
2	00	2	6
	01	1	6
	10	1	5
	11	0	5
3	000	3	3
	001	2	3
	010	2	3
	100	2	3
	011	1	2
	101	1	3
	110	1	2
	111	0	2

3.3 Bit Distribution of the Proposed Binary Sequence

Let $\mathcal{S}_\lambda$ be a binary sequence of having a period of λ. Again, let $b^{(n)}$, $Z(b^{(n)})$, and $D_{\mathcal{S}_\lambda}(b^{(n)})$ denotes a bit pattern of length n, number of 0's in $b^{(n)}$, and number of appearance of $b^{(n)}$ in $\mathcal{S}_\lambda$, respectively. Then, the distribution of bit patterns in the binary sequence which defined over the sub extension field can be given by the following theorem.

$$D_{S_\lambda}\left(b^{(n)}\right) = \begin{cases} q^{M-(n\cdot m')} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)-1} \cdot \left(\frac{q+1}{2}\right)^{Z\left(b^{(n)}\right)} & \text{(11a)} \\ \qquad \text{when } 0 \le Z\left(b^{(n)}\right) < n, & \\ \lambda - \sum_{u=0}^{Z\left(b^{(n)}\right)} {}_nC_u \cdot D_{S_\lambda}\left(b^{(n)}\right) & \text{(11b)} \\ \qquad \text{when } Z\left(b^{(n)}\right) = n. & \end{cases}$$

Let ω be a primitive element in the extension field $\mathbb{F}_{q^M}$, where $M = m/m'$, m be a composite number which denotes the extension degree of the primitive polynomial, and m' be one of the factors of m. Then, utilizing the trace function and Legendre symbol one period of a binary sequence is generated as follows.

$$\mathcal{S}_\lambda = \{s_i\}, s_i = f_2\left(\left(\mathrm{Tr}_{q^M|q}\left(\omega^i\right)\Big/\!\!\Big/ p\right)\right), i = 0,1,2,\ldots,\lambda-1,\ldots, \quad (12)$$

Here λ be the period of the sequence and it is given by the following equation as,

$$\lambda = \frac{2(q^M - 1)}{q - 1}. \quad (13)$$

At first, a primitive polynomial is used, then the trace value is calculated, then the Legendre symbol outputs zero, QR or QNR in $\mathbb{F}_q$, and finally the sequence coefficients s_i is given by the mapping function $f_2(\cdot)$.

The authors of this paper observe the distribution of n-bit patterns in a binary sequence. It should be noted that here n satisfies $1 \leq n \leq (m/m')$ relation. The distribution of n-bit patterns evaluated by observing the consecutive sequence coefficients $(s_i, s_{i+1}, \ldots, s_{i+(n-1)})$. Particularly,

$$\begin{aligned} s_{i+0} &= f_2\left(\left(\mathrm{Tr}\left(\omega^i \cdot \omega^0\right)\right\rVert_p\right), \\ s_{i+1} &= f_2\left(\left(\mathrm{Tr}\left(\omega^i \cdot \omega^1\right)\right\rVert_p\right), \\ &\vdots \\ s_{i+(n-1)} &= f_2\left(\left(\mathrm{Tr}\left(\omega^i \cdot \omega^{n-1}\right)\right\rVert_p\right), \end{aligned}$$

where $0 \leq i \leq (q^M - 2)$. By observing the above sequence coefficients, the distribution of bit patterns $D_{\mathcal{S}_\lambda}$ is determined by the following trace values.

$$\mathrm{Tr}\left(\omega^i \cdot \omega^0\right), \mathrm{Tr}\left(\omega^i \cdot \omega^1\right), \ldots, \mathrm{Tr}\left(\omega^i \cdot \omega^{n-1}\right). \quad (14)$$

Let $\mathcal{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{m-1}\}$ be a basis, ω be a primitive element and with this basis ω^i is represented as,

$$\omega^i = \sum_{j=0}^{m-1} a_{i,j}\alpha_j, \text{where } a_{i,j} \in \mathbb{F}_q \text{ and } 0 \leq i \leq q^M - 2. \quad (15)$$

Again let $\mathcal{B} = \{\omega^0, \omega^1, \ldots, \omega^{n-1}, \beta_n, \ldots, \beta_{m-1}\}$ be a dual basis of $\mathcal{A}$ in $\mathbb{F}_q$ over $\mathbb{F}_{q^M}$. Then we also have

$$\omega^t = \omega^t + \sum_{j=0}^{\frac{m}{m'}-1} 0 \cdot \beta_j, \text{where } 0 \leq t < n. \quad (16)$$

Since $\mathcal{A}$ and $\mathcal{B}$ are dual bases to each other, then $\mathrm{Tr}\left(\omega^i \cdot \omega^t\right)$ be calculated as follows.

$$\begin{aligned} \mathrm{Tr}\left(\omega^i \cdot \omega^t\right) &= \mathrm{Tr}\left(\sum_{j=0}^{m-1} a_{i,j}\alpha_j \cdot \left(\omega^t + \sum_{j=0}^{\frac{m}{m'}-1} 0 \cdot \beta_j\right)\right) \\ &= a_{i,t}. \end{aligned}$$

Therefore, by using the dual basis, the distribution of bit patterns $D_{\mathcal{S}_\lambda}\left(b^{(n)}\right)$ determined by the trace values becomes as follows.

$$\begin{aligned} &\mathrm{Tr}\left(\omega^i \cdot \omega^0\right), \mathrm{Tr}\left(\omega^i \cdot \omega^1\right), \ldots, \mathrm{Tr}\left(\omega^i \cdot \omega^{n-1}\right) \\ &= (a_{i,0}, a_{i,1}, \ldots, a_{i,n-1}). \end{aligned}$$

Thus, instead of using sequence coefficients $(s_i, s_{i+1}, \ldots, s_{i+(n-1)})$, we can consider the dual basis representation of these coefficients as $\left(a_{i,0}, a_{i,1}, \ldots, a_{i,(n-1)}\right)$. Additionally, all the above trace values belong to the sub extension field $\mathbb{F}_q$.

Furthermore, $\omega^i\left(0 \leq i \leq q^M - 2\right)$ in (15) represents every non-zero vectors in the extension field $\mathbb{F}_{q^M}$ as,

$$\left\{\mathrm{Tr}\left(\omega^0\right), \mathrm{Tr}\left(\omega^1\right), \mathrm{Tr}\left(\omega^2\right), \mathrm{Tr}\left(\omega^3\right), \ldots, \mathrm{Tr}\left(\omega^{q^M-2}\right)\right\}. \quad (17)$$

According to the trace property, non-zero $\mathbb{F}_q$ elements appear $q^{M-m'}$ times and zero appears one less than the other elements i.e. $q^{M-m'} - 1$ times in the above equation.

3.3.1 Relation Between the Sequence Coefficients With the Trace Values and Legendre Symbol Calculation

Depending on the three different types of trace values (0, QR, and QNR), the Legendre symbol outputs three different values (0, 1, and $p-1$), and finally the mapping function outputs 0 and 1 as sequence coefficients s_i. This dependency between the trace and Legendre symbol is explained as follows.

Table 3: Relation between the sequence coefficients with trace and Legendre symbol calculation -I.

s_i	$\mathrm{Tr}\left(\omega^i\right)$
0	0 or QR in $\mathbb{F}_q^*$
1	QNR in $\mathbb{F}_q^*$

According to the above table, the sequence coefficient 0 comes from the two cases: one is for the $\mathrm{Tr}(0)$ case and another one is for the QR in $\mathbb{F}_q^*$ case. To deal with this two cases uniquely, let us denote **0** and 0 for the first and second cases, respectively. In addition, 1 comes for QNR in $\mathbb{F}_q^*$ case. Thus the above table can be further modified as follows.

Table 4: Relation between the sequence coefficients with trace and Legendre symbol calculation -II.

s_i	$\mathrm{Tr}\left(\omega^i\right)$
0	0
0	QR in $\mathbb{F}_q^*$
1	QNR in $\mathbb{F}_q^*$

To distinguish the appearance of 0, this paper uses the notation **0**, when zero comes from $\mathrm{Tr}(0)$ and 0 when zero comes from QR. Let the number of **0** be denoted by u and $T_{u,n}$ denotes the number of bit patterns including u times **0** and $Z(b^{(n)}) - u$ times 0. Thus, T_n can be considered as,

$$\sum_{u=0}^{Z\left(b^{(n)}\right)} T_{u,n}. \quad (18)$$

In the following section, the distribution of bit patterns in the binary sequence defined over the sub extension field theoretically proven.

3.3.2 Proof of (11a)

The period of the binary sequence is given by the following equation as,

$$\lambda = \frac{2(q^M-1)}{q-1}. \tag{19}$$

After rewriting the above equation we obtain,

$$q^M - 1 = \lambda \cdot \left(\frac{q-1}{2}\right). \tag{20}$$

To observe the distribution of bit patterns, the above relation becomes as follows.

$$D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right) = D_{\mathcal{S}_\lambda}\left(b^{(n)}\right) \cdot \left(\frac{q-1}{2}\right). \tag{21}$$

Thus, we must consider two cases of the sequence length such as $\mathcal{S}_{q^M-1}$ and $\mathcal{S}_\lambda$. Hence, we will observe the distribution of bit patterns in $\mathcal{S}_{q^M-1}$ as $D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right)$ and $\mathcal{S}_\lambda$ as $D_{\mathcal{S}_\lambda}\left(b^{(n)}\right)$.

In the previous section, we explained that n-bit patterns can be considered as $b^{(n)} = (a_{i,0}, a_{i,1}, \ldots, a_{i,n-1})$. On the other hand, the remaining $(m-(n\cdot m'))$-bit patterns are composed of $(a_{i,nm'}, a_{i,nm'+1}, \ldots, a_{i,m-1})$ coefficients of ω^i, which is given by the (16). In addition, the number of combinations of $(a_{i,nm'}, a_{i,nm'+1}, \ldots, a_{i,m-1})$ becomes $q^{M-nm'}$. It should be noted that here ω^i represents all of the non-zero coefficients in the extension field $\mathbb{F}_{q^M}$.

As mentioned previously, when the trace value is equal to 0 or QR, then the sequence coefficients becomes **0** and 0, respectively. In addition, if the trace value is equal to QNR, then the sequence coefficients becomes 1. Additionally, u denotes the number of **0** in $b^{(n)}$ $\left(\text{where } 0 \le u \le Z\left(b^{(n)}\right)\right)$ from $\mathrm{Tr}(0)$, then the other 0's comes from $Z\left(b^{(n)}\right) - u$ QR's, and finally 1's comes from $n - Z\left(b^{(n)}\right)$ QNR's. Therefore, by separating 0, $T_{u,n}$, and T_n the combination of n-bit patterns can be given as follows.

$$\begin{aligned} T_{u,n} = {}& {}_nC_u \cdot {}_{n-u}C_{Z\left(b^{(n)}\right)-u} \cdot \left(\frac{q-1}{2}\right)^{Z\left(b^{(n)}\right)} \\ & \times {}_{n-Z\left(b^{(n)}\right)}C_{n-Z\left(b^{(n)}\right)} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \end{aligned} \tag{22}$$

Furthermore, T_n can be derived as,

$$\begin{aligned} T_n &= \sum_{u=0}^{Z\left(b^{(n)}\right)} T_{u,n} \\ &= \sum_{u=0}^{Z\left(b^{(n)}\right)} {}_nC_u \cdot {}_{n-u}C_{Z\left(b^{(n)}\right)-u} \cdot \left(\frac{q-1}{2}\right)^{Z\left(b^{(n)}\right)} \\ &\quad \times \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \end{aligned} \tag{23}$$

According to the above equation, T_n can be calculated by $Z\left(b^{(n)}\right)$. In addition, there are ${}_nC_{Z\left(b^{(n)}\right)}$ possible bit patterns that have the same $Z\left(b^{(n)}\right)$. To calculate the $D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right)$ for each $b^{(n)}$, T_n needs to be divided by ${}_nC_{Z\left(b^{(n)}\right)}$.

$$\begin{aligned} D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right) &= q^{M-(n\cdot m')} \cdot \frac{T_n}{{}_nC_{Z\left(b^{(n)}\right)}} \\ &= q^{M-(n\cdot m')} \sum_{u=0}^{Z\left(b^{(n)}\right)} \frac{{}_nC_u \cdot {}_{n-u}C_{Z\left(b^{(n)}\right)-u}}{{}_nC_{Z\left(b^{(n)}\right)}} \\ &\quad \times \left(\frac{q-1}{2}\right)^{Z\left(b^{(n)}\right)-u} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \end{aligned} \tag{24}$$

The above equation can be further modified as follows.

$$\begin{aligned} & \sum_{u=0}^{Z\left(b^{(n)}\right)} {}_nC_u \cdot {}_{n-u}C_{Z\left(b^{(n)}\right)-u} \\ &= \frac{n!}{\left(n-Z\left(b^{(n)}\right)\right)!} \cdot \sum_{u=0}^{Z\left(b^{(n)}\right)} \cdot \frac{1}{u!\left(Z\left(b^{(n)}\right)-u\right)!} \\ &= \frac{n!}{\left(n-Z\left(b^{(n)}\right)\right)!} \cdot \sum_{u=0}^{Z\left(b^{(n)}\right)} \cdot \frac{1}{u!\left(Z\left(b^{(n)}\right)-u\right)!} \\ &\qquad \times \frac{\left(Z\left(b^{(n)}\right)\right)!\left(n-Z\left(b^{(n)}\right)\right)!}{n!} \\ &= \sum_{u=0}^{Z\left(b^{(n)}\right)} \frac{(Z(b^{(n)}))!}{u!(Z(b^{(n)}))!} = {}_{Z\left(b^{(n)}\right)}C_{Z\left(b^{(n)}\right)-u}. \end{aligned} \tag{25}$$

Thus, (24) becomes as follows:

$$\begin{aligned} D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right) &= q^{M-(n\cdot m')} \sum_{u=0}^{Z\left(b^{(n)}\right)} {}_{Z\left(b^{(n)}\right)}C_{Z\left(b^{(n)}\right)-u} \\ &\quad \times \left(\frac{q-1}{2}\right)^{Z\left(b^{(n)}\right)-u} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \end{aligned} \tag{26}$$

By using the bilinear theorem, the above equation can be rewritten as,

$$\begin{aligned} D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right) &= q^{M-(n\cdot m')} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \\ &\quad \times \left(\frac{q-1}{2}+1\right)^{Z\left(b^{(n)}\right)} \\ &= q^{M-(n\cdot m')} \cdot \left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \cdot \left(\frac{q+1}{2}\right)^{Z\left(b^{(n)}\right)}. \end{aligned} \tag{27}$$

From the (21), $D_{\mathcal{S}_\lambda}\left(b^{(n)}\right)$ holds the following relation as follows,

$$D_{\mathcal{S}_\lambda}\left(b^{(n)}\right) = D_{\mathcal{S}_{q^M-1}}\left(b^{(n)}\right) \cdot \left(\frac{q-1}{2}\right)^{-1}. \tag{28}$$

Therefore, using the (27), $D_{\mathcal{S}_\lambda}\left(b^{(n)}\right)$ can be given by the following relation as,

$$\begin{aligned} D_{\mathcal{S}_\lambda}\left(b^{(n)}\right) &= q^{M-(n\cdot m')}\cdot\left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)} \\ &\quad\times\left(\frac{q+1}{2}\right)^{Z\left(b^{(n)}\right)}\cdot\left(\frac{q-1}{2}\right)^{-1} \\ &= q^{M-(n\cdot m')}\cdot\left(\frac{q-1}{2}\right)^{n-Z\left(b^{(n)}\right)-1}\cdot\left(\frac{q+1}{2}\right)^{Z\left(b^{(n)}\right)}. \end{aligned} \tag{29}$$

Thus, the first part of the (11a) is proven.

3.3.3 Proof of (11b)

Let us consider the case that $Z\left(b^{(n)}\right)=n$. Therefore, the combination of n-bit patterns except the all-zero patterns is given as follows:

$${}_nC_{Z\left(b^{(n)}\right)}. \tag{30}$$

Thus, the distribution of all-zero patterns becomes

$$D_{\mathcal{S}_\lambda}\left(b^{(n)}\right)=\lambda-\sum_{u=0}^{n-1}{}_nC_u\cdot D_{\mathcal{S}_\lambda}\left(b^{(n)}\right). \tag{31}$$

Thus, the second part of the (11b) is proven. In addition, the theorem in (11) is also proven.

4 Result and Discussion

This section explains the distribution of bit patterns in the binary sequence which generated over the sub extension field based on some experimental results. Then, a comparison between the binary sequence defined over the sub extension field and our previous geometric sequence [13] also introduces in terms of the distribution of bit patterns property. Here, H_{wt} denotes the hamming weight.

4.1 Experimental Results

Let us consider the distribution of bit patterns in the binary sequence, introduced in this paper which generated over the sub extension field in the following examples.

Example 1 *Let $p=5, m=4$, and $m'=2$, then the sequence having a period of* 52 *becomes as follows its distribution of n-bit patterns is shown in Table 5.*

$$\begin{aligned} \mathcal{S}_{52} = \{&0100110100000010101111001 1 \\ &1011001011111001010000110 0\}. \end{aligned} \tag{32}$$

Table 5: Bit distribution of the binary sequence $\mathcal{S}_{52}$ with $p=5, m=4$, and $m'=2$.

n	$H_{wt}(b^{(n)})$	$Z(b^{(n)})$	$D_{S_{52}}(b^{(n)})$
1	0	1	27
	1	0	25
2	0	2	14
	1	1	13
	2	0	12

Example 2 *Let $p=3, m=6$, and $m'=2$, then the sequence having a period of* 182 *becomes as follows its distribution of n-bit patterns is shown in Table 6.*

Table 6: Bit distribution of the binary sequence $\mathcal{S}_{182}$ with $p=3, m=6$, and $m'=2$.

n	$H_{wt}(b^{(n)})$	$Z(b^{(n)})$	$D_{S_{182}}(b^{(n)})$
1	0	1	101
	1	0	81
2	0	2	56
	1	1	45
	2	0	36
3	0	3	31
	1	2	25
	2	1	20
	3	0	16

Example 3 *Let $p=7, m=9$, and $m'=3$, then the sequence having a period of* 235986 *becomes as follows its distribution of n-bit patterns is shown in Table 7.*

Table 7: Bit distribution of the binary sequence $\mathcal{S}_{235986}$ with $p=7, m=9$, and $m'=3$.

n	$H_{wt}(b^{(n)})$	$Z(b^{(n)})$	$D_{S_{235986}}(b^{(n)})$
1	0	1	118337
	1	0	117649
2	0	2	59341
	1	1	58996
	2	0	58653
3	0	3	29757
	1	2	29584
	2	1	29412
	3	0	29241

4.1.1 Observation

It was found that the experimental results explicitly support the (11). In addition, the number of appearance of each bit pattern is related to the number of zeros contained in each bit pattern. Moreover, $D_{\mathcal{S}_\lambda}(b^{(n)})$

increases in proportion to $Z(b^{(n)})$. To confirm this, let us check the **Example** 2 with $n = 3$.

$Z(b^{(3)}) = 0$:

$$D_{S_{182}}(b^{(3)} = 111) = 3^{6-(3\times2)} \cdot 4^{3-0-1} \cdot 5^0 = 16.$$

$Z(b^{(3)}) = 1$:

$$D_{S_{182}}(b^{(3)} = 011) = 3^{6-(3\times2)} \cdot 4^{3-1-1} \cdot 5^1 = 20,$$
$$D_{S_{182}}(b^{(3)} = 101) = 3^{6-(3\times2)} \cdot 4^{3-1-1} \cdot 5^1 = 20,$$
$$D_{S_{182}}(b^{(3)} = 111) = 3^{6-(3\times2)} \cdot 4^{3-1-1} \cdot 5^1 = 20.$$

$Z(b^{(3)}) = 2$:

$$D_{S_{182}}(b^{(3)} = 001) = 3^{6-(3\times2)} \cdot 4^{3-2-1} \cdot 5^2 = 25,$$
$$D_{S_{182}}(b^{(3)} = 010) = 3^{6-(3\times2)} \cdot 4^{3-2-1} \cdot 5^2 = 25,$$
$$D_{S_{182}}(b^{(3)} = 100) = 3^{6-(3\times2)} \cdot 4^{3-2-1} \cdot 5^2 = 25.$$

$Z(b^{(3)}) = 3$:

$$D_{S_{182}}(b^{(3)} = 000) = 182 - (1 \times 16 + 3 \times 20 + 3 \times 25) = 31.$$

4.2 Comparison With Our Previous Work

By combining the features of the M-sequence and Legendre sequence our previous work [13] proposed a geometric sequence, namely NTU (Nogami-Tada-Uehara) sequence. According to our previous research work, NTU sequence always holds long period, low correlation, high linear complexity properties which are the important considerations to use any sequence in cryptographic applications. Another crucial consideration before utilizing them in any secure applications, is to judge the randomness of a sequence. To do so, we need to evaluate the distribution of bit patterns property in a sequence. After the experimental observation, it was found that in terms of distribution of bit patterns NTU sequence is not uniformly distributed. In other words, in case of binary NTU sequence, there is much difference in appearance between the 0 and 1. To improve this drawback, instead of prime field (which used in the NTU sequence generation procedure), we focused on the sub extension field during the sequence generation procedure in this research work. As a result, after utilizing the sub extension field, the distribution of bit patterns becomes close to uniform. This comparison is shown in the following tables (Table 8 and Table 9).

It should be noted that the NTU sequence is controlled by 2 parameters (p and m), on the other hand the sequence over the sub extension field is controlled by 3 parameters (p, m, and m'). Therefore, it is not possible to make the comparison between these two sequences in terms of the same length (in other words, the same period λ). The authors kept the difference as minimum as possible.

One of the most notable outcomes of this comparison result is the NTU sequence holds higher difference in terms of the appearance between the 'all zero' and 'all one' patterns. In other words, it also confirms the ununiform distribution of bit patterns. On the other hand, sequence defined over the sub extension field minimizes this difference to make it close to uniform. This comparison graphically shown in Figure 3.

Table 8: Comparison in bit distribution between the sub field binary sequence and NTU sequence -I.

n	$H_{wt}(b^{(n)})$	$D_{S_{182}}(b^{(n)})$	%	$D_{NTU_{242}}(b^{(n)})$	%
1	0	101	55.49	161	66.52
	1	81	44.51	81	33.48
2	0	56	30.76	107	44.21
	1	45	24.72	54	22.31
	2	36	19.78	27	11.15
3	0	31	17.03	71	29.33
	1	25	13.73	36	14.87
	2	20	10.98	18	7.43
	3	16	8.79	9	3.71

Table 9: Comparison in bit distribution between the sub field binary sequence and NTU sequence -II.

n	$H_{wt}(b^{(n)})$	$D_{S_{240200}}(b^{(n)})$	%	$D_{NTU_{275514}}(b^{(n)})$	%
1	0	122551	51.02	156865	56.93
	1	117649	48.98	117649	43.07
2	0	62526	26.03	89637	32.53
	1	60025	24.98	67228	24.40
	2	57624	23.99	50421	18.30
3	0	31901	13.28	51221	18.59
	1	30625	12.74	38416	13.94
	2	29400	12.23	28812	10.45
	3	28224	11.75	21609	7.84
4	0	16276	6.77	29269	10.62
	1	15625	6.50	21952	7.96
	2	15000	6.24	16464	5.97
	3	14400	5.99	12348	4.48
	4	13824	5.75	9261	3.36

Recently, there are lots of considerations to use a long period pseudo-random sequence in cryptographic applications. The use of binary sequence in a stream cipher is one of the most common application. Before applying a sequence in such applications, the linear complexity and distribution of bit patterns are considered as the most important properties regarding a sequence to check its randomness. Among these two, the authors observed the linear complexity property in their previous work [16] and it always holds a maximum value of the linear complexity. As a continuation, the authors focused on the distribution of bit patterns in this paper. According to the comparison results, the binary sequence generated over the sub extension field holds much better (close to uniform) compared to our previous binary sequence in terms of distribution of bit patterns. Therefore, the binary sequence defined over the sub extension field can be a suitable candidate for some cryptographic applications.

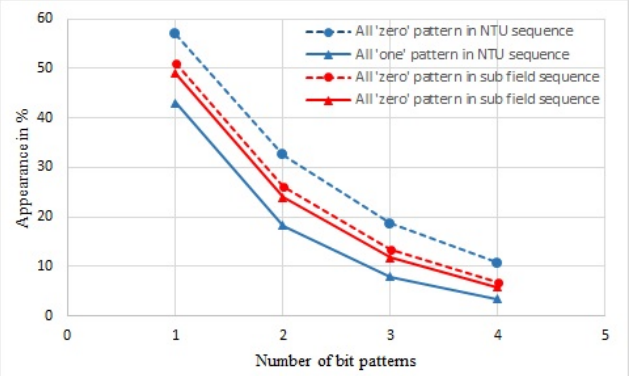


Figure 3: Appearance of 'all zero' and 'all one' bit patterns in the NTU and sub field sequence.

5 Conclusion

In this paper, the authors observed the distribution of bit patterns in a binary sequence which defined over the sub extension. The number of appearances is related to the number of zeros contained in each bit pattern. Furthermore, the authors theoretically prove the distribution of bit patterns property. In addition, they also made a comparison between the binary sequence defined over the sub extension field and our previous work on binary sequence based on distribution of bit patterns property. According to the comparison results, the binary sequence generated over the sub extension field holds much better (close to uniform) compared to our previous binary sequence. As a future work, we would like to consider an efficient implementation to enhance the usability of our proposed sequence a Cryptographically Secure Pseudo Random Number Generator (CSPRNG).

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This work has been supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (A) Number 16H01723.

References

[1] J. Daemen and V. Rijmen, The Design of Rijndael, Springer-Verlag Berlin Heidelberg, Germany, 2002.

[2] Richard A. Mollin, RSA and Public-Key Cryptography, Chapman & Hall CRC, 2002.

[3] H. Cohen and G. Frey, Handbook of Elliptic and Hyperelliptic Curve Cryptography, Discrete Mathematics and Its Applications, Chapman & Hall CRC, 2005.

[4] V. Edemskiy, On the Linear Complexity of Interleaved Binary Sequences of Period 4p Obtained from Hall Sequences or Legendre and Hall Sequences, IEEE Electronic Letter, 50, 8, 604–605, 2014.

[5] M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator", ACM Trans. on Modeling and Computer Simulation, **8**(1), 3–30, 1998. http://doi.org/10.1145/272991.272995

[6] L. Blum, M. Blum, and M. Shub, "A simple Unpredictable Pseudo-Random Number Generator", SIAM Journal on Computing, **15**(2), 364–383, 1986. http://doi.org/10.1137/0215025

[7] X. Tang and G. Gong, "New Constructions of Binary Sequences with Optimal Autocorrelation Value/Magnitude", IEEE Trans. Inf. Theory, **56**(3), 1278–1286, 2010. http://doi.org/10.1109/TIT.2009.2039159

[8] J. S. No, H. K. Lee, H. Chung, H. Y. Song, and K. Yang, "Trace Representation of Legendre Sequence of Mersenne Prime Period", IEEE Trans. on Inform. Theory, **42**(6 PART 2), 2254–2255, 1996. https://doi.org/10.1109/18.556617

[9] J. Ren, "Design of Long Period Pseudo-Random Sequence from the Addition of m-sequences over $\mathbb{F}_p$", EURASIP Journal on Wireless Communication and Networking, **1**, 12–18, 2004. http://doi.org/10.1155/S1687147204405052

[10] C. Ding, "Pattern Distributions of Legendre Sequences", IEEE Transactions on Information Theory, **44**(4), 1693–1698, 1998. http://doi.org/10.1109/18.681353

[11] B. Z. Moroz, "The distribution of power residues and non-residues", Vestnik Leningrad Univ. Math, **16**, 164–169, 1961. (In Russian with English summary)

[12] T. Helleseth, "Maximal-length sequences", Encyclopedia of Cryptography and Security, 2nd Ed. Springer, 5–14, 2011. https://doi.org/10.1007/978-1-4419-5906-5_359

[13] Y. Nogami, K. Tada, and S. Uehara, "Geometric Sequence Binarized with Legendre Symbol over Odd Characteristic Field and Its Properties", IEICE Trans. on Fund. of Electronics and Computer Sciences, **E97.A**(12), 2336–2342, 2014. https://doi.org/10.1587/transfun.E97.A.2336

[14] Y. Nogami, S. Uehara, K. Tsuchiya, N. Begum, H. Ino, and R. H. Morelos-Zaragoza, "A Multi-value Sequence Generated by Power Residue Symbol and Trace Function over Odd Characteristic Field", IEICE Trans. on Fund. of Electronics, Communications and Computer Sciences, **E99.A**(12), 2226–2237, 2016. https://doi.org/10.1587/transfun.E99.A.2226

[15] A. M. Arshad, Y. Nogami, H. Ino, and S. Uehara, "Auto and Cross Correlation of Well Balanced Sequence Over Odd Characteristic Field", Fourth International Symposium on Computing and Networking (CANDAR), Hiroshima, Japan, 2016. http://doi.org/10.1109/CANDAR.2016.0109

[16] A. M. Arshad, T. Miyazaki, S. Heguri, Y. Nogami, S. Uehara, and R. H. Morelos-Zaragoza, "Linear Complexity of Pseudo Random Binary Sequence Generated by Trace Function and Legendre Symbol Over Proper Sub Extension Field", Eighth International Workshop on Signal Design and Its Applications in Communications (IWSDA), Sapporo, Japan, 2017. http://doi.org/10.1109/IWSDA.2017.8095741

[17] A. M. Arshad, T. Miyazaki, Y. Nogami, S. Uehara, and R. H. Morelos-Zaragoza, "Multi-value Sequence Generated by Trace Function and Power Residue Symbol Over Proper Sub Extension Field", IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW), Taipei, Taiwan, 2017. http://10.1109/ICCE-China.2017.7991089

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 380-387 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

An Analysis of Social Media Usage in Winery Businesses

Constantina Costopoulou[1,*], Maria Ntaliani[1], Filotheos Ntalianis[2]

[1]*Informatics Laboratory, Department of Agricultural Economics & Rural Development, Agricultural University of Athens, 75 Iera Odos, 118 55 Athens, Greece*

[2]*Department of Business Administration, University of Piraeus, Karaoli & Dimitriou 80, 185 34 Piraeus, Greece*

A R T I C L E I N F O

Article history:
Received: 31 May, 2019
Accepted: 23 July, 2019
Online: 16 August, 2019

Keywords:
Social Media
Social Media Analytics Tools
Wine Industry
Greece
Germany
Facebook

A B S T R A C T

Although social media is used by an ever-increasing number of businesses, its benefits have not yet been fully exploited. Social media analytics tools can prove helpful in this regard, since they comprise a unique information source for business intelligence. As the social media environment rapidly evolves, various social media analytics tools are emerging. This research examines the usefulness of social media analytics tools for assessing corporate websites and Facebook pages. Two social media analytics tools, Website Grader and LikeAlyzer, are selected to evaluate the wine industry in Greece and Germany. A comparative study of the two countries is subsequently presented. The paper suggests that researchers and practitioners can easily use these analytics tools to acquire valuable quantitative data for a fast, effective and efficient assessment of corporate websites and Facebook pages and thus improve their social media usage. The results show that the two countries present similarities, but there is great potential for improvement in their use of social media.

1. Introduction

Digital marketing is transforming the very nature of business through the implementation of digital technologies which serve the marketing goals of businesses. Digital marketing is classified into various categories, including search engine marketing (e.g. Bing, Google), influencer marketing, social media (SM) marketing, online display advertising, mobile marketing, affiliate marketing (e.g. Amazon), email marketing, and content marketing.

SM marketing provides an innovative way of communicating with the public, attracting new and maintaining old customers, improving customer relationships, creating brand awareness, increasing market share and return of investment, gaining business revenue from sales, and engaging more customers in charities, corporate social responsibility activities and other issues [1]. Nowadays, more firms are gradually adopting SM, for example, YouTube, Twitter, Pinterest, Google+, Facebook, Instagram, and Flicker, setting aside their initial doubts regarding their effectiveness [2].

As the digital environment evolves, SM hold a substantial position in the next generation business intelligence platform. Effective SM use can have a decisive impact on a digital marketing campaign [3]. Thus, more and more SM analytics tools are emerging. Understanding SM usage in the context of digital marketing through analytics tools is critical for both academic and business communities. These tools have become a unique information source for assessing a business's presence in the SM environment as they collect, monitor, analyze, summarize and visualize a vast amount of SM data [4]. However, businesses from various sectors, including the food and beverage industry, have not yet exploited the full potential of these tools.

Wine may be considered a social product, as it is included in the top eight product categories involved in discussions over the internet. Thus, the internet provides an ideal medium for wine promotion and sale. Surveys have shown that in comparison to many other consumer products, wine appears more often as a search engine keyword [5]. In the past wine consumers sought information from specialized magazines, books, and newsletters. Lately more and more consumers are consulting SM connoisseurs, sommeliers and the wine-loving audience in order to help them decide on a satisfactory wine purchase. On the other hand, businesses have a lot to gain from the feedback provided by customers through SM. Professionals in the wine industry are

*Corresponding Author: Constantina Costopoulou, tina@aua.gr, +302105294183

trying to understand consumer behavior and preferences through the opinions and criticism diffused through various SM [6].

Today competition in the wine market is at its highest level to date. Around the world the vineyard surface area covers 7.4 million hectares and the production of wine has reached 292.3 million hectoliters. The exact number of wine producers cannot be safely estimated since, in addition to professional winemakers, there are hobbyists, family micro-businesses and single producers. In 2018 Italy, France and Spain were the dominant global wine exporters. Specifically, in terms of volume Spain is the largest exporter and the three countries together represent half of the world market. In terms of value, France – maintaining a successful strategy for high quality products – is the leader; France and Italy together hold a share of the global market of approximately 30% [7].

However, this intense competition requires new strategies which go beyond the production of high quality wine, leading wineries to strive for differentiation [8]. Towards this end, SM have succeeded the geographic decoupling of wine customers and wineries affecting data collection and sharing, communication, and the overall wine purchasing experience.

Despite the significant impact of SM on wine customers and winery businesses [9, 10] digital marketing for wine is still complex [11]. Wineries are increasingly adopting SM [12] but at a slow pace, as most lack the knowledge to incorporate and exploit SM effectively in their marketing strategy [13-16]. A study of Italian wineries has shown that it is principally large wineries with a number of employees with SM knowledge and skills that are highly motivated for adopting SM [17].

The current study aims to shed light on the usefulness of SM analytics tools for helping businesses to evaluate SM usage, plan their SM efforts and revise them regularly. It focuses on the wine industry in Greece and Germany. Two freely available online tools are used to extract useful data for both countries. The German market holds the largest share of wine imports in Europe, becoming a target for wine exporting countries, such as Greece. In addition, Greek wine exporters are trying to become more active in the German market [18].

The article is structured as follows. We begin with a discussion of SM analytics tools. We then outline the methods and tools used for assessing German and Greek wineries' websites and Facebook pages. The SM analytics tools, Website Grader and LikeAlyzer, are selected as these offer an accurate/comprehensive and quick assessment. Next, we present the evaluation for the two countries – 105 Greek and 100 German wineries. Finally, we provide the survey results and draw some conclusions.

2. SM Marketing

SM regard communication means among people, allowing content to be created and shared in the digital networking environment [19]. SM can be described as an online platform for participation, discussion, and sharing; its use encourages the exchange and diffusion of opinion and information among users [20]. [21] argues that the revolutionary role of content publishers provided to SM users contributes to a democratization of information. Although SM comprise sources of information, they also provide social and emotional support, as within the SM context there is emphasis on the interaction rather than on the content itself [22, 23]. [24] distinguishes five main categories of SM: collaborative projects (e.g. Wikipedia); social networking sites (e.g. Facebook); content communities (e.g. YouTube, Pinterest, and Flickr); blogs (e.g. Blogger, Tumblr, and Twitter); and virtual worlds (e.g. Second Life). SM reflect daily life discussions, occurring at work or at home, at a restaurant or supermarket, which travel far away from the place they happen and are communicated to the public through the digital world [25]. According to the definition by [26], SM can be considered an interactive dialogue. This "dialogue is often triggered by consumers/businesses and/or products/services that circulate among them to set in motion a revealing communication on promotional information, so that it allows learning from one another's use and experiences, eventually benefitting all of the involved parties" [p. 290].

While for individuals SM serve as a means of communication for users with similar interests, businesses can leverage user conversations about their products, services and brand, and engage in discussions for using the relevant information to make better business decisions [27]. [28] designates SM marketing as an interdisciplinary and interoperable concept that uses SM to achieve business aims by adding value for the parties involved. [29] defines it as the use of SM technologies (e.g. Web 2.0, Web 3.0) and software for creation, communication, delivery and sharing of offers that have value for business owners.

SM marketing focuses on producing attractive content for customers and promoting the use and sharing of this content through the various social networks they follow. In this way "word-of-mouth" marketing is practiced, establishing a strong relationship between consumers and products, with a limited cost for businesses [5]. Although the general public uses SM primarily to develop social contacts rather than for shopping, the success of SM marketing is based on a more indirect approach. More specifically, instead of offering products for sale, the company encourages members of a community to discuss and disseminate the use and benefits of these products [30]. Several businesses try to exploit consumers' opinions expressed through SM by involving them in product design choices [31] using SM analytics tools. In light of this, a significant number of data services, platforms and tools have been developed [32]. Some of the most popular examples are: Buffer, Hootsuite, Google Analytics, Talkwalker, TapInfluence, Brand24, Klear, Keyhol, Quintly, IBM Watson Personality analytics, BuzzSumo, Funnel.io, nTuitive.social, Sprinklr, Website Grader, MeetEdgar, MavSocial, ZohoSocial, Lithium, Social report, SumAll, Social Mention, Cyfe, Sendible, Eclincher, SproutSocial, AgoraPulse, Bit.ly, Everypost, Socedo, Socialoomph, and Fanpage Karma. However, there are also analytics tools for specific SM, for example weepi, Followerwonk, Postchup, Twitonomy and Tweetdeck for Twitter; Komfo, Sociograph.io, LikeAlyzer and AdEspresso for Facebook; Owlmetrics, Storyheap and SocialDrift for Instagram; and ViralWoot and Tailwind for Pinterest. Table 1 presents a literature review of the use of SM analytics in various business sectors. The review was based on the Google Scholar search engine.

3. Methods and Tools

The approach used for the evaluation of the corporate wineries' websites and their Facebook pages can be analyzed in three phases:

Table 1. Social media analytics tools review

Social Media Analytics Tools	Paper	Sector
Facebook: Sociograph.io	[33]	Traded firms
	[34]	Radiology
LikeAlyzer	[35]	Hotel
	[36]	Tourism
Pinterest: ViralWoot	[37]	Academic libraries
	[38]	Data management & analytics of big data
Twitter: Followerwonk	[39]	Most influential emergency physicians
	[40]	Hospital pharmacy
Twitonomy	[41]	Communication tool in congress
	[42]	SM marketing
All SM Hootsuite	[43]	SM marketing
Google Analytics	[44]	Usability of e-commerce sites
	[45]	Libraries
	[46]	Tourism
Talkwalker	[47]	Business and market research
	[48]	Fashion
Brand24	[49]	Start-up companies
Quintly	[50]	Tourism
BuzzSumo	[51]	Health
	[52]	SM marketing in higher education institutions
Website Grader	[53]	Website evaluation tools
	[54]	Universities' websites
MavSocial	[55]	Tourism
SumAll	[56]	Engaging library users
Social Mention	[48]	Fashion
	[57]	SM networking
Sendible	[58]	SM management tools
	[59]	SM marketing
AgoraPulse	[60]	Fitness
Fanpage Karma	[61]	Brand communication
	[62]	SM marketing
	[63]	Wine tourism

- Determination of the sample of winery businesses.
- Selection of appropriate SM analytics tools for assessing websites and Facebook pages.
- Data collection and analysis.

The survey was conducted in 2019 and the sample for Greek and German wineries was identified through the Google search engine. For the Greek market the keywords used were: "winery", "retail sale winery", "Greek winery", "wine", "wine trade", "winemaking", "wine in Greece". For the German market the keywords used were: "Wein", "Weingut", "Deutscher Wein", "Wein Online-Shop", "German wineries", and "German wines". 105 Greek and 100 German wineries were identified, which have a corporate website and/or a Facebook page.

The selection of the appropriate SM analytics tools was based on the aforementioned literature review (Table 1) and free online availability of tools. Website Grader (website.grader.com) was selected for corporate websites and the LikeAlyzer (likealyzer.com) was selected for Facebook pages [35, 54, 64-67].

Website Grader is a comparative online tool, which evaluates a website based on a standard rating (from 0 to 100) in relation to other sites. The rating indicates whether the website works well for the company and helps to improve website quality. More specifically, the features that Website Grader evaluates are:

- *Performance:* evaluates the overall appearance of a website, the size of web pages and its speed. Performance is probably one of the most important features, as it is associated with the visitor's first impression and experience of the website. This can help increase the visibility and impact of the business. The value of this feature ranges from 0 to 30. The *performance* feature includes the following characteristics: page size, page requests, page speed, browser cashing, page redirects, compression, and render blocking.
- *Mobile readiness:* refers to the capacity to view the website on any mobile device, for example smart phones, in terms of response and viewer settings. As a large portion of the purchasing community visits pages and makes extensive use of the internet and SM from their mobile device, the flexibility of a website to work efficiently on these devices is important. Its value ranges from 0 to 30. Responsiveness and viewport are the characteristics included in this feature.
- *Search Engine Optimization (SEO):* relates to the ranking of the website by search engine users. The short title of a website, short descriptions and keywords can help improve the rankings. Its value ranges from 0 to 30. The *SEO* feature includes the following characteristics: page titles, sitemap, headings, and meta description.
- *Security:* checks the existence of a security certificate, such as a Secure Sockets Layer (SSL), for ensuring security and authentication of information submission. The value of this feature ranges from 0 to 10.

The LikeAlyzer tool was selected for the evaluation of wineries' Facebook pages. It is a popular analysis tool [66], which provides monitoring and feedback on an organization's Facebook presence and impact, allowing the comparison and exploration of Facebook potential. More specifically, it assesses any Facebook page, provides suggestions for improvement and compares it with other Facebook pages. The features that the LikeAlyzer evaluates are the following:

- *Frontpage:* refers to the visitor's first impression of the page. This includes information about the business, for example, about, username, profile picture and cover photo, and call to action.
- *About:* regards more detailed business information, such as milestones, contact information, location and products.
- *Activity:* concerns the activity of the page and how often it occurs. It evaluates the number of pages liked, native

Facebook videos, the number of posts per day, and the average post length.

- *Response:* refers to the response to visitors, either through comments on posts or through personal messages. It evaluates whether visitors can post.
- *Engagement:* evaluates visitors' activity on the page, and some features that can only be accessed by the page developer. The number of people who have interacted with the page and the total page likes are measured.

It is worth noting that LikeAlyzer does not provide an overall score, but each feature is graded from 0 to 100%.

4. Results

4.1. The Greek Case

Europe is a leader in global wine exports. In Greece the importance of the winemaking sector is reflected in its potential for promoting traditional products abroad [18]. As one of Greece's main export products, wine comprises more than 50% of the total export value of the Greek beverage industry. Fragmentation of domestic production and the large number of businesses that are exclusively active in winemaking, mainly small and medium sized companies, are the main characteristics of the sector [3]. There are also large wineries with extensive distribution networks throughout Greece, which are responsible for a significant proportion of wine production. Moreover, although the Greek economic crisis has resulted in a decrease in demand, recently it has generated an increased interest in non-bottled wines [68].

Of the 105 Greek wineries in the sample, 86% have a corporate website for promoting and advertising their products/services, 72% support a Facebook page, 37% use Instagram, 38% use Twitter, 17% use YouTube, and 10% use Pinterest. This leads to the conclusion that Greek wineries use and trust SM, and especially Facebook, Instagram and Twitter, for marketing activities [3]. It is evident that Greek wineries have turned to the digital environment for success in the market. The wine industry has been among the slowest to adopt SM [69]. As regards the year the winery established its internet presence, 24% established between 1999 and 2005, 22% between 2006 and 2011, while 54% between 2012 and 2018. Turning to the languages supported by corporate websites, 82% are both in Greek and English and 18% only in Greek, indicating that the majority are interested in export activities.

Greek winery websites were evaluated with Website Grader. The overall mean score was 66.42%. The overall mean score for each of the Website Grader features is: (a) *performance:* 19.45/30; (b) *mobile readiness:* 23.62/30; (c) *SEO:* 17.06/30 and (d) *security:* 6.12/10. Figure 1 shows the scores of *performance* characteristics of a Greek winery website.

Figure 2 shows the overall recommendations of the Website Grader tool for the improvement of the particular website. The recommendations refer only to *performance* and *SEO* features, which scored lower than satisfactory values.

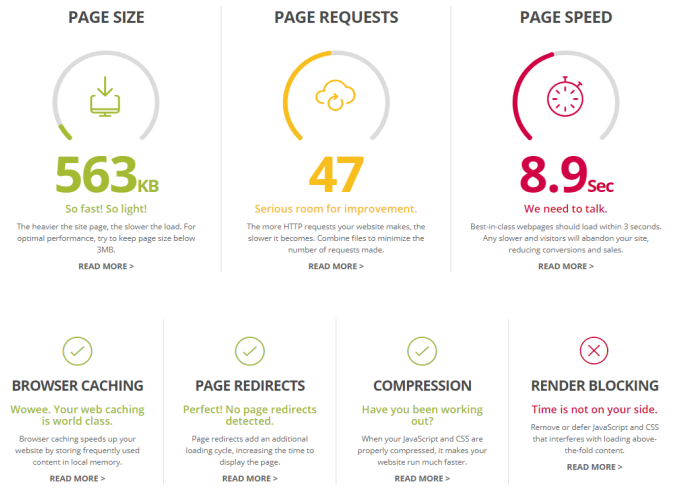


Figure 1. Website Grader Results for performance characteristics of a Greek winery website

Figure 2. Website Grader recommendations for performance and SEO characteristics of a Greek winery website

Next, the results from the LikeAlyzer evaluation are presented. The *Frontpage* feature mean score for the sample is 97% and is analyzed according to the following attributes:

• Profile picture and cover photo: 92.81% of the wineries use a profile picture/cover photo in their profile.

• Username: 93.75% use a shortened page address to promote their business more easily.

• About: 89.06% provide information about the page and what someone can expect when starting to follow the page.

• Call to Action: 100% use "Call to Action" buttons (e.g. "Message Now", "Call Now", "Sign Up Now", "Shop Now") to support interaction with users and promote the business' main objectives.

The *About* feature mean score is 72% and is analyzed according to the following attributes:

• Milestones: only 3.13% use milestones to tell the story of their business.

• Contact information: 64.06% of all contact information (telephone, email, website) is provided on the Facebook page. The percentage is quite low compared to the usefulness and convenience of this information for users. In addition, the lack of

such information has a negative impact on the reliability and safety of the page.

• Location: 70.31% provide the location of their headquarters.

• Products: 26.56% promote their products on their page.

The *Activity* feature mean score is 54% and is analyzed according to the following attributes:

• Posts per day: 0.24 are the daily posts made by businesses, which needs improvement to enhance user engagement.

• Average post length: 192.88 characters is the average length of the posts, which can be improved as it has been noted that posts with 40-100 characters generate more user engagement.

• Pages liked: 42.06 is the average number of "Likes" made by the businesses in the sample to other businesses. Increasing this average will positively affect the cross-sector dialogue and interest, as well as the overall engagement.

• Native Facebook videos: 1.56 is the average number of videos published. Given that video is considered the most engaging format on Facebook, this feature needs to be improved.

The *Engagement* feature mean score is 3% and is analyzed according to the following attributes:

• People talking about this: on average 163 users accessed the page or interacted with its content within the past week.

• Total page likes: 5,840 is the average number of "Likes" by users per business. A larger number of likes corresponds to the business' wider reach.

The *Response* feature mean score is 85% and is analyzed according to the following attribute:

• Users can post: 52.38% of businesses offer their users the option of posting on their page.

4.2. The German Case

German businesses have a long history in the wine sector. For the last forty years Germany has been a leader in global wine imports. Most of the Greek wine imports are made by Germany. Despite the maturity of the German wine sector, it is currently undergoing structural changes. In the near future it is expected that expenditure for marketing will increase, as consumer habits change and new marketing approaches and channels (e.g. blogs and electronic auctions) are required on the basis of a strategic path for sustainability [71,72]. The German Wine Institute, which comprises the country's wine industry marketing organization for the generic promotion of German wines domestically and abroad, already uses SM tools, including Twitter, Instagram, YouTube and Facebook [73].

In 2012 German wineries' SM usage was explored in two research studies [6, 74]. In [6], a sample of 321 German wineries was studied to identify winemakers' SM preference. Facebook was found in this study to be German winemakers' favorite platform. In [74] the SM usage of German wineries and its impact on a sample of Facebook and non-Facebook users was assessed. The turnover of purchases of Facebook users was higher than that of non-users as they are interested in sales offers from wineries they support through SM. 87% of the sample purchased wine from a specific winery and 15% purchased wine from this winery on a regular basis.

All 100 German wineries have a corporate website and 53% also have a Facebook page. Concerning the year the winery established its internet presence, 61% established between 1998 and 2005, 15% between 2006 and 2011, and 24% in 2018. As regards the languages supported by corporate websites, 29% are only in German, 71% are available both in German and English, and 7% support a third language.

Data collected from the German wineries indicate that they engage in SM marketing practices to a high degree. Regarding SM usage, 96% of wineries use Facebook, 28% use Instagram, 8% use Twitter, and 1% use YouTube and Pinterest.

German winery websites were evaluated with Website Grader, achieving an overall mean score of 70.58%. Specifically, the scores for each feature are presented as follows: (a) *performance:* scores reached an average of 23.61/30; (b) *mobile readiness*: 23.27/30; (c) *SEO*: 16.79/30; and (d) *security*: 7.65/10.

Next, the results from the LikeAlyzer evaluation are presented. The *Frontpage* feature mean score of the sample is 94.30% and is analyzed according to the following attributes:

• Profile picture and cover photo: 86.79% of the wineries use a profile picture/cover photo in their profile.

• Username: 94.34% use a shortened page address.

• About: 81.13% provide information about the page.

• Call to Action: 100% use "Call to Action" buttons.

The *About* feature mean score is 82.90% and is analyzed according to the following attributes:

• Milestones: only 9.43% use milestones.

• Contact information: 85% of all contact information is provided on the Facebook page. The percentage is not extremely high, offering scope for improvement for the German wineries.

• Location: 90.56% provide the location.

• Products: 9.43% promote their product, showing low exploitation of Facebook for promotion.

The *Activity* feature mean score is 42.50% and is analyzed according to the following attributes:

• Posts per day: 0.09 are the daily posts.

• Average post length: 186.74 characters is the average text size of the posts.

• Pages liked: 28.91 is the average number of "Likes".

• Native Facebook videos: 0.91 is the average number of videos published.

The *Engagement* feature mean score is 3% and is analyzed according to the following attributes:

• People talking about this: on average 70 people have interacted with a page.

• Total page likes: 4,515 is the average "Likes" made by users per business.

The *Response* feature mean score is 90.56% and is analyzed according to the following attribute:

• Users can post: 90.56% of businesses allow users to post on their page.

In the following figures an example of a German winery assessed with LikeAlyzer is presented. Figure 3 presents the overall impression of the Facebook page, and Figures 4 and 5 show the assessment of Frontpage and Activity features respectively.

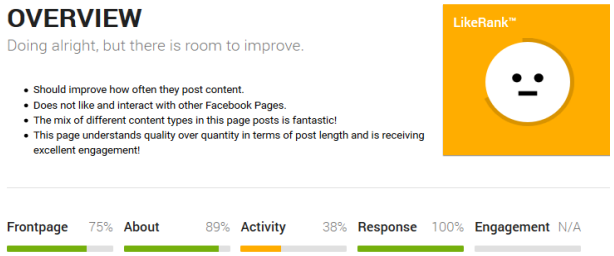


Figure 3. Overview of the Facebook page of a German winery assessed by LikeAlyzer

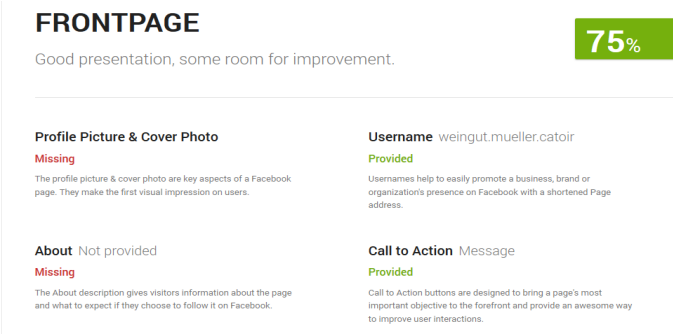


Figure 4. Frontpage feature assessment by LikeAlyzer

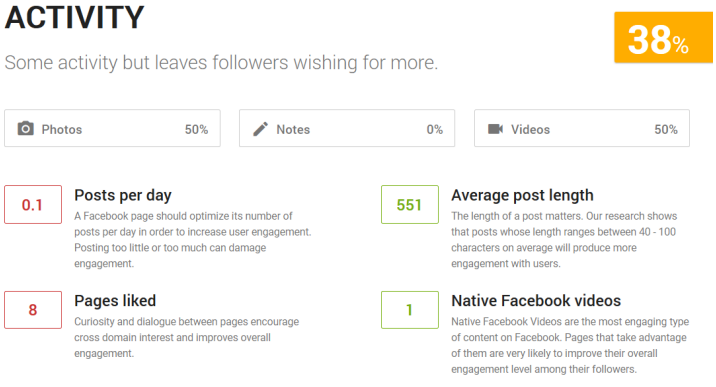


Figure 5. Activity feature assessment by LikeAlyzer

4.3. Discussion

The assessment results of the Greek and German wine industry using Website Grade and LikeAlyzer tools are presented in Table 2. According to the Website Grader evaluation, the *performance* rating for Greece is 19.45 out of 30, and for Germany 23.61. This slight difference in *performance* of German corporate websites shows that they create a better first impression and experience for users. However, Greek winery websites enrich their content with many images and videos in addition to a standard description of their products, which significantly reduces their speed. Regarding *mobile readiness*, Greek and German websites have similar scores, though flexibility and responsiveness for mobile devices have to be further improved. Moreover, according to [4], 60% of users are unlikely to return to a website with accessibility problems from a mobile device, while 40% of users choose to visit a competitor's site. It is therefore essential for winery websites to offer customers an optimized experience for their smart phones and tablets. SEO was 17.06 for Greek wineries and 16.79 for German. Greece seems to be ranked more highly when customers search through a search engine. As regards security, Greek websites scored 6.12, while German websites are ahead with a score of 7.65. Security seems to be moderate for both countries; Greek websites in particular need immediate improvement with the implementation of security protocols such as SSL.

Table 2. Greek and German winery assessment results from Website Grader and LikeAlyzer

Features	**Countries**	
	Greece	**Germany**
Website Grader		
Overall	66.42 %	70.58 %
Performance	19.45/30	23.61/30
Mobile readiness	23.62/30	23.27/30
SEO	17.06/30	16.79/30
Security	6.12/10	7.65/10
LikeAlyzer		
Frontpage	97.00	94.30
About	72.00	82.90
Activity	54.00	42.50
Engagement	3.00	3.00
Response	85.00	90.56

Well-designed corporate websites provoke customers' emotional reactions while browsing [76]. From the study of German and Greek wineries' corporate websites, there is an overall impression that the website design presents some differentiation. In particular, Greek wineries' homepages have rich, impressive and colorful photos with people celebrating food and wine, experiences from product use, and information on the winery's history, showing a better understanding of how to project customers' ideal experience. In contrast, the appearance of German wineries' websites is less meticulous, generating less positive or even negative comments/impressions among users [3]. A German study conducted in 2016 showed that websites with large-sized, high-quality photos of joyful people or inspired moments, such as a panoramic view of the vineyard, can potentially create an emotional impact on users/customers [75]. Customers' emotional influence and engagement can be taken forward when wineries try to tell their own business story, allowing customers to get to know the winemakers better and identifying similar values that connect them.

Our research also provides data on extroversion to foreign customers. Specifically, 83% of Greek wineries and 71% of German wineries provide their websites in at least two languages. This indicates that Greece constitutes a dynamic player in wine exports and endeavors to develop its export activities further, and highlights Germany's position as the largest importer of wine.

The LikeAlyzer evaluation shows that Greek and German winery Facebook pages are in general terms at similar levels. However, both countries have prospects for further improvement.

Regarding the *Frontpage* feature that shapes visitors' first impressions, Greece's score of 97% is higher than Germany's score of approximately 94%. Concerning the *About* feature, German pages provide a satisfactory amount of business information. However, it is worth mentioning that Greek wineries exploit Facebook more effectively than Germany by providing more product information. The *Activity* feature scores are low for both countries although Greek pages have more posts per day, pages liked and native videos. The *Engagement* feature has the same score (3%) for both countries since only Facebook page administrators have access to the specialized data. Finally, Germany scores more highly for *Response*, showing that German wineries allow more user interaction.

5. Conclusions

Enterprises recognize the value of SM analytics in innovation and product development, and customer service operations and strategy. Social media analytics are recently used by business managers and professionals.

SM analytics are easier to use than traditional media analysis and can provide a fast, effective and efficient assessment of corporate websites and Facebook pages. This is verified by the current research. Two freely available SM analytics tools were used to assess the corporate websites and Facebook pages of the winery industry in Greece and Germany. Few studies to date have focused on the implementation of SM analytics in the food and beverage industry. In this research the use of such tools has proved easy and effective, providing specialized data which can offer valuable insights into the wine industry in both countries. The websites of German wineries are of higher quality – from a technical perspective – than those of Greek wineries. However, the latter have a stronger Facebook presence. This can be explained by the overall attitude of Greek internet users and businesses towards SM usage. According to the Digital Economy and Society Index 2019 [77], the percentage of internet users that uses SM is 73%, 57% and 65% for Greece, Germany and the European Union respectively; the percentage of businesses that uses SM is 21%, 16% and 21% for Greece, Germany and European Union respectively.

The key contribution of this paper lies in the use of simple standard analytics to help winery managers and practitioners to plan their SM efforts and revise them regularly, using inexpensive open source tools. A limitation of this paper regards the exclusive focus on Facebook. Assessing the overall SM strategy of wineries requires the study of other SM, for example, Pinterest, Instagram, and Twitter. Such platforms can have additional potential for online visibility and direct sales by linking to corporate websites or electronic shops [75]. Future work will focus on the use of free SM analytics tools for Twitter and Instagram for retrieving data that can be used to build an overall assessment of the wine industry, as well as other food and beverage sectors. It is expected that this research will give researchers and practitioners an assessment of the wine industry and motivate them to use SM analytics tools to improve digital marketing strategies in the wine and other industries, for example, tourism, health and entertainment. This in turn can help to create a roadmap for exploitation of SM benefits.

References

[1] L. Weber, Marketing to the social web: How digital customer communities build your business. John Wiley & Sons, 2009.

[2] M. Armara, Digital Marketing: the effectiveness of advertising through Social Media. Master Thesis. University of Piraeus, Department of Industrial Management and Technology, 2015.

[3] C. Costopoulou, M. Ntaliani, F. Ntalianis, "Social Media in Greek and German Wineries' Websites" in 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, 2018. https://ieeexplore.ieee.org/abstract/document/8633596

[4] D. Zeng, H. Chen, R. Lusch, S. H. Li, "Social media analytics and intelligence" IEEE Intelligent Systems, 25(6), 13-16, 2010.

[5] I. Fabiatou, "Greek and German Wineries' Websites", B.Sc. Thesis, Supervisor: C. Costopoulou, Dept. of Agricultural Economics and Development, Agricultural University of Athens, 2018.

[6] G. Szolnoki, H. Dieter, "Online, face-to-face and telephone surveys-Comparing different sampling methods in wine consumer research" Wine Economics and Policy, 2(2), 57-66, 2013. https://doi.org/10.1016/j.wep.2013.10.001

[7] OIV, Organization Internationale de la Vigne et du Vin, State of the Vitiviniculture World Market, State of the sector in 2018. http://www.oiv.int/public/medias/6679/en-oiv-state-of-the-vitiviniculture-world-market-2019.pdf

[8] A. Morrison, R. Rabellotti, "Gradual catch up and enduring leadership in the global wine industry" Research Policy, 46(2), 417-430, 2017. https://doi.org/10.1016/j.respol.2016.09.007

[9] L. Thach, T. Lease, M. Barton, "Exploring the impact of social media practices on wine sales in US wineries" Journal of Direct, Data and Digital Marketing Practice, 17(4), 272-283, 2016. https://doi.org/10.1057/dddmp.2016.5

[10] D. Kolb, L. Thach, "Analyzing German winery adoption of Web 2.0 and social media" Journal of wine research, 27(3), 226-241, 2016. https://doi.org/10.1080/09571264.2016.1190324

[11] N. A. Viana "Digital wine marketing: Social media marketing for the wine industry" in BIO Web of Conferences 7, 03011, EDP Sciences, 2016. https://doi.org/10.1051/bioconf/20160703011

[12] G. Szolnoki, D. Taits, M. Nagel, A. Fortunato, "Using social media in the wine business: An exploratory study from Germany" International Journal of Wine Business Research, 26(2), 80-96, 2014. https://doi.org/10.1108/IJWBR-09-2013-0031

[13] D.A. Laverie, W.F. Humphrey, N. Velikova, T.H. Dodd, J.B. Wilcox "Building wine brand communities with the use of social media: a conceptual model" in 6th AWBR International Conference, Bordeaux, France, 2011.

[14] R. Dolan, S. Goodman, C. Habel "How (and why) are wineries using Facebook for marketing?" Grapegrower & Winemaker, 595, 85-86, 2013.

[15] G. Szolnoki, R. Dolan, S. Forbes, L. Thach, S. Goodman, "Using social media for consumer interaction: An international comparison of winery adoption and activity" Wine Economics and Policy, 7(2), 109-119, 2018. https://doi.org/10.1016/j.wep.2018.07.001

[16] T. Pucci, E. Casprini, C. Nosi, L. Zanni, L. "Does social media usage affect online purchasing intention for wine? The moderating role of subjective and objective knowledge" British Food Journal, 121(2), 275-288, 2019. https://doi.org/10.1108/BFJ-06-2018-0400

[17] A. Galati, S. Tinervia, A. Tulone, M. Crescimanno, "Drivers affecting the adoption and effectiveness of social media investments: The Italian wine industry case" International Journal of Wine Business Research, 2019. https://doi.org/10.1108/IJWBR-04-2018-0016

[18] I. Eleftheriadis, "The competitiveness of Greek exports of wine to the markets of Germany and China" Master Thesis, School of Agriculture, Aristotle University of Thessaloniki, 2012.

[19] T. Ahlqvist, A. Bäck, M. Halonen, S. Heinonen, Social media roadmaps, Helsinki: Edita Prima Oy, 2008.

[20] R. Jones, "Social Media Marketing 101, Part 1", 2009. http://searchenginewatch.com/article/2064413/Social-Media-Marketing-101-Part-1

[21] W. D. Evans, "Social marketing campaigns and children's media use" The Future of Children, 18(1), 181-203, 2008. https://doi.org/10.1353/foc.0.0009

[22] D. Nazan, I. Menevi, R. Eyyam, "What is the motivation for using Facebook?." Procedia-Social and Behavioral Sciences 15, 2642-2646, 2011. https://doi.org/10.1016/j.sbspro.2011.04.162

[23] J. Zhang, "Social Media and distance education", 2010. http://deoracle.org/online-pedagogy/emerging-technologies-social-media-and-distance-education.html.

[24] A. M. Kaplan, M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media" Business Horizons, 53(1), 59-68, 2010. https://doi.org/10.1016%2Fj.bushor.2009.09.003

[25] S. Tsartsaflis, E. Kletsios, "How social media influence the branding of mass catering and entertainment companies in a period of economic crisis", B.Sc. Thesis, Alexander Technological Educational Institute of Thessaloniki, Department of Marketing, 2013.

[26] Y. K. Dwivedi, K.K. Kapoor, H. Chen, "Social media marketing and advertising" The Marketing Review 15(3), 289-309, 2015. https://doi.org/10.1362/146934715X14441363377999

[27] S. Falkow, "Social media strategy" E+ White paper 2400 (2009).

[28] R. Felix, P. A. Rauschnabel, C. Hinsch, "Elements of strategic social media marketing: A holistic framework" Journal of Business Research, 70, 118-126, 2017. https://doi.org/10.1016%2Fj.jbusres.2016.05.001
[29] T. L. Tuten, M. R. Solomon, Social Media Marketing, Sage, 2017.
[30] G. Schneider, Electronic Commerce, Cengage Learning Editions, 2015.
[31] J. Mahoney, S. Lawson, R. Stone, "What do you Think of the Return of Dunga-Rees?" Social Media Interactions Between Retail Locations and their Customers" in the 32nd annual ACM conference on Human factors in computing systems, ACM, Toronto, Canada 2014. http://dx.doi.org/10.1145/2559206.2581299
[32] B. Batrinca, P. C. Treleaven, "Social media analytics: A survey of techniques, tools and platforms" AI & Society, 30(1), 89-116, 2014. doi:10.1007/s00146-014-0549-4
[33] B. Kaushik, H. Hemani, P.V. Ilavarasan, "Social media usage vs. stock prices: an analysis of Indian firms" Procedia computer science, 122, 323-330, 2017. https://doi.org/10.1016/j.procs.2017.11.376
[34] R. L. Seidel, A. Jalilvand, J. Kunjummen, L. Gilliland, R. Duszak, "Radiologists and social media: do not forget about Facebook" Journal of the American College of Radiology, 15(1), 224-228, 2018. https://doi.org/10.1016/j.jacr.2017.09.013
[35] V. S. Kumbhar, "Hotel Websites Facebook Data Analysis Using Weka" in Second International Conference of Commerce and Management (ICCM) on Development through Transformation: Prospects for Inclusive Growth, Shivaji University, Kolhapur, 2017.
[36] R. McGee, "The Use of Facebook to Promote Tourist Destinations: A Case of the Middle East" Doctoral dissertation, University of New Haven, 2016.
[37] L. Bingle, "Pinterest in Academic Libraries: Social media policy on visual social networks". https://ojs.library.ubc.ca/index.php/seealso/article/view/188985/186518
[38] R. Saravanakumar, C. Nandini, "A survey on the concepts and challenges of big data: Beyond the hype" Advances in Computational Sciences and Technology, 10(5), 875-884, 2017.
[39] J. Riddell, A. Brown, I. Kovic, J. Jauregui, "Who are the most influential emergency physicians on Twitter?" Western Journal of Emergency Medicine, 18(2), 281, 2017. https://dx.doi.org/10.5811%2Fwestjem.2016.11.31299
[40] J. M. Delgado, J. T. De Pourcq, E. M. Boquet, J. M. Sesmero, F.M. Alonso, "Social authorities concerning# hospital pharmacy on twitter", Eur J Hosp Pharm, 2019.
[41] J. Straus, R. Williams, C. Shogan, M. Glassman, "Social media as a communication tool in Congress: Evaluating Senate usage of Twitter in the 113th Congress" in APSA 2014 Annual Meeting Paper, Washington, D.C., 2014.
[42] P. E. Lutu, "Data analytics to support social media marketing: challenges and opportunities" in CONF-IRM, Cape Town, South Africa, 54, 2016.
[43] P. R. Geho, J. Dangelo, "The evolution of social media as a marketing tool for entrepreneurs" The Entrepreneurial Executive, 17, 61, 2012.
[44] L. Hasan, A. Morris, S. Probets, "Using Google Analytics to evaluate the usability of e-commerce sites" in International Conference on Human Centered Design, Springer, Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-642-02806-9_81
[45] W. Fang, "Using Google Analytics for improving library website content and design: A case study" Library Philosophy and Practice, 1-17, 2007. https://doi.org/doi:10.7282/T3MK6B6N
[46] B. Plaza, "Google Analytics for measuring website performance" Tourism Management, 32(3), 477-481, 2011. https://doi.org/10.1016/j.tourman.2010.03.015
[47] S. A. Parnsup, D. Tancharoen, N. Phaphoom, A. Kongthon, "An Adoption of Social Media Monitoring Tools for Business and Market Research" TNI Journal of Business Administration and Languages, 4(1), 1-4, 2016.
[48] S. Ranfagni, M. Faraoni, "Be Social and be Tuned: Evaluate your Brands in Online Communities" in 2017 Global Fashion Management Conference at Vienna, Austria, 2017. https://doi.org/10.15444/GFMC2017.03.06.01
[49] E. Gheribi, A. Kotnicka, A. Klemens, M. Klepacz, R. Kwiatkowski "Analysis of Financing Sources for start-up Companies" Management and Education, 14(1), 22-27, 2018.
[50] S. Asongu, N. Odhiambo, "Tourism and social media in the world: An empirical investigation" African Governance and Development Institute, Working Paper, 2018.
[51] P. M. Waszak, W. Kasprzycka-Waszak, A. Kubanek, "The spread of medical fake news in social media–the pilot quantitative study" Health Policy and Technology, 7(2), 115-118, 2018. https://doi.org/10.1016/j.hlpt.2018.03.002
[52] K. Aman, N. Hussin, "The Effectiveness of Social Media Marketing in Higher Education Institution" International Journal of Academic Research in Business and Social Sciences, 8(9), 827-834. http://dx.doi.org/10.6007/IJARBSS/v8-i9/4657
[53] S. S. Khandare, S. Gawade, V. Turkar, "Survey on website evaluation tools" in 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE), IEEE, Bhopal, India, 2017.
[54] S. Kaur, K. Kaur, P. Kaur, "An empirical performance evaluation of Universities Website" International Journal of Computer Applications, 146(15), 10-16, 2016. https://doi.org/10.5120%2Fijca2016910922
[55] R. Ribeiroae, T.Florentino, "Digital Transformation in Tourism: a high level analysis of the impact that social networks and mass collaboration concept is having at tourism service providers" in 7th International Conference on Cinema and Tourism-ICCT, Porto, 2016.
[56] E. Ramsey, A. Vecchione, "Engaging library users through a social media strategy" Journal of Library Innovation, 5(2), 71-82, 2014.
[57] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media" Business horizons, 54(3), 241-251, 2011. https://doi.org/10.1016/j.bushor.2011.01.005
[58] S. Thapa, G. Skinner, "Review of Social Media Management Tools and Related Literature" International Journal of Advanced Trends in Computer Applications, 2(1), 89-99, 2015.
[59] I. Gasparini, "Social Media Marketing: Facebook e le Banche", Ph.D Thesis, Università di Modena e Reggio Emilia, 2015.
[60] J. García-Fernández, A. Elasri-Eijaberi, F. Pérez-Tur, X. M. Triadó-Ivern, L. Herrera-Torres, P. Aparicio-Chueca, "Social networks in fitness centres: the impact of fan engagement on annual turnover" Journal of Physical Education and Sport, 17(3), 1068-1077, 2017.
[61] A. Huertas, E. Marine-Roig, "Destination brand communication through the social media: What contents trigger most reactions of users?" Information and Communication Technologies in Tourism, 295-308, Springer, 2015.
[62] S. A. Movsisyan, "Social media marketing strategy of Yerevan brandy company", Annals of Agrarian Science, 14(3), 243-248, 2016. https://doi.org/10.1016/j.aasci.2016.08.010
[63] M.M. Rodriguez-Fernandez, E. Sanchez-Amboage, V.A. Martinez-Fernandez, "The emergent nature of wine tourism in Ecuador and the role of the Social medium Facebook in optimising its positioning" Revista ESPACIOS, 38(14), 24-37, 2017.
[64] S. Kaur, K. Kaur, P. Kaur, "Analysis of website usability evaluation methods" in 3rd International Conference on Computing for Sustainable Global Development (IEEE), New Delhi, India, 2016.
[65] B. F. Canziani, D. H. Welsh, "Website quality for SME wineries: Measurement insights", Journal of Hospitality and Tourism Technology, 7(3), 266-280, 2016. http://dx.doi.org/10.1108/JHTT-02-2016-0009
[66] M. M. Rodríguez-Fernández, E. Sánchez-Amboage, C. Rodríguez-Vázquez, M. D. Mahauad-Burneo, "Galician Spas in Facebook, Media and Metamedia Management", Advances in Intelligent Systems and Computing, 503, Springer, 2017. https://doi.org/10.1007/978-3-319-46068-0_43
[67] M. Teijeiro-Álvarez, C. Rodríguez-Vázquez, F. Blázquez-Lozano, "The Importance of Social Capital in Higher Education. A Study of The Facebook Fan Pages", New Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing, 445, 2016. https://doi.org/10.1007/978-3-319-31307-8_48
[68] S. Tsitos, "Empirical investigation of Greek winemaking enterprises in international markets", Master Thesis. Department of Economics, University of Piraeus, 2016.
[69] O. Notta, A. Vlachvei, "Web site utilisation in SME business strategy: The case of Greek wine SMEs" in 6th International Business and Social Sciences Research Conference, Dubai, UAE, 2013.
[70] M. Machill, C. Neuberger, F. Schindler, "Transparency on the Net: functions and deficiencies of Internet search engines", Info 5(1), 52-74, 2003.
[71] M. Dressler, "The German wine market: A comprehensive strategic and economic analysis" Beverages, 4(4), 92, 2018. https://doi.org/10.3390%2Fbeverages4040092
[72] A. G. Goncharuk, "Wine business performance benchmarking: A comparison of German and Ukrainian wineries" Benchmarking: An International Journal, 25(6), 1864-1882, 2018. https://doi.org/10.1108%2Fbij-06-2017-0131
[73] DWI, Deutsches Weininstitut/ The German Wine Institute. www.germanwines.de, 2019
[74] C. Hoffmann, G.Szolnoki, L. Thach, "Cross-Cultural Comparison of Social Media Usage in the Wine Industry: Differences between the United States and Germany" Successful Social Media and Ecommerce Strategies in the Wine Industry. Palgrave Macmillan, New York, 154-166, 2016.
[75] P. Merdian, E. Rüger-Muck, G. Raab, "Emotional impact of wine selling websites: An investigation of the online perception of wine web stores" in 9th Academy of Wine Business Research Conference, Adelaide, Australia, 2016.
[76] S. Wilhelm, "Erfolgsfaktor Online-Handel: Tipps für die E-Commerce-Praxis - von der Gründung bis zur Expansion, Frankfurt am Main, Deutscher Fachverlag, 2012.
[77] DESI, "The Digital Economy and Society Index (DESI) for 2019", European Commission, Brussels, 2019 https://ec.europa.eu/digital-single-market/en/de

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, 387-394 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Formal Ontology-based Framework Towards an Intelligent Temporal Scheduler

Patience Usoro Usip*, Udoinyang Godwin Inyang, Daniel Ekpenyong Asuquo, Enobong Umoren

University of Uyo, Computer Science Department, Faculty of Science, PMB 1017, 520003, Uyo, Nigeria

ARTICLE INFO

Article history:
Received:04 February, 2019
Accepted:05 April, 2019
Online: 26 April, 2019

Keywords:
Time Complexities and Scheduling
Formal Representations
Qualitative Reasoning
Timetabling Process
Allen's Interval Logic

ABSTRACT

Time scheduling as seen in timetabling processes with few and/or competing resources has exposed complex interoperable time scheduling. Attempts to resolving these time scheduling processes has been undertaken, using several classical methods, with difficulty due to inherent complexities, constraints and conflicting issues. The use of ontology-based approaches to resolve time complexity is recently adopted due to its ease in interoperability and reuse of data. The probability weighted ontology provides the various types of complexity as a requirement for the complexity reduction. To determine the optimality of the resulting timetable required the evaluation of its criteria using the analytical hierarchy process. The need for a formal representation to explicate the intelligent behaviour of the ontology-based framework of the temporal scheduler arises. Hence, this work is aimed at providing that formal representation of the logical part of the ontology. The basic rules handling the constraints in the timetabling process are outlined with the corresponding formal representations of the interval-based logic using first-order logic. The semantic model of the temporal scheduler is further described following Guarino's formal ontology model. The unified modeling language (UML) design of a system framework prototype that adopts the formal model is also given. Through the formal ontology-based framework, all constraints that will give optimality are explicated and incorporated into the allocation reasoned, which results in an optimal formal ontology-based model. This will ensure reliability, ease of use and the likelihood of re-usability of the resulting timetable.

1. Introduction

This paper is an extension of work originally presented in a conference [1]. Scheduling is a constrained allocation of resources to objects in space and time, aimed at minimizing the total cost of a set of used resources [2]. This involves creating schedules by temporarily assigning activities to resources considering one or more objectives and some constraints. Project scheduling in [3] with emphasis on testing the dependencies of ordered activities using Pearson's Correlation Coefficient (PCC) is not centered on temporal scheduling which is the focus of this paper. Basically, PCC is used in [3] to obtain weights for activities and then classify them as dependent or independent so as to avoid frequent usage of same resources multiple times thereby reducing expenses and project duration. Temporal reasoning is a suitable methodology to synchronize operations, make allocations for resources, determine sequences of operations, and evaluate average earliness or tardiness of a schedule. Time, which is modeled by quantitative values and constraints, is used to represent every schedule [4, 5]. Formalizing time scheduling process is a complex task that involves different reasoning strategies, structures and elements from multiple languages. Several attempts of formalizing time schedules using classical techniques exist in the literature (simulated annealing [6], genetic algorithm [7], Tabu search [8, 9], case-based reasoning [10], and graph coloring approach [11, 12, 13, 14]. However, these approaches lack formalization and are not simple to implement needed semantics during problem solving. Moreso, their solutions, though functional, may be error prone when considering new sets of constraints. Ontology-based approaches are emerging to resolving such complexities due to their ease in interoperability and reuse of data. The identified time complexities in [1] can be resolved in the allocation reasoner

*Daniel Ekpenyong Asuquo: Department of Computer Science, University of Uyo, Nigeria; +2348035680549; danielasuquo@uniuyo.edu.ng

(AR) by implementing the represented formal notations that describe the embedded semantics in the given rules applicable for inference in the weighted ontology framework. In this work, the heuristics are formalized through ontology [15], by explicitly incorporating all constraints that will give valid answers to the competency questions in the AR and to enable their execution by ontology reasons.

The focus of this work is to construct the basic rules handling the constraints in the timetabling process with the corresponding formal representations of the interval-based logic [16, 17] using first-order logic. Furthermore, the semantic model of the temporal scheduler is further described following Guarino's formal ontology model.

The rest of the paper is organized as follows. Section 2 reviews related works and the use of ontology in scheduling problems including publishing meetings and event, resource allocation and timetabling while section 3 presents a formalization of the complexity reduction rules using first order logic and Allen's interval relations for optimal time scheduling process. Section 4 presents the semantic model of the formal time scheduling ontology with different components for obtaining optimality of time, with a mathematical description of the conceptual model. In section 5, unified modeling language (UML) diagrams are used for structural and behavioral description of the time scheduling ontology (TSONT) while section 6 concludes the paper.

2. Related works

Ontology is a model building tool for specifying the scope of an application domain as well as it entities with the attributes or desired system features. Formally, it represents knowledge as a hierarchy of concepts within a domain and uses shared vocabulary to denote the concepts' types, properties, functionalities and interrelationships. Ontology has been used as a structural tool to organize information and represent knowledge about the world in many areas, such as Artificial Intelligence (AI), Semantic Webs, Information Architecture, System Engineering, Library Science and so on. The formalization of rules enables the complexities found in the production of timetable to be managed and optimized. Ontology is capable of uniquely combining inherent techniques such as semantic search, ontology matching and ontology mapping to reduce the timescale and cost complexity in data and resolve conflicts that always occur. A typical complexity reduction case between two departmental ontologies is shown using the Anchor-PROMPT mapping technique [18]. The representation of the complexity types using the probability weighted ontology (PWO) [19] and the evaluation of its gold standard criteria using analytic hierarchy process (AHP) model [1] provide users with the window to reschedule any time based on the level of identified complexity to attain optimality. The various users' windows for rescheduling are formally represented in the following sections.

Anchor-PROMPT is a conventional ontology mapper that automatically finds semantically similar features (described by terms) between two ontologies, taking a set of pairs of related terms from the source ontologies as input. Commonalities in two different departmental ontologies are mapped together to expose the causes of conflict or clash and resolve them. Consider the mapping of two departmental ontologies shown in figure 1 with Computer Science department in Faculty of Science as an anchor department, owning a certain course, CSC211. This course, CSC 211, is also being offered by students of the departments of Mathematics, Statistics and Physics in the same faculty, Vocational Education in faculty of Education, and Geo-Informatics in faculty of Environmental Studies. Ontology mapping involves linking the interactions around the course in the anchor department with that of the departments it is providing services to. In case of any clash or conflict with other courses offered by the students of any cooperating departments, the Time Scheduler or Allocation Reasoner (AR) along with the interval-based reasoning formally represented in this paper will re-schedule either of the courses concerned to resolve the conflict.

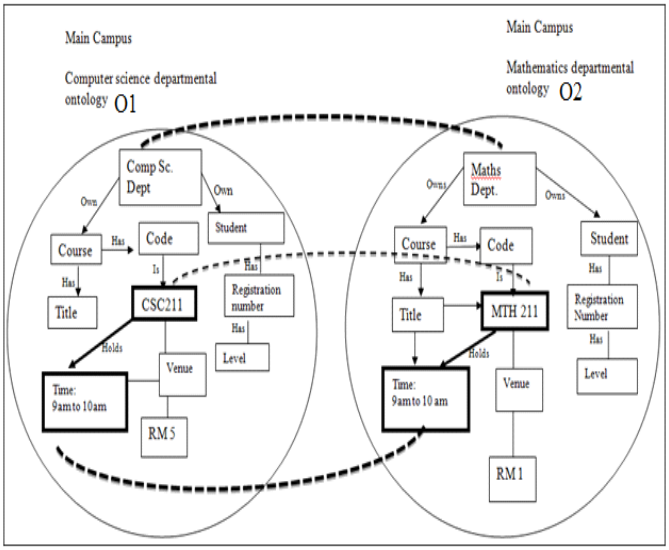


Figure 1. Mapping of two departmental ontologies

As shown in Figure 1, the mapping of ontology O_1 and ontology O_2 by Anchor- PROMPT involves accepting pairs of related terms called anchors, defined by the user or automatically identified by lexical and syntactic matching, as input, from the source graph (ontology); and traversing the path between the anchors in the corresponding ontology. The link between the classes (nodes) defined by the hierarchical relations forms a path. The Anchor-PROMPT then compares the terms along these paths to find similar terms frequently appearing in similar positions on similar paths. In a nutshell, once the sets of similar terms are generated, e.g. {CSC211, MTH211}, scheduled for the same time $\{t_1, t_1\}$ and in the same venue $\{v_1, v_1\}$ which is also scheduled, the AR will be notified, and with the help of the interval-based reasoning, the AR will automatically reschedule either CSC211 or MTH211 for another time and venue respectfully giving priority to CSC211 on account that the high number of cooperating departments produces more complexities, as identified in PWO [1].The identified time complexities as resolved in the AR are expressively represented using formal notations to describe the embedded semantics in the given rules applicable for inference in the weighted ontology framework [1].

2.1 Ontology in scheduling and timetabling

Ontology has been applied in several scheduling task. The ontological framework in [20] specifies a meta-model that provides a vocabulary for formulating application model in a problem domain as well as a set of constraints on what can be expressed. The authors proposed the use of ontology to structure and simplify the process of constructing domain-specific problem-solving tools. Their work focused specifically on the task of scheduling, which ultimately led to the development of OZONE - a toolkit for configuring constraint-based scheduling systems.

With OZONE, it was possible to define domains, constraints as well as reusable and extensible concepts for describing and representing scheduling problem. In [21], a brief review of ontology for scheduling and publishing meetings and events was presented with a framework having a single consistent representation with a hierarchical structure capable of capturing some uncertainty and complexity that occur in real world system. The work attempts to present a balance between logical complexity in the ontology and the content language.

In [22], a generic task ontology for scheduling problems was proposed, expressing that ontology is generic if it is both domain and application independent. The work clearly described the class of scheduling tasks, independently of various ways by which these tasks can be solved. The given ontological framework provided a fairly fine-grained structure needed to build the scheduling system. The given cost related axioms ensure that an optimal solution or result is constructed by subsuming the various preferences in a scheduling task specification. Also, resource availability axiom was used to tackle the conflict between various jobs for the use of same resources depending on their overlapped time range.

An ontology-driven system for solving resources-constrained scheduling problems, orders-oriented and lean mass customization-based manufacturing was presented in [23]. This aimed at getting an easily customizable scheduling system that could be richly exploited by different manufacturers. Within the project Model Based Java Software Development Technology, it assumed the commencement of system development from the ontological conceptualization of the domain, constructing the models of the problem and implementing the system through transformation of models into codes.

In [24], a generic library problem solving method for scheduling application, following the limited coverage provided by different attempts made in the past in developing the libraries of scheduling problem-solvers was proposed. They subscribed to the Task-Method-Domain-Application knowledge modeling framework which provides a structured organization for the different components of the library and approached the work at two different levels – the task level and the method level. At the task level, a generic scheduling task ontology was constructed to formalize the space of scheduling problem. At the method level, a generic problem solving model of scheduling that generalize from the variety of approaches to scheduling problem-solving was constructed. In their work, they subscribed to a top down approach of scheduling construction and analyzed the importance of scheduling research both from the theoretical and engineering perspectives.

The authors in [25] proposed a framework which brought a step toward a generic semi-automatic timetabling tool. They investigated different types of timetabling problems and developed a framework for timetabling application with a central domain ontology tagged “ontology for timetabling”. The work developed a local search component that can deal with the search space based on problem characteristics passed on by the semantic components and was simplified into steps in order to generate good quality timetables in a reasonable amount of time. The first step was towards solving general problem that consists of mapping their data representation onto the ontology. The next step was using the tools to assists in determining the constraints and object of the problem. Custer *et al*. [25] based their timetabling ontology on the general OZONE scheduling ontology in [20] and developed a generic algorithmic component to generate a solution for real world problems.

The formal relationship between different components in a particular universe of discourse is described using formal languages. These descriptions are the basis for any ontology. An explicit description of the domain of discourse is formally represented in the formal ontology model in Figure 2 [26]. This model is adapted in this work for the semantic model of the formal time scheduling ontology in section 4, as it gives an explicit description of the domain of discourse**.**

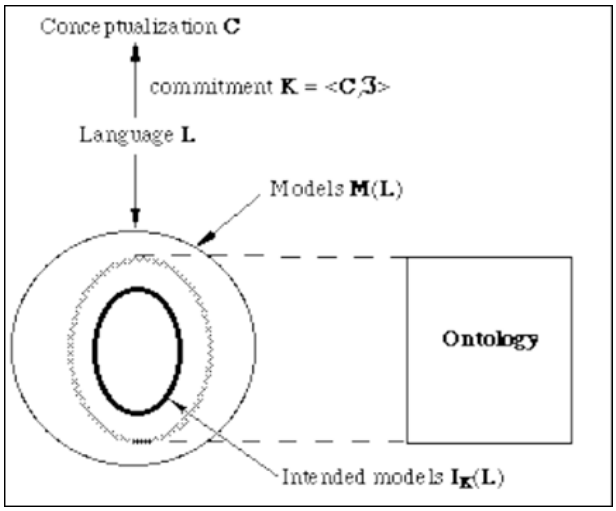


Figure 2. Formal ontology model [26]

3 Complexity reduction rules and axiomatization

The set of competency questions - which are constraints to be satisfied, introduces the time complexities under consideration. Some of the commonly applied competency questions include: Are students from other cooperating departments or faculty offering this course with students of the

host department? Is the said course scheduled for the same time with course(s) in the serviced or cooperating department? Is the said course scheduled to hold in the same venue with course(s) in the other department? Are there students in other levels in the cooperating departments who enroll for this course? Is a particular lecturer scheduled to teach more than one course at the same time? Is the venue for the lecture on different campuses from that of students offering the course? Is more than one course allotted a particular venue at the same time? The Rule set comprises rules that address the constraints /complexities and also proposes axioms for their resolutions through the application of Allen's interval relations. The assumption that a faculty has all its constituent (departments) within the same locality (campus) hold for this work: A sub-set of the rules is given as follows:

R1: IF student of $dept_1$ *offers* $course_1$ AND $dept_1$ NOT *owns* $course_1$ THEN $time_1$ for $course_1$ **NOT *overlaps*** $time_2$ for $course_2$ in $dept_2$.

R2: IF $course_1$ *holds* at $time_1$ and $course_2$ holds at $time_2$ AND $time_1$ ***equals*** $time_2$ AND students offer $course_1$ and $course_2$ THEN $time_1$ and $time_2$ ***overlaps***.

R3 IF $course_1$ *holds* at $time_1$ and $course_2$ holds at $time_2$ AND time1 is ***during*** $time_2$ AND students offer $course_1$ and $course_2$ THEN $time_1$ and $time_2$ ***overlaps***.

R4: IF $course_1$ *holds* at $time_1$ and $course_2$ holds at $time_2$ AND $time_1$ ***starts*** with $time_2$ AND students offer $course_1$ and $course_2$ THEN $time_1$ and time2 ***overlaps***.

R5: IF $course_1$ *holds* at $time_1$ and $course_2$ holds at $time_2$ AND $time_1$ ***finishes*** at $time_2$ AND students offer $course_1$ and $course_2$ THEN $time_1$ and $time_2$ ***overlaps***.

R6: IF $course_1$ and $course_2$ are of the same department, $dept_1$, THEN reschedule $course_2$ to a different time from $course_1$ (**NOT *overlaps***).

R7: IF $course_1$ and $course_2$ are in the same faculty but different departments ($dept_1$ and $dept_2$), THEN $time_1$ and $Time_2$ for $course_1$ and $course_2$ respectivelyshould **NOT *overlaps*** or $course_1$ **NOT *during*** $course_2$.

R8: IF $course_1$ and $course_2$ are not in the same faculty, THEN time for $course_1$ and time for $course_2$**NOT *overlaps*** and **NOT *meets***.

R9: IF $time_1$ for $course_1$ and $time_2$ for $course_2$ ***meet***, THEN $time_2$ for $course_2$ or $time_1$ for $course_1$ will be re-scheduled for next available time (**NOT *meets***)

R10: IF time for $course_1$ and $course_2$ are ***equal*** AND venue for $course_1$ is the same venue as venue for $course_2$, THEN $course_2$ should be re-scheduled for the next available time or $course_2$ re-scheduled for the next available venue

R11: IF time for $course_1$ and $course_2$ are ***equal*** or ***meet***, AND the lecturer teaching $course_1$ in $campus_1$ is also scheduled to teach $course_2$ in $campus_2$ within the space of a period, THEN $course_2$ should be res-scheduled for another time greater than a period.

R12: IF $course_1$ is fixed to hold at $time_1$ at $campus_1$ AND $course_2$ $time_2$ at $campus_2$ THEN $time_1$ **NOT *meets*** $time_2$.

R13: IF students who previously failed $course_1$ offer $course_1$ again, and are taking lectures at $time_1$ THEN $time_2$, for their current level course, **NOT *overlaps*** $time_1$.

R14: IF $course_1$ in $campus_1$ and $course_2$ in $campus_2$, THEN $time_1$ for $course_1$ and $time_2$ for $course_2$ **NOT *meets*** or **NOT *overlaps.***

R15: IF venue capacity is less than students' enrolment in $course_1$ THEN re-schedule $course_1$ another suitable venue.

Formalizing the complexity reduction rules in the previous section explicitly and expressively using First-Order Logic(FOL) yields the following AX1 to AX15 as axioms. The conventional approach of conceptualizing domain concepts; **lecturer, student, venue and time** with relations specific to the domain under consideration: *owns, has, offers, holds* and *teaches,* results in a non-optimal time scheduling. This therefore calls for the inclusion and utilization of the Allen's interval relations; ***overlaps, meets, equals, during*** and their negations. Axioms AX1 to AX15 are rule instances that form the basis for the AR based on ontology and produces an optimal time-scheduling process.

Let S,V,T,C, L represent students, venue, time, course and levels respectively. Let $S=\{s_1, s_2, \ldots s_n\}$ be the set of students, $V=\{v_1, v_2, \ldots v_n\}$ be the set of venues, $CS=\{cs_1, cs_2, \ldots cs_n\}$ be the set of courses, $T=\{t_1, t_2, \ldots t_n\}$ be the set of time slots , P = {p1, p2, …, p3} be the set of campuses and $L=\{L_1, L_2, \ldots L_n\}$ be the set of students level

AX1: *s, d1, cs1, t1, cs2, d2, t2,*
offers(s,cs1)owns(d1,cs1)offers(s,cs2)
owns(d2,cs2)
cooperates(d1, Host, d2, Serviced).

AX2: *s, cs1, t1, cs2, t2.*
holds(cs1,t1)holds(cs2,t2)
equals(t1,t2)offers(s, cs1) offers(s, cs2)
overlaps(t1, t2).

AX3: *s, cs1, t1, cs2, t2.*
holds(cs1,t1)holds(cs2,t2)
during(t1,t2) offers(s, cs1)
offers(s, cs2)overlaps(t1, t2).

AX4: *s, cs1, t1, cs2, t2.*
holds(cs1,t1)holds(cs2,t2)
starts(t1,t2) offers(s, cs1) offers(s, cs2)
overlaps(t1, t2).

AX5: *S, cs1, t1, cs2, t2.*
holds(cs1,t1)holds(cs2,t2)
finishes(t1,t2)offers(s, cs1) offers(s, cs2)overlaps(t1, t2).

AX6: *cs1, t1, d1, cs2, t2.*
owns(d1,cs1)owns(d1,cs2)
holds(cs1,t1)holds(cs2,t2)equals(t1,t2)
equals(l1, l2)overlaps(t1, t2).

AX7: *cs1, t1, d1, fac1, cs2,t2, d2. fac2.*
owns(d1,fac1,cs1)owns(d2, fac2, cs2)
overlaps(t1,t2) during(holds(cs1, cs2)) .

AX8: *cs1, t1, cs2, t2, t.*
holds(cs1,t1)holds(cs2,t2) finishes(t1,t)
starts(t2,t) meets(t1, t2).

AX9: *cs1, t1, d1, fac1, cs2, t2, d2. fac2.*
owns(d1, fac1, cs1) owns(d2, fac2, cs2)
equals(f1,f2)
meets(t1,t2) overlaps(t1, t2).

AX10: *cs1, t1, v1, cs2, t2, v2.*
holds(cs1,t1)holds(cs2, t2) equals (v1, v2)
equals(t1, t2).

AX11: *cs1, t1, p1, cs2, t2, p2.*
holds(cs1,t1, p1) holds(cs2, t2, p1) equals (t1, t2) equals(p1,p2) teaches(lect,p2) teaches (lect, p2)
overlaps (t1, t2)meets (t1, t2)

AX12: *cs1, t1, p1, cs2, t2, p2.*
holds(cs1,t1,p1) holds(cs2, t2, p2) equals (t1, t2) equals (p1, p2) meets (t1, t2)

AX13: *s, cs1, t1, d1, cs2, t2.*
offers(s,cs1,t1) offers(s, cs2, t2) owns(d1, c1)
equals(t1, t2) equals(l1, l2) overlaps(t1, t2).

AX14: *s, cs1, t1, d1, cs2, t2.*
offers(s,cs1,cs2)
holds(cs1, p1) holds(cs2, p2)
meets(t1,t2)overlaps(t1, t2).

AX15: *s, cs1,v1, t1, nos, v2, vcap.*
offers(s, cs1, v1) holds(cs1, v1, t1) (nos>vcap)
holds(cs1, v2)

As an interpretation, axiom *AX6* can be read thus*: 'For all students, courses and departments, there exist course cs2, such that cs2 is offered by student and department owns student and cs2 is not own by department. It therefore follows that time$_1$ for cs1 should not overlap with time$_2$ for cs2".*

This formalism utilizes *offers and owns as* domain-specific relations. It also explored a reverse relation to *owns* between students and departments. The power of Allen's interval relation, ***overlaps***, is also brought into consideration not forsaking the expressive tools in FOL. The use of the universal and the existential qualifiers gave life to the axiom. Putting all these together in Protégé ontology tool where the departmental timetable ontology has been developed and mapped together with reasoning to make the task of responding to the earlier proposed competency questions an easy one. A general and optimal lecture timetable with clear semantics as described in the formal representation is thus achieved.

4 Semantic modeling of the formal time scheduling ontology

Figure 3 represents the semantic model of the formal time scheduling ontology with different components which are related accordingly to show how an optimality of time can be obtained. The semantic model of the formal time scheduling ontology is adapted from the semantic model by Guarino [26], where, Time Scheduling Ontology (TSONT) is defined as the conceptualization of the domain concepts expressed mathematically in equation (1). Further description of the conceptual model is explicated in equations (2) to (4), where a set of conceptual relations, R is on a structure, S being the model of the language L with vocabulary V.

$$C_{TS} =< D_T, T_C, R > \quad (1)$$

$$<D_T, T_C> = DSR + AIR \quad (2)$$

The resulting model is given as < S, I > w*here,*

$$S = <D_T, R> \quad (3)$$

and the usual interpretation function, I is expressed as $I = V \rightarrow D_T R$. Fixing the intentional Time Schedule by means of a structure such as $< C_{TS}, P>$ where,

$C_{TS} = < D_T, T_C, R>$ and $P:V \rightarrow D_T R$ is an intentional interpretation, hence, $K = < C_{TS}, P >$ is an Ontological commitment for L. That is, L commits to C_{TS} by means of K and C_{TS} is the underlying conceptualization of K. With the commitment,

$$C_{TS} =< D_T, T_C, R, I > \quad (4)$$

where, C_{TS} is the time scheduling conceptualization, D_T is the timetabling domain, T_C *is the* interdepartmental allocation complexities, R is the relations, *DSR* is the Domain Specific Relation, *AIR* is the Allen Interval Relations, I is the Interpretation, P is the Intentional Interpretation, S is the Structure, and K is the Ontological Commitment. The mathematical model in equation (1) is used to give a semantic interpretation that will further explicate the logic behind ontology.

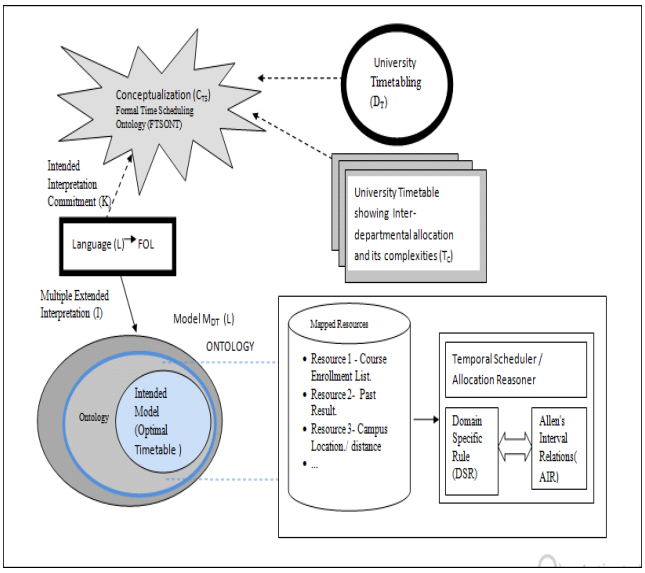


Figure 3. Semantic model of the formal time scheduling Ontology

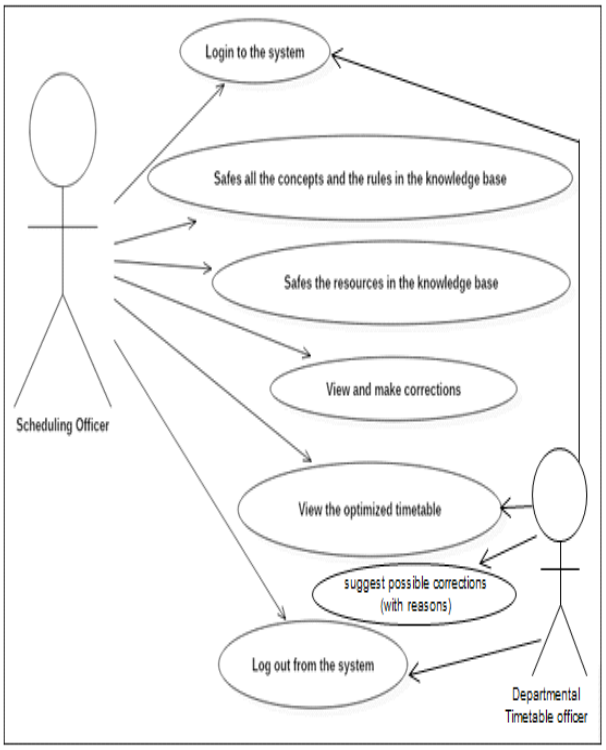


Figure 4. Use case diagram for the TSONT framework

5. System design for TSONT model

5.1 Use case description

The Use Case diagram in figure 4 gives a clear understanding of the developed time scheduling ontology (TSONT) system. The concepts and the resources of the timetable are saved into the knowledge base by the scheduling officer, after a successful log in. The scheduling officer then sends the resources to the allocation reasoner or time scheduler for proper allocation of time. Furthermore, the officer can view the system outcomes, make corrections where necessary, and then review for optimized timetable, and then logs out. The departmental timetable officers can in turn log in to view the optimized timetable, suggest possible corrections for consideration and effect by the scheduling officer, and logs out.

5.2 Class diagram for TSONT framework

Figure 5 represents the class diagram of the TSONT framework, which shows the interactions between classes that constitute the domain of discourse. The student offers courses and the courses are taught by the lecturers. Courses are held in venues, and each venue has an interval of time assigned for it. The venue also has its capacity which informs the number of students that can be allocated to the venue.

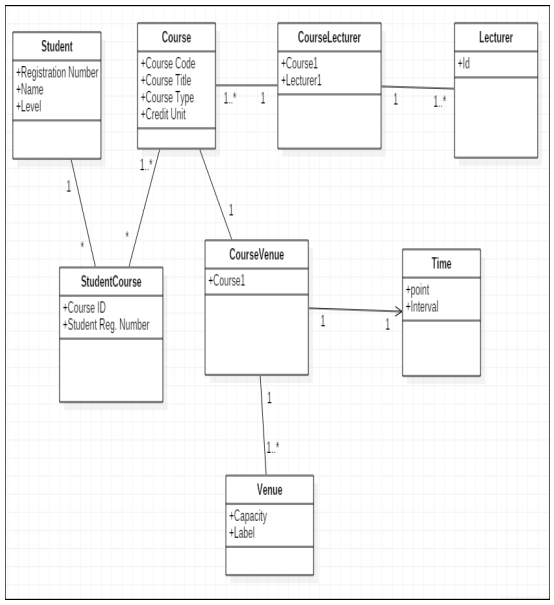


Figure 5. Class diagram for TSONT framework

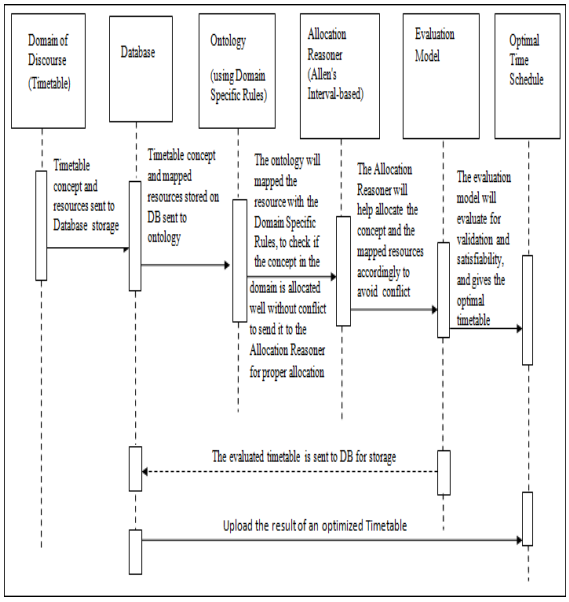


Figure 6. Sequence diagram for the TSONT framework

5.3 Sequence diagram

Figure 6 represents the sequence diagram of how the proposed TSONT system works. The resources used for time schedule and the concepts of the domain of discourse are sent to the database (DB) for storage. The ontology picks the concepts and resources needed for time schedule and sends them to the AR, which helps the rule-based and the interval-based relations, check and allocate the courses, lecturers, venue to the time-slots and to a particular set of students, considering the students who failed or dropped course at a lower level.

The AR sends the results to be evaluated by an ontology evaluation tool, for validation and satisfiability, after which the conclusion of a time complexity free time schedule is sent back to the DB for storage and future reuse. The DB uploads the optimized time schedule for printing and distribution.

6. Conclusion

The university timetabling, an instance of space-time related conflicts or complexities, requires the design of an interval-based temporal ontology to help in its resolution demands full understanding of how the ontology handles the constraints that will answer the stated competency questions. The formal axioms from the stated rules update the constraints in the ontology using the interval based relations. The implementation of the represented formal notations to describe the embedded semantics in the given rules is applicable for inference in the weighted ontology framework. It aims at formalizing the heuristics through ontology, by explicitly incorporating all constraints that will give valid answers to the competency questions in the AR and enabling their analysis and resolution by the ontology-based reasoners. The formal semantic model further explicates the ontology and shows how it is applicable for intelligent complexity reduction and resolution in time scheduling process. The time scheduling ontology based on this formal semantic model will provide an optimal time table schedule.

Conflict of Interest:

The authors declare no conflict of interest

References

[1] P. U. Usip, U. G. Inyang, D. E. Asuquo, E. M. Umoren, "An evaluation model for a probability weighted ontology of temporal complexities", in Proceedings of 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions (CONTISS 2017), Bengaluru, India, 159-165, 2017.

[2] A. Wren, "Scheduling, timetabling and rostering - A special relationship"? International Conference on the Practice and Theory of Automated Timetabling (PATAT 1995), 46-75, 1995.

[3] J. Shafi, A. Waheed and S. Sanober (2016). Dependency Test: Portraying Pearson's Correlation Coefficient Targeting Activities in Project Scheduling. International Journal of Advanced Computer Science and Applications. 7(9). 351-356. 10.14569/IJACSA.2016.070948.

[4] A. Modupe, O. Olusayo, and O. Olantunde, "Development of a University lecture timetable using modified genetic algorithm approach", International Journal of Advanced Research in Computer Science and Software Engineering, (4), 163-168, 2014.

[5] S. Kristiansen and T. R. Stidsen, "A comprehensive study of educational timetabling: A survey", Department of Management Engineering, Technical University of Denmark, DTU Management Engineering Report, (8), 2013.

[6] D. Abramson, "Constructing school timetable using simulated annealing: sequential and parallel algorithm", Management Science, 37(1), 98-113, 1994.

[7] A. Colorni, M. Dorigo, and V. Manniezzo, "A genetic algorithm to solve the timetable problem", Journal of Computational Optimization and Applications, 1, 90-92, 1992.

[8] N. Ibaraki, "Tabu Search for the constraint satisfaction problem as a general problem solver", European Journal of Operational Research, 599-623, 1998.

[9] E. K. Burke and J. Newall, "Enhancing timetable solutions with local search methods", In: Burke, E. K., DeCausmaecker, P. (eds.): Selected papers from the 4th International conference on the Practice and Theory of Automated Timetabling. Lecture Notes in Computer Science, Vol. 2740, Springer-Verlag, Berlin Heidelberg New York, 344- 354, 2003.

[10] Q. Rong, "Case-base reasoning for course timetabling problem", PhD Thesis, University of Nottingham, 2002.

[11] J. P. Rickman, "The design of a course-timetabling system using graph-coloring and artificial intelligence", Honors Program Theses, Paper 15, 2014. https://scholarship.rollins.edu/cgi/viewcontent.cgi?article=1014 HYPERLINK "https://scholarship.rollins.edu/cgi/viewcontent.cgi?article=1014&context=honors"& HYPERLINK "https://scholarship.rollins.edu/cgi/viewcontent.cgi?article=1014&context=honors"context=honors

[12] R. Ganguli and S. Roy, "A study on course timetable scheduling using graph coloring approach", International Journal of Computational and Applied Mathematics, 12(2), 469-485, 2017.

[13] R. K. Bania and P. Duarah, "Exam time table scheduling using graph coloring approach", International Journal of Computer Sciences and Engineering, 6(5), 84-93, 2018.

[14] R. P. Badoni, and D. K. Gupta, "A graph edge colouring approach for school timetabling problems", International Journal of Mathematics in Operational Research, 6(1), 117–132, 2014.

[15] J. Roa, E. Reynares, C. M. Laura and P. D. Villarreal, "Ontology-based heuristics for process behavior: formalizing false positive scenarios", 2017. Available from: https://www.researchgate.net/publication/316743617 [accessed Jan 13 2019].

[16] J. Allen, "Maintaining knowledge about temporal intervals", Communications of the ACM. 26(11), 832-843, 1983.

[17] M. Hemalatha, V. Uma and G. Aghila, "Time ontology with reference event based temporal relations", International Journal of Web and Semantic Technology, 3(1), 23-31, 2012.

[18] N. Noy and M. Musen, "Anchor-PROMPT: using non local context for semantic matching", in Proceedings of the Workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI), Seattle, Washington, 2001.

[19] A. Formica, M. Missikoff, E. Pourabbas and F. Taglino, "Bayesian approach for weighted ontologies and semantic search", in Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (2), 171-178, 2016.

[20] S. Smith and A. Becker, "An ontology for constructing scheduling system", AAAI, Spring, 120 Symposium on Ontology Engineering, Newport Beach, California, 129, 1997.

[21] C. Sayers and R. Letsinger, "An ontology for publishing and scheduling events and the lessons learned in developing it", in Workshop on Ontologies in Agent Systems, 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002), Bologna, Italy, 2002.

[22] D. Rajpathak, E. Motta and R. Roy, "A generic task ontology for scheduling applications", in Proceedings of the International Conference on Artificial Intelligence, Las Vegas, 2001.

[23] J. Sanko and V. Kotkas, "Ontology-driven scheduling system for manufacturing", Baltic Journal of Modern Computing, 4(3), 508 – 522, 2016.

[24] D. G. Rajpathak, E. Motta, Z. Zdrahal and R. Roy, "A generic library of problem solving methods for scheduling applications", IEEE Transactions on Knowledge and Data Engineering, 18(6), 815-828, 2006.

[25] N. Custers, P. De Causmaecker, P. Demeester, and G. Berghe, "Semantic component for timetabling", International Conference on the Practice and Theory of Automated Timetabling (PATAT 2004), 17-33, 2004.

[26] N. Guarino, (Ed.), "Formal ontology", in Proceedings of the First International Conference on information systems (FOIS'98), 46, 3-15, Trento, Italy, 1998.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 388-393 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Proposal of Control Method Considering the Path Switching Time of SDN and Its Evaluation

Kosuke Gotani[1], Hiroyuki Takahira[1], Misumi Hata[1, 2], Luis Guillen[1], Satoru Izumi[*, 1], Toru Abe[1, 3], Takuo Suganuma[1, 3]

[1]*Graduate School of Information Sciences, Tohoku University, 980-8577, Japan*

[2]*Japan Society for the Promotion of Science, 980-8577, Japan*

[3]*Cyberscience Center, Tohoku University, 980-8577, Japan*

ARTICLE INFO

Article history:
Received: 17 June, 2019
Accepted: 31 July, 2019
Online: 16 August, 2019

Keywords:
Software Defined Network
OpenFlow
Path Switching Time

ABSTRACT

Recently, communication demands often change because of the various network services in companies and individuals. Software Defined Networking (SDN) has emerged as a viable control paradigm that allows flexible communication, using OpenFlow as its default standard and enabler. However, when changes happen frequently in SDN networks due to unforeseen reasons -such as a network failure or topology changes- it takes a long time to perform all the operations. For instance, to change a routing path, first new paths must be calculated, then the controller must transmit the commands to the network elements, which has to process those commands. This process can cause a delay, or even disruption, in the communication service. Therefore, this paper proposes a network control method to reduce the time to change a path using OpenFlow.

1. Introduction

1.1. Background and Overview

This paper is an extension of a previous work originally presented at the 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM2018) [1]. Recently, the spread of cloud services has led to the diversification of communication [2], as well as the demand for frequent changes on the communication demand.. Thus, it is necessary to respond promptly to such demands, e.g., unexpected network failure. In particular, a technology that can control the network flexibly is needed.

Therefore, SDN [3,4] and OpenFlow [5,6] have recently attracted attention. SDN can control the network flexibly by software and OpenFlow is one of the most popular southbound implementation protocols of SDN. The OpenFlow architecture consists of an controller and OpenFlow-enabled switches which are centrally managed by the controller. In this OpenFlow network, when a failure or server configuration change occurs, all the paths of the traffic that is passing by through the involved switches from a source to a destination (also known as flow) must be changed.

To change the path of a flow, a new flow entry must be installed in all the switches' memory. A flow entry is represented by a condition part and an action part. The condition part includes specific values on fields within the flow header, such as MAC address or IP address. The action part includes how to route the flow when the values of its header match the values of the condition part.

However, it takes a long time to change various paths of the several flows.Specifically, it takes around 10 ms to change a single path due to the influence of the transmission time of a flow entry form the controller to the switch, and the processing time to add flow entry into the switch [7, 8, 9]. Therefore,in time-sensitive situations, such as disasters, in which several paths must be changes in a short period of time, not all the flow changes will be successful . For example, SONET (Synchronous Optical Network) needs to be recovered within 50 ms, but, it will be challenging to change more than six flows within that time [10]. This time constraint can lead to delays or even service outage. In this way, a technique to switch many flows in a short time is needed.

The main goal of this research is to reduce the delay or service outage in OpenFlow networks, with particular focus on the time needed to change the path of a flow. The proposal consists of a network control method that considers the time needed to change a path in each switch. In this method, the initial assumption is that

[*]Satoru Izumi, 2-1-1, Katahira, Aoba-ku, Sendai, 980-8577, +81-22-217-5453, izumi@ci.cc.tohoku.ac.jp

the network model consists of OpenFlow-enabled switches with different processing time, linked by connections with different bandwidth. Then, the paths with the shortest switching time and the least bandwidth requirements are selected. The basic design was described in [1], and the preliminary simulation experiments presented in [11]. In this paper, the authors evaluate the effectiveness of the proposed method using emulated environments. Based on the obtained results, we confirmed that the proposed method can reduce the switching time for all paths, so that it can realize the reduction of communication delay and network services outages.

1.2. Novelty and Contribution

In this work, a novel approach which chooses communication paths considering not only the processing time to add flow entry of the switch, but also the available bandwidth in the links. Existing approaches mainly focus on the reduction of the calculation time of the paths, or using a backup path in advance. However, they do no deal with the processing time of the switch.

By considering both the processing time and the bandwidth, it can shorten the path switching time and improve the amount of data transferred in a short period. Moreover, tt can also contribute to the reduction of the packet-loss due to the communication interruption during the network failure. This method is effective for services that require lowdelay communication. Furthermore, it is possible to improve the amount of data transfer between two sites in the situation where path change frequently, such as is the case in a disaster situation.

1.3. Paper Organization

In the following Section 2, we explain the related work on fast path switching and the target problem. In Section 3, we describe the proposed control method. The design of the algorithm is shown in Section 4. Then, Section 5 shows the evaluation of the proposed method and its effectiveness. Finally, we conclude this paper in Section 6.

2. Related Work

2.1. Related Work on Fast Path Switching

There are several authors that studied fast path switching in OpenFlow networks, which are mainly categorized in two approaches. The first approach, consists of switching to a backup path registered in advance in a proactive manner [13,14,15], while the second one reactively calculates a new path when a request occurs [10,16,17,18]. These authors select the paths considering a variety of single parameters, e.g., switching time, bandwidth of the link between the controller and the switch, and power consumption. However, all of them largely depend on the available amount of memory (TCAM) to hold the flow table. In this research, we deal with methods for calculating and switching the path reactively.

Cascone et al. [13] proposed a recovery mechanism based on fast reroute of paths in disaster situations. Mohan et al. [14] investigated algorithms to choose a backup path to decrease the number of flow entries. Stephens et al. [15] showed a mechanism to compress the flow table for fast recovery from link failure.

Astaneh et al. [10] proposed a path selection method to reduce the path switching time while considering the communication bandwidth. They reduced the cost of the path switching compared to the traditional approach of using Dijkstra's algorithm. Sharma et al. [16] proposed an in-band based path switching method for failure recovery. Paris et al. [17] showed a dynamic control scheme for network reconfiguration. Malik et al. [18] proposed a method to reduce the path calculation and switching time by partially reusing the route before the change of path. Therefore it can reduce the number of added flow entries.

2.2. Target Problem

The related work presented in the previous section did not take into account environments in which there are heterogeneous switches. Many organizations often build their network by combining switches with various specifications because of budget constraints and differences on the time of purchase [12]. Moreover, the processing time is also different in these switches, causing large delays when there is a heavy traffic on low-processing time switches. Therefore, it is necessary to create a network control method to cope with the difference in processing time when adding flow entries.

3. A Proposal of Control Method Considering Path Switching Time

3.1. Overview

In this section, to solve the problems mentioned above, the authors of this paper propose a network control method that considers the path switching time of each switch. In this method, the network model is designed with different processing time for each switch, and the proposed algorithm selects paths with the shortest time for all paths. Our proposed method considers both the processing time of the switch and the network bandwidth.

The proposed method chooses a path with enough bandwidth, and that goes through high-performance switches to add a flow entry, so that the overall path switching time is reduced when there are several flows. In addition, the proposed method also selects paths sequentially without considering the combination of paths, so that the computational time per path is reduced.

3.2. Path Selection Considering Path Switching Time

This section presents the basic concept of the path-selection mechanism that considers the path switching time as shown in the example depicted in Figure 1. In this example, there are four switches (S_1 to S_4) and the processing time to add a flow entry in each switch is different. Now, there is a communication flow from S_1 to S_4, and for some reason the path must be changed. In this case, there are two candidate paths; p_1 ($s_1 \rightarrow s_2 \rightarrow s_3$) and p_2 ($s_1 \rightarrow s_4 \rightarrow s_3$). The overall path switching time of the flow is the processing time for the switch on the path that has the maximum processing time. For instance, the path switching time of p_1 is 4.0 ms and that of p_2 is 6.0 ms. Thus, the proposed method chooses p_1 which has the shortest switching time.

3.3. Path Selection Considering Available Bandwidth

The path cost, which is used to choose the path with enough communication bandwidth, is calculated by the available bandwidth of the links on the path. Such that when the link with a larger communication bandwidth is used in the path, the path cost

is smaller. Conversely, when the link bandwidth of the path is smaller, the path cost is larger. In the example shown in Figure 2, the path cost of the upper path is smaller because the bandwidth of the links on that path are larger than the ones in the lower path, and therefore the former is selected.

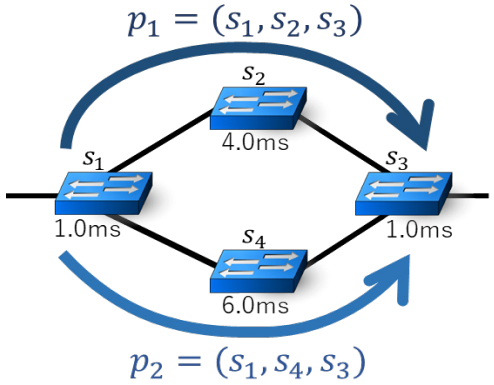


Figure 1: Example of Path Selection Considering Path Switching Time

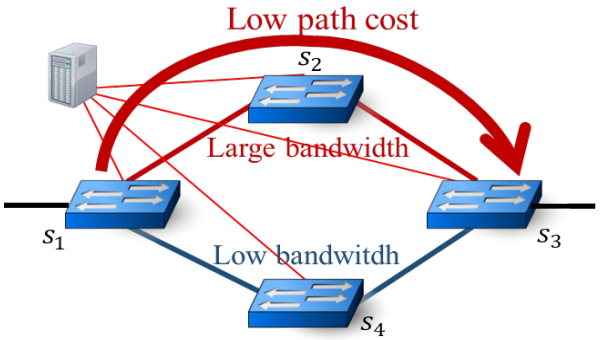


Figure 2: Example of Path Selection Considering Available Bandwidth

4. Design

4.1. Network Model

The network model is defined as below.

- $s_i \in S\ (i = 1, 2, ..)$: Switch
- t_i : Processing time for adding one flow entry into s_i
- $e_{i,j} \in E$: Link between s_i and s_j
- $B_{i,j}$: Available bandwidth of $e_{i,j}$
- p_k : Path (list of switches)
- $c_{i,j}$: Link cost of $e_{i,j}$ (= 1Gbps/$B_{i,j}$)

4.2. Definition of Path Switching Time

The path switching time is defined as:

"The time from the arrival of the first additional flow entry from the controller to the switch until the completion of the path switching on all switches"

Figure 3 shows an example. Suppose that tthere are three flows (flow 1, 2, and 3) assigned to the paths shown in Figure 3. In this case, t_1 is 1 ms, and s_1 needs to switch the three flows. Thus, the time to complete the processing of the additional flow entry is 1 ms * 3 = 3 ms. Then, t_2 is 4 ms and the s_2 needs to switch two flows (flow 1 and 2). Therefore, the time to complete the processing of the additional flow entries is 8 ms (4 ms * 2). For the other switches, the time is calculated in the same way. Finally, the path switching time in this example is 8 ms, which is the longest time to complete the process.

The path switching time $T(P_m)$ for a set of paths (P_m) is defined as in (1)

$$T(P_m) = \max_{\{i|s_i \in S\}} \{t_i \times \sum_{p_k \in P_m} z_i(p_k)\} \quad (1)$$

Here, $z_i(p_k)$ means whether the list of p_k includes s_i.

$$z_i(p_k) = \begin{cases} 1\ (s_i \in p_k) \\ 0\ (otherwise) \end{cases} \quad (2)$$

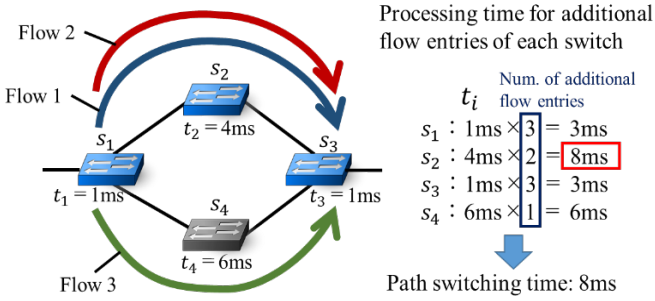


Figure 3: Example of Path Switching Time

4.3. Link Cost

The proposed method prioritizes the path switching time rather than the path cost, and chooses paths sequentially. The overview of the sequential path selection method is as follows:

- Choose a path for one flow at a time.
- Substract the required bandwidth of the flow from the link bandwidth on the path, and update the link cost every time the path is selected.
- Assign a flow to the path until there is no available bandwidth left for the flow.

Figure 4 shows an example of the link cost update. If the link bandwidth is 1 Gbps, the link cost is 1. Then, when a flow requires 300 Mbps of bandwidth, it is inserted into the link. The available bandwidth of the link becomes 700 Mbps, and the link cost increases to about 1.4.



Figure 4: Example of Link Cost Update

4.4. Flow of Path Selection

This section explains the process of path selection. In order to consider both the processing time of the switch and the path cost, the proposed method chooses paths as detailed below.

1. Compute the paths with the smallest increase in the path switching time.
2. Choose the path with the lowest path cost form the computed paths.
3. Repeat process 1 and 2 for the number of flows that needs to change.

5. Experimentation

5.1. Overview

The proposed method was evaluated in an emulated environment. The network used in the experiment is shown in Figure 5. Here, the available bandwidth $B_{i,j}$ of all links is set as 1 Gbps. Also, the processing time of the source switch, and destination switch was set to 1 ms; while remaining switches were randomly set from 1 to 10 ms.

Several flows with different required bandwidth were inserted, and randomly shutdown links. After the link is down the paths of the flows are switched. To compare the effectiveness of the proposal, the obtained throughput is obtained and compared with an existing method, which selects paths considering only path cost.

In this experiment, Mininet (ver. 2.2.2) [19] was used as a network emulator on a single computer, whose specification is as below:

- CPU: Intel (R) Xeon (R) CPU E5-2650 v4 @ 2.20 GHz) x 8 cores
- Memory: 16 GB

We also used OpenDaylight (0.3.3-Lithium-SR3) [20] as the OpenFlow controller, and OpenVSwitch (ver. 2.5.5) [21] as the OpenFlow switch.

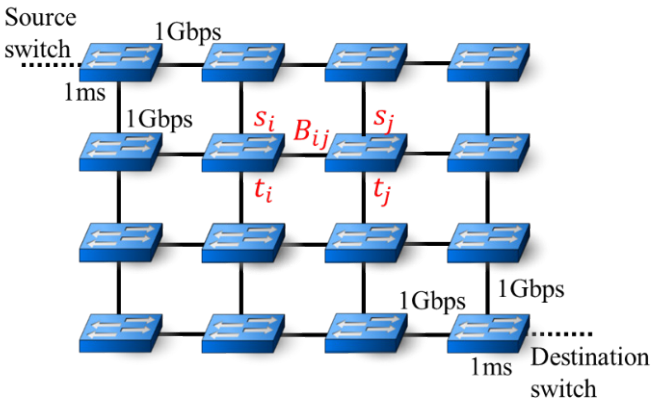


Figure 5: Network Topology Used in the Experiment

5.2. Results

The authors categorize the throughput comparing the proposed method with the existing method into the following three patterns:

- The proposed method always dominates (Win).
- The proposed method dominates until a certain time (Win[time]).
- The proposed method is equal to the existing method (Eq).

Examples of graphs of the three patterns are shown in Figures 6, 7, and 8. The comparison of the data transfer amount between the proposed method and the existing method is shown in Table 1. From the table, it is observed that when the total bandwidth of the flows is 1,800 Mbps or more, and the number of disconnected links is large, the data transfer amount of the proposed method is always better. However, when the total bandwidth of the flows is 1,800 Mbps or more, and the number of link disconnections is small, then the amount of transferred data of the proposed method is larger within 0.1 to 0.2 seconds.

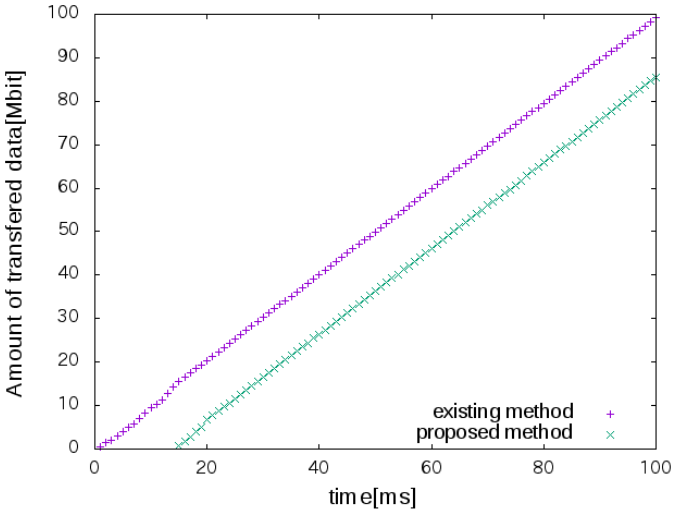


Figure 6: The Proposed Method Dominates Always (Win)

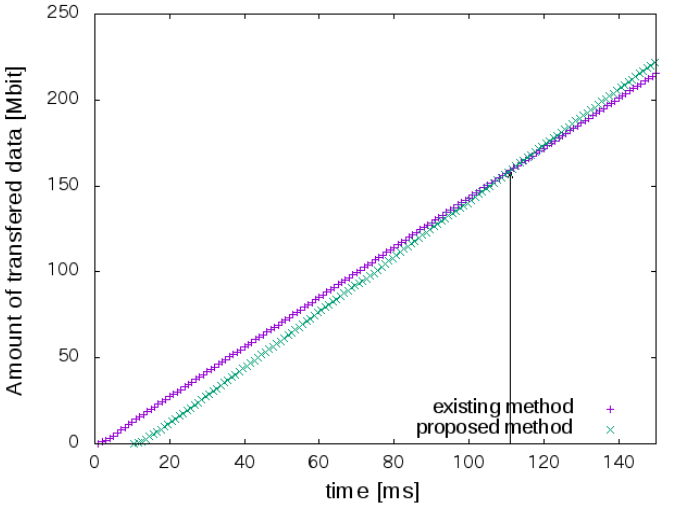


Figure 7: The Proposed Method Dominates Until A Certain Time (Win[112s])

Moreover, when the total bandwidth of flows is less than 1,400Mbps, and the amount of transferred data of the proposed method is larger is within 1.5 seconds, the proposed method is larger regardless of the number of disconnected links. Finally, when the total bandwidth of flows is less than 1,200 Mbps, the proposed method is always better.

Table 1: Comparison of the data transfer amount between the proposed method and the existing method

		Total bandwidth of the flows [Mbps]										
		1,000	1,200	1,400	1,600	1,800	2,000	2,200	2,400	2,600	2,800	3,000
Number of down links	0	Win	Win	Win [1.5ms]	Win [0.3ms]	Win [0.1ms]	Win [0.1ms]	Win [0.1ms]	Win [0.1ms]	Win [0.1ms]	Win [0.1ms]	Win [0.1ms]
	4	Win	Win	Win [1.5ms]	Win [0.3ms]	Win [0.2ms]	Win [0.2ms]	Win [0.2ms]	Win [0.1ms]	Win	Win	Win
	8	Win	Win	Win	Win [0.2ms]	Win [0.1ms]	Win	Win	Win [0.1ms]	Win	Win	Win
	12	Win	Win	Win	Win [0.1ms]	Win	Win	Win	Win	Win	Win	Win
	16	Win	Win	Win	Win	Win	Win	Win	Win	Win	Win	Eq
	18	Eq	Eq	Eq	Eq	Eq	Eq	Eq	Eq	Eq	Eq	Eq

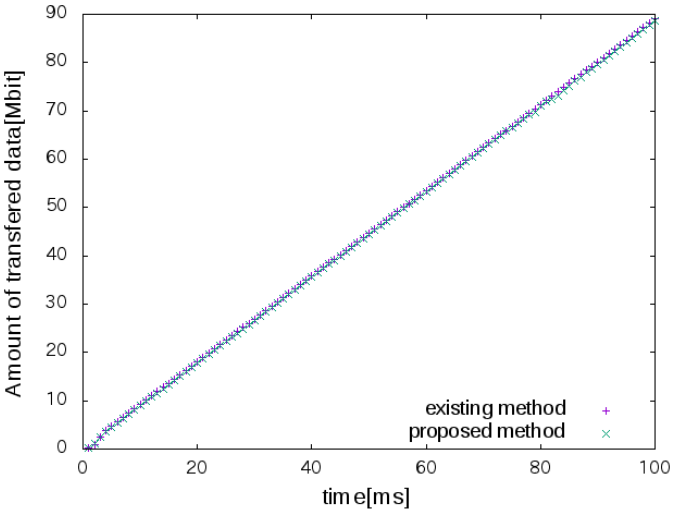


Figure 8: The Proposed Method Is Equal To The Existing Method (Eq)

5.3. Discussion

From the experimental results in the previous sections, the authors of this paper confirm that the proposed method is more effective than the existing method, since the amount of transferred data is equal or larger to the existing method while the path switching time is shortened. Morever, the path switching process is more effective than the existing method when these events occurs frequently and there is not enough bandwidth.

6. Conclusion

This paper presents a network control method considering the path switching time in SDN. We designed the path selection method and evaluated with an emulated experimentation. From the experimental results, it is confirmed that the proposed method is effective in situation where changes occurs frequently in a short time due to link disconnection and restoration (e.g., disasters).

As future work, we will extend the proposed method by introducing other parameters and experiment with real networks. Currently, we focused on the time from the arrival of the first additional flow entry from the controller to the switch until the completion of the path switching on all switches. Thus, we will evaluate the total time of changing paths from the moment some events occur until they complete changing the paths in comparison with related work.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] K. Gotani, H. Takahira, M. Hata, L. Guillen, S. Izumi, T. Abe, and T. Suganuma, "OpenFlow Based Information Flow Control Considering Route Switching Cost," Proc. the 5th International Conference on Informationand Communication Technologies for Disaster Management (ICT-DM 2018), pp.1-4, 2018.

[2] Ministry of Internal Affairs and Communications, ahttp://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd252140.html (Accessed 2019).

[3] S. H. Yeganeh, A. Tootoonchian, and Y. Ganjali, "On Scalability of Software-Defined Networking," IEEE Communication Magazine, Vol. 51, No. 2, pp.136-141, 2013.

[4] "Software-Defined Networking (SDN) definition," Open Networking Foundation, https://www.opennetworking.org/sdn-resources/sdndefinition, (Accessed 2019).

[5] N. McKeown, T. Anderson, H. Balakrishna, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," ACM SIGCOMM Computer Communication Review, Vol. 38, Iss. 2, pp.69-74, 2008.

[6] Open Networking Foundation, "OpenFlow Switch, Specification (Version 1.5.1)," https://www.opennetworking.org/wp-content/uploads/2014/10/openow-switch-v1.5.1.pdf (Accessed 2019).

[7] A. Nguyen-Ngoc, S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, and M. Jarschel, "Performance Evaluation Mechanisms for FlowMod Message Processing in OpenFlow Switches," Proc. of IEEE Sixth International Conference on Communications and Electronics (ICCE2016), pp.40-45, 2016.

[8] C. Rotsos, N. Sarrar, S. Uhlig, R. Sherwood, and A. Moore, "OFLOPS: An open framework for openflow switch evaluation," Proc. of Passive and Active Measurement (PAM2012), LNCS 7192, pp.85-95, 2012.

[9] R. Bifulco and A. Matsiuk, "Towards Scalable SDN Switches: Enabling Faster Flow Table Entries Installation," Proc. of 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM 2015), pp.343-344, 2015.

[10] S. A. Astaneh and S. Shah Heydari, "Optimization of SDN Flow Operations in Multi-Failure Restoration Scenarios," IEEE Transactions on Network and Service Management, Vol.13, No.3, pp.421-432, 2016.

[11] K. Gotani, H. Takahira, M. Hata, L. Guillen, S. Izumi, and T. Abe, "OpenFlow Based Information Flow Control Considering Route Switching Cost," Proc. the 2nd IEEE International Workshop on Information Flow Oriented Approaches in Internet of Things and Cyber-Physical Systems (InfoFlow 2019), pp.527-530, 2019.

[12] A. Feldmann, P. Heyder, M. Kreutzer, S. Schmid, J.P. Seifert, H. Shulman, K. Thimmaraju, and M. Waidner, "NetCo: Reliable Routing With Unreliable Routers," Proc. of the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop (DSN-W), pp.128-135, 2016.

[13] C. Cascone, L. Pollini, D. Sanvito, A. Capone, and B. Sanso, "SPIDER: Fault resilient SDN pipeline with recovery delay guarantees," Proc. of IEEE NetSoft Conference Workshops (NetSoft), pp. 296-302, 2016.
[14] P. M. Mohan, T. Truong-Huu, and M. Gurusamy, "TCAM-Aware Local Rerouting for Fast and Efficient Failure Recovery in Software Defined Networks," Proc. of 2015 IEEE Global Communications Conference (GLOBECOM 2015), pp.1-6, 2015.
[15] B. Stephens, A. L. Cox, and S. Rixner, "Scalable Multi-Failure Fast Failover via Forwarding Table Compression," Proc. ACM SIGCOMM Symposium on SDN Research Article (SOSR 2016), no. 9, pp.1-12, 2016.
[16] S. Sharma, D. Staessens, D. Colle, M. Pickavet, and P. Demeester, "In-Band Control Queuing and Failure Recovery Functionalities for OpenFlow," IEEE Network, vol. 30, no. 1, pp. 106-112, 2016.
[17] S. Paris, G. S. Paschos, and J. Leguay, "Dynamic Control for Failure Recovery and Flow Reconfiguration in SDN," Proc. International Conference on the Design of Reliable Communication Networks (DRCN 2016), pp.152-159, 2016.
[18] A. Malik, B. Aziz, M. Adda, and C. Ke, "Optimization Methods for Fast Restoration of Software-Defined Networks," IEEE Access, Vol.5, pp.16111-16123, 2017.
[19] Mininet, http://mininet.org/
[20] OpenDaylight, https://www.opendaylight.org/
[21] Open vSwitch, http://openvswitch.org/

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 394-400 (2019

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Transmission Line Restoration Using ERS Structure

Sushri Mukherjee[1,*], Sumana Chattaraj[1], Md. Irfan Khan[1], Dharmbir Prasad[2], Pradip Barua[1], Harish Agarwal[1]

[1]*Supreme & Co. Pvt. Ltd., 700020, West Bengal, India*

[2]*Asansol Engineering College, 713305, West Bengal, India*

ARTICLE INFO

Article history:
Received: 14 April, 2019
Accepted: 25 September, 2019
Online: 28 October, 2019

Keywords:
ERS, Bypass line
Reconductoring
Modular structure, Emergency Restoration
Power transmission services

ABSTRACT

In the past decades, various ideas have proposed for reconductoring of overhead lines to enhance power transmission capacity considering growing energy demand. Conventional method of reconductoring is use to take couple of weeks but availability of shutdown for a prolong period is a major constraint in line uprating. Emergency restoration structure (ERS) is effective to deliver such challenging task within hardly 4-5 hours line shutdown only. This quick restoration and easiness is mobility features are making ERS versatile for restoration even after any natural disasters. In this article, a case study of 132 kV restorations is described as a conceptual proof of concepts. Successful execution of this project resolved a 400 kV transmission line hindrance between Atur and Indira Puram since March 2017 and was likely to remain idle due to a 132 kV single circuit transmission line crossing its route.

Nomenclature

A_e	effective surface area of structure
B	base width of ERS
b	arm length of ERS
C_a	amplification factor
C_f	overall force co-efficient
C_F	shape co-efficient
C_g	gust response factor
E_i	expenditures
F	wind force
H	structure height
H_G	wind gradient
i	years count
k_0	factor of conversion
k_1	probability factor of risk
k_2	factor of terrain roughness
k_3	factor of site topography
m	overturning moment
p	surface contact pressure
P_s	ultimate pull out force
P_t	anchor surrounding force
P_u	shearing stress
r	discounting rate
R_i	factor of probabilistic cost
S	conductor stringing load
S_T	total structural area
V_G	wind speed
V_h	wind speed at certain height
V_H	wind speed above ground level
V_b	regional basic wind speed
V_d	design wind speed
V_r	reference wind speed
w	failure zone effective soil weight
α	wind speed inequality co-efficient
φ	structural solidity ratio
ρ	air density
$1/\lambda$	increase in velocity exponent with height

[*]Sushri Mukherjee, Research Engineer, Supreme & Co. Pvt. Ltd., W.B., India, Contact No: 8372834629 & Email ID: sushri.engg@supreme.in

ERS	emergency restoration structure
FoS	factor of safety
MDD	maximum dry density
MSPT	maximum bearing test of soil
NPV	yearly expenditures net present value
OMC	optimum moisture content of soil
RoW	right of way
R&D	research and development
SPT	standard penetration test
T&D	transmission and distribution
UTS	ultimate tensile strength

1. Introduction

1.1 General

In the era where power became basic necessity, it becomes essential for utilities to strengthen their infrastructure for accommodating the changes posed by deregulation and transfer reliable power to the consumers [1]. Pressure on the existing network to transfer more power has led to transmission line congestion. These issues could be addressed by uprating as it allows considerable increase in power transfer capability within the existing infrastructure. It's troublesome as shut down is required for reconductoring as these lines are invariably over loaded. The primary concern in emergency restoration is to revive the transmission network and get the line charged and back into operation within the stipulated time frame. These are the key objectives that compelled the development of an Emergency Restoration Structure (ERS).

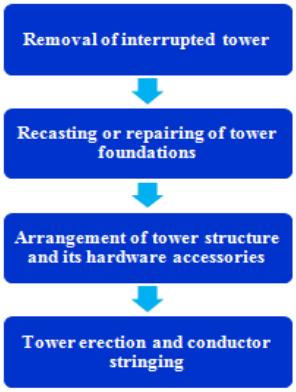


Figure 1: Flowchart of power restoration by conventional method

The Indian Power Grid has developed more complex after the resolution of "One Nation One Grid" by the Central Government. The T&D industries are enduring with R&D challenges like - reliability maintenance, useful life expansion, enhancement of transmission capacity by upgrading/uprating and public & employee safety [2-4]. The overhead transmission lines are covering miles of distance thus; it is exposed to several weather related calamities (*e.g.,* speedy wind, heavy rain, cyclone *etc*) throughout the wide area [5-7]. Unavailability of electricity due to any of these reasons or may be due to regular maintenance purpose may cause inconvenience to the consumers' along with financial loss to the service provider [8-16]. These issues could be resolved by improving grid resiliency using ERS bypass arrangement. The proposed ERS usage less time and RoW in compare conventional method of restoration considering following steps (as presented in Figure 1).

Now-a-days, most of the transmission lines are functioning inefficiently due to decay in the transmitting property. Rejuvenation of aging transmission infrastructure is required to boost its efficiency. The key challenges faced by any power transmission system are as below:

- *Transmission line congestion*
- *Fewer private players*
- *Lack in investments*
- *Strategic planning unavailability*

Transmission line utilities are always trying to reduce customer minutes lost and to avoid any concurrent financial losses [17]. They are always investing in quick restoration technology; even they are getting their manpower trained to counter devastation on emergency basis caused by natural disasters. Previously, such investment was required on a case specific critical components inventory. This led to a major chunk of financial reserves availability to handle such situations [18, 19]. The interchangeable modular ERS structure could be easily transported to even difficult locations and can be reused in future exigencies. The ERS could reduce the inventory of spare towers substantially by using it as temporary service. The installation of ERS was being optimized to overcome site challenges and reduce labour effort with rigorous installation and minutely observing the safety parameter as well enhancing the speed of the process. In this paper, a case study of 400 kV transmission line between Atur and Indira Puram, abandoned since March 2017 and was likely to remain idle due to a 132 kV Single Circuit Transmission line crossing its route.

1.2 Motivation behind the Present Study

Pressure on the existing network to transfer more power has led to transmission line congestion. These issues could be addressed by uprating as it allows considerable increase in power transfer capability within the existing infrastructure. It's troublesome as shut down is required for reconductoring as these lines are invariably over loaded. The primary concern in emergency restoration is to revive the transmission network and get the line charged and back into operation within the stipulated time frame. These are the key objectives that compelled the development of an ERS tower [20-22].

1.3 Paper Layout

In continuation of aforementioned discussions, remaining study has been presented in succeeding sections. Section 2 presents development of modular ERS structure; while Section 3 deals with reconductoring using ERS, Section 4 focused on overhead transmission line restoration; Section 5 presents observation of the Case Study and finally, the paper has been concluded in Section 6.

2. Development of ERS Module

A quick restoration of power using ERS tower could be made based on certain aspects [23-24], which are as under:

2.1 ERS Design Considerations

A case study has been done on an executed project for UPPTCL. In this paper, the guyed ERS towers were deployed for

reconductoring purpose under influence of varying wind speed in between the wind zone 30-55 m/s up to30 m height of tower as per IS:875-1987(Part III).The proposed structure sustained a major bending moment of lateral forces under influence of both tensile and compressive forces. There is optimal allocation of guy wires are required to gain stable workability of the structure. The project was successfully delivered on 5th June, 2017. The whole study could be co-related with following formulations:

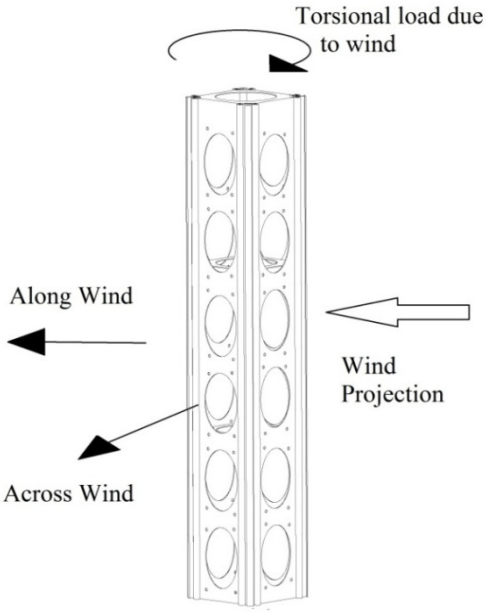


Figure 2: ERS structures' isometric view

According to metrological requirements of the site, the referred wind speed [25-28] could be given by (1),

$$V_r = \frac{V_b}{k_0} \tag{1}$$

Designed wind speed [25-31] of the ERS set up could be given in (2),

$$V_d = V_r k_1 k_2 k_3 \tag{2}$$

As per Devenport formulations for wind gradient is given by (3) [25-28]. Torsion wind load is shown in Figure 2,

$$V_H = V_G \left(\frac{H}{H_G} \right)^{\alpha} \tag{3}$$

The sectional view shows the side view of the single plane which has oval hollow section to optimize the weight and reduce effective wind pressure on the section. In order to prevent deformation of the structure several construction optimization was done. Significant features of ERS have been briefed as below:

- Erection is very quick due to modular and interchangeable section.
- Minimum numbers of component and identical fittings to reduce identification and segregation process.
- Quick installation with minimum power interruption and cost effective.
- No need of special foundation or tools.
- Erection is possible in any kind of soil and terrain.
- Compatible with 66 kV-765 kV stringing.

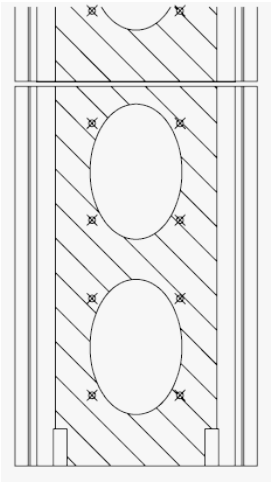


Figure 3: Sectional front view of the ERS module

Wind force [29] imposing on ERS module is given by (4),

$$F = C_f p A_e \tag{4}$$

Dynamic pressure [29], considering design variables is formulated by (5),

$$p = \frac{C_g C_F C_a \rho V_h^2}{2} \left(\frac{H}{h} \right)^{\frac{1}{\lambda}} \tag{5}$$

As per Ryle's formula, the ERS base width could be presented by equation (6) [29],

$$B = 0.013\sqrt{m} \tag{6}$$

For maintain FoS at $32^{o}\mathbf{C}$, the conductors' tension without any external load should not exceed 35% of the specified UTS. However, this percentage shrinks to 25% under loaded condition. Aspect ratio is 65.93 of the ERS modular tower as shown Figure 3 and it can be formulated as in (7) [29],

$$\varphi = \frac{S_T}{Hb} \tag{7}$$

Pull out forces to evaluate anchoring of ERS can be formulated by equation (8),

$$P_u = P_s + P_t + w \tag{8}$$

As per CIGRE, the costing NPV of the undertaken project could be formulated by (9) [30, 31],

$$NPV = \sum_{i=0}^{n} \frac{(E_i + R_i)}{(1+r)^i} \tag{9}$$

2.2 ERS towers' Hardware Integration

The ERS tower comprises of multiple components, each of these (as portrayed in Figure 4) are equally important and designed for meeting emergency requirements of the overhead

line. Each modular columns (height: 3 m) are identical and fabricated of high strength steel, so they can be easily mounted without following any particular sequence of the complete tower. Foundation plates are designed such it can allow distribution of weight to the ground with minimum load bearing capacity. There was one articulated gimbals joint, which allow linking between foundation plate and sections. The composite insulators were used to reduce RoW issues although at higher voltage levels.

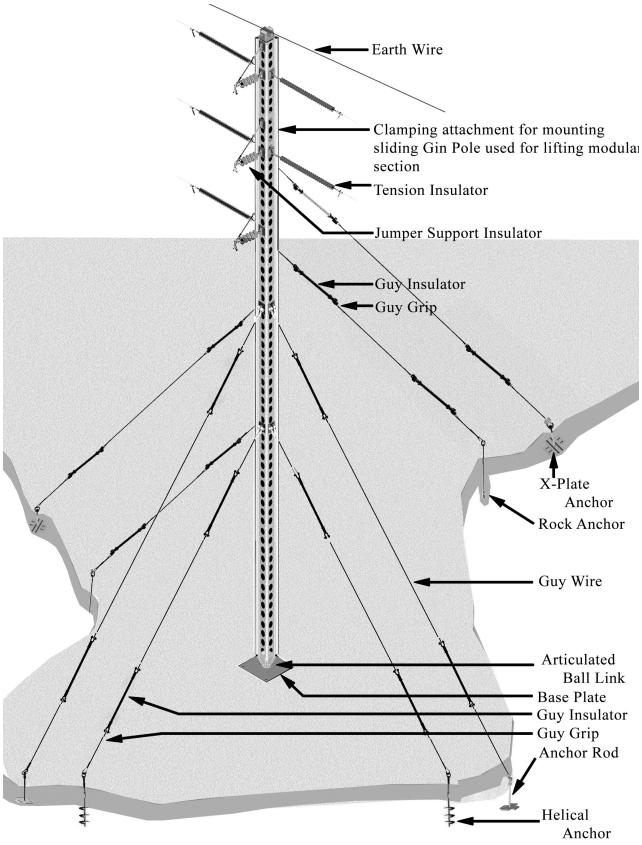


Figure 4: Generalized ERS Set Up.

3. Overhead Transmission Line Restoration

This paper briefly presents a case of electricity transmission from Loni to Mohan Nagar division (UP, India) under following subsections:

3.1 GIS Mapping of the Site

The ERS tower has been utilized at UPPTCL site to re-build 132 kV lines confronting flyover, water passage and street intersections, separately (depicted in Figure 5). However, 132 kV overhead lines was thwarting hanging of 400 kV transmission lines, as height of that line was 51 m from the top most earth wire.

Consequently, a restructuring was made to curtail height of 132 kV by de-amassing second tower (ref to Figure 5) such that, the 400 kV line can be set back on it. Total three ERS towers have been utilized to bypass. Usually, three weeks supply interruption could have been there with conventional method for restoration but application of ERS has delivered this project by 5 hours only. Before initiating execution work of this project, a extensive site survey and planning was done to accomplish this project.



Figure 5: Ghaziabad site restoration layout

3.2 Deployment of Restoration Works

In the current work, three ERSs (height: 30m) were deployed based on the requirements of the specific site UP state, in order to meet resilience and sustainability of bypass arrangement [28]. The analysis of stringing arrangement was performed using PLS CADD software tools. The complete solutions for line restoration or construction of any kind of bypass line by using ERS tower on rental or end to end services basis is being provided, as we have a dedicated team of execute the work. It involves several steps for the enactment of the ERS installation:

Step-1: The initial step involves site survey of the installation site. There are various factors that are taken into account before the installation. A dedicated team is at the first analyses and collects information regarding site. It includes the information such as accessibility to the site, existing tower to ERS tower clearance, RoW clearance, layout of the site, span length *etc*. Site accessibility also helps to pre-determine the anchor to be used for the type of soil as well kind of tools. After the examination process and collection of all the details, the project engineers proceed towards the next step. The load assessment is done by the design team as per the side condition and feasibility.

Step-2: The second step involves material transportation such as the modular sections and all the equipment that are required for the proper guying and installation of the tower.

Step-3: The third step includes the installation part. Most importantly the major advantage of our ERS towers is that civil foundation is not required. The tower is entirely supported using guy wires. Hence, it is essential that all the guy wires are properly anchored to the soil. Our team took this as a challenge and developed different kind of anchors that would support the structure for different soil conditions.

Step-4: After the installation of ERS base the gimbals joint is then attached to it and with the help of it modular section is installed. Gin poles are used for lifting the next modular sections one by one.

Step-5: For stabilizing the tower guy arrangement has been used. After that all the hardware are installed and conductor stringing is done. Some extra wires are removed to ease the congestion along the axis of the conductor since the

tensioned conductor support the tower followed by inspection and joint commissioning to charge the line.

The overhead transmission line has been circumventing from tower-1 to that of -2 by a new root as mapped in Figure 6(a) and erected line was shown in Figure 6(b).

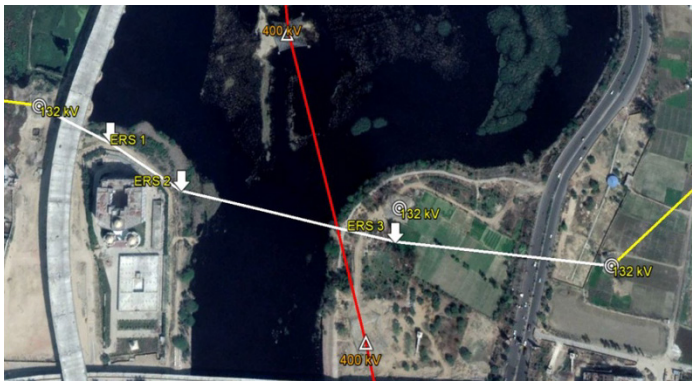


Figure 6(a): Bypass arrangement routing layout



Figure 6(b): Flyover and river crossing

The range existences between re-steered lines are customary cross section tower-1 to ERS-1: 111 m, ERS-1 to -2: 93 m, ERS-2 to -3: 229 m lastly ERS-3 to regular lattice tower-3: 250 m. In continuation with site overview, skilled team with their tools deployed to pursue the project.

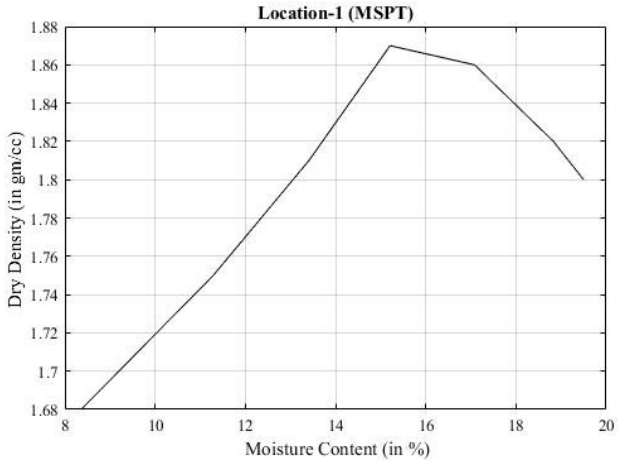


Figure 7: Characteristics of MSPT test for soil sample-1.

4. Observations from Case Study

The heavy compaction test has been tested for soil samples taken from location-1 and -2 while; light compaction test is found to be appropriate for location-3. These types of test were carried out to check MDD and OMC of soil as per IS: 2720 (Part VIII) 1983. The field dry density core cutter testing method complying IS:2720 (Part VII & XXIX) has been performed across the field at each location as shown in Table 1.

Table 1: Field test findings from core cutter method.

Parameters	Sample- 1	Sample- 2	Sample-3
Field BD (gm/cc)	1.93	1.91	1.92
Field MC (%)	4.79	6.16	4.82
Field DD (gm/cc)	1.84	1.80	1.83
DD (% of MDD)	98.39%	96.25%	97.86%

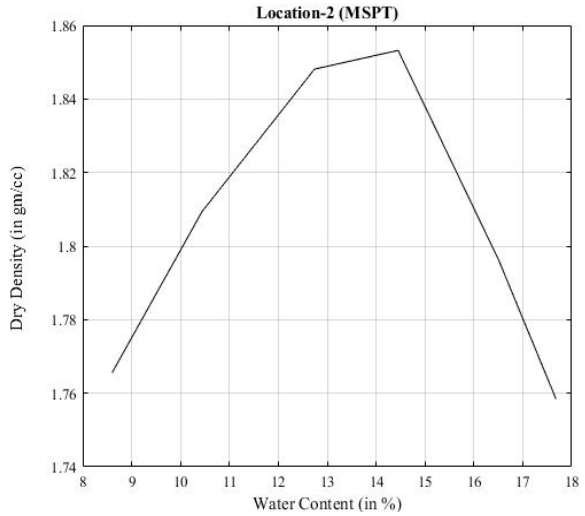


Figure 7: MSPT test characteristics of soil sample-2.

The recorded test findings of moisture content and dry density are provided in Table 2 and their graphs are plotted in Figure 7-9. The interesting outcomes (OMC & MDD) are highlighted by **boldfaced** in the subsequent column of Table 1. The compaction test observations exceeding 95% of the MDD; thus, the site was viable to erect the ERS tower.

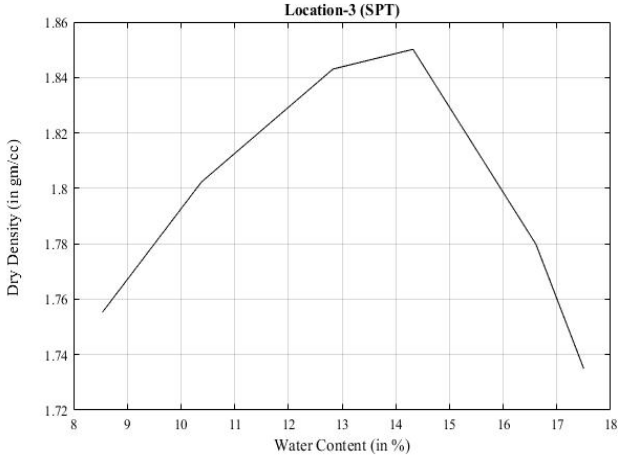


Figure 8: SPT test characteristics of soil sample-3

Table 2: Results obtained after compaction test

Locations	Testing methods	Parameters	
		Content of moisture (%)	Dry density (gm/cc)
1	MSPT	8.360	1.68
		11.29	1.75
		13.41	1.81
		15.21	**1.87**
		17.08	1.86
		18.83	1.82
		19.50	1.80
2	MSPT	8.600	1.766
		10.44	1.809
		12.74	1.848
		14.45	**1.853**
		16.51	1.796
		17.68	1.758
3	SPT	8.540	1.76
		10.38	1.80
		12.83	1.84
		14.32	**1.85**
		16.61	1.78
		17.50	1.73

Table 3: Site specific benefit comparison

S. No.	Particulars	Restoration using ERS	Conventional restoration
1	Structural footprint	<1 m^2	36-64 m^2
2	Foundations of structure	Not required.	Depth: 2-4 m
3	Angle of protection	65º	30º
4	Down time of supply	5 Hours	20 Days
5	Average value of power flow	5 MW	5 MW
6	Generation loss/day	120000 Units	120000 Units
7	Unit price	INR 2.5	INR 2.5

After execution of this project, a cost-benefit analysis was made (given in Table 3) to ascertain future scope of ERS in service customers.

5. Conclusion

Indeed, even in open vitality markets, electricity transmissions are normally considered as a usual monopoly and sometimes bring contests among investors. This starts leapfrogging of new ideas so as to get by in the aggressive power industry. Thusly, development of ERS gives an elective potential answer for detour the customary transmission tower for rebuilding just as line reconductoring. Line uprating proves to be a smart solution to cater the increased loads on existing transmission lines. However, this exercise is problematic as the shutdown required for reconductoring cannot be provided as those lines are already critical. This exercise helped in completing a project which would evacuate 580 MW lines where substantial investment was already made on 98% stretch of the line and associated substation would have remained idle for indefinite period. Project milestones and achievements are concluded as below:

- *Line Crossing:* Stringing was challenging as the line had a flyover, a river and a highway crossing
- *Delay in Shutdown:* Even after installing the ERS, re-tapping was frenetic due to shutdown required for 2 hours from UPPTCL was difficult.
- *No Heavy Machinery:* No heavy machinery or earth movers were introduced for execution of above, other than for transportation, which depicts the agility of the system.
- *Challenging Feat:* The bypass line across the highway was done without any disruption of highway traffic even for a minute.
- *Rough Weather:* There was stormy wind after erection but no damage was made to the ERS.

Conflict of Interest

Authors are declaring that there is no conflict of interest with any other individual or organization.

Acknowledgement

The project shown as a case study was done for UPPTCL, India. Authors are thankful to Mr. Arup Das (SCPL) and Mr. Prakash Reddy (SCPL) for their active co-operation.

References

[1] "World Energy Outlook Special Report : India Energy Outlook" International Energy Agency, 2015.

[2] O. H. Abdalla, T. M. Alkhusaibi, M.A. Thani, M. N. Alkhusaibi, "Restoration of a 132 kV Over Head Transmisssion Line affected by Tropical Cyclone Gonu in Oman" in IEEE Proc. Transm. Distrib. Conf. and Expo., 1-5, 2008. DOI: 10.1109/TDC.2008.4517063

[3] L. N. Agrawal, P. E. Erickson, "Planning and Training reduce Restoration time for damaged transmission lines in Inida" in Int. Proc. on Transm. and Distrib. Constr., Oper.and Live-Line Maint., 101-108, 2000. DOI: 10.1109/TDCLLM.2000.882807

[4] M.I. Khan, D. Prasad, K. Sarkar, S. Mukherjee, P. Barua, H. Agarwal, 2019, March. Finite Element Modeling of 400 kV ERS Bypass Transmission Line Quantifying Installation Difficulty. In 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia) (pp. 526-531). IEEE. 10.1109/GTDAsia.2019.8715871

[5] D. Prasad, R. Kumaran, G. Agarwal, P.K. Reddy, Md. I. Khan, 2016. Emerging Solutions for Line Compaction. CIGRE India Journal, 5(2), pp. 21-26.

[6] S. Mukherjee, D. Prasad, Md. I. Khan, P. Barua, H. Agarwal, 2019. Hexa Spacer Damper for Vibration Energy Decaying of 765 kV Transmission Line. 2nd International Conference on Innovations in Power and Advanced Computing Technologies, i-PACT-2019, on 22nd - 23rd March, 2019 at Vellore Institute of Technology, Vellore, India.

[7] D. Prasad, V. Mukherjee, 2016. A novel symbiotic organisms search algorithm for optimal power flow of power system with FACTS devices. Engineering Science and Technology, an International Journal, 19(1), pp.79-89. 10.1016/j.jestch.2015.06.005

[8] D. Prasad, A. Mukherjee, V. Mukherjee, 2017. Application of chaotic krill herd algorithm for optimal power flow with direct current link placement problem. Chaos, Solitons & Fractals, 103, pp. 90-100. 10.1016/j.chaos.2017.05.037

[9] D. Prasad, A. Mukherjee, G. Shankar, V. Mukherjee, 2017. Application of chaotic whale optimisation algorithm for transient stability constrained optimal power flow. IET Science, Measurement & Technology, 11(8), pp.1002-1013. 10.1049/iet-smt.2017.0015 10.1049/iet-smt.2017.0015

[10] D. Prasad, A. Mukherjee, V. Mukherjee, 2017. Transient stability constrained optimal power flow using chaotic whale optimization algorithm. In Handbook of neural computation (pp. 311-332). Academic Press. 10.1016/B978-0-12-811318-9.00017-X
[11] D. Prasad, V. Mukherjee, 2018. Solution of optimal reactive power dispatch by symbiotic organism search algorithm incorporating FACTS devices. IETE Journal of Research, 64(1), pp.149-160. 10.1080/03772063.2017.1334600
[12] V. Mukherjee, A. Mukherjee and D. Prasad, 2018. Whale Optimization Algorithm with Wavelet Mutation for the Solution of Optimal Power Flow Problem. In Handbook of Research on Predictive Modeling and Optimization Methods in Science and Engineering (pp. 500-553). IGI Global. 10.4018/978-1-5225-4766-2.ch023
[13] D. Prasad, A. Banerjee, R.P. Singh, 2019. Optimal Reactive Power Dispatch Using Modified Differential Evolution Algorithm. In Advances in Computer, Communication and Control (pp. 275-283). Springer, Singapore. 10.1007/978-981-13-3122-0_26
[14] A. Banerjee, D. Prasad, R.P. Singh, 2015. Smart reactive power control of hybrid power system model by Seeker Optimized Algorithm. In ETES-2015 conference organized by IET on 16-17th July, 2015 at Asansol, West Bengal, India.
[15] R.P. Singh, A. Banerjee, D. Prasad, 2014. Solution of optimal power flow problem using modified differential evolution algorithm. In ETES-2014 conference organized by IET at Asansol, West Bengal, India.
[16] R.P. Singh, D. Prasad, P. Kumar, and M.I. Khan, 2017, August. Modified differential evolution algorithm for optimal power flow with FACTS devices. In 2017 International Conference on Smart grids, Power and Advanced Control Engineering (ICSPACE) (pp. 48-52). IEEE. 10.1109/ICSPACE.2017.8343404
[17] N. Ozog, E. Desjardins, J. Jatskevich, "Bulk power system restoration interdependency risk modelling" in IEEE Proc. Elect. Power Energy Conf., 1-7, 2008. DOI: 10.1109/EPC.2008.4763347
[18] R. J. Schweiner, K. E. Twomey, K. E. Lindsey, "Transmission Line Emergency Restoration Philosophy at Los Angeles Department of Water and Power," in IEEE Proc. Transm. and Distrib. Constr., Oper. Live-Line Maint., 11-17, 2003. DOI: 10.1109/TDCLLM.2003.1196462
[19] J. C. Pohlman, "The case for Universal Standards to cover the structural components and practices used in Emergency Restoration Systems" in Int. Proc. on Transm. and Distrib. Constr., Oper.and Live-Line Maint., 109-115, 2000. DOI: 10.1109/TDCLLM.2000.882808
[20] J. F. Nolasco, P. Nefzger, U. Kaintzyk, Overhead power lines: Planning, Design, Construction, Springer, 2003.
[21] T. S. Kishore, S. K. Singal, "Optimal economic planning of power transmission lines: A review" Ren. and Sust. Energy Rev., 39(5), 949-974, 2014.https://doi.org/10.1016/j.rser.2014.07.125
[22] L. Bauer, P. Ulardic, J. Muller, "Reinforcement Strategies for Extending the Service Life of 110 kV Overhead Transmission Lines" in Int. Conf.and Exhibition on Electr. Distr., 1997. DOI: 10.1049/cp:19970547
[23] S. G. Krishnasamy, G. L. Ford, C. I. Orde, "Predicting the structural performance of transmission lines uprated by reconductoring" IEEE Trans. Power Appar. Syst., 100(5), 2271-2277, 1981. DOI: 10.1109/TPAS.1981.316740
[24] "Conductors for the Uprating of Existing Overhead Lines,"CIGRÉ WG B2.55 Brochure (Ref. No. TB 244), 2003.
[25] A. Marin, J. Salmeron, "Electric capacity expansion under uncertain demand:decomposition approaches" IEEE Trans. Power Syst., 13(2), 333-339, 1998. DOI: 10.1109/59.667347
[26] I. Albizu, A. J. Mazón, I. Zamora, "Methods for Increasing the Rating of Overhead Lines" in Power Tech IEEE Proc., Russia, 1-6, 2005. DOI: 10.1109/PTC.2005.4524481
[27] M. T. Chay, F. Albermani and H. Hawes, "Wind loads on transmission line structures in simulated downbursts," in First World Congress on Asset Management, Gold Coast, Australia, 2006.
[28] S. S. Murthy, A. R. Santhakumar "Transmission line structures," McGRAW-HILL Book Co., 1990.
[29] P. J. Ryle, "Steel tower economics" J. Inst. Electr. Eng. - Part II: Power Eng., 93(33), 263-274, 1946. DOI: 10.1049/ji-2.1946.0063
[30] H. Niroumand, Soil reinforecement for anchor plates and uplift response, Butterworth-Heinemann, 2017.
[31] J. C. Kaltenbach, J. Peschon, E. H. Gehrig "A mathematical optimization technique for the expansion of electric power transmission systems" IEEE Trans. Power Appar. Syst., 89(1), 113-119, 1970. DOI: 10.1109/TPAS.1970.292677

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 4, 404-413 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Supporting Better Physical Activity in a Smart City: a Framework for Suggesting and Supervising Walking Paths

Mário Rodrigues[*,1], Rita Santos[2], Alexandra Queirós[1], Anabela Silva[3], João Amaral[1], Patrícia Simões[3], Jorge Gonçalves[4], Ciro Martins[1], António Pereira[5], Nelson Pacheco da Rocha[6]

[1]*ESTGA & IEETA, University of Aveiro, 3810-193 Aveiro, Portugal*

[2]*ESTGA & DigiMedia, University of Aveiro, 3810-193 Aveiro, Portugal*

[3]*ESSUA & CINTESIS.UA, University of Aveiro, 3810-193 Aveiro, Portugal*

[4]*ESTGA & GeoBioTec, University of Aveiro, 3810-193 Aveiro, Portugal*

[5]*ESTG & CIIC, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal*

[6]*DCM & IEETA, University of Aveiro, 3810-193 Aveiro, Portugal*

ARTICLEINFO

Article history:
Received: 01 June, 2019
Accepted: 27 July, 2019
Online: 20 August, 2019

Keywords:
mHealth
Smart City
Walk Paths
Mobile App
Health Professional Back-Office

ABSTRACT

The increase of elderly population creates the need to promote healthy aging, with autonomy and independence, for preserving the functional capacity and quality of life as much as possible. To achieve this goal the recommendations include walking a certain amount of steps daily, given that the exact amount of steps changes with age, lesions, and chronic health conditions that can affect health and well-being.
Understanding how much exercise is adequate for an individual requires specialized knowledge and training. It is important to avoid risks that include: (1) be too aggressive in exercising and increase injuries, create new ones, and later refrain in exercising more; or (2) be too conservative and in the long run have an activity level under each one's potential and below what is recommended.
In this article is proposed and discussed a framework that aims to support people having an adequate level of activity for its particular condition and to do it without disrupting daily routines. The framework is supported by a software system built for monitoring well-being and physical activity in the context of Smart Cities. The system is composed by: (1) smartphone applications that interact with the end-user for showing possible exercises and walk routes, and that collects some relevant data; (2) a back-office application that collects and presents data obtained from the smartphone applications. The back-office application is designed for health professionals to follow users' progression over time, to recommend new exercises or to correct less optimal situations.
The first tests with a system implementing this framework show that the solution is robust and able to be used in a large scale.

1 Introduction

This paper is an extension of work originally presented in 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW) [1].

Active and healthy aging is important as individuals should have the best life possible and this usually means a long and healthy life. It is important for individuals as well as for communities since a political concern is the mounting pressure on national health services of aging societies due to the ever larger proportion of elderly population. Therefore it is important to minimize costs associated to assistance to people and digital technologies can give an important contribution. As healthier aging benefits both individuals and societies, there is a growing awareness of the importance of promoting it, with autonomy and independence, and preserving each one's functional capacity.

The question is thus: how to help people aging with good health? Several studies suggest that monitoring people's level of activity has the potential to foster health and well-

[*]Mário Rodrigues, IEETA – Campus Universitário de Santiago, Aveiro, Portugal, Tel. +351 234 370 500 & mjfr@ua.pt

being by transforming it from an individual goal towards a collective and social goal [2, 3]. Some experiments have concluded that the use of information and communication technology (ICT) enabled devices to collect individual performance data, a communication network for sharing those data, can contribute to increase the activity levels and to promote community building. Other studies go further ahead and infer that connecting residents and municipal health professionals through digital applications can decrease ansiety levels thus increasing well-being and promoting a sense of comfort as for instance to reassure young parents that a nurse is as close as a click of a button [2]. Mobile health (mHealth) is a paradigm that refers to medical and health practice supported by mobile devices. Despite the rapid growth of mHealth, mainly due to the significant growth in mobile device penetration [4], there are still a number of challenges including the need of having a high degree of security, reliability, quality, and effectiveness. Technological advances (e.g. accelerometers) have allowed for the objective measurement and characterization of physical activity and sedentary behavior in terms of volume, duration, intensity and frequency. This helps to personalize strategies and interventions aiming to increase physical activity and decrease sedentary behavior as well as goal setting [5]. Most mobile applications (apps) related to health, well-being, and physical exercise are not properly validated or under the supervision of health professionals, even though some are quite popular and well rated by users [6]. Privacy concerns also arise when apps collect and store data. The importance and potential impact on people's life of health related advices and personal information makes it a quite sensitive topic to be explored without proper validation and supervision.

In this paper we discuss and present a framework for a monitoring system of wellbeing and physical activity in the context of Smart Cities. The framework is more than a single mHealth app as it is designed for health professionals take control and operate the system, having access to user specific information as well as to contextual information about users' environment.

Although the proposed framework is more than an app, as it will have health professionals operating the system and will take into consideration contextual and user specific information, the app is a key component as it will be the interface for end users. The back-office component is designed to work with a range of alternative apps however it is important to provide a default app that, at least, works as a showcase of good practices and helps to illustrate the framework potential.

The framework is designed for the global Smart City, in the sense that its technical solutions and its conceptual approach take advantage of sensor and city state information. The framework is deployed in a concrete Smart City: the Smart City of Águeda that is a member of the Open and Agile Smart Cities. Águeda was selected for two reasons: (1) it is recognized nationally and internationally as one of the most dynamic Smart Cities of Portugal; (2) it belongs to a region where the proportion of elder population is increasing, as in many other Portuguese and European areas. The Portuguese census of 2011 showed that 22,5% of the population of in this region was 65 years old or more.

From the city authorities perspective, collecting information where people usually practice outdoor exercise, or stroll, is valuable for planning public spaces. More and better infrastructures like fitness equipment, benches for people to seat and rest for a while, children's playgrounds, sports equipment as Basketball fields, among other, can be better located if city officials posses such information. Also, concerning mobility, is important to know here people concentrate when deciding the location of bicycle parking structures, electrical vehicle's chargers, bus stops, and so on and so forth [7].

After the introduction in this section, section II presents the relevant background and related work on smart city projects and on mHealth and related mobile applications. The third section explains the developed work, discussing the relevant implementation details and presenting the interface of the system. The following section, Section IV, discusses the main features of the proposed framework and explains how it helps to mitigate known shortcomings in current mHealth solutions. The paper ends with the conclusions in section V.

2 Background and Related Work

This section will first provide some background information on smart cities followed by recent relevant projects on this topic. Then, the section introduces background information on mHealth ending with a description of three of the most popular apps for mHealth.

2.1 Smart Cities

Smart City has become a key marketing term for researchers and politicians to justify technological investments in urban areas. Several cities and less dense urban areas in quite different stages of development claim to be smart, with no clear definition of what a Smart City is. One universally accepted aspect is that ICT are the smartness enabler. With this reality in mind, some researchers have identified a minimal set of features that must fulfilled for an urban area be able to claim its smartness [8]. It is widely accepted that to be considered smart an urban area must: (1) have an explicit agenda about the development of its smartness; (2) providing open data about itself; (3) provide software services or mobile applications that may be used widely; and (4) have some kind of digital infrastructure that allow data communication in its area of actuation.

As a measurement of the order of magnitude of smart city initiatives, a study from 2017 [8] found that "more than 300 cities of all types and sizes participate in approximately 40 coalitions, thinking tanks and forums or are being advised by organizations, in their attempts to co-define their innovative future, while most of them claim to be smart, and almost all cities want to engage in this era". Smart cities aim to improve the quality of life for residents and visitors. As such, most of the city's intelligence indicators are defined with respect to the potential impact on people's lives. The study conducted in [9] lead to the observation that "a Smart City is quintessentially enabled by the use of technologies (especially ICT) to improve competitiveness and ensure a more sustainable future by symbiotic linkage of networks of peo-

ple, businesses, technologies, infrastructures, consumption, energy and spaces".

The European Union (EU) adopted an indicator that measures the city's smartness according to six dimensions [9, 10]. This indicator is also adopted here because of its relevance and because the proposed framework will be implemented in an EU city, and the dimensions are:

1. Smart Governance - relates to the degree of citizen participation in public life, transparency of governance, and the quantity and quality of public and social services.
2. Smart Economy - how competitive is the economy measured by its innovative spirit, entrepreneurship, international insertion, among other.
3. Smart Mobility - addresses concerns such as the existence of transport infrastructure considering (inter-)national and local accessibility, availability of IT infrastructure, and sustainability of the transportation system.
4. Smart Environment - what is the concern level about natural resources, including environmental conditions, air quality, sustainable resource management and ecological awareness.
5. Smart People - is about human and social capital, including the level of qualification, the promotion of lifelong learning and ethnic plurality, and the openness of minds.
6. Smart Living - is relative to the quality of life, including cultural and educational facilities, health conditions, quality of housing and tourist attractiveness, among other.

Smart City initiatives can address any combination of these six dimensions, and can be quite distinct in terms of goals and technical solutions adopted. For example, innovation in Smart Living includes transformations in construction, urban planning and renovation of facilities whereas innovation in Smart Mobility may involve tracking systems, and in Smart Governance may just entail a change in official procedures without further material investment. Overall, solutions that drive initiatives may involve cyber-physical integration of sensors and other hardware, or be primarily a software engineering challenge to generate new applications and services, or be more closely related to data science and analysis of big data. Considering the six dimensions of the EU smart city indicator, the framework proposed is aligned with the Smart Living dimension. As such, here will be described works that contribute to Smart Living and that are ready to be deployed in a real context.

2.1.1 i-SCOPE

The i-SCOPE project – interoperable Smart City services through an Open Platform for urban Ecosystems [11] covers four dimensions of the Smart City indicator: Living, Environment, Mobility, and People. In the context of the project was developed an open platform that supports three services: (1) optimization of buildings energy consumption through the precise evaluation of the potential and loss of solar energy; (2) environmental monitoring through real-time noise mapping, leveraged by the involvement of citizens, as citizens and their mobile devices can act as sensors, thus being able to measure noise levels throughout the city in a distributed way; and (3) a service more related to this work, which aims to improve the inclusion and personal mobility of elderly citizens and mobility difficulties. In this context, the platform presents a detailed urban layout, with features and barriers, to allow a precise and personalized routing service that takes into account the characteristics of individuals and the city [11]. The i-SCOPE had pilots implemented in the cities of Wien (Austria), Newcastle (United Kingdom), Indjija (Serbia), Lázio Region (Italy), and Zadar County (Croatia).

2.1.2 Amsterdam Smart City Platform

The Amsterdam Smart City Platform [12, 13] is an initiative that includes several projects developed and implemented in Amsterdam (The Netherlands). Being a comprehensive initiative, it contributes to the various profiles of the Smart City indicator. From the set of projects that contribute to the profile of Living, two are related more related to the framework discussed in this work: VITAMINE [14] and PAUL - Active Urban Living based on playful data [15].

The VITAMINE project, dubbed the Dutch word for vitamin, was created to help the elderly who are looking for a suitable lifestyle [14]. The idea is to increase the frequency of training activities by providing and supporting an additional home workout program. The absence of an exercise supervisor or instructor is mitigated by the availability of online support. In addition, it also focuses on the diet of its users and the example given is to examine whether the use of extra protein has or does not effect on exercise.

The PAUL project aims to understand how the physical activity of city dwellers can be increased using custom apps [15]. It is observed that different apps to stimulate people to exercise exercise use different strategies [6]. The way to encourage more exercise, or at least maintain the same level of activity, varies and can be motivational messages, games, or peer comparison. The project assumes that the type of app that works best for a particular person is a highly individual choice. Thus, the project studies how to make the best match between the user and the app.

2.2 mHealth

Health Apps are a good option for the elderly especially because of its mobility and portability. However, some important issues need to be addressed since older people often find it difficult to use new technologies including smartphones and respective apps [16]. However, the Internet and the use of smartphones among the elderly population will increase in the near future, as the elderly in the future will be more likely to use the Internet to search information than the elderly today. In addition, they are concerned about health and are likely to use the Internet for health-related issues [17]. Thus, there seems to be a potential interest in older adults to use mHealth apps, which is essential for these systems to be adopted by that group.

The development of new ICT applications, especially those that need mobile devices or the Internet, should help reduce the digital divide. This suggests that, in order to have an application accessible to a large group of older adults, it is appropriate to opt for a human-centered design process whose key premises are the active involvement of users in the development process as well as in the evaluation of interactive systems [18]. In addition, a human-centered design choice also appears to be a valuable approach when developing a mobile health application for other stakeholders, such as health professionals.

Experiments exist with context-aware recommender system that offers personalized recommendations of exercise routes to people according to their medical condition, and real-time information from a smart city [19]. This recommender has predefined routes and selects the best route based on a memory-based method that employs neighborhood search (to determine groups of similar users) and on information such as air quality, ultraviolet radiation, wind speed, temperature, and precipitation. The user can then select the best course according to their profile (age, effort, distance) and can inform the system about unexpected situations that could affect other users. Users can also propose new routes [19].

There are dozens of fitness apps and discussing a representative set here would not bring further insights into discussion. Nevertheless, here will be described three popular and feature-full apps selected from dozens due to their higher ratings voted by hundreds of users, for better understanding of what is being made and the real potential of such applications. The first application was developed for iOS and Android, the second one just for iOS, and the third one just for Android.

2.2.1 Pacer Pedometer & Step Tracker

This app uses a set of parameters based on the smartphone sensors and usage patterns to classify the user's lifestyle from sedentary to highly active [20]. It contains a step counting algorithm, and allows to check the performance over a certain period of time, presenting results using graphics for easier reading. The historical data is recorded and can be used to check whether or not the user has progressed against their goal, in terms of number of steps per day as well as calorie and weight loss. A set of predefined objectives are defined to give users a sense of progression. The app encourages users to exercise in case of a long period of poor activity, and it has a social component in which users, besides being able to share the number of steps taken, can also encourage friends to exercise as well.

2.2.2 The Walk: Fitness Tracker Game

This app was created in conjunction with United Kingdom's National Health Service and essencially is a game to stimulate people to walk more often [21]. The apps monitors the activities of the user to verify if the user is walking, provides several stories and users are required to move for progressing in the story. So this app is masked in a game for encouraging people to reach the adequate number of steps per day thus improving their health. As such, it motivates users to follow the story and not so much to take into account the number of steps or distance traveled.

2.2.3 Pedometer - Step Counter

The main feature of this app is a step counter. As a secondary functionality, it computes several metrics acquired through specifically designed exercises [22]. The application records the evolution of the metrics over time. Users can have training plans focused on each part of the body with the purpose of achieving their own goals more efficiently, addressing more specifically the part of the body they intend to exercise. It includes several other tests possible using the smartphone sensors such as hearing test and color blindness test.

3 Developed Work

The objective of this framework is to explain how mHealth can be further improved by making sure that health professionals are at the core of the decisions, without hindering the vibrant existing mHealth app market. Having health professionals at the core of the decisions means that users are followed by health professionals, and not just by algorithms or predefined generic reference values, and that they are in control of the recommendations passed to users having the ability to take into account the individual health history or recent changes in its health conditions. Health professionals also have access to contextual information such as weather conditions, points of interest and sidewalk conditions, among others, to obtain an accurate perspective of the real environment where users will develop their walks. They are able to recommend (new) exercises and walk paths, or introduce the necessary changes to correct less optimal situations.

3.1 Framework Description

Figure 1 depicts the proposal overview. Distinct type of mHealth apps are able to interact with the back-office application. At the top of the figure is represented a health professional analyzing the users' data and assuring that they have the best possible walk program, at the bottom of the picture are represented some users doing their exercises, and between them is a system composed by the communication networks and by the back-office.

Two health professionals helped defining the requirements of the framework, of the back-office application, and helped to define personas to guide the user experience requirements. The requirements identified for the back-office application are discuss in detail in [1]. The back-office application is composed by five main modules:

1. Health professional interface - is a web based graphical user interface that integrates information from the other components, presenting it to health professionals in an intuitive and organized way. The interface is multi-user as distinct health professionals can follow distinct sets of end-users. Also there is the ability to share users to accommodate events such as vacations or planned leaves. It is considered particularly important providing an area where health professionals can

find all essential information to support their decision-making process, making it easier, faster and less prone to errors. The two interface panels presented as Figures 2 and 3 will be explained next.

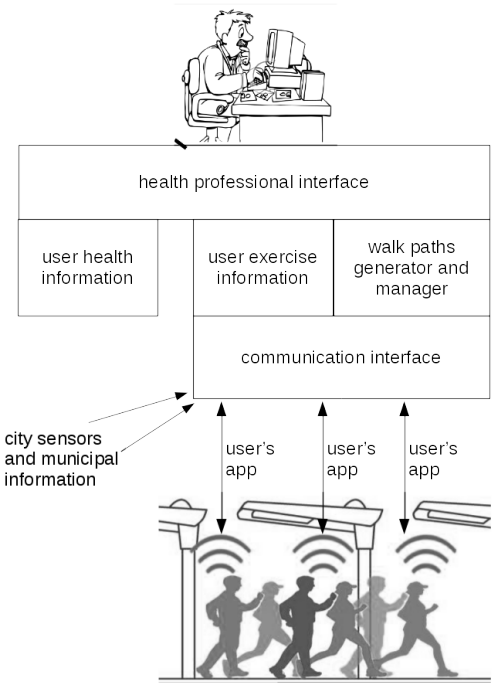


Figure 1: Depiction of the core actors of the system with details regarding the dataflow and the main system modules.

2. User health information module - manages health information that is private and sensitive. This information is never accessible to the mobile app or to any other module than the health professional interface. The informaton is stored in a specific and isolated database and it is inserted by the health professional when the user subscribes the service. This information is updated whenever the user meets with the health professional responsible for its follow up. Fig. 2 presents a screen-shot of the health professional interface relative to this module. The left panel contains generic information such as user's name, contacts, weight and height, and education. The right panel contains the user's medical record that includes heart rate, blood pressure, six minute walking test (6MWT), regular medication, and a medical record of some relevant conditions. No special concern about data security – other than stock data encryption - was taken care in the prototype development as, when deployed, these part of the system needs to comply to the specific norms of each instituion and/or can just be a reference to an already existing database.

3. Exercise information module - manages user exercise information that, being private information, is not as sensitive as the one managed by the user information module. Since this module interacts with the communication interface that, by its turn, interacts with the (third party) apps, there exists an increased factor of vulnerability. For that reason this module does not explicitly have information about the user. The user is referred by a code that does not encode personal information. This code is for the health professional interface query the user health information module about which user corresponds to a given code. When this happens the user health information module just replies if that health professional is authorize to view such data. Fig. 3 presents the health professional interface relative to the exercise information. In the left panel is visible the resume of the user, at the center a map showing the walk routes suggested to the user and, at the right panel, the amount of times the user walked those courses.

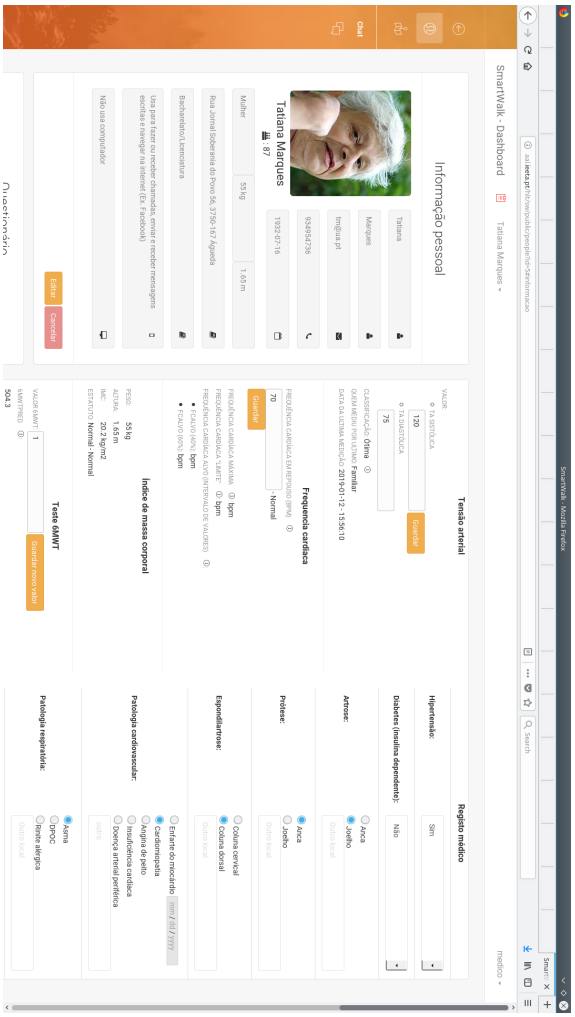


Figure 2: The back-office screen to insert and visualize health information of one user.

4. Walk paths generator and manager - this module allows to generate paths using multi-criteria [23]. It is possible ask for path of a given distance, to force paths to include given places - like rest areas or streets without cars - and also to exclude some areas such as more polluted areas or less safe. Then the algorithm optimizes the rest of the path by preferring streets with larger sidewalks, avoid the busiest roads, etc [23]. The interface of this module is a contextual menu that

pops up by right-clicking at the map of the exercise interface (see menu inside the map in Figure 3).

5. Communication interface - is handles all data received from or passed to the apps, receives data from the city sensors (temperature, wind, air quality,...) and from the municipality information systems (roads/streets under maintenance, social events,...). This interface is implemented as a RESTful Web Service and so it does not start a communication by its initiative. When the system has to pass information to someone, that information is stored and later sent when the mobile app initiates a communication with the system and ask for such information. If the health professional needs to contact the user the contact request goes by alternative ways such as short message service (SMS) or e-mail.

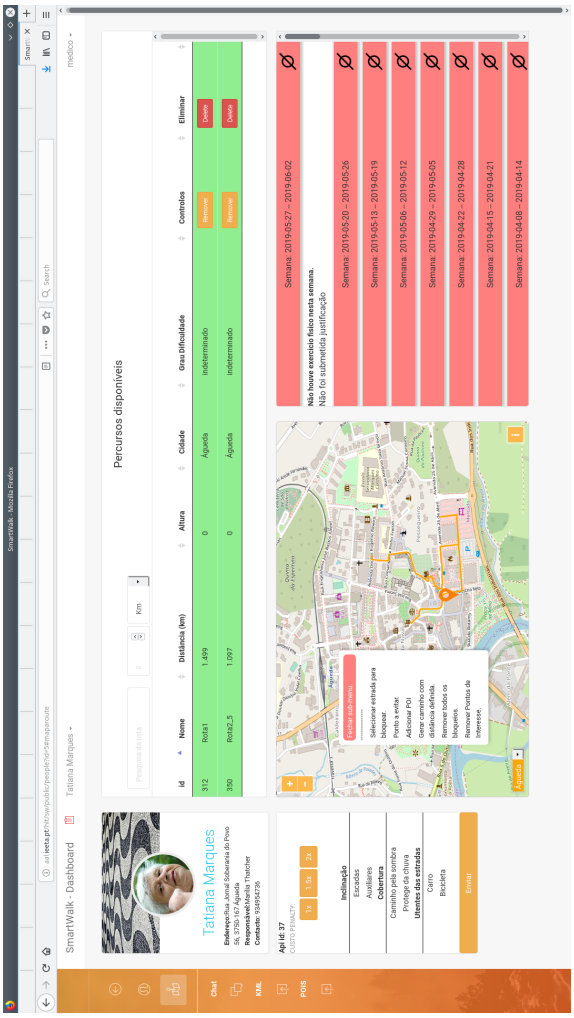


Figure 3: The back-office screen to define a walk plan for a given person, and to see how often that plan was executed.

3.2 Data Transport

The ability to allow distinct apps to interact with the back-office application implies a public definition of the data to be passed between the apps and back-office, besides the implementation of the communication interface.

The back-office sends to the app free text messages for users' information in JavaScript object notation (JSON). To send the walk paths, the back-office application generates a file in keyhole markup language (KML) format, places that file in an externally acessible universal resource location (URL), and sends that URL in the JSON message.

The data transport is made by JSON and a data package can contain one or more of the following sections: (1) user messages, (2) physical activity, (3) heart data, (4) inactivity justification.

All data packages include a code that represents the user. This code does not contain personal information as mentioned before, is used for the back-office associate the received data with the respective user. As such, in case of communication interception is not possible to identify the user just by looking to the data package.

All pieces of information have a timestamp representing when that information was generated. These timestamps are used for correctly build the events timeline at the back-office. As users can use apps without an active data link the time when the information arrives at the back-office can be quite distinct of when it was generated.

The section "User Messages" contains information written by users. It can have any free text message the user wants to share with the health professionals, namely questions regarding any concern he may have. This specific part of the package is bi-directional in the sense that the healt professional can use it to send messages (questions) to users. It also includes information about medications the user is taking. Table 1 presents an example of a message.

Table 1: Example of a "user messages" section.

```
"User Messages": [{
  "Message": {
    "Text": "Everything is OK",
    "Time Stamp": "2019-01-09  15:56:10"
  }},{
  "Medications": [{
      "Name": "Placebo 1",
      "Time Stamp": "2019-01-09  15:56:10",
      "Amount": 10,
      "Reason": "Arrhythmia"
    }, {
      "Name": "Placebo 2",
      "Time Stamp": "2019-01-09  15:56:10",
      "Amount": 4,
      "Reason": "Headache"}]}]
```

The section "Physical Activity" includes information of routes walked by users and how was the experience in terms of satisfaction, polution, etc. It includes information about how many stops the user has made during the course because it is an indicator of problems - it should be considered a stop if the user remains at the same place for a minute or more. It is also collected information about if any pain was felt and its location, the total duration of the walk, and a list of global positioning system (GPS) coordinates representing the actual path made. Table 2 presents an example of this data section.

Table 2: Example of a "physical activity" section representing that the route 305 was walked during 40 minutes with two stops, the first during 1 minute and the second for 10 minutes.

```
"Physical Activities": [{
  "Physical Activity": {
    "Active": {
      "Start": "2019-01-09 15:56:10",
      "Stop": "2019-01-09 16:36:10",
      "Route": {
        "Other": "No Comments",
        "Satisfaction": {
          "Pleasure of Course": 3,
          "Pollution Level": 2,
          "Luminosity": 8,
          "State of the tours": 6,
          "Population Density": 4
        },
        "Difficulty": 4,
        "RouteId": 305 },
      "StopDuringActivity":[{
          "Start": "2019-01-09 16:06:10",
          "Stop": "2019-01-09 16:07:10",
          "Coords": "40.575423,-8.446624"
        }, {
          "Start": "2019-01-09 16:09:10",
          "Stop": "2019-01-09 16:19:10",
          "Coords": "40.575612,-8.446191"
        }],
      "Pain Condition": {
        "Hands": 0,
        "Feet": 0,
        "Ribcage": 0,
        "Thigh": 4,
        "Head": 7,
        "Shoulders": 0,
        "Lumbar": 3,
        "Neck": 1,
        "Knee": 0,
        "Arms": 0,
        "Hip": 0 },
      "Duration": 40,
      "Path": ["40.575320,-8.446634",
               "40.575320,-8.446639",
               ...]}}}]
```

Two very important measures of effort, that have implications on the helth and well-being, are the heart rate and the blood pressure. This are reported in a "heart data" section illustrated in Table 3. These values are stored in the "exercise information" module and not in the "user health information" module since there is no guarantee that is was correctly obtained. These values are used just for monitoring and, if a consistent anomaly is verified, the next time the health professional meets the user he will try to understand the reason of the anomaly.

When some user is inactive for a long period of time, the back-office issues an alert for the health professional. If he finds appropriate it can question the user the reason of its inactivity. An example is presented in Table 4.

Table 3: Example of a "heart data" section representing that the values were obtain by a familiar. Other options could be a drug store or a health professional.

```
"User Heart Data": {
  "Blood Pressure": {
    "Diastolic": 70,
    "Systolic": 120
  },
  "Time Stamp": "2019-01-09  15:56:10",
  "Heart Rate": 68,
  "Who Measured": "Familiar" }
```

3.3 Mobile App

As mentioned before, the data transport of the framework is open and standardized for allowing third party apps to integrate the framework seamlessly. Nevertheless it was found important to provide a default app that, at least, helps to illustrate the potential of the framework.

Table 4: Example of a "inactivity" section. The main purpose is to allow helath professional understand if users were just busy and did have time to use the system or if it is a situation that need proper attention.

```
"Justification of inactivity": [{
  "Justification": {
    "Other": "None",
    "Time Stamp": "2019-01-09  15:56:10",
    "Absence": true,
    "Disease": false
  }
}, {
  "Justification": {
    "Other": "None also",
    "Time Stamp": "2019-01-09  15:56:10",
    "Absence": false,
    "Disease": true }}]
```

The devices that run the app are required to have wifi capabilities in order to send the data for the back-office and, to receive from it, the walk paths in KML format. They also need to have a GPS sensor for helping users complete the walk paths assigned to them, and acquiring information about the paths users actually did for subsequent analysis by health professionals.

The devices are not required but are strongly encouraged to have a heart beat sensor. This recommendation is due to the heart beat rate is strongly correlated with physical effort, and thus should be monitored by health professionals to assess if exercises are adequate to the given persons. It is not mandatory as heart beat rate and blood presure can be measured after the exercise by a health professional, or in a publicly available machine or at home, and then the values can be inserted manually. Even if the app is running on a device with a heart beat sensor, such as in the case of smartwatches, current technology is not accurate if a person is moving, thus the exercise should be finished before the value is measured.

3.4 Default App Implementation

Apps can run on devices such as smartphones and/or smartwatches. Apps running on smartphones have the advantage of benefiting from a bigger screen for the grphical user interface and longer energy autonomy, alongside these devices are well widespread troughtout population. Apps running on smartwatches have the advantage of benefitins from a heart beat sensor and being less intrusive as these devices are smaller. The default app runs combined on both types of devices: it runs on a smartwatch that acquires heart beat data and passes it by bluetooth low energy (BLE) to the smartphone, which combines that data with its own (GPS, user inputs, ...) and sends alltoghether to the back-office.

A default app was thus designed taking usability considerations adequate for older people [24, 25], and fully developed to show how each framework feature is accionable from the app. The user interface was designed with the support of three personas and two usage scenarios specifically developed to foster discussion and reflection [25]. Those three personas were characterized in terms of age, personality behaviour, and interests. One usage scenario was elaborated including one of the three personas while in the other scenario the other two personas interacted with which other.

The findings obtained from the personas and scenarios development lead the designed decisions of the app including: the way it was designed, how it works, and why it is easy to use for an elderly user who may not be used to the technology or has health impediments that make its use difficult [26]. The automatically app provides help if it is waiting for user's input and no activity is detected. Figure 4 presents a screen capture of such event in the home screen.

When the app does not detect an available heart beat sensor it will prompt users about the heart rate. Such form is presented in Figure 5. The fields DATE and HOUR will have as default values the current date and hour, respectivelly.

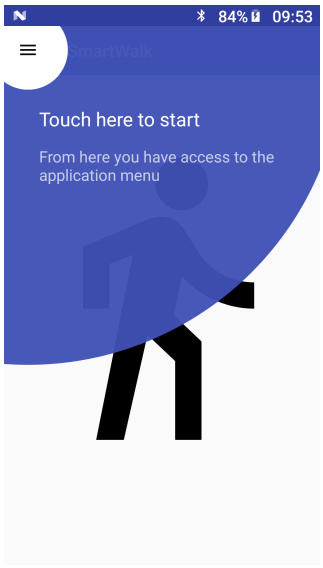


Figure 4: Screen capture of the default app home screen. The app provides contextual help when user action is required and no user activity is detected.

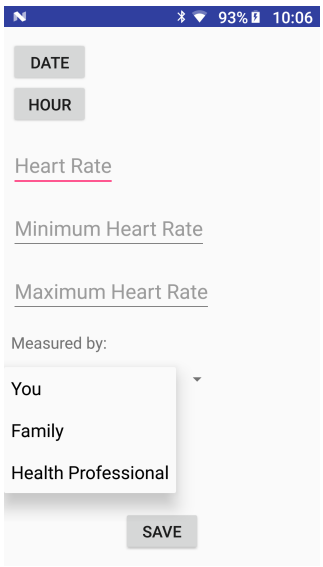


Figure 5: Screen capture of the default app. The app allows manual insertion of data if sensors are not available or if users prefer to measure themselves or ask someone else.

4 Discussion

Projects that conform to this framework can contribute to the challenges associated with the Smart Living profile. This profile is about lifestyles, behavior and consumption made through ICT, healthy and safe living in a culturally vibrant city with diverse cultural facilities, and good quality accommodation. Smart Living also promotes tourist attractiveness and high levels of social cohesion and social capital [9].

In particular, these types of systems described here can contribute to a healthy ICT-based lifestyle, as it aims to support people to maintain an adequate level of physical activity by advising them with a walk plan defined by a professional and transmitted using an ICT-based solution. In addition, health professionals, by knowing the places where people walk, can propose the most popular places to walk, thus being safer and more appealing. As it includes contextual information such as climatic conditions, points of interest, and temporary events among other, it has the potential to contribute to increase tourist attractiveness. One possibility that this type of system offers is also to propose joint walking plans, thus contributing to increase social cohesion.

Placing health professionals at the center of the decision allows individualized and quite different paths for individuals who, even if at first glance may seem to have the same profile, may have very different needs among themselves. [27] states that physical activity among the older population differs widely. According to studies that gathered information from ethnographic studies and interviews with health professionals, it is necessary to create different profiles of the elderly, with respect to physical activity and health conditions, and consequently create and adapt different routes to a specific profile. Adequate physical activity levels vary with health conditions and lifestyle [28]. Considering this, it is necessary to evaluate the health of users, namely the elderly. Thus it is of utmost importance to allow health professionals evaluate and define the profile of the user, and adapt the path of the walk to their state of health, physical activity level,

and personal preferences.

Regarding health conditions, the first thing to evaluate is whether the user needs to see a doctor before becoming more physically active. This can be reported using the Physical Activity Readiness Questionnaire - PAR-Q [29]. Then the 6-minute walk test - 6MWT [30] will be used to assess the user's level of physical activity. Other health information should be requested, such as an existing medical diagnosis, location and intensity of pain, usual medication, heart rate, and usual physical activity. With this package of health professional information can define the profile of the elderly.

A characteristic highlighted by one of the personas created was the need to make periodic evaluations face to face. For example, if a user as a health problem as breaking a leg, there is a need to make a new health assessment, or if the user leaves the city for a period of time and does not engage in any physical activity. Health information is needed to define the profile of the elderly and to customize the routes for the health conditions of the users, so it is necessary to have different routes. This also allows us to adapt the routes to personal information about the user, for example, a user who prefers routes integrated in a natural environment or if he prefers to pass in a local market. Finally, to induce and motivate the practice of physical activity, it is necessary to maintain face-to-face contact through periodic health assessment and to maintain the adaptability and personalization of the routes.

A system conforming to the proposed framework was implemented and tested with simulated users. The simulated users asked for walking courses and sent information regarding their physical activities (see Table 2), and sometimes other data relative to the other parameters, using distinct combinations of the four information sections. Users also sent information sections that were not defined and that should be silently ignored. The system handled efectivelly the distinct combinations of information sections and the information sections not defined. These results show that this design should cope well with upgrades since one major difficulty for upgrading systems is when distinct versions with distinct features coexist in the same environment. Although just one instance (and thus one version) of the back-office is expected in one usage environment, regarding the mobile applications one should expect a variety of implementations and with different versions. The system was also tested for temporary loss of connectivity and didn't show problems in such cases. The option of having a back-office connectivity module implementing a RESTful web service proved quite robust to the loss of connectivity events.

To extend the system in the future is necessary to define the JSON format of the message to be passed between the back-office and the apps, and implement the respective code and extend the database if necessary. These are the miminum things to do in any case. There is no necessity to change the lower level communication mechanisms between the back-office and the apps as the extension is just in the semantic part of the content.

In this stage of the work there is no automation (e.g. no artificial intelligence methods involved). This work was about designing and validating a monitoring platform that could be used by mobile apps with the benefit of having health professionals supervising the exercises. Without this stage consolidated this framework would be like alternatives already available: fully automated solutions not necessarily validated by health professionals.

A possible future extension would be provide some statistics as the places most visited by users, for instance for city planning purposes. This can be readilly available and it was not addressed since since third parties didn't yet identified those needs.

5 Conclusion

This article presented a framework to help people to have a more adequate physical activity, using the infrastructure of smart cities when it exists. As people age, because of their individual health history, it is more difficult to know what the proper level of physical exercise is, and therefore they should have the support of a professional with specialized knowledge and training. If people are left alone by deciding their own level of activity, they may be too aggressive and aggravate their injuries, create new ones and refrain from exercising more, or be very conservative and, in the long run, have a level of activity under their potential.

For this, information about chronic conditions, previous surgeries and current injuries, among others, is needed, raising concerns about privacy, and people are often reluctant to share these details. As such, the framework is designed to be operated by people who already know this information, or who are already legally authorized to own it, health professionals and probably the local public health unit. Also, special care was taken to make sure that health sensitive information remains private.

The back-office application was designed for health professionals be able to make fast, error-free decisions and minimize the risk of overloading them with much more work. In addition, the end user application can be made by third parties. One default app was designed and developed considering that some people are able to easily operate the smarphone and some have difficulties in using ICT, some may have problems of vision, etc. The best guidelines for the interfaces of the elderly were considered.

The project is implemented in Águeda, a city of about fifty thousand inhabitants, and ready for large-scale testing. The municipality of Águeda is a stakeholder in the project and it is expected that the project will be implemented definitively after all the tests and eventual corrections. All the technology planned for use is available in the market, such as smarphones and wristband models. This is important because the project is being deployed locally, but it has been carefully designed to be deployed (easily) in any other city.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment Not displayed due to anonymity purposes.

References

[1] M. Rodrigues, R. Santos, A. Queirós, A. G. Silva, J. Amaral, L. J. Gonçalves, A. Pereira, N. P. Rocha. "Meet smartwalk, smart cities

for active seniors." In 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW). IEEE, 2018. https://doi.org/10.1109/TISHW.2018.8559493

[2] G. Trencher, A. Karvonen, "Stretching ŝmart: advancing health and well-being through the smart city agenda", Local Environment, Taylor and Francis Group, 2017. https://doi.org/10.1080/13549839.2017.1360264

[3] T. Miura , K. Yabu, A. Hiyama, N. Inamur, M- Hirose, T. Ifukube, "Smartphone-Based Gait Measurement Application for Exercise and Its Effects on the Lifestyle of Senior Citizens". in 2015 Human-Computer Interaction, INTERACT, 2015. https://doi.org/10.1007/978-3-319-22698-9_7

[4] S. R. Steinhubl, E. D. Muse, E. J. Topol, "The emerging field of mobile health", Science Translational Medicine, **7**(283), 2015. https://doi.org/10.1126/scitranslmed.aaa3487

[5] P. Husu, J. Suni, H. Vähä-Ypyä, H. Sievänen, K. Tokola, H. Valkeinen, T. Mäki-Opas, T. Vasankari, "Objectively measured sedentary behavior and physical activity in a sample of Finnish adults: a cross-sectional study", BMC Public Health, **16**(1), 920, 2016. https://doi.org/10.1186/s12889-016-3591-y

[6] P. Simões, A. G. Silva, J. Amaral, A. Queirós,N. P. Rocha, M. Rodrigues, "Features, Behavioral Change Techniques, and Quality of the Most Popular Mobile Apps to Measure Physical Activity: Systematic Search in App Stores", JMIR Mhealth Uhealth, **6**(10), 2018. https://doi.org/10.2196/11281

[7] M. Batty, "Big data, smart cities and city planning", Dialogues in Human Geography, **3**(3), 274-–279, 2013. https://doi.org/10.1177/2043820613513390

[8] L. Anthopoulos, "Smart utopia VS smart reality: Learning by experience from 10 smart city cases", Cities, **63**, 128 – 148, 2017. https://doi.org/10.1016/j.cities.2016.10.005

[9] C. Manville, G. Cochrane, J. Cave, J. Millard, J. K. Pederson, R. K. Thaarup, A. Liebe, M. Wissner, R. A. Massink, B. Kotterink, "Mapping smart cities in the EU", European Parliament; Directorate General for Internal Policies. Policy Department Economic and Scientific policy A, 2014.

[10] M. Kullman, J. Campillo, E. Dahlquist, C. Fertner, R. Giffinger, J. Grosse, N. B. Groth, G. Haindlmaier, A. Kunnasvirta, F. Strohmayer, et al., "Note: The PLEEC Project-Planning for Energy Efficient Cities", Journal of Settlements and Spatial Planning, **5**, 89–92, 2016. https://doi.org/10.19188/09JSSPSI052016

[11] D. Patti, R. De Amicis, F. Prandi, E. D'Hondt, H. Rudolf, P. Elisei, I. Saghin, "iScope smart cities and citizens", in REAL-CORP, 2013.

[12] R. P. Dameri, "Comparing Smart and Digital City: Initiatives and Strategies in Amsterdam and Genoa. Are They Digital and/or Smart?", Smart City: How to Create Public and Economic Value with High Technology in Urban Space, Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-06160-3_3

[13] Amsterdam project, "Amsterdam Smart City platform", https://amsterdamsmartcity.com, accessed 20 May 2019

[14] J. Helder, M. Schets, "VITAMINE – Active senior citizens in Amsterdam", https://amsterdamsmartcity.com/projects/vitamine, accessed 17 May 2019.

[15] Y. Slaats, "PAUL – Playful data-driven Active Urban Living", https://amsterdamsmartcity.com/projects/playful-data-driven-active-urban-living, accessed 17 May 2019.

[16] R. Gudjonsdottir, "Personas and Scenarios in Use", PhD dissertation, KTH, Stockholm, 2010.

[17] A. Blomquist, M. Arvola, "Personas in Action: Ethnography in an Interaction Design Team", in 2002 Nordic Conference on Human-computer Interaction, NordiCHI '02, 2002. http://doi.acm.org/10.1145/572020.572044

[18] J. Pruitt, A. Tamara, "The persona lifecycle: keeping people in mind throughout product design". Elsevier, 2010.

[19] F. Casino, C. Patsakis, E. Batista, F. Borràs and A. Martínez-Ballesté, "Healthy Routes in the Smart City: A Context-Aware Mobile Recommender", IEEE Software, **34**(6), 42–47, 2017. https://doi.org/10.1109/MS.2017.4121209

[20] Pacer Health, "Pacer Pedometer & Step Tracker (Play Store)", https://play.google.com/store/apps/details?id=cc.pacer.androidapp&rdid=cc.pacer.androidapp, accessed 17 May 2019

[21] Six to Start, "The Walk: Fitness Tracker Game", https://itunes.apple.com/us/app/the-walk-fitness-tracker-game/id678971662?mt=8& app=itunes& ign-mpt=uo%3D4, accessed 17 May 2019.

[22] Guava Studio, "Pedometer – Step Counter", https://play.google.com/store/apps/details?id=com.guava.pedometer.stepcounter&rdid=com.guava.pedometer.stepcounter, accessed 17 May 2019

[23] J. Amaral, M. Rodrigues, L. J. Gonçalves, C. Teixeira, "Customized Walk Paths for the Elderly", in 2019 Information Technology and Systems, ICITS, Quito, Peru, 2019. https://doi.org/10.1007/978-3-030-11890-7_71

[24] A. I. Martins, A. Queirós, A. G. Silva, N. P. Rocha, "ICF Based Usability Scale: Evaluating Usability According to the Evaluators' Perspective About the Users' Performance", in 2016 International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI, Vila Real, Portugal, 2016. http://doi.acm.org/10.1145/3019943.3019997

[25] A. Queirós, A. G. Silva, P. Simões, C. Santos, C. Martins, N. P. Rocha, M. Rodrigues. "Smartwalk: personas and scenarios definition and functional requirements." In 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW). IEEE, 2018. https://doi.org/10.1109/TISHW.2018.8559574

[26] D. Bastos, J. Ribeiro, F. Silva, M. Rodrigues, R. Santos, C. Martins, N. Rocha, A Pereira. "SmartWalk Mobile–A Context-Aware m-Health App for Promoting Physical Activity Among the Elderly." In World Conference on Information Systems and Technologies. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-16184-2_79

[27] F. Sun, I.J. Norman, A.E. While, "Physical activity in older people: a systematic review", BMC public health, **13**(1), 2013. https://doi.org/10.1186/1471-2458-13-449

[28] I. Bautmans, M. Lambert, T. Mets, "The six-minute walk test in community dwelling elderly: influence of health status", BMC geriatrics, **4**(1), 2004. https://doi.org/10.1186/1471-2318-4-6

[29] D. Warburton, V. Jamnik, S. Bredin, D. McKenzie, J. Stone, R. Shephard, N. Gledhill, "Evidence-based risk assessment and recommendations for physical activity clearance: an introduction", Applied Physiology, Nutrition, and Metabolism, **36**(S1), 2011. https://doi.org/10.1139/h11-060

[30] R. E. Rikli, C. J. Jones, "The reliability and validity of a 6-minute walk test as a measure of physical endurance in older adults", Journal of aging and physical activity, **6**(4), 363–375, 1998.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 414-420 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Development of Wavelet-Based Tools for Event Related Potentials' N400 Detection: Application to Visual and Auditory Vowelling and Semantic Priming in Arabic Language

Nadia Mountaj[1], El-Mehdi Hamzaoui[2,*], Mohamed Majid Himmi[1], Mireille Besson[3]

[1]*LIMIARF Laboratory, University Mohammed V, Faculty of Sciences, Rabat 10000, Morocco.*

[2]*National Centre for Nuclear Energy, Science and Technology (CNESTEN), Rabat 10001, Morocco.*

[3]*Cognitive Neurosciences Laboratory, CNRS — Aix-Marseille University, Pôle 3C, Marseille 14000, France.*

ARTICLE INFO

Article history:
Received: 04 May, 2019
Accepted: 31 July, 2019
Online: 16 August, 2019

Keywords:
Daubechies
ERP signals
N400
Wavelets
Vowelling priming
Semantic priming

ABSTRACT

Neurological signals are generally very weak in amplitude and strongly noisy. As a result, one of the major challenges in neuroscience is to be able to eliminate noise and thus exploit the maximum amount of information contained in neurological signals (EEG...). In this paper, we aimed at studying the N400 wave of the Event-Related Potentials (ERPs) that may reflect the effects of vowelling and semantic priming in Arabic language. To improve the quality of the recorded ERP signals, we considered a nonlinear filtering method based on 10th order Daubechies discrete wavelet transform combined to principal component analysis (PCA). Among all tested wavelets, the Daubechies one showed high values of the used signal processing metrics. Thus, it allowed a significant enhancement of the signal to noise ratio while using only 10 ERP trials. In addition, we confirm its effectiveness while comparing the filtered outputs to those obtained using the averaging technique implemented in the conventional EEGLab toolbox. In a second step, the Mexican Hat function was used to achieve continuous wavelet analysis of the filtered signals. This time-scale analysis method permitted to get an alternative representation of the ERPs and to detect the N400 wave with significantly greater accuracy.

1. Introduction

The Event Related Potential (ERP) N400 wave is a negative deflection or component elicited by the brain as a reaction to semantically unexpected words in sentence contexts [1]. The N400 component has been found in many languages, such as English, French, Mandarin Chinese..., but to our knowledge, only a few studies have examined the occurrence of an N400 in the Arabic language [2].

The shallow or deep character (i.e vowelled or unvowelled words respectively) of written depends on how its phonology is reflected by the orthography of the language. Indeed, Arabic written words are composed from consonant and long vowels in addition to diacritics. These last reflect the vowellisation of the written word in order to enable inferring specific pronunciation. Also, Arabic is characterized by a non-concatenative morphology whereby every surface form is analyzable into a consonantal root, that conveys semantic meaning, and a word pattern (made up of vowels and of a subset of consonants) conveying morphosyntactic and phonological information.

In the present experiment, we recorded and analyzed ERP, in particular the N400 component, while participants performed a semantic judgment task with Arabic words. As many neural signals, ERPs are very weak signals and strongly corrupted by noise. Thus, previous studies aimed at improving the quality of ERP signals using statistical methods [3,4], linear and nonlinear adaptive filtering [5], neural network based techniques [6] and wavelets denoisng techniques [7,8].

In the present study, our aim was to go one step further to improve the quality of ERP signals. We used discrete wavelet combined with principal component analysis (PCA) as nonlinear filtering tools. This allowed us to enhance the signal to noise ratio and thereby to highlight the N400 component. In addition, we used

* El-Mehdi Hamzaoui, Email: hamzaouielmehdi@gmail.com

the Mexican hat function to achieve the time-scale analysis of the filtered ERPs in order to detect the N400 with more accuracy.

2. Material and methods

2.1. Experiment

In our experiment, which was approved by the Ethics Committee of the Mohammed V University, a total of 20 Master MSc and PhD students (10 women), aged between 20 and 34 years old, were tested after giving their written consent to participate in the experiment. They were all right-handed and without neurological disorders. They all use Arabic daily.

Each participant was comfortably seated in a Faradized room and was asked to silently read two words that were successively presented at the center of a computer screen. A total of 256 Arabic prime-target word pairs were used as stimuli with 128 pairs presented in the vowelled condition and 128 pairs presented in the unvowelled condition. For each condition, 64 pairs were semantically related and 64 pairs were semantically unrelated. For both vowelled and unvowelled pairs, two lists were constructed so that across lists, the same target word was paired once with a semantically related prime and once with a semantically unrelated prime [2]. The order of presentation of the two lists was balanced across participants.

EEG data was continuously recorded using 24 electrodes (impedance < 5kΩ) mounted on an elastic head cap according the 10/20 International EEG System [9]. The signals were amplified using SAI amplifiers (San Diego) and recorded at a sampling frequency of 250 Hz. The Electro-Occulogram (EOG) was recorded from an electrode placed under the right eye to detect eye blinks. Finally, two reference electrodes were placed on the left and right mastoids. The experiment was conducted in a Faraday's cage in order to reduce external interference [10].

The analyzed ERP signals correspond to the electrodes F3, F4, C3, C4, P3, P4, Fz, Cz and Pz. Previous results have shown that the N400 component is larger over centro-parietal regions of the right hemisphere than over frontal regions.

At the end of the experiment, three participants have been excluded because they present too many ocular and muscular artifacts contaminating the EEG signal.

2.2. Data analysis

Since the creation of wavelets, scientific and technical applications based on this mathematical tool have continued to be developed [11,12], exploiting their power and their efficiency for to perform multiresolution data analysis [13,14]. In the present work, we used a filtering technique developed by AminGhafari M. et al., [15] in order to denoise multivariate signals. This method combines both univariate wavelet decomposition of the signal and the principal component analysis (PCA) of the resulting wavelet coefficients in order to evaluate the correlation structure of the noise.

According to AminGhafari M. et al., [15], the algorithm performs the filtering task in four main steps. First, for a matrix X (nxp) of p observed signals, we achieve the wavelet decomposition at a defined level K. This results into two matrices D_k and A_k that contain respectively the details and approximation coefficients up to the level K of the p signals. Second, using the matrix Dj of finest details, a minimum covariance estimation is calculated and used for the diagonalization of a robust estimate of the noise covariance matrix. The obtained diagonal matrix is then used for changing the basis at each level 1<i<K. In the third step, the matrix D_k undergoes a classical one-dimensional soft thresholding. We then apply the PCA to both the detail and approximation coefficients matrices in order to choose the appropriate number of useful principal components. The best number is automatically defined using the Kaiser distance criterion which retains components associated with eigenvalues higher than the mean of all eigenvalues.

From the simplified matrices D and A, inverting the wavelet transform provides a new matrix containing the filtered signals. These lasts correspond to the main features of the original matrix X.

In our study, we tested 8 wavelet functions belonging to 5 wavelet families which are Daubechies of order 10 (db10) [16,17], Coiflets of order 5 (coif5) [18], Symlets of order 4 and 8 (sym4 and sym8 respectively) [19], Biorthogonal (bior1.1) [20] and Reverse Biorthogonal (rbio1.5) [18].

To evaluate the performance of the different wavelets, we computed the structural similarity index (SSIM), which is usually used to evaluate image quality on the basis of its luminance, contrast and structure characteristics. It is based on the comparison of an image I with a reference image [21,22]. It is known that the closer the SSIM value is to 1, the stronger the structural similarity between the evaluated image and the reference image is. By contrast, an SSIM value close to 0 indicates that there are no similarities between the 2 images [21]. To use this metric, we considered the matrix containing the wavelet based filter's output signals as an image (30 matrices of 9 analyzed electrodes x 2200 samples for each participant) and we compared it to the matrix that contained the averaged trials for each electrode.

In addition, we computed the signal-to-noise ratio (SNR) via a MATLAB® routine. This signal processing metric is commonly used to assess the performance of signal processing methods. It is often expressed in decibels as [23,24]:

$$SNR = 20.log_{10}(\frac{P_x}{\sigma^2}) \quad (1)$$

where P_x and σ^2 denote the power density of the original and the noise signals respectively.

In the case of closer SNR values, we also used the mean square error (MSE), given by the following equation 2, as second metric to evaluate the accuracy of the chosen wavelet. The value of the MSE, the closer the filtered signal is to the original one and thus the better filtering method is [6,10,21,23].

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(x(n) - \hat{x}(n))^2 \quad (2)$$

where $x(n)$ and $\hat{x}(n)$ denote the original and the filtered signals respectively. N is the length of x(n).

In the second step, we used the continuous wavelet transform (CWT) as an alternative method to the classical signal time representation. This is based on the idea that wavelet analysis can provide accurate and specific time-frequency decomposition of

neurologic signals. This method has already been applied to EEG denoising [18,24,25], ERP component separation [26], spindle and spike detection [27,28,29], etc. It allowed an automatic processing of the signal and provided both qualitative and quantitative information.

The continuous wavelet transform converts a continuous signal into extremely redundant signal of dual continuous variables which are the translation and the scale. The resulting changed signal is easy to interpret and valuable for time-frequency or time-scale analysis [30]. In general, CWT of a signal s(t) is defined as:

$$S(\alpha,\beta) = \frac{1}{\sqrt{\alpha}}\int_{-\infty}^{+\infty} \psi\left(\frac{t-\beta}{\alpha}\right)^{*} . s(t)dt$$

where α and β represent the scaling and the dilation factors respectively, whereas ψ* is the complex conjugate of the mother wavelet function.

The scale α is associated to the wavelet's central frequency Fc and varies according to frequencies (in Hz) by:

$$\alpha = \frac{F_c}{f * \Delta T} \quad (3)$$

ΔT is the sampling period.

For time-scale analysis, we represented the modulus of the CWT coefficients, which corresponds to the energy density of the analyzed signal, according to $\log_2(\alpha)$ (ordinate axis) and time (abscissas axis) [12]. In this graphical representation, called scalogram, a color map is used to quantify the energy density of the transformed signal. The highest value of the energy corresponds to the white color whereas the lowest is represented by black color [31,32,33].

In our application, ψ is chosen to be the Mexican Hat (equation 4), which is the second-order derivative of the Gaussian function. This last is not a wavelet, but all its derivatives can be used as wavelets, particularly the first and the second derivatives [34,35]. In practice, the Mexican Hat is expressed by the real function of equation (5) and represented by the figure 1 below.

$$\psi(t) = \frac{2}{\frac{1}{\pi^4}\sqrt{3\delta}}\left(\frac{t^2}{\delta^2} - 1\right) exp\left(-\frac{t^2}{2\delta^2}\right) \quad (4)$$

where 0<δ<1 represents the wavelet width [35]

$$\psi(t) = (1-t)exp\left(-\frac{t^2}{2}\right) \quad (5)$$

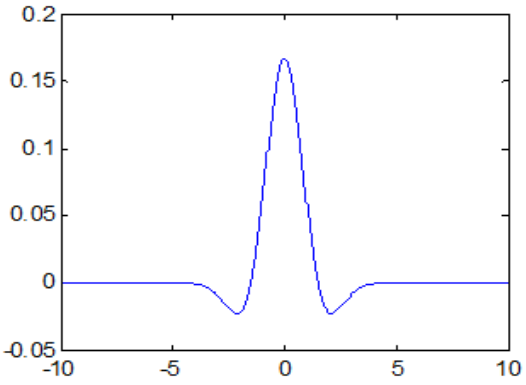


Figure 1. Plot of real function of the Mexican Hat function.

As illustrated by this figure, the Mexican Hat waveform looks like most of the waves that compose the ERPs (Figure 2). This was the reason to use it to perform the time-scale analysis of our data set. In addition, this function is easy to implement under the MATLAB environment.

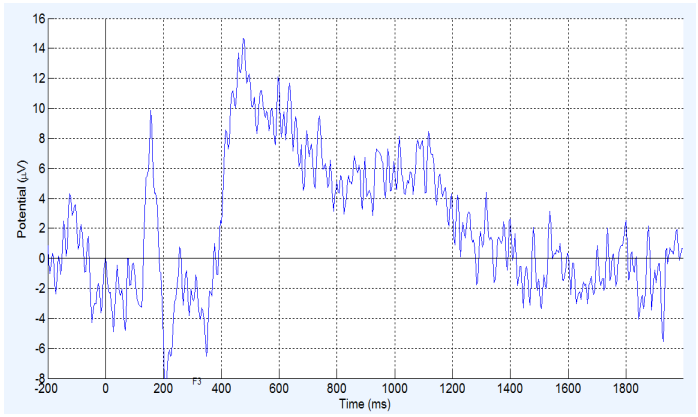


Figure 2. Example of recorded ERP.

3. Results and discussion

The averaged values of the performance metrics, SNR and MSE, corresponding to the application of the Aminghafari's algorithm with different DWT functions to our signals are presented in Table 1. As described in the Materials and Methods section, we recorded 30 ERP signals per electrode from each participant (9 electrodes and 20 participants). We processed 5, 10 and 15 trials to evaluate the efficiency of the filtering algorithm and thus to define the most accurate DWT function that allowed a good filtering of the ERPs. Results showed that the 10th-order Daubachies wavelet (db10) yielded the best results in terms of accuracy of de-noising ERP signals. In fact, this DWT function shows the highest SNR values and the lowest MSE values in all test conditions.

Importantly, and as presented in Table 2, db10 showed the highest values of the structural similarity index for all electrodes. In addition, and as illustrated on Figure 3, its application to filter the ERPs showed good improvement of the visual quality of waveforms' plots.

Visual comparison of these results to those obtained using the classical averaging method implemented in the EEGLab Toolbox showed that, for each electrode, the same visual quality of plots is obtained via our method and via EEGLab when averaging all recorded signals for SNR values of about 3.5. However, using our db10-PCA filtering method, only 5 to 10 ERP trials are sufficient to improve the SNR value and to highlight the occurrence of an N400 component. This can be explained by the fact that our filtering method is a nonlinear technique that takes into consideration some statistics of the signal and how the noise is affecting the signal. By contrast, the averaging method is a linear method that assumes that the noise is additive, white and Gaussian.

In order to still improve accuracy in N400 waveform detection, a time-scale representation of the ERP signals was performed on the basis of the Mexican hat CWT function. An example of results obtained for the 9 recorded electrodes is illustrated on Figure 4.

Table 1: Averaged values of metrics obtained using 5, 10 and 15 EEG trials.

Wavelet Name		5 ERP trials		10 EEG trials		15 EEG trials	
		MSE	SNR	MSE	SNR	MSE	SNR
Daubachies	db10	62,84	7,90	74,96	7,10	68,61	6,13
Coiflets	coif5	70,19	6,64	92,70	6,50	79,73	5,66
Symlet	sym4	67,63	7,19	86,94	6,77	78,32	5,66
	sym8	68,64	7,20	89,95	6,63	78,05	5,75
Mery	dmey	102,60	5,42	141,43	5,82	114,99	4,49
Biorthogonal wavelet	bior 1.1	73,68	6,59	84,64	6,38	85,34	5,08
Reverse biorthogonal	rbio 1.5	67,19	7,26	88,92	6,65	77,26	5,77

Table 2: Averaged values of structural similarity index metric.

Wavelet Name		Structural similarity index metric								
		F3	F4	C3	C4	P3	P4	Fz	Cz	Pz
Daubachies	db10	0,97	0,99	0,95	0,98	0,93	0,94	0,99	0,98	0,96
Coiflets	coif5	0,54	0,50	0,52	0,50	0,58	0,49	0,55	0,60	0,59
Symlet	sym4	0,28	0,26	0,23	0,28	0,3	0,31	0,29	0,23	0,31
	sym8	0,25	0,30	0,27	0,31	0,27	0,28	0,31	0,29	0,32
Mery	dmey	0,08	0,11	0,09	0,10	0,07	0,14	0,08	0,15	0,08
Biorthogonal wavelet	bior 1.1	0,18	0,22	0,20	0,17	0,21	0,22	0,26	0,22	0,23
Reverse biorthogonal	rbio 1.5	0,27	0,35	0,25	0,33	0,28	0,31	0,36	0,33	0,27

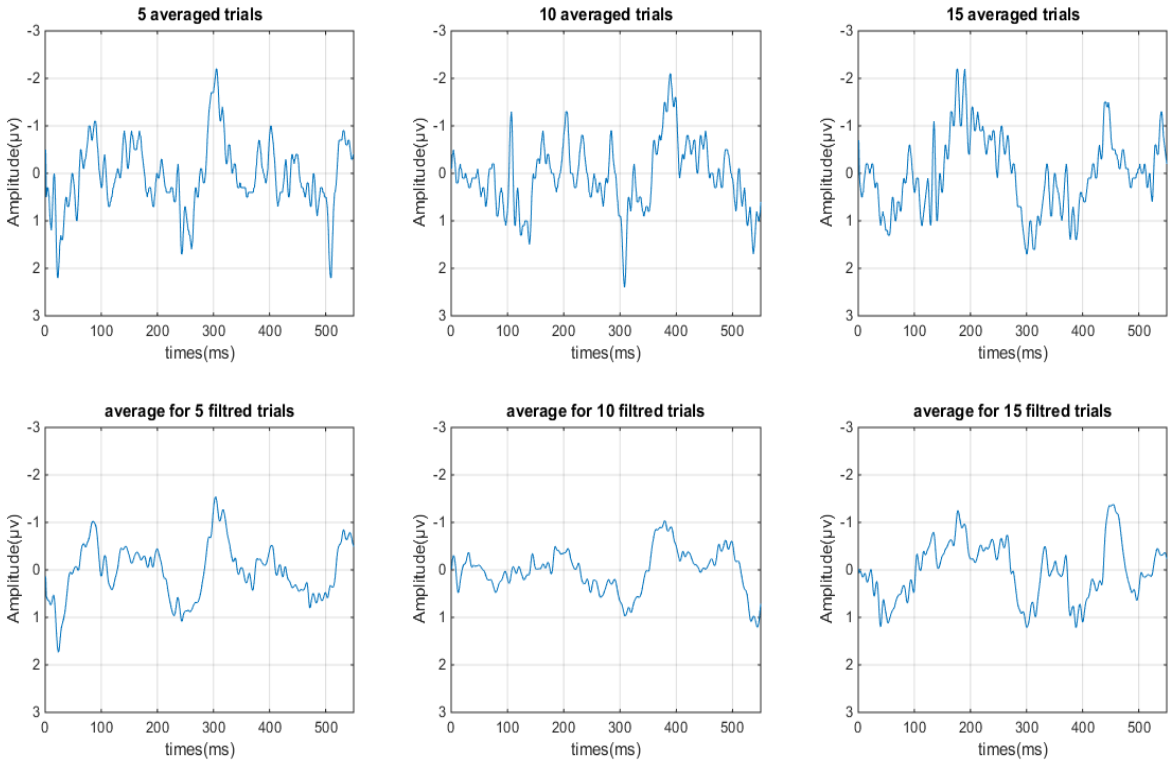


Figure 3. 5, 10 and 15 ERP trials filtered using both EEGLab Averaging technique (Up) and the developed db10-PCA filtering method (down).

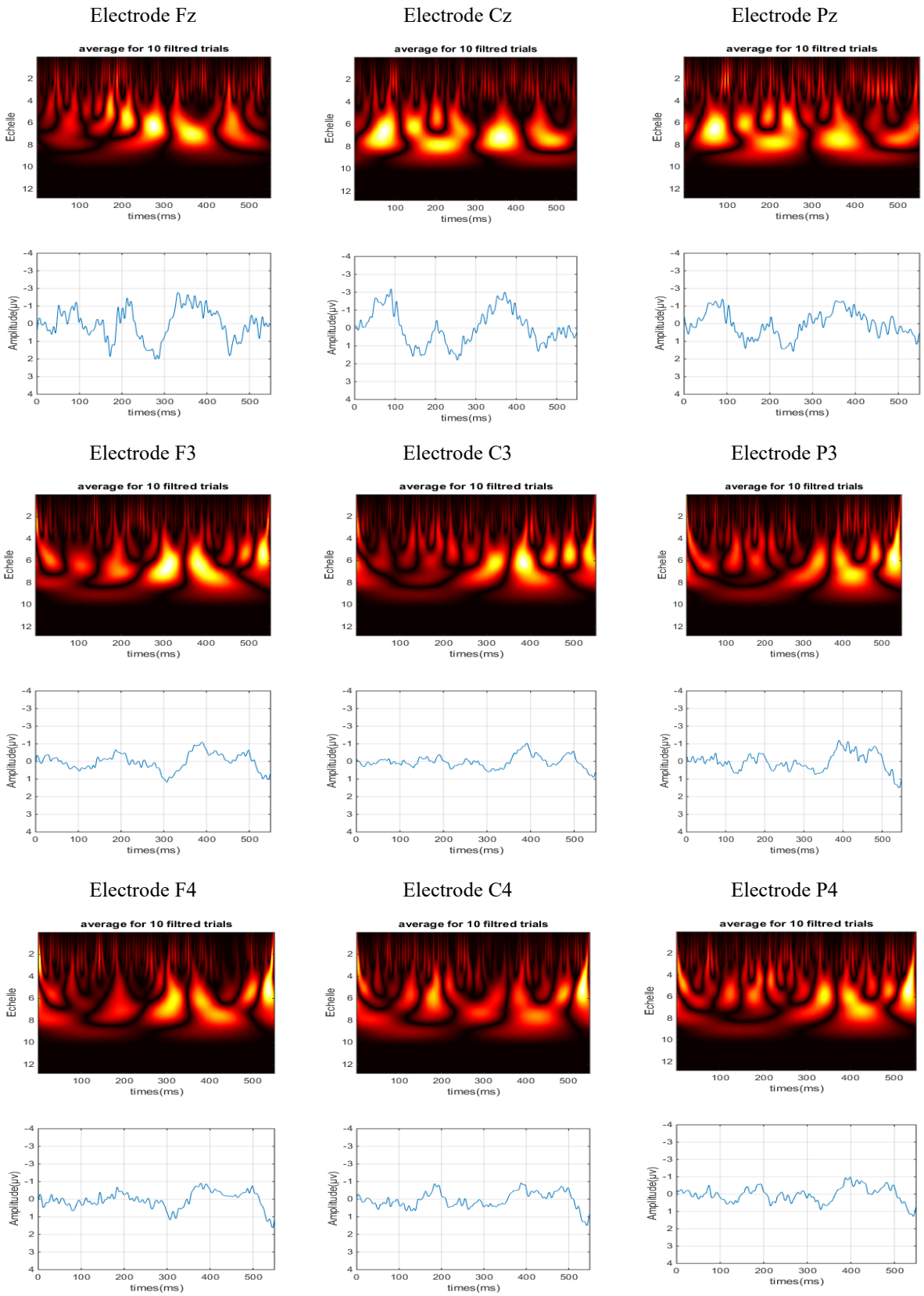


Figure 4. Scalograms of the average of 10 filtered ERP trials using db10-PCA.

Visual reading of these scalograms revealed the presence of light-colored vertical bands for medium and small scales, depicting smaller changes throughout the ERP signal. In particular, high-energy concentrations are observed in the time range of the N400 component [360; 470msec] and for scales between 3 and 8. Importantly, these regions correspond to a maximum energy around the position of the N400 component in the temporal representation of the ERP signal.

Based on these results and similar ones obtained by processing the entire set of data available for this study, the maximum energy localization in time and scale allowed us to define a qualitative criterion to detect the N400 occurrence and position. Moreover, we have demonstrated in our previous scientific works, that the primed Arabic words elicit smaller N400 components than unprimed Arabic words [2]. This result is confirmed by the use of the CWT technique. Indeed, for unprimed Arabic words, a high energy band is detected around 380 to 410 millisecond, that is, in a narrow range of about 30 milliseconds, whereas for primed words, this energy region is of very low intensity.

4. Conclusion

In this study we exploited the wavelets, as signal processing tools, to improve the quality of the ERP signals recorded during a semantic priming task in Arabic. We used the discrete wavelet transform to denoise the recorded signals. Moreover, the filtering procedure combined the DWT and the PCA methods to form a nonlinear filter that allowed us to improve the visual quality of ERP plots based on a few trials only. Importantly, using the SSIM, the SNR and MSE metrics, we demonstrated that 10th-order Daubachies wavelet of order 10, was the most efficient to improve the SNR and thus to reveal the occurrence of the N400 component. Finally, visual comparison with the results obtained using EEGLab tools, showed that our procedure allowed a clear improvement of the quality of the ERP plots.

In a second part, we used the continuous wavelet transform based on the Mexican Hat function to perform the time-scale analysis of the filtered ERPs. The resulting scalograms allowed us to define qualitative and quantitative criteria to detect the presence of the N400 component in the auditory and visual evoked signals. The qualitative criterion consists of visual reading of the energy density representation whereas the quantitative criterion is based on defining with acute precision the maximum of the N400 energy within time and scale positions. In this case, we found that, for unprimed Arabic words, a high energy band is detected between 3 and 8 on scale axis and in the time range [380; 410msec]. This last corresponds to the normal position of the N400 wave in the temporal representation of an ERP. In the case of primed Arabic words, a very low intensity energy region is present in the time range [360; 470msec].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by the French-Moroccan Research Consortium (GDRI-Neuro), coordinated by Dr Driss Boussaoud and supported by the French Government and the French Institute of Life Sciences.

References

[1] M. Kutas, S. A. Hillyard, “Reading senseless sentences: brain potentials reflect semantic incongruity” Science, 207(4427), 203-205, 1980.
[2] N. Mountaj, R. El Yagoubi, M. Himmi, F.L. Ghazal, M. Besson, S. Boudelaa, “Vowelling and semantic priming effects in Arabic” Int. J. Psychophysiol., 95(1), 46-55, 2015.
[3] G. Inuso, F. La Foresta, N. Mammone, F.C. Morabito, “Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings” In 2007 IEEE International Joint Conference on Neural Networks, Orlando, FL, USA, 2007.
[4] D. Kang, L. Zhizeng, “A Method of Denoising Multi-channel EEG Signals Fast Based on PCA and DEBSS Algorithm” in 2012 IEEE International Conference on Computer Science and Electronics Engineering (ICCSEE), Hangzhou, Zhejiang, China, 2012.
[5] A. Garcés Correa, E. Laciar, H.D. Patiño, H.D. Valentinuzzi, “Artifact removal from EEG signals using adaptive filters” in 2007 16th Argentine Bioengineering Congress and the 5th Conference of Clinical Engineering, San Juan, Argentina, 2007.
[6] G. Kaushik, H. Sinha, L. Dewan, “Biomedical signals Analysis by Dwt Signal Denoising with Neural Networks” J. Theor. & App. Inf. Tech., 62(1), 184-198, 2014.
[7] M. Balamareeswaran, D. Ebenezer, “Denoising of EEG signals using Discrete Wavelet Transform based Scalar Quantization” Biomed. & Pharmacol. J., 8(1), 399-406, 2015.
[8] R. Princy, P. Thamarai et B. Karthik, “Denoising EEG signal using wavelet transform” Int. J. of Adv. Res. in Comp. Eng. & Tech., 4(3), 1070-1074, 2015.
[9] H. Jasper, “The ten-twenty electrode system of the International Federation” Electroenceph. Clin. Neurophysiol., 10, 371-375, 1958.
[10] N. Fathima, K. Umarani, “Reduction of Noise in EEG Signal using Faraday’s Cage and Wavelets Transform: A comparative Study” Int. J. Eng. Sci. & Comp., 6(7), 8566-8569, 2016.
[11] S.P. Suhas, K.P. Minal, “Quality advancement of EEG by wavelet denoising for biomedical analysis” in 2012 IEEE International Conference on Communication, Information and Computing Technology (ICCICT), Hammamet, Tunisia, 2012.
[12] P. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, Bristol: Institute of Physics Publishing, 2002.
[13] H. Li, H. Hu, T. Kobayashi, T. Saga et N. Taniguchi, “Wavelet Multi-Resolution Analysis of Dual-plane Stereoscopic PIV Measurement Results in a Lobed Jet” in 2001 4th International Symposium on Particle Image Velocimetry, Göttingen, Germany, 2001.
[14] A. Drissi, E. Maliani, M. El Hassouni, Y. Berthoumieu, D. Aboutajdine, “multi-model approach for multi component texture classification” in 2012 5th International Conference on Image and Signal Processing (ICISPP12), 2012.
[15] M. Aminghafari, N. Cheze, J. Poggi, “Multivariate denoising using wavelets and principal component analysis” Comput. Stat. Data An., 50(9), 2381-2398, 2006.
[16] D. Sundararajan, Discrete wavelet transform: a signal processing approach, John Wiley & Sons, 2016.
[17] I. Daubechies, Ten lectures on wavelets, SIAM, 1992.
[18] S. Mallat, A wavelet tour of signal processing, Academic Press, 1999.
[19] M.S. Chavan, N. Mastorakis, M.N. Chavan, M.S. Gaikwad, “Implementation of symlet wavelets to removal of gaussian additive noise from speech signal” in 2011 Joint WSEAS International Conferences on Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics, Cambridge, UK, 2011.
[20] C. Uyulan, T.T. ERguzel, “Comparison of Wavelet Families for Mental Task Classification” J. Neurobehavioral Sci., 3(112), 2016.
[21] R. Dosselmann, X.D. Yang, “A comprehensive assessment of the structural similarity index” Signal, Image Video P., 5(1), 81-91, 2011.
[22] A. Hore, D. Ziou, “Image quality metrics: PSNR vs. SSIM” In 2010 IEEE 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010.
[23] A. Sundar, V. Pahwa, C. Das, M. Deshmukh, N. Robinson, “A Comprehensive Assessment of the Performance of Modern Algorithms for Enhancement of Digital Volume Pulse Signals” Int. J. Pharma. Med. Biolo. Sci., 5(1), 91-98, 2016.
[24] A. Dixit, S. Majumdar, “Comparative analysis of coiflet and daubechies wavelets using global threshold for image de-noising” Int. J. Adv. Eng. Tech., 6(5), 2247, 2013.
[25] P. Khatwani, A. Tiwari, “Removal of Noise from EEG Signals Using Cascaded Filter – Wavelet Transforms Method” Int. J. Adv. Res. Electrical, Electronic Instr. Eng., 1(21-23,12), 2014.

[26] C. Yongjian, "Neural Network Based EEG Denoising" in 2008 IEEE 30th Annual International EMBS Conference, Vancouver, British Columbia, Canada, 2008.
[27] L. Yu, "EEG De-Noising Based on Wavelet Transformation" In 2009 IEEE 3rd International (ICBBE 2009), Beijing, China, 2009.
[28] J. Raz, L. Dickerson, B. Turetsky, "A wavelet packet model of evoked potentials" Brain and Language, 66(1), 61-88. 1999.
[29] T. Kalayci, O. Ozdamar, N. Erdol, "The use of wavelet transform as a preprocessor for the neural network detection of EEG spikes" in 1994 IEEE Southeastcon Conference, Miami, FL, USA, 1994.
[30] S.J. Schiff, J. Heller, S.L. Weinstein, J. Milton, "Wavelet transforms and surrogate data for electroencephalographic spike and seizure detection" Optical Eng., 33(7), 2162-2170, 1994.
[31] Z. W. Tang, and N. Ishii, "The recognition system with two channels at different resolu-tion for detecting spike in Human's EEG" IEICE Trans. Info. Sys., 76(3), 377-387, 1993.
[32] T. Kristjansson, J. Hershey, "High Resolution Signal Reconstruction" in 2003 IEEE Workshop on Automatic Speech Recognition and Understanding, St Thomas, VI, USA, 2003.
[33] T. Perez-Muñoz, J., Velasco-Hernandez, E. Hernandez-Martinez, "Wavelet transform analysis for lithological characteristics identification in siliciclastic oil fields" J. Appl. Geophys., 98, 298-308, 2013.
[34] Z. Zhou, H. Adeli, "Time-frequency signal analysis of earthquake records using Mexican hat wavelets" Comput-Aided Civ. Inf., 18(5), 379-389, 2003.
[35] E-M. Hamzaoui, F. Regragui, "Discrimination of Visual Evoked Potentials Using Image Processing of Their Time-Scale Representations" in 2013 2nd Conference in Electronics, Telecommunications, and Computer Engineering (CETC'13), Lisbon, Portugal, 2013
[36] R. M. Rao, AS. Bopardikar, Wavelet Transforms, Addison-Wesley, 1998.
[37] S.J. Schiff, "Resolving time-series structure with a controlled wavelet transform" Optical Engineering, 31(11), 2492–2496.
[38] G. Samorodnitsky, M. Taqqu, Stable non-Gaussian random processes, Chapman and Hall, 1994.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 421-430 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Balanced RF Power Amplifier using the Improvement 3-dB Quadrature Hybrid Couplers for Mitigation of the Reverse Intermodulation in FM Broadcasting Systems

Panya Hantula, Rangsan Tongta*

School of Telecommunication Engineering, Institute of Engineering, Suranaree University of Technology, 111 University Avenue, Muang District, Nakhon Ratchasima, Thailand.

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 29 July, 2019
Online: 16 August, 2019

Keywords:
Reverse Intermodulation
FM Power Amplifiers
Balanced RF Power Amplifiers
3-dB Quadrature Couplers
Broadside-coupler Striplines
Printed Circuit Broads (FR-4)

ABSTRACT

FM broadcast stations generate reverse intermodulation signals that cause communication problems between ATC controllers and pilots in the Air Traffic Control systems (ATC) in Thailand. In this paper proposes a balanced RF power amplifier using the improvement 3-dB quadrature hybrid couplers to reduce the reverse intermodulation of FM broadcasting systems. The mathematics analysis of the balanced RF power amplifier for two closely located FM stations shows that can reduce the reverse intermodulation products. A reverse signal from nearby FM stations travels into two amplifiers of a balanced amplifier that create intermodulation signals. The quadrature hybrid coupler in a balanced RF amplifier to provide intermodulation signals of two amplifiers have a 180˚ out-of-phase which cause reducing intermodulation signals. In this paper designs 3-dB quadrature hybrid couplers for a proposed RF balanced amplifier. The low-cost PCBs are used to make prototype couplers. The proposed couplers are provided experimental results of return losses that are lower than -15 dB. The directed port has a coupling coefficient -3.3 dB and the coupled port has a coupling coefficient -3.4 dB. The prototype couplers are applied to a balanced amplifier that obtains a maximum output power 210 watts with 17.20 dB gain. A proposed balanced RF amplifier is measured the reverse intermodulation products of with a situation of two closely located FM broadcast stations to compare with a single-stage class-C amplifier. The experimentation demonstrates of the transmitter using a balanced RF power amplifier that can reduce an intermodulation product 18.47 dBc.

1. Introduction

The RF power amplifiers are important components for wireless communication systems. The design techniques for RF amplifier can use classes A, B, C, and AB, etc. Each type of class-amplifiers has different features and performances. The class-C amplifiers are suitable for FM broadcasting systems because it has the output power and efficiency higher than other types. The disadvantage of class-C power amplifiers is non-linearity which creates harmonic and intermodulation signals. Harmonic signals are multiple of the fundamental frequency. It can be eliminated by using the Low Pass Filters. Intermodulation signals are unwanted signals of non-linear RF power amplifiers. When the power amplifier of FM station is injected by two frequencies which are generated intermodulation signal to fall into the other frequency ranges. The intermodulation products can be calculated from $2f_2$-f_1 and $3f_2$-$2f_1$ [1].

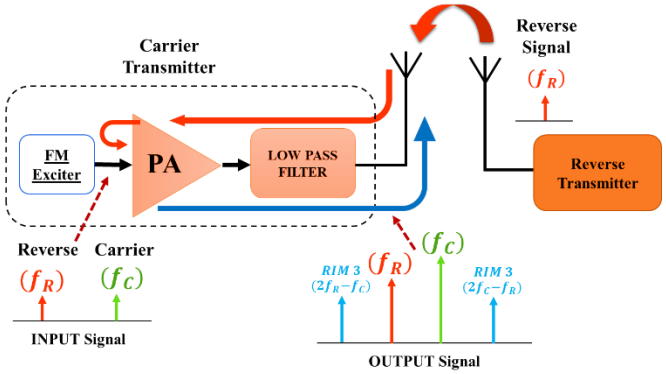


Figure 1: The reverse intermodulation products of FM broadcast stations

*Rangsan Tongta, Engineering, Institute of Engineering, Suranaree University of Technology. 111 University Avenue, Muang District Nakhon Ratchasima, Thailand. Email: tongta@sut.ac.th

Although, the FM broadcast station uses an RF power amplifier with a single frequency. In practice, the signals from nearby FM broadcast stations can radiate to the own station. In Figure 1, two closely located FM broadcasting stations, the signal from a reverse transmitter radiates into a carrier transmitter via an antenna and transmission line in the reverse direction, called "Reverse Signal". The reverse signal travels until it present at the input of a carrier transmitter. So, the station own signal and the reverse signal create intermodulation signals, called "Reverse Intermodulation (RIM)".

The FM broadcast station uses a low pass filter in the final stage to block unwanted signals, such as harmonic and intermodulation signals. A low pass filter has a wide transition band, the frequency response to change from a pass-band to a stop-band. Which frequencies of intermodulation products are very close to a cut-off frequency point of the filter that cannot eliminate intermodulation products with a low pass filter. The FM broadcast station using the frequency band is 88-108 MHz that are close to the frequency band of Air Traffic Control (ATC) systems, using the frequencies between 108-137 MHz. Therefore, reverse intermodulation signals from FM station generates may be interferences communication between air traffic controllers and pilots, as shown in Figure 2. [2]

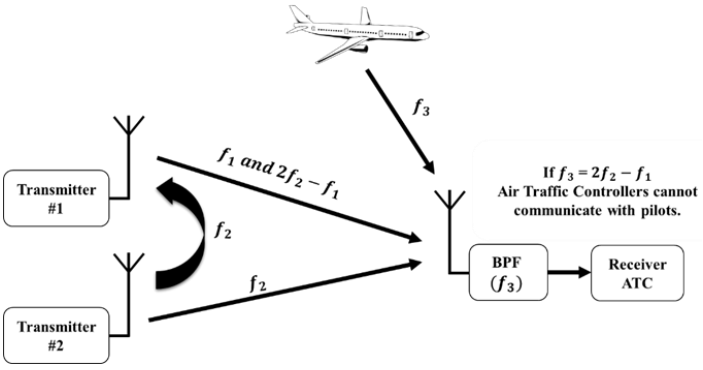


Figure 2: Intermodulation products interference to Air Traffic control systems

To overcome this problem, circulators or cavity bandpass filters are alternatives to reduce intermodulation products. It can block reverse signals from nearby stations and unwanted signals from the own station. Disadvantages of circulators and cavity bandpass filters are adding considerable cost, complexity and power loss to a system. As will be shown, none of the techniques available to reduce reverse intermodulation are particularly attractive.

This paper presents a balanced RF power amplifier using the improvement 3-dB quadrature hybrid couplers to reduce reverse intermodulation signals of FM broadcast station. The concept of balanced RF power amplifiers is structured from two identical quadrature couplers and two signal-stage amplifiers. The reverse signal from nearby FM stations radiates to the own station until it present at the input of two signal-stage amplifiers that generated intermodulation signals. The second quadrature hybrid coupler provides a 180˚ out-of-phase of two intermodulation signals from two amplifiers that can cancel intermodulation signals. The mathematics analysis of a balanced RF power amplifier for two closely located FM stations shows that can reduce the reverse intermodulation products, will be explained in section 3. In the previous work presents a design and implementation of the prototype 3-dB quadrature couplers [3]. In this work, prototype 3-dB quadrature couplers are included in a structure of a balanced RF power amplifier to reduce reverse intermodulation signals of FM broadcast station.

The quadrature hybrid couplers are an important component for a balanced RF power amplifier. Due to it can separate an input signal into two identical amplifiers and combines both output signals. The quadrature hybrid coupler has a very beneficial feature that can reduce a reverse intermodulation signal of FM broadcasting systems. The branch-line hybrid couplers are applied to traditional couplers for the balanced rf power amplifier. The branch-line is a simple type of quadrature couplers that are realized by using four transmission lines to construct a single-box coupler [4]-[5]. Each line in the single-box is a quarter-wavelength transmission line. The disadvantage of branch-line couplers is a large physical size because a quarter-wavelength transmission line determines the physical size of branch-line couplers. Especially, FM broadcasting systems use a low-frequency range that is a long transmission line that increases physical size. The large physical size is unsuitable to use for an RF power amplifier of the FM stations [6]-[10]. The quadrature hybrid couplers in this paper are structured by using the broadside-coupler stripline technique to reduce the size of couplers. Section 4 describes the design, implementation of quadrature hybrid couplers.

The prototype quadrature couplers are applied to a balanced RF power amplifier. The two identical single-stage amplifiers use a class-C RF amplifier with two MRF151 N-channel MOSFETs and operate at the frequency range of 88-108 MHz. Experimental results of the total output power, a frequency response of prototype balanced power amplifier will be presented in section 5. The situation of two closely located FM broadcast stations is used to create the reverse intermodulation signal. Two carrier transmitters that use a single-stage power amplifier and a balanced RF power amplifier have been measured. The FM broadcast station using a balanced power amplifier eliminates the intermodulation signals by using the phase-shifting property of the quadrature hybrid coupler. The reverse signals (f_R) from nearby FM stations and the intermodulation signals have a 180˚ out-of-phase which can be canceled the reverse signals and RIM signals without the cavity bandpass filters and circulators. Experimental results of the FM broadcast station using a balanced RF power amplifier comparing with single-stage RF power amplifier are presented in sections 6.

2. RF Power Amplifier for FM Broadcasting Systems

2.1. Class-C Power Amplifiers

The RF power amplifiers are an electronic circuit that transfers an input power to load such as the antenna by amplifying. FM broadcasting systems require high output power and efficiency. The Class-C amplifier offers the highest efficiency more than Class-A, Class-B, and Class-AB. The efficiency of Class-C amplifiers can be approximated about 70-80%. The disadvantage of Class-C amplifiers is non-linearity which has poor distortion qualities and generates harmonics [11]. When two signals are the input of an amplifier that creates intermodulation components (IM). The several techniques are presented reducing the intermodulation distortion by improved the RF power amplifiers, such as the feed-forward techniques [12], etc. They are

suitable to the multicarrier amplifiers which cannot eliminate the reverse signal from nearby FM broadcast stations.

2.2. Non-linear Distortion and Intermodulation Products

Non-linear distortions can be described in terms of a non-linear relationship between the input and output signals. The non-linearity creates several effects such as harmonic and intermodulation signals. Harmonic signals are unwanted signals which are included a number of integer multiples of an input frequency. Intermodulation signals occur when an input of non-linear systems are composed of two frequencies. FM broadcast stations use a Class-C amplifier that is a nonlinear system. Therefore, we can express by:

$$y(t) = \alpha_1 x(t)+\alpha_2 x(t)^2+\alpha_3 x(t)^3+\alpha_3 x(t)^4+\ldots \quad (1)$$

Where $y(t)$ represents the output, $x(t)$ represents the input and α_n represents the gain coefficient.

$$x(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) \quad (2)$$

If an input signal consists of two signals with different frequencies that are applied to a nonlinear system (1). The resultant of the output can be expressed as:

$$\begin{aligned} y(t) = {} & \left[\frac{\alpha_2 A_1^2}{2}+\frac{\alpha_2 A_2^2}{2}\right]+ \\ & \left[\alpha_1 A_1+\frac{3\alpha_3 A_1^3}{4}+\frac{3\alpha_3 A_1 A_2^2}{2}\right]\cos(2\pi f_1 t)+ \\ & \left[\alpha_1 A_2+\frac{3\alpha_3 A_2^3}{4}+\frac{3\alpha_3 A_1^2 A_2}{2}\right]\cos(2\pi f_2 t)+ \\ & \left[\frac{\alpha_2 A_1^2}{2}\right]\cos(2\pi 2f_1 t)+\left[\frac{\alpha_2 A_2^2}{2}\right]\cos(2\pi 2f_2 t)+ \\ & \left[\frac{3\alpha_3 A_1^2 A_2}{4}\right]\cos\left(2\pi t(2f_1+f_2)\right)+ \\ & \left[\frac{3\alpha_3 A_1 A_2^2}{4}\right]\cos\left(2\pi t(2f_2+f_1)\right)+ \\ & \left[\frac{3\alpha_3 A_1^2 A_2}{4}\right]\cos\left(2\pi t(2f_1-f_2)\right)+ \\ & \left[\frac{3\alpha_3 A_1 A_2^2}{4}\right]\cos\left(2\pi t(2f_2-f_1)\right)+\ldots \end{aligned} \quad (3)$$

From (3), the output that does not contain only the fundamental frequency. The nonlinearity gives other frequencies adding to output components. The second term of this expression shows the amplitude of the fundamental frequencies. The third term shows the second-order harmonics. The fourth and fifth terms show third-order intermodulation distortions. The terms of $2f_1$-f_2, $2f_2$-f_1 are third-order intermodulation products. The intermodulation signals are very close to fundamental frequencies which are cannot be easily filtered.

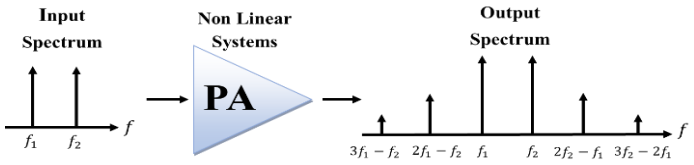


Figure 3: Intermodulation Products of nonlinear systems

3. Balanced RF Power Amplifier

The balance RF power architecture shows in Figure 4. Two identical signal-stage amplifiers are parallel connecting with two quadrature hybrid couplers. Both amplifiers share the same input signal ($x_{in}(t)$) which is fed through the quadrature hybrid coupler. The first quadrature hybrid coupler on the input is the splitter that divides an input signal (from port 1) into two equal amplitude outputs (port 2 and port 3) with a 90˚ phase difference. The input signal of the first amplifier (PA1) has been 90˚ phase-shifted and the signal of the second amplifier (PA2) has been 180˚ phase shifted. Two amplifiers have the same bias voltage that provides equal output power capability. The second quadrature hybrid coupler on the output is a combiner that is connected to the output stages of two amplifiers. which an output signal of PA1 connect to port 1 and an output signal of PA2 connect to port 4. Both output signals are in-phase on port 3 that means output signals from two amplifiers are combined. While output signals are out-of-phase at port 2 that mean the output signals from two amplifiers are canceled. A port 2 of a hybrid coupler is an isolation port that is terminated by match load 50Ω.

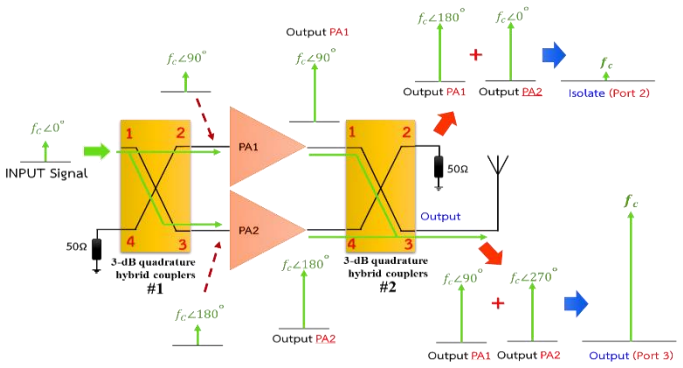


Figure 4: The balanced RF power amplifier structures

3.1. Analysis of the Balanced RF Power Amplifier to Reduce Reverse Intermodulation Products

This paper proposes the balanced RF power amplifier that reduces the reverse intermodulation products. From Figure 5, an input signal is $x_i(t)$ which represents the FM modulation signal. The input signal is fed through the first quadrature hybrid coupler which separates an input signal into two equally signals with a 90˚ phase difference, we can express by:

$$x_{i1}(t) = \frac{1}{2}\cos(2\pi f_C t+90^\circ) \quad (4)$$

$$x_{i2}(t) = \frac{1}{2}\cos(2\pi f_C t+180^\circ) \quad (5)$$

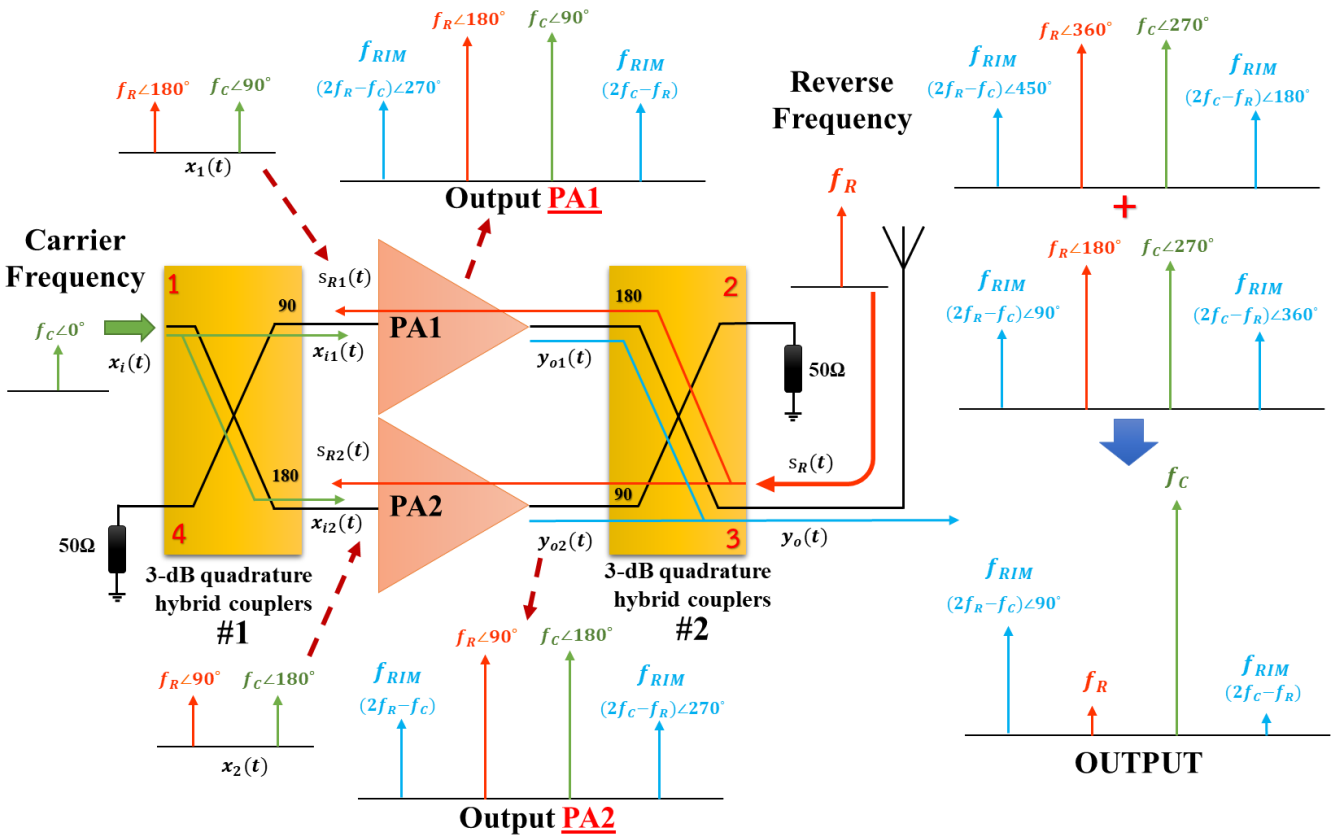


Figure 5: The analysis of a balanced RF power amplifier to reduce reverse intermodulation products

Where $x_{i1}(t)$ is an input signal for the first power amplifier and the $x_{i2}(t)$ is an input signal for the second power amplifier. The f_C represent a carrier frequency. When the reverse signal ($S_R(t)$) from nearby FM station travels in a reverse direction until it presents at the output stage, the port 3 of a second quadrature hybrid coupler. A similar expression for the output stage of a balanced amplifier, a second quadrature hybrid coupler is the symmetric 4-port microwave device that separates a reverse signal to two equally signals with a relative phase difference of 90˚. We can express both reverse signals by:

$$S_{R1}(t) = \frac{1}{2}\cos(2\pi f_R t + 180^\circ) \quad (6)$$

$$S_{R2}(t) = \frac{1}{2}\cos(2\pi f_R t + 90^\circ) \quad (7)$$

Where $S_{R1}(t)$ and $S_{R2}(t)$ are the respective reverse signals traveling into two power amplifiers and the f_R is a frequency of a reverse signal. Both reverse signals travel in the reverse direction through the coupler until they present at the input of each amplifier. Input signals ($x_i(t)$) together with the reverse signals ($S_{Ri}(t)$) present at an input of each power amplifiers, can be written as follows:

$$x_1(t) = x_{i1}(t) + S_{R1}(t) \quad (8)$$

$$x_2(t) = x_{i2}(t) + S_{R2}(t) \quad (9)$$

The $x_1(t)$ and $x_2(t)$ are the respective input signals of two power amplifiers. We use $x_1(t)$ and $x_2(t)$ to a nonlinearity equation of power amplifiers (1). Therefore, a resultant of the output can be expressed as:

$$\begin{aligned} y_{o1}(t) = {} & [\alpha_2] + \left[\alpha_1 + \frac{9\alpha_3}{4}\right]\cos(2\pi f_R t + 180^\circ) + \\ & \left[\alpha_1 + \frac{9\alpha_3}{4}\right]\cos(2\pi f_C t + 90^\circ) + \\ & \left[\frac{\alpha_2}{2}\right]\cos(2\pi 2f_R t + 180^\circ) + \\ & \left[\frac{\alpha_2}{2}\right]\cos(2\pi 2f_C t + 90^\circ) + \\ & \left[\frac{3\alpha_3}{4}\right]\cos(2\pi t(2f_R + f_C) + 90^\circ) + \\ & \left[\frac{3\alpha_3}{4}\right]\cos\left(2\pi t(2f_C + f_R)\right) + \\ & \left[\frac{3\alpha_3}{4}\right]\cos(2\pi t(2f_R - f_C) + 270^\circ) + \\ & \left[\frac{3\alpha_3}{4}\right]\cos\left(2\pi t(2f_C - f_R)\right) + \\ & \left[\frac{\alpha_3}{4}\right]\cos(2\pi 3f_R t + 180^\circ) + \\ & \left[\frac{\alpha_3}{4}\right]\cos(2\pi 3f_C t + 90^\circ) + \ldots \end{aligned} \quad (10)$$

$$y_{o2}(t) = [\alpha_2] + \left[\alpha_1 + \frac{9\alpha_3}{4}\right]\cos(2\pi f_R t + 90^\circ) + \quad (11)$$

$$\left[\alpha_1 + \frac{9\alpha_3}{4}\right]\cos(2\pi f_C t + 180^\circ) +$$

$$\left[\frac{\alpha_2}{2}\right]\cos(2\pi 2f_R t + 90^\circ) +$$

$$\left[\frac{\alpha_2}{2}\right]\cos(2\pi 2f_C t + 180^\circ) +$$

$$\left[\frac{3\alpha_3}{4}\right]\cos\left(2\pi t(2f_R + f_C)\right) +$$

$$\left[\frac{3\alpha_3}{4}\right]\cos(2\pi t(2f_C + f_R) + 90^\circ) +$$

$$\left[\frac{3\alpha_3}{4}\right]\cos\left(2\pi t(2f_R - f_C)\right) +$$

$$\left[\frac{3\alpha_3}{4}\right]\cos(2\pi t(2f_C - f_R) + 270^\circ) +$$

$$\left[\frac{\alpha_3}{4}\right]\cos(2\pi 3f_R t + 90^\circ) +$$

$$\left[\frac{\alpha_3}{4}\right]\cos(2\pi 3f_C t + 180^\circ) + \ldots$$

The $y_{o1}(t)$ and $y_{o2}(t)$ are respective output signals of two power amplifiers. Both output signals are several effects of a nonlinear distortion that contains the term of the carrier frequency, the reverse frequency, the harmonics, and the third-order intermodulation products. Both output signals are combined by a second quadrature coupler which is connected to an output signal $y_{o1}(t)$ at port 1 and an output signal $y_{o2}(t)$ at port 4. The term of a reverse frequency (f_R) and the term of intermodulation product ($2f_C$-f_R) are out-of-phase on port 3 of a second quadrature coupler. Thus, the reverse signal and the intermodulation signal are canceled. A similar expression, a term of a carrier frequency (f_C) is in-phase on port 3. So that the total output power can be increased and can be expressed as:

$$y_o(t) = 2\alpha_2 + 2\left[\alpha_1 + \frac{9\alpha_3}{4}\right]\cos(2\pi f_C t + 270^\circ) + \quad (12)$$

$$\alpha_2\cos(2\pi 2f_C t + 270^\circ) +$$

$$\left[\frac{6\alpha_3}{4}\right]\cos(2\pi t(2f_C + f_R) + 180^\circ) +$$

$$\left[\frac{6\alpha_3}{4}\right]\cos(2\pi t(2f_R - f_C) + 90^\circ) +$$

$$\left[\frac{\alpha_3}{4}\right]\cos(2\pi 3f_C t + 270^\circ) + \ldots$$

The math model of balanced RF power amplifiers with a carrier frequency (f_C) 100 MHz and a reverse frequency (f_R) 95 MHz is simulated. The line spectrums of simulation results show in Figure 6. The intermodulation signals in term of ($2f_C$-f_R) and a reverse signal are eliminated.

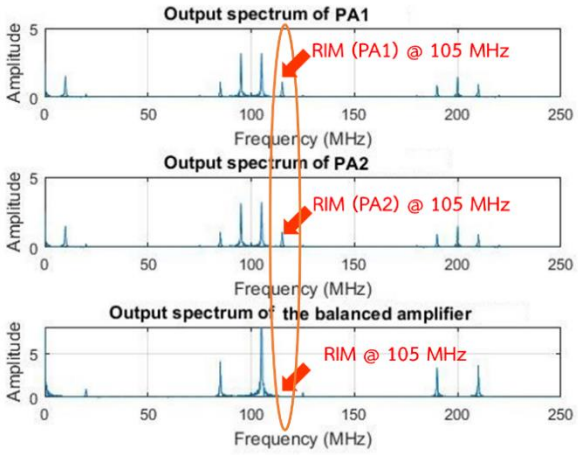


Figure 6: Simulation results of the math model based on a balanced RF power amplifier

3.2. Intermodulation Measurements

The third-order intermodulation distortion creates additional frequencies in terms of $2f_1 - f_2$ and $2f_2 - f_1$. The measurement is described by the power ratio between the power level of a fundamental frequency and intermodulation products, can be expressed as:

$$P_{IM3}(\text{dBc}) = P_C(\text{dBm}) - P_S(\text{dBm}) \quad (13)$$

Where the P_C is a power of a fundamental frequency and the P_s is the power of a third-order intermodulation product. The P_{IM3} is the intermodulation level (dBc) of an intermodulation product relative to a power level of the fundamental frequency.

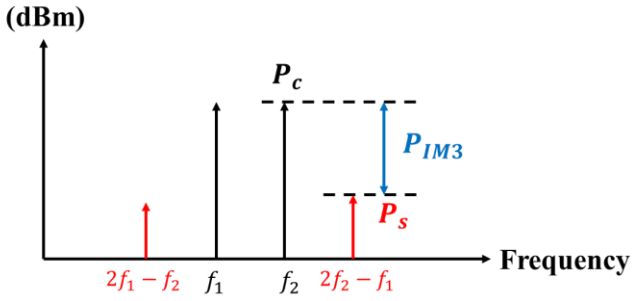


Figure 7: Intermodulation Distortion Measurement

4. Design and Implementation of 3-dB Quadrature Hybrid Couplers

The conventional couplers for balance power amplifiers use branch-line hybrid couplers because it is simple and easy to make. The size of the branch-line couplers is increased when using in an FM frequency range. The reducing of the physical size of branch-line couplers, many methods have been suggested. The slow-wave transmission line technique can make short-wavelength transmission lines by reducing phase velocity [13]-[17]. The increasing of inductances and capacitances with the same ratio causes reducing phase velocity while a characteristic impedance of the coupler is unchanged. Also, lumped-element hybrid couplers use for reducing the circuit size of branch-line hybrid couplers [18]-[19]. Lumped-element circuits are realized by using

LC-microstrip resonance circuits to replace the line sections of branch-line couplers. The lumped-element circuits have a characteristic impedance equivalent to the quarter-wavelength transmission line while the circuit size of lumped-element circuits is shorter than the branch-line couplers.

Lange couplers are two parallel microstrip lines with alternate lines on the same surface [20]- [21]. The alternate lines can reduce the occupied area of couplers. However, after the circuit of couplers have been processed, the circuit requires wire bonders for bonding the conductor spacing of coupler lines [22]. Wire bonders increase the manufacturing processes which are difficult to realize. Broadside-coupler striplines are the hybrid couplers without bonding wires [23]. The broadside-coupler structure comprises two parallel quarter-wavelength transmission lines on the opposite surface. A coupled line is a floating transmission line overlay on a signal line. The length of a coupled line determines the physical size of the hybrid couplers. It can be significantly reduced the occupied area of couplers by using symmetrical circuits of two parallel striplines with the opposite surface and using tightly stripline. The coupling coefficient of couplers is increased by using printed circuit broads with a small substrate thickness, e.g., 0.1mm substrate thickness, which has expensive cost and is difficult to find [24]-[29].

The broadside-coupled striplines in this paper consist of two parallel lines which are embedded in a dielectric between two ground planes [26]-[28], as shown in Fig. 8. The energy transfer from a signal line to a coupler line occurs at the through parallel striplines.

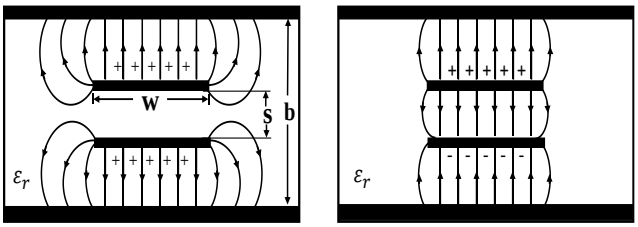


Figure 8: voltages and currents on two striplines.
(a) Even-mode field distribution.
(b) Odd-mode field distribution.

The configuration as shown in Fig. 8 is used to design a broadside coupler. The coupling coefficient can be calculated in terms of a characteristic impedance (Z_0). The voltage coupling coefficient (C) is calculated from the characteristic impedances of even-mode (Z_{0e}) and odd-mode (Z_{0o}) which can be expressed as [30]:

$$Z_{0e} = Z_0\sqrt{\frac{1+C}{1-C}} \tag{14}$$

$$Z_{0o} = Z_0\sqrt{\frac{1-C}{1+C}} \tag{15}$$

According to Cohn [31]-[33], the even-mode and odd-mode characteristic impedances with respect to ground planes can be calculated by:

$$Z_{0e} = \frac{188.3}{\sqrt{\varepsilon_r}}\frac{K(k^{\square})}{K(k)} \tag{16}$$

$$Z_{0o} = \frac{296.1}{\sqrt{\varepsilon_r}\frac{b}{s}\,tan^{\square-1}(k)} \tag{17}$$

Where $K(k)$ and $K(k^{'})$ are complete elliptic integrals of the first kind and $k' = \sqrt{1-k^2}$. The ground plane spacing (b) and the spacing between two parallel striplines (s) are used to calculate the k parameter which can be calculated by:

$$\frac{K(k')}{K(k)} \approx \frac{2}{\pi}\ln\left[\frac{4}{k}\right] \tag{18}$$

Thus, we can calculate the stripline width (w) from:

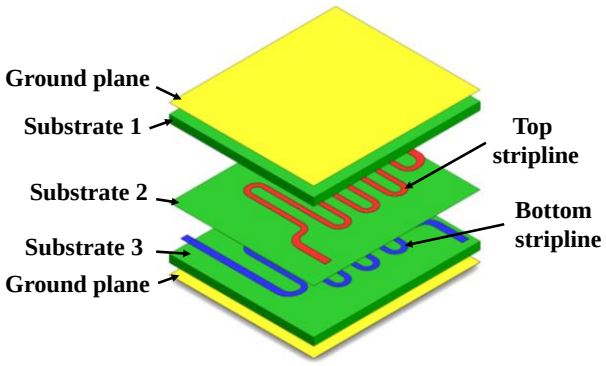
$$\frac{w}{b} = \frac{1}{\pi}\left[\ln\left(\frac{1+R}{1-R}\right) - \frac{s}{b}\ln\left(\frac{1+\frac{R}{k}}{1-\frac{R}{k}}\right)\right] \tag{19}$$

$$R = \sqrt{\left[\frac{\left(k\frac{b}{s}-1\right)}{\left(\frac{1}{k}\frac{b}{s}-1\right)}\right]} \tag{20}$$

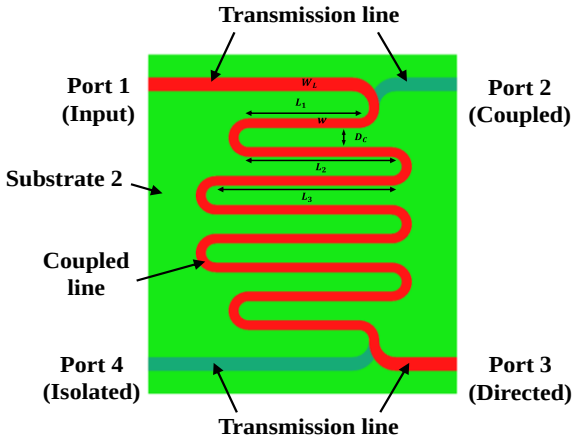
The circuit design of a 3-dB quadrature hybrid coupler, as shown in Figure 9. The top stripline is embedded between the substrate 1 and the top ground plane. The bottom stripline is embedded between the substrate 3 and the bottom ground plane. The space between the top stripline and the bottom stripline is inserted by a substrate 2, as shown in Figure 9 (a). Assume, the input signal is injected into port 1, called "input port". The bottom stripline passes an input signal to port 3, called "directed port". Also, an input signal is coupled via the top stripline to port 2, called "coupled port". An input signal is divided by a 3-dB quadrature hybrid coupler into two equal amplitude outputs with a 90˚ relative phase difference. The input and coupled signals are canceled at the port 4 that has no signal going out from this port, called "isolated port".

The hybrid coupler is implemented at the frequency of 98 MHz, It is a center frequency of FM broadcast systems. The FR-4 printed circuit broads (PCB) are used to make prototype couplers. A dielectric constant (ε_r) of an FR-4 PCB is 4.6 and a substrate thickness (s) is 0.8 mm. Assuming the characteristic impedance (Z_0) is 50 Ω, coupling coefficients (C) is 0.707 The summarize calculated parameters from (14) – (20) are the stripline width (w) 2.54 mm. The ground plane spacing (b) is 10.4 mm. The length of a coupled line (L) is 385 mm. The circuit of a 3-dB quadrature hybrid coupler using tightly microstrip line, as shown in Figure 9(b). This circuit designed is simulated with the CST software to verify a proposed hybrid coupler.

Due to the general FR-4 PCBs have 1.6 mm and 0.8 mm substrate thickness while the dimension of ground plane spacing (b) of the proposed coupler is 10.4 mm. Thus, we choose an FR-4 PCB with 1.6 mm substrate thickness to make substrate 1 and substrate 3 which can be made by using PCBs piling up to three layers. The constructed of a prototype coupler is shown in Fig. 9. The overall physical dimension is 75×85 mm.

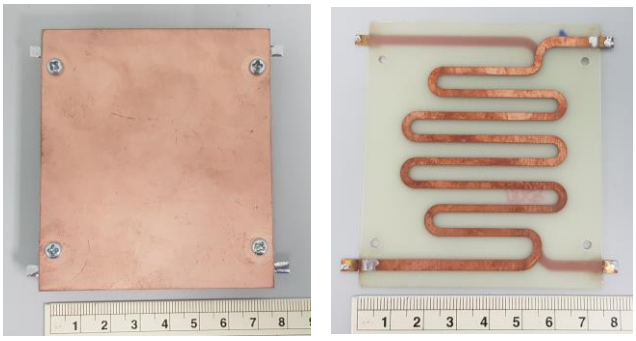


(a) Layers of the 3dB quadrature coupler



(b) Circuit layouts were designed

Figure 9: The configuration of a 3-dB quadrature hybrid coupler
(a) Layers of the 3-dB quadrature hybrid coupler
(b) Circuit layout of the 3-dB quadrature hybrid coupler



(a) Physical size of the coupler (b) Circuit layout of the coupler

Figure 10: Photographs of the prototype coupler

4.1. Measurement Results of the Quadrature Hybrid Couplers

In this section presents the experimental results of the prototype 3-dB hybrid coupler. The frequency range between 25 MHz to 150 MHz has return losses lower than -15dB, as shown in Figure 11. The center frequency of an FM broadcasting systems (98MHz) provides a maximum coupling coefficient, as shown in Figure 12. The coupling coefficient between a signal line and a coupler line is -3.4 dB (S21) and The transmitted coefficient between a signal line and a coupler line is -3.3 dB (S31). The relative phase difference between the two output ports is 90 ± 2˚, as shown in Figure 13. The measured isolations (S41) are lower than -17 dB. The input and coupled signals are canceled at the port 4.

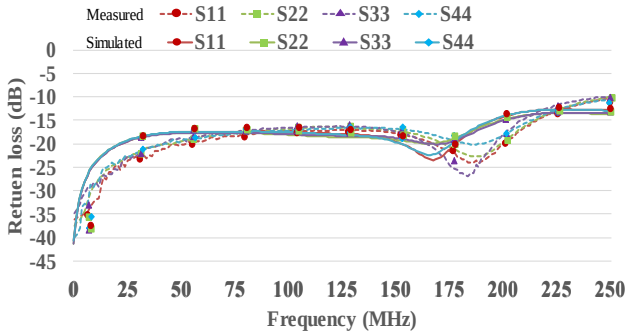


Figure 11: Measured and simulated results of the return loss

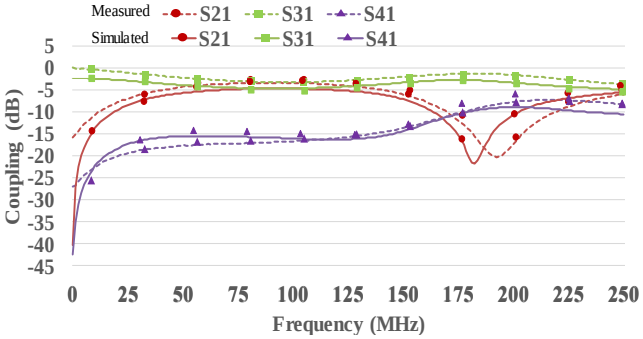


Figure 12: Measured and simulated results of the coupling coefficient

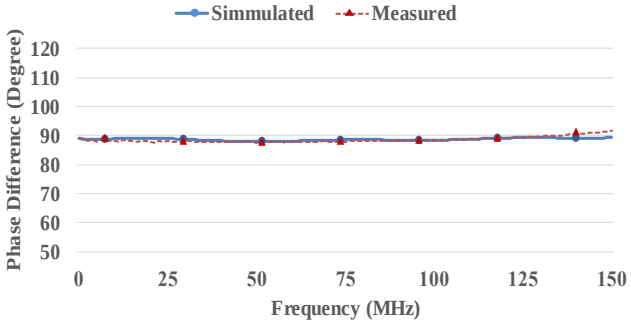


Figure 13: Measured and simulated results of the relative phase difference

5. Design and Implementation of Balanced RF Power Amplifiers

The balanced power amplifier was implemented by using two prototype 3-dB quadrature hybrid couplers. The first coupler divides an input signal into two identical single-stage amplifiers. The amplifier modules were designed for FM broadcasting systems, at 88-108 MHz frequency range, based on class-C push-pull amplifiers. The single-stage amplifier consists of 2 x MRF151 N-channel MOSFET. The bias configuration set a drain-source voltage (Vds) = 48V, drain-source current (Ids) = 0.3A and gate-source voltage (Vgs) = 2.9V. Maximum output power is 250-W with gain 14 dB. Two amplifiers have the

same bias voltage which provides an equal output power capability. The output signals from two amplifiers are combined by the second quadrature coupler. The total power is transmitted to the antenna systems.

The prototype of a balanced power amplifier shows in Figure 14. The port 1 of the first couplers is an input port and the port 3 of the second couplers is an output port. Isolated ports, port 4 of first couplers and port 2 of seconded couplers, are terminated by a dummy load 50 Ω. When the input and output impedance is poorly matched with the RF power amplifiers that are propagated signals reflected into a dummy load.

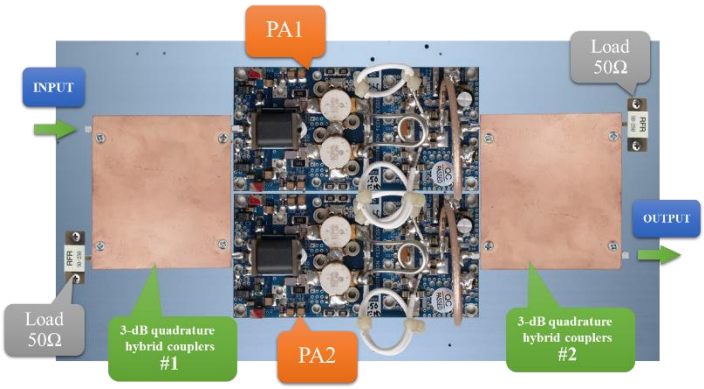


Figure 14: The configuration of a prototype balanced RF power amplifier

5.1. Measurement Results of the Balanced RF Power Amplifiers

The ROHED&SCHWARZ NRP2 Power meter is used to measure a balanced RF power amplifier. The Agilent E4421B signal generator creates an FM modulation signal and injects to the input port of a balanced amplifier. The first measurement focuses on the output power of each single-stage amplifier. The proposed balanced amplifier was tested at frequency 105 MHz that is a reverse frequency. Output powers of each amplifier module compared to the input power level displays in Figure 15. when the input power level is 4 watt that provides the output power of two amplifiers approximately 116.20 watts and 118.90 watts respectively. The line curves of two amplifiers have been overlapped. The result shows that two amplifiers are equal output power capability which can be combined.

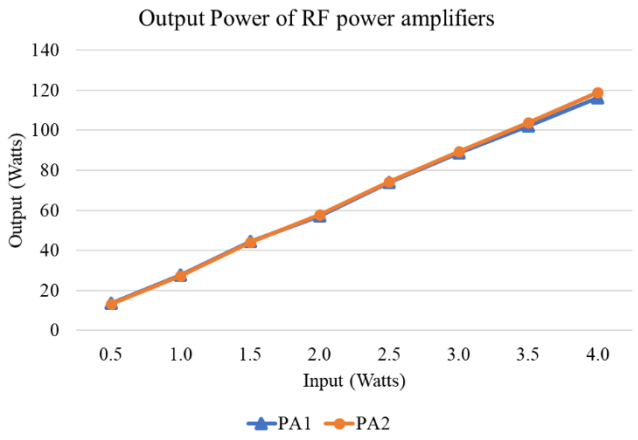


Figure 15: Measured results of the output power of each amplifier module

The second measurement focuses on the total output power of a balanced RF power amplifier. The total power is realized by the sum of two signals from two single-stage amplifiers that use the second quadrature hybrid coupler. The prototype balanced amplifier obtains an output power of 210 watts with a gain of 17.20 dB when the input power level is 4 watts. The output power has been dropped off 25.1 watts from sum signals of two amplifiers due to the insertion loss of the coupler at the output stage.

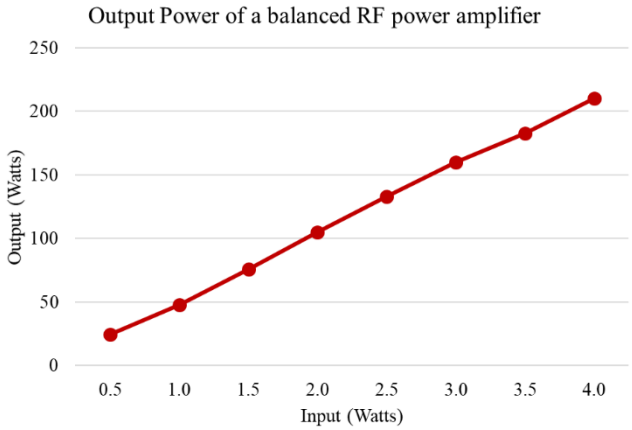


Figure 16: Measured results of the balanced RF power amplifier

6. Measurement Results of a Balanced RF Power Amplifier using to Reduce Reverse Intermodulation Products

6.1. Measurement Setup

This section presents experimental results of the FM broadcast station using a balanced RF power amplifier to reduce the reverse intermodulation. A practical setup, as shown in Figure 17, is used to demonstrate the situation of two closely located FM broadcast stations. The one FM transmitter generates a reverse frequency (f_R) that is transmitted 500 watts output power at 99.50 MHz. Another FM transmitter generates a carrier frequency (f_C) at 105.00 MHz. The two parallel antenna systems for a reverse frequency and a carrier frequency are installed in the same tower which has a height 60 meter. This configuration provides a maximum level of reverse intermodulation products. The reverse transmitter uses a cavity bandpass filter in the output stage which ensures to block unwanted signals from this transmitter.

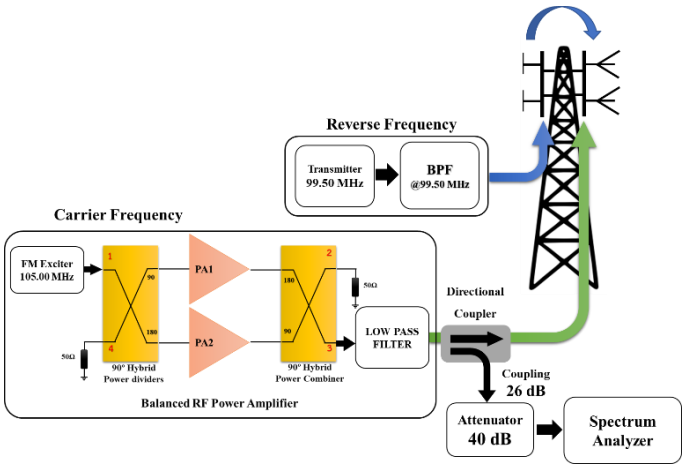


Figure 17: Measurement setup to demonstrate the situation of two closely located FM broadcast stations.

Two transmitters are measured in this research. The first transmitter uses a single-stage RF power amplifier, based on a class-C push-pull, and the second transmitter uses a prototype balanced amplifier in this research. Two transmitters are used to generate a carrier frequency (f_C) that is transmitted 200 watts output power at 105.00 MHz. Almost output signals from a transmitter is measured via a directional coupler with a coupling value of 26 dB and an attenuator of 40 dB. Intermodulation products are provided in terms of $2f_1 - f_2$ and $2f_2 - f_1$ that are a frequency of 94.00 MHz and 110.50 MHz respectively. Since the frequency of 110.50 MHz is out of an FM frequency band and falls into the frequency range of Air Traffic Control systems. Therefore, the measurement results consider the intermodulation product in terms of $2f_2 - f_1$ only.

6.2. Measurement Results of the FM transmitters using a single-stage power amplifier.

Measurement results of a carrier transmitter using a single-stage power amplifier present in Figure 18. The carrier transmitter transmits an FM modulation signal that is the frequency of 105.00 MHz without cavity bandpass filters. The signal strength of a reverse frequency, at the frequency of 99.50 MHz, is -42.35 dBm that is a signal traveling until it present at the output state of the carrier transmitter, port 3 of the second quadrature coupler. The output power of a carrier frequency at 105.00 MHz is -17.55 dBm and the intermodulation level at 110.50 MHz is 30.87 dBc.

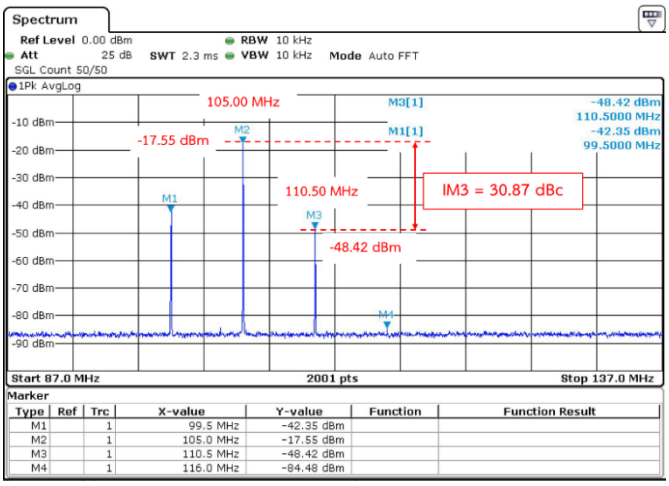


Figure 18: Intermodulation level of the FM transmitter using a single-stage RF power amplifier

6.3. Measurement Results of the FM transmitters using a Prototype Balanced RF Power Amplifier

The FM transmitter using a prototype of a balanced RF power amplifier is measured that is compared with a single-stage RF power amplifier. The output power of a balanced amplifier must be adjusted equal to a single-stage RF power amplifier. From Figure 17, the frequency spectrum shows that the signal strength of a reverse frequency at 99.50 MHz is -50.89 dBm which is decreased by 8.54 dBm. The output power of a carrier frequency at 105.00 MHz is -16.50 dBm and the intermodulation level is 49.34 dBc. Table 1 shows a comparison of the measurement results of two FM transmitters that use a single-stage RF power amplifier and a balanced RF power amplifier. It shows that both transmitters provide the same power level which is approximate -17 dBm. A power level of a reverse frequency is less than a single-stage RF power amplifier. The balanced RF power amplifier provides a good performance that improvements an intermodulation level from 30.84 dBc to 49.34 dBc. Measurement results are according to the analysis in section 3. The results show the mitigation of an intermodulation signal that is approximate 18.47 dBc without cavity bandpass filters. The mitigation intermodulation signal is difficult to obtain a maximum reduction due to the quadrature hybrid couplers have slightly different amplitudes and phases.

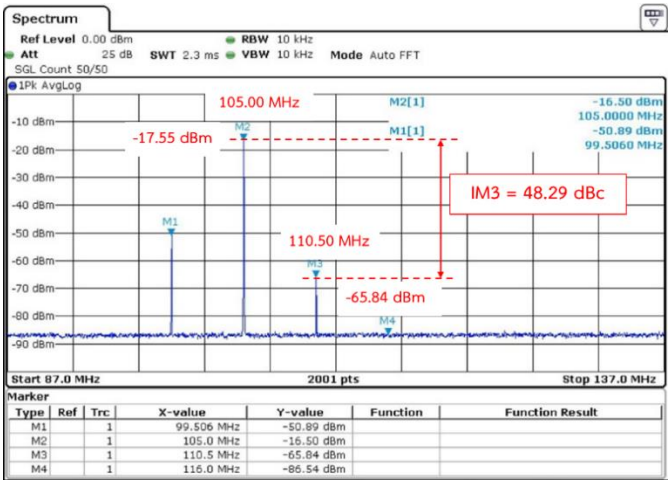


Figure 19: Intermodulation level of the FM transmitter using an FM transmitter using a Balanced RF Power Amplifier

Table 1: Measurement results of two FM transmitters.

	Signal Strength		
Frequency (MHz) / Amplifier types	99.50 (dBm)	105.00 (dBm)	110.50 (dBc)
The single-stage RF power amplifier	-42.35	-17.55	30.87
The balanced RF Power Amplifier	-50.89	-16.50	49.34
Δ	-8.54	1.05	18.47

7. Conclusions

This paper proposes a balanced RF power amplifier using the improvement 3-dB quadrature hybrid couplers to reduce the reverse intermodulation of FM broadcast stations. The balanced RF power amplifier was implemented by using the prototype 3-dB quadrature hybrid couplers. The FR-4 low-cost PCB with a broadside-coupled striplines technique is chosen to construct prototype couplers that can reduce the overall size of a hybrid coupler and suitable to use for balanced RF power amplifies. Experimental results of the prototype coupler, The return loss is lower than -15 dB. The coupled port has a coupling coefficient -3.4 dB and the directed port has a coupling coefficient -3.3 dB. The phase difference between a coupled and directed port is $90 \pm 2^{\circ}$. Isolation values are lower than -17 dB. The overall size of the prototype couplers is 75x 85 mm. We apply a prototype 3-dB quadrature hybrid couplers to a balanced power amplifier which are structured from two identical quadrature couplers and two signal-stage amplifiers. The two identical single-stage amplifiers use a class-C RF amplifier with the MRF151 N-channel MOSFET and operate at the frequency range of 88-108 MHz. A balanced

amplifier obtains an output power of 210 watts with a gain of 17.20 dB when the input power level is 4 watts.

The situation of two closely located FM broadcast stations is used to create the reverse intermodulation signal. The one transmitter generates a reverse signal (f_R) at 99.50 MHz that is transmitted 500 watts output power and another transmitter generate a carrier signal (f_C) at 105.00 MHz. The antenna systems for two frequency are installed in the same tower which has height 60 meters. Two carrier transmitters which use a balanced RF power amplifier and a single-stage RF power amplifier are measured. The carrier transmitter using a balanced RF power amplifier improvements an intermodulation level from 30.84 dBc to 49.34 dBc. The results show the mitigation of an intermodulation signal approximate 18.47 dBc without cavity bandpass filters. The mitigation intermodulation signal is difficult to obtain a maximum reduction due to the quadrature hybrid couplers have slightly different amplitudes and phases.

Acknowledgment

This work was supported by The National Broadcasting and Telecommunications Commission (NBTC) of Thailand and the Research Department Institute of Engineering, Suranaree University of Technology (SUT), Nakhon Ratchasima, Thailand.

References

[1] David Hand C.Eng MIET, “THE RF CHALLENGES OF ATC COMMUNICATIONS” Consultant Engineer - Park Air Systems Ltd.

[2] Jaiyen, N., Hantula, P., & Tongta, R. (2017, December). Real-time audio similarity comparison algorithm. In 2017 IEEE 15th Student Conference on Research and Development (SCOReD) (pp. 477-480). IEEE.

[3] Hantula, P., Jaiyen, N., & Tongta, R. (2018, August). A 3-dB Quadrature Coupler Using Broadside Striplines for FM Power Amplifiers. In 2018 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM) (pp. 1-2). IEEE.

[4] Alcon Garcia, P., Esparza Lopez, N., Herran Ontanon, L. F., & Las Heras Andres, F. (2016). Complex Impedance Transformers Based on Branch-Line Hybrid Couplers. Progress In Electromagnetics Research, 69, 147-157.

[5] Jung, S. C., Negra, R., & Ghannouchi, F. M. (2008). A design methodology for miniaturized 3-dB branch-line hybrid couplers using distributed capacitors printed in the inner area. IEEE Transactions on Microwave Theory and Techniques, 56(12), 2950-2953.

[6] Hussein, Ehab Abdul Razzaq, and Mohammed A. Abdulkadhim. "Performance improvement of BER in OFDM system using feed forward technique on power amplifier." International Journal of Computer Applications 75.4 (2013).

[7] Amanpreet Kaur, Rajbir Kaur “Reducing the Third-Order Inter Modulation Distortion by Feed Forward Linearization of Power Amplifier” International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016

[8] Parsons, K. J., R. J. Wilkinson, and P. B. Kenington. "A highly-efficient linear amplifier for satellite and cellular applications." Global Telecommunications Conference, 1995. GLOBECOM'95. , IEEE. Vol. 1. IEEE, 1995.

[9] Grebennikov, Andrei. "Linearity Improvement Techniques for Wireless Transmitters: Part 2." High frequency electronics (2009): 16-26.

[10] Katz, Allen, et al. "Sensitivity and mitigation of reverse IMD in power amplifiers." Power Amplifiers for Wireless and Radio Applications (PAWR), 2011 IEEE Topical Conference on. IEEE, 2011.

[11] Berglund, Bo, Thorsten Nygren, and Karl-Gosta Sahlman. "RF multicarrier amplifier for third-generation systems." ERICSSON REV (ENGL ED) 78.4 (2001): 184-189.

[12] Ma, Hongbo, and Quanyuan Feng. "An improved design of feed-forward power amplifier." PIERS Online 3.4 (2007): 363-367.

[13] Tsai, K. Y., Yang, H. S., Chen, J. H., & Chen, Y. J. E. (2011). A miniaturized 3 dB branch-line hybrid coupler with harmonics suppression. IEEE Microwave and Wireless Components Letters, 21(10), 537-539.

[14] Sun, K. O., Ho, S. J., Yen, C. C., & van der Weide, D. (2005). A compact branch-line coupler using discontinuous microstrip lines. IEEE Microwave and Wireless Components Letters, 15(8), 519-520.

[15] Caillet, M., Clenet, M., Sharaiha, A., & Antar, Y. M. (2009). A compact wide-band rat-race hybrid using microstrip lines. IEEE Microwave and Wireless Components Letters, 19(4), 191-193.

[16] Li, H., Zhang, H. L., Hu, B. J., Wei, X. D., & Zeng, W. (2015, September). Novel wideband quadrature hybrid coupler with tunable power dividing ratio. In Signal Processing, Communications and Computing (ICSPCC), 2015 IEEE International Conference on (pp. 1-4). IEEE.

[17] de Paco, P., Verdu, J., Menendez, Ó., & Corrales, E. (2008). Branch-line coupler based on edge-coupled parallel lines with improved balanced response. IEEE Transactions on Microwave Theory and Techniques, 56(12), 2936-2941.

[18] Vijayaraghavan, A. S., & Dunleavy, L. (2011). Design and optimization of lumped element hybrid couplers. High Frequency Electronic.

[19] Moubadir, M., Aziz, H., Touhami, N. A., & Mohamed, A. (2018). A miniaturized branch-line hybrid coupler microstrip for long term evolution applications. Procedia Manufacturing, 22, 491-497.

[20] Breed, G. (2009). Transmission line and lumped element quadrature couplers. High Frequency Electronics.

[21] Kajfez, D., Paunovic, Z., & Pavlin, S. (1978). Simplified design of Lange coupler. IEEE Transactions on Microwave Theory and Techniques, 26(10), 806-808.

[22] Mao, S. G., & Wu, M. S. (2007). A novel 3-dB directional coupler with broad bandwidth and compact size using composite right/left-handed coplanar waveguides. IEEE Microwave and wireless components letters, 17(5), 331-333.

[23] Gruszczynski, S., Wincza, K., & Sachse, K. (2006). Design of compensated coupled-stripline 3-dB directional couplers, phase shifters, and magic-T's—Part I: Single-section coupled-line circuits. IEEE transactions on microwave theory and techniques, 54(11), 3986-3994.

[24] Javadzadeh, S. M. H., Majedi, S. M. S., & Farzaneh, F. (2010, September). An ultra-wideband 3-DB quadrature hybrid with multisection broadside stripline tandem structure. In International Conference on Mobile Multimedia Communications (pp. 672-681). Springer, Berlin, Heidelberg.

[25] Chang, C. P., Chiu, J. C., Chiu, H. Y., & Wang, Y. H. (2008). A 3-dB quadrature coupler using broadside-coupled coplanar waveguides. IEEE Microwave and Wireless Components Letters, 18(3), 191-193.

[26] Chiu, J. C., Lin, C. M., & Wang, Y. H. (2006). A 3-dB quadrature coupler suitable for PCB circuit design. IEEE transactions on microwave theory and techniques, 54(9), 3521-3525.

[27] Abdolhamidi, M., Mohammad-Taheri, M., & Ali-Abad, M. A. A "Design Equations for Broadside and Edgewise Stripline Couplers".

[28] May, J. W., & Rebeiz, G. M. (2008). A 40-50-GHz SiGe 1: 8 differential power divider using shielded broadside-coupled striplines. IEEE Transactions on Microwave Theory and Techniques, 56(7), 1575-1581.

[29] Bhartia, P., & Pramanick, P. (1988). Computer-aided design models for broadside-coupled striplines and millimeter-wave suspended substrate microstrip lines. IEEE transactions on microwave theory and techniques, 36(11), 1476-1481.

[30] Bahl, I. J., & Bhartia, P. (1981). The design of broadside-coupled stripline circuits. IEEE Transactions on Microwave Theory and Techniques, 29(2), 165-168.

[31] Cohn, S. B. (1960). Characteristic impedances of broadside-coupled strip transmission lines. IRE Transactions on Microwave Theory and Techniques, 8(6), 633-637.

[32] Hilberg, W. (1969). From approximations to exact relations for characteristic impedances. IEEE Transactions on Microwave Theory and Techniques, 17(5), 259-265.

[33] Raab, F. H., Asbeck, P., Cripps, S., Kenington, P. B., Popovic, Z. B., Pothecary, N. and Sokal, N. O. (2003). RF and microwave power amplifier and transmitter technologies Part 3. High Frequency Electronics, 2(3), 22-36.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 431-439 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Effectiveness and Comparison of Digital Substations Over Conventional Substations

Aashir Waleed[*,1], Umar Siddique Virk[2], Muhammad Tanveer Riaz[2], Shaikh Bilal Mehmood[1], Saeed Ahmad[2], Muhammad Rameez Javed[1], Ali Raza[1]

[1]*Department of Electrical Engineering, University of Engineering & Technology, Lahore (Faisalabad Campus), 38000, Pakistan*

[2]*Department of Mechatronics & Control Engineering, University of Engineering & Technology, Lahore (Faisalabad Campus), 38000, Pakistan*

ARTICLE INFO

Article history:
Received: 06 May, 2019
Accepted: 07 August, 2019
Online: 16 August, 2019

Keywords:
Digital substations
Intelligent electronic device
Smart grids
Grid topology
Substation

ABSTRACT

The entire electrical system is changing at an abrupt speed since its beginning. With the rapid advance of renewable energies (which are constantly subject to fluctuations), the electricity generation is distributed to countless locations. The traditional model of the unidirectional flow of electricity is leaning towards multidirectional flows. This situation needs significant changes in the transport infrastructure of the electrical power system and mode of operation. Technological advancements are making a smarter and more organized management system of electricity supply and demand. This development needs to incorporate modern communication, monitoring, and control capabilities, along with the intelligent transport, generation, and storage value chain. Digital substations integrate thousands of sensors by providing greater comfort, self-reliance, availability, and safety while at the same time reducing costs, risks, and environmental damage. Conventional power system lacks digital communication, remote monitoring & control, and consumer participation. Fiber-optic cabling in digital substations will attain not only real-time data transmission but also remote maintenance. The comparison between conventional and digital substations has also been presented in this work along with latest trends in smart protection. Electronic protection and control devices based on a microprocessor, communicating with cloud servers, will enhance the flexibility to control and monitor the digital substations. A significant amount of data generated by components in the electrical substations will allow greater sophistication of the monitoring, diagnostic, protection and optimization capabilities of the facilities. The goal for the review of said advancements is to have a smart electrical system with greater efficiency and better user experience.

1. Introduction

The existing grid infrastructure seems to meet part of the expectations (from traditional and renewable generation to transport and part of the distribution), but it must improve significantly from the end user and the functionalities expected from it. The forecasts indicate moderate growth in demand, a substantial increase in renewable energies and a need for firm and flexible power [1]. Consequently, a new concept of an electrical network, i.e., "intelligent network" emerges. The basic definition of "intelligent networks" may correspond to "Electric networks that intelligently integrate the behavior and actions of all the actors connected to them (those who generate electricity, those who consume it and who perform both actions) to provide a safe, economical and sustainable electricity supply"[2]. In such networks, very advanced integrated systems are used in the process of supplying electricity at all levels. The function of these integrated systems is to protect the components of the electrical system, control of the power flow and the monitoring of the process. For example, electrical system automation devices are integrated into communication networks to exchange information between various devices, as well as with supervisory systems. The technological trends of the integrated systems in the automation of electrical systems will be determined by three different technological trends such as electronic Integration, integration of switching equipment and data communication [3].

[*]Dr. Aashir Waleed, Department of Electrical Engineering, University of Engineering & Technology, Lahore (Faisalabad Campus), 38000, Pakistan, Email:aashir.walid@uet.edu.pk

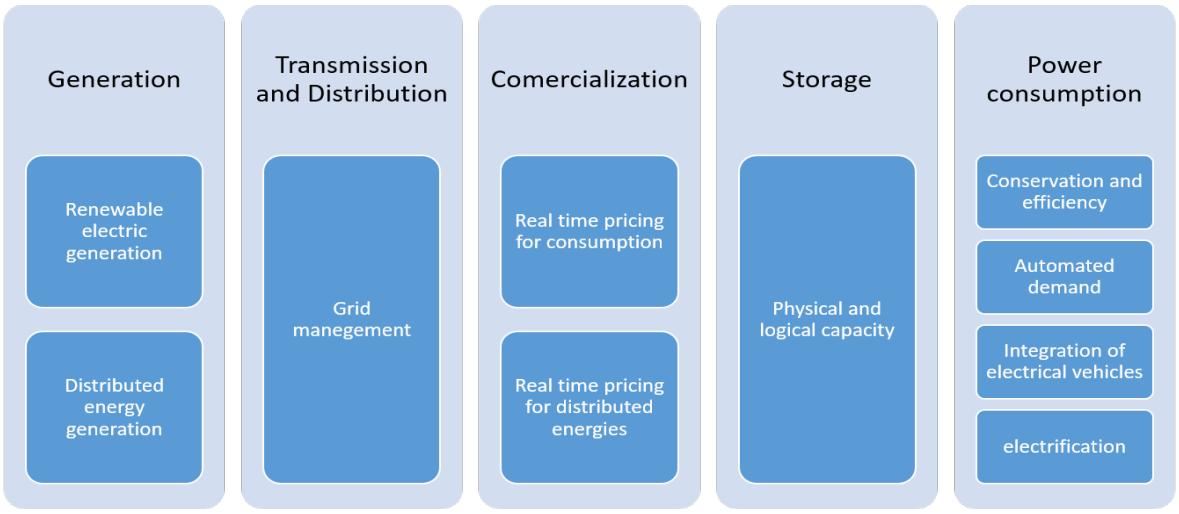
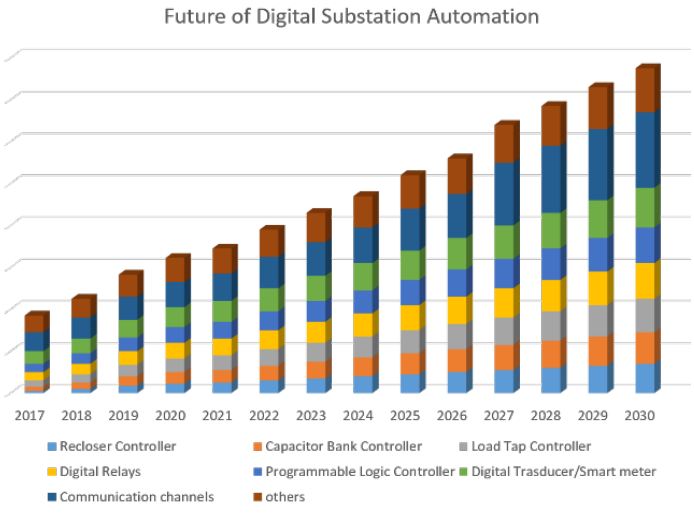


Figure 1. Challenges for current grid topology.

In the last years, an extensive research of energy production based on fossil energy sources has been performed. Fossil fuel sources have been exploited believing that there is an infinite availability of resources, but these sources cause huge damage to environment. The old energy model based on fossil fuel energy involved a chain of processes like generation, distribution, transportation, and consumption. However, to meet the consumer demand, it is necessary to upgrade the above energy model to distributed energy generation. This new energy model (Figure 1) will support variety of energy sources and an increase in the use of renewable energy and efficiency.

The main advantage of the recent energy model is distributed energy generation system. This system allows individuals to generate energy at micro-level and link to the national grid [4], which drastically reduces losses caused by the energy transport. This system facilitates the connections of all kinds of renewable energies to the system, which will ultimately improve the power system. This is possible by promoting the integration of non-manageable energies such as wind or solar power and by using energy storage devices that will be able to charge and discharge in the network. Keeping in mind the future growth and energy trends, the future projection of digital substations is provided in Figure 2.



To carry out all the actions mentioned previously, the grid topology must [5], [6] have following features.

- Allow self-management of incidents, to deal with the defects that have occurred in the system and guarantee the electric flow in all the points.
- Be equipped with strength against attacks and destabilizations.
- Increase the flow of consumer participation, promoting the local generation process and providing excess energy to the network during peak hours.
- Have the capacity to supply power of adequate quality to the digital era. The higher number of generation points will allow the delivery of different energy qualities for each type of application.
- Fix a variety of generation and storage modes, thanks to the micro-grids and the generation of the energy that is distributed.
- Facilitate the flowering of markets, due to the inclusion of new elements in the network such as the electric vehicle, a more sizeable number of renewable energies, etc.
- Perform a more competent optimization of your assets and operation; this is achieved by automating all the elements that are involved.

2. Composition of Smart Grids

The Smart Grid concept refers to intelligent management of electricity using computer and communication technologies. The smart grid concept encompasses many interconnections between the different elements of the generation, transmission and distribution of electricity. In addition, the concept is strongly linked to the inclusion of unconventional renewable energies (both large-scale with large generators and household level) in the electricity grid, thus changing the current paradigm of uniqueness in generation towards a distributed generation concept.

Smart grid is smart energy network that integrates all the parameters of generation, transmission, distribution and

consumption. It also provides advanced control processes to achieve an efficient and sustainable supply (Figure 3) [7]. The motivation behind creation of smart grid is: care for the environment, reliability and quality of supply and improvement in the operation of networks.

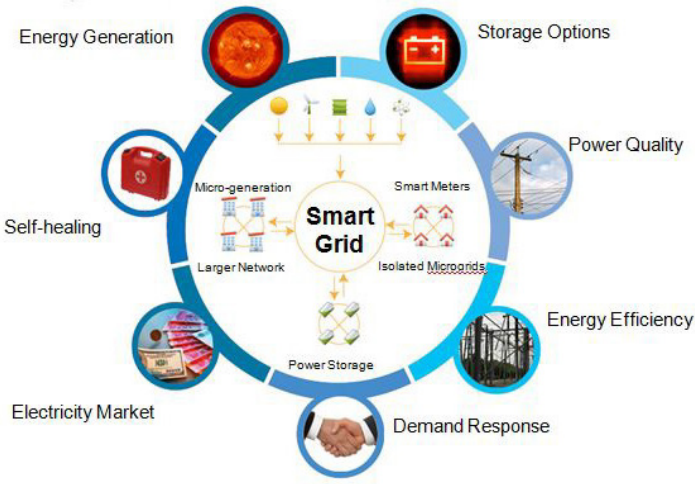


Figure 3. Different aspects that make up the smart grid

The following components must be developed to meet the above-mentioned objectives objective of smart grid:

2.1. Smart Metering:

The smart metering helps consumers to generate their own energy and inject surplus generated energy into the network at a certain price. This increases the dynamism of the system and decreases the prices of electricity [8].

2.2. Load Management:

The data delivered by the smart meters helps in effective management of load. The hourly consumption of energy as delivered by meters is the main deciding factor for the connection or disconnection management of generation sources and loads. This connection or disconnection can be automatic through intelligent controllers, which entails awareness and education of the population in general. Conclusively, this helps in developing a culture of low consumption and energy saving.

2.3. Inclusion of Renewable Energies and Electric Mobility

The smart grid also covers the issue of microgeneration in households. For that purpose, it is necessary to study the natural resources that can be used in the houses of different geographical sectors of the country and improve the technologies in their connection to the distribution lines.

2.4. Renewal of The Current Transmission Networks:

A fundamental point to achieve the above points requires the inclusion of an effective transmission network. The distributed generation and a strong increase in the amount of non-conventional renewable energies injected into the network require a transmission system that supports these bidirectional flows.

So, conclusively smart grid requires different technologies that can ensure a correct operation of this new network due to the complexity that is added to become intelligent [9].

3. Latest Developments in Smart Protection

Microgrid is normally operated parallel with utility grid and e.g. during faults in upstream network it can be separated quickly from utility grid and operated independently as an island grid. MMS will be responsible from the overall economic and energy effective operation of microgrid taking account the technical boundary conditions in both normal and island operation. In this paper only technical aspects related to LV microgrids are discussed. Technical choices made in the microgrid concept must be such that they can be justified by the needs of normal operation, but at the same time allowing and supporting the solutions needed

for implementation of island operation.

In overall the operation and control of LV microgrid is a very complex issue because there are number of things that will have influence to the behavior of microgrid in different ways. For example the dynamics of islanded microgrid, e.g. lack of inertia, is totally different when compared to the normal operation of the microgrid parallel with utility grid. Islanded microgrid is much more sensitive to disturbances and successful island operation requires fast, accurate and stable control.

In addition the protection of the future microgrid is a challenging issue and very strongly connected to the control and operation issues of a microgrid. The conventional protection in distribution networks is designed to operate for high fault current levels in radial networks, but during island operation of the microgrid high fault currents from the utility grid are not present. Also most of the DG units that will be connected to the LV microgrid in the future are converter interfaced and have limited fault current feeding capabilities. This means that the traditional fuse protection of LV network is no longer applicable and new protection methods must be developed. The developed protection scheme for microgrid must be supported by the technical choices made in the microgrid operation and control issues.

In the development of the new protection scheme for LV microgrids many things must be considered including amount of protection zones in LV microgrid, speed requirements for microgrid protection in different operation states and configurations and protection principles for parallel and island operation of the microgrid.

4. Technological Trends

Among the new technological trends of smart grids, substations are intelligent networks, which are designed to manage the consumption of each element connected to it in an "intelligent" way [10]. These networks will use innovative equipment and services, along with new communication, control, monitoring, and self-diagnosis technologies that will help to achieve the following objectives [11]:

- Reinforce and automate the network, achieving an improvement in the functionality of the system, having better quality values and less loss in the network.

- Optimize the connection between the areas that have renewable energy sources, improving the connection capacities and reducing connection costs.
- Study the decentralized generation architectures, making smaller installations (distributed generation) in balance with the system [12].
- Develop intermittent generation integration and new storage technologies
- To advance in the development of the electricity market, allowing new functionalities and services to the market.
- Allow consumers to manage their energy efficiency better through active demand management.
- Enable the penetration of the electric vehicle, accommodating these new mobiles and dispersed loads to the grid, minimizing the development of new infrastructure and enabling the energy storage features they possess [13].

4.1. Advanced Automated Distribution

Advanced distribution automation allows the electrical power distribution system to be completely controllable to increase efficiency, sustainability and reliability. Some applications for advanced automation of distribution are monitoring, data acquisition and control, voltage and reactive power control, fault localization, automatic reconfiguration, isolation and restoration service, user information, time management, asset management and remote control. The main reason for the advanced distribution automation is to guarantee the safety of the service, the quality of the energy and the efficiency of the energy system. The above is possible once the next three control processes of the distribution operation are automated [14]. These control processes are: (1) the data is arranged almost in real time, (2) the optimal decisions and (3) the coordination with the services of the different generation and distribution systems to control the distribution operations.

An automated system of advanced distribution will be able to perform data collection, along with the verification and correction of consistency of that data. Besides, the system should reconfigure the multilevel feeder. It should be pre-armed with the corrective action schemes and the organization of urgency actions in distribution. It should take care of the previous assembly of recovery schemes and coordination of corrective actions in the distribution, registration, and reports [15].

4.2. Distributed Energy Resources

Distributed energy source technologies consist of storage and power generation resources labeled as "behind the meter." These resources are installed at customer's end and are operated to supply partial or total energy consumed by the client [16]. These energy resources may also be able to inject energy into the distribution system and the transmission system or into a local network in parallel with the public network. Moreover, certain technologies such as combined heat and power (CHP) systems, photovoltaic solar energy (PV), microgrids, wind turbines, microturbines, standby generators and energy storage are included [17]. These resources represent immense importance because they can contribute to the reliability of the network. They may be able to produce energy to support the load of the host or the system. These technologies have evolved at a great pace in the years that have passed. However, the current challenges require unique planning, operation, and market management [18].

Moreover, the term that characterizes it as "behind the meter" means that it represents certain resources that are not directly connected to the commercial electric power system but are connected to the customer's access point. In this case, it is possible that the resources are acting to meet the customer's internal electrical loads, or the resources may be working to sell to the electric power system [19].

4.3. Information and Communication Technologies

Advancement in information and communication technologies (ICT) has been helping to maintain the distribution of electricity while limiting the environmental consequences. In addition to this, ICTs allow greater incorporation of renewable energy sources, allowing the transport of little carbon. ICT also includes electric vehicles and inducing changes in the structure with respect to excessive consumption of electricity [20]. Among its specific political implications, ICTs include convergent telecommunications services. These services can be the changes in connectivity requirements, changes in the roles of ICT companies as partners in the electricity sector and the development of new ability for personnel of the industry [21]. Smart electrical substations guarantee the interoperability between elements connected to the network. They ensure privacy to the critical infrastructures and commercial use of power-sharing applications.

The communication between different devices of the Transformation Center and the Control and Supervision Center has not yet had a standard that allows efficient data communication. This limitation makes it difficult to implement the SCADA systems in the transformation centers [22]. Moreover, the interface with the relays was in many cases non-existent. To resolve these problems, the International Electrotechnical Commission launched some standard protocols called as IEC 61850. This protocol defines the communication between different devices connected to a local area network. Intelligent Electronic Devices (IEDs) are replacing the old wired communication techniques.

IEDs can integrate communication elements for tele-control [23]. Out of the three main layers of smart grid technologies (energy layer, communication layer, and application layer), the communications layer enables the existence of the smart grid (Figure 4). Communication technologies act as a circulatory system to interconnect the different systems and devices. It links the energy layer with the application layer, to communicate both with the energy supply chain [24].

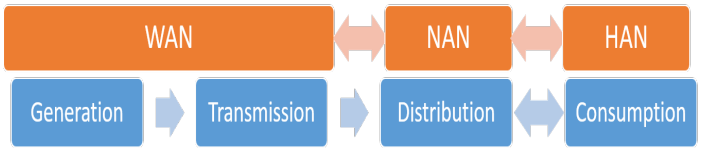


Figure 4. Smart grid topology for power flow in blue and information flow in orange

4.4. Supervisory System

The components of a supervisory system range from detection equipment, fault protection, communication, measurement, and data processing by the software. The software integrates all the intelligent devices (IED) that are part of the system [25]. However, the most relevant components of a supervisory system are the sensors of state and measurement. These sensors are installed in the field and are the base of the generation information for the supervising system. The status sensors have only two possible values zero or a logical one. These logical values reflect the states of the control elements of the system. For example, the status of a power switch can be estimated by the logical values of these sensors [26]. The protection relays that are currently manufactured are mainly IEDs. A typical IED (Intelligent Electronic Device) can contain around five to twelve protection functions and five to eight control functions. Control functions include separate devices control, an "auto-recloser" function, the self-monitoring function, and the communication functions etc. [27]. Currently, some IEDs work with the standard for the automation of IEC61850 electrical substations. This standard allows interoperability and good communications capacity (Figure 5) [28].

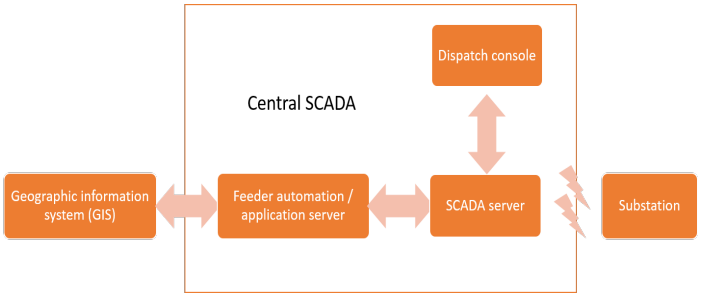


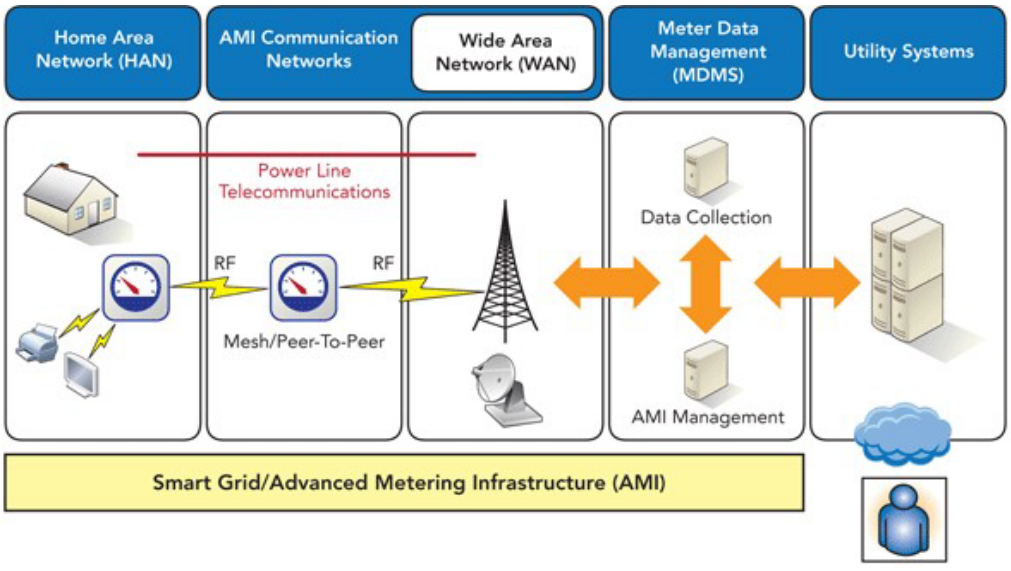
Figure 5. Feeder automation based on SCADA, current proposals in execution.

The communications network is the backbone of the entire supervisory system in real time. Its efficiency will depend on the speed and response time of the system to external events. The software is the last component that is required in the supervisory system; it is the way of representing the system [29].

4.5. Protection and Automation of the Network

The automation systems in substations for interconnecting a series of devices have existed for about 20 years, using proprietary protocols. These methods have been mainly responsible for the supervision of elements. Nowadays, this type of systems has evolved and continue to do so by following the protocols and actions declared in IEC 61850 [30]. This can be achieved by using peer-to-peer communications and enabling the exchange of data between systems at various levels and with various tools. This system also allows the supervision, control a series of devices or variables [31]. On the other hand, the increase in renewable energy generation and co-generation require the application of technologies in such a way that their management and protection is possible [32]. This impact on electrical networks is currently manifested, both in the transmission and distribution. The effect on substations, both in their protection and in their control, has been profound. This is due to the need of managing the intermittent electricity and several voltage levels [33].

The automation systems in substations for interconnecting a series of devices have existed for about 20 years, using proprietary protocols. These methods have been mainly responsible for the supervision of elements. Nowadays, this type of systems has evolved and continue to do so by following the protocols and actions declared in IEC 61850. This can be achieved by using peer-to-peer communications and enabling the exchange of data between systems at different levels and with various tools. This system also allows of supervisory control of devices or variables [34]. On the other hand, the increase in renewable energy generation and co-generation require the application of technologies in such a way that their management and protection is possible. The effect on substations, both in their protection and in their control, has been profound. This is due to the need of managing the intermittent electricity and several voltage levels. Therefore, for the development and supervision of the elements of the network, there are many emerging solutions for energy monitoring and parameters associated with it [35].



5. Digital Substations Control

To carry out and be able to develop all the functionalities that are expected from smart grids, this capacity (primarily redundant bi-directional data communications and the possibility of remote management) must be reproduced throughout the entire distribution network. It is already beginning to work in the section of the electrical consumption measurement infrastructure [36]. It must be implemented and improved in all sections of the power system.

The systems are based on bidirectional communications, technologies are associated with RF (radio frequency) wireless communications or power line carrier communications or broadband options (Figure 6) [37].

The desired functionality of smart meters could be the bidirectional communications with the electric company and with other devices (such as possible local energy managers). The other possible use of smart meter is alarms associated with quality of supply, personalization of the contract, planning, and possibility of bidding and purchase of electricity at the desired moments and control of loads, etc. [38]. This communication and monitoring capacity must be expanded to all measurement points and equipment in the electrical infrastructure. Also, it should be able to manage all distributed energy resources, considering not only the amount of energy flow but also the direction of it.

The technologies that affect smart grids are those that try to optimize and rationalize the management of electricity demand. As already mentioned, electrical energy cannot be stored (at least in enormous quantities), so it is necessary to generate it at the same time as it is needed. This situation means that at the very moment when a dedicated appliance is turned on, there must be a power plant (or any other type of electricity generator) that produces the necessary electricity [39].

5.1. Transport Network

To guarantee the electrical supply, in the absence of a line, the topology can be reconfigured to redirect the energy flows and thus avoid cutting off the supply. In this case, the systems of fault detection and reconfiguration of the network must act in small time spans for the control to be carried out in real time.

5.2. Distribution Network

The trend towards Smart Grids implies a replication or expansion of the existing capacities in the transport network to the distribution network. Considering that, unlike the conventional transport network, there are several owners and managers of the distribution networks. This situation will involve the definition of standard instruments and the creation of tools based on proven technology solutions. These tools will allow the integration of all types of generating plants, decentralized management of energy and automation of distribution and measurement services, supported by a system of communications that reaches the end user [40].

The electrical distribution companies can analyze and remotely control the electrical network by replacing meters with smart equipment. This equipment can take hourly measurements with Tele-management capacity [41]. In most cases, to guarantee interoperability, it is likely to install concentrators in the transformation centers with communication capabilities to control center of the corresponding distributor [42]. On the contrary, for the communication between the transformation center and the meters, the tendency in Europe is to use the power line carrier. The evolution of current electricity grids towards intelligent electricity networks implies the start of new intelligent equipment. This equipment will have local decision-making capacity and new communication technologies or the adaptation of existing ones. The current map of communication protocols used in the electrical system is very extensive and is regulated by the TC57 technical committee of the International Electrotechnical Commission (IEC). Thus, IECTC57 brings together several working groups to standardize the communications in the electrical system through the development of data models and generic interfaces. Each of these working groups has been responsible for defining and maintaining a communication standard based on the communication needs of each point of the electrical network [42].

Thus, it is worth highlighting:

- IEC60870-5 to communicate SCADA masters and electrical substations for the control and acquisition of data on serial lines or TPC / IP
- IEC60870-6, also known as TASE-2 for communications between powers centers over WAN networks.
- IEC61970 to interconnect energy management applications or EMSs in the environment of the control centers.
- EC61968 to communicate the control centers with the systems of the distribution network.
- IEC61334 for communications on PLC distribution lines.
- IEC62325 that defines a new interface between local utilities and the liberalized energy market.
- IEC62351 to determine security profiles to be used in all the previous ones at the TCP / MMS / 61850 level.
- IEC61850 (see fig 5) for automation in the environment of electrical substations (station and process buses) and communication between their IEDs (Intelligent Electronic Devices).

With the reference IEC61850 (see Figure 7 [43])), other similar regulations have been developed in other areas of application [44] such as:

- IEC61400-25 inherits a subset of communications services defined in IEC61850. It provides a new mapping of communications to Web Services and extends the data model modeling the functionalities, data, and attributes present in a wind turbine.
- IEC61850-7-420, which extends the data model by modeling the features, data, and attributes present in distributed generation systems such as photovoltaic systems, storage systems, diesel generators and heat exchange systems.
- IEC61850-7-410 is identical to IEC61850-7-420 for hydroelectric power plants.

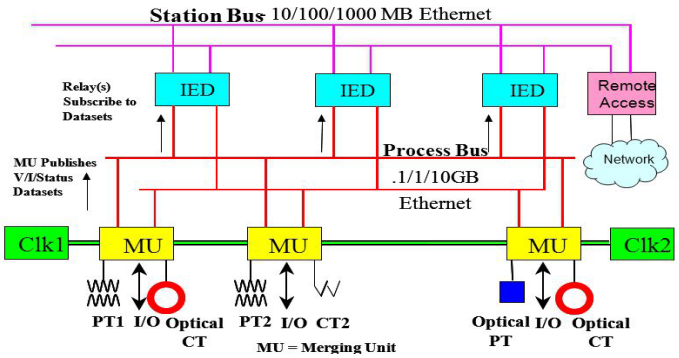


Figure 7. IEC61850 substation architecture.

6.1. Less Copper

Digital Substations utilize less copper by (1) replacing copper cables between switchyard and relay house by fiber optics, (2) replacing horizontal wiring between protection and control IEDs with IEC 61850, (3) reducing the number of connections between primary apparatus and redundant process interface modules and (4) reduction in cables for power supply and other connection for primary equipment.

6.2. Less Transport

In digital substation, more than 30 tons of material can be saved [45]. This helps in less transportation. The fiber optic installation reduces the cabling under 90% and by using optical instead of conventional CTs achieves an 80% weight reduction on CTs.

6.3. Space Reduction

IED installations require less space than conventional IOs. A space reduction around 30% to 60% for panels is possible. This results in higher integration of control and protection systems and reduction of switchyard footprint.

6.4. Less Installation and Outage Time

Digital substations cause faster installation of control systems due to reduction in number of panels for installation. There are fewer cables to be pulled, connected and tested. This causes reduction of feeder outage time by 40 to 50% during secondary system upgrades. This helps in full system test from process IO to protection, control and SCADA system offsite and Installation of new FO based system while the station is in service

6.5. Operational Cost Reduction

The Supervision of all exchanged data reduces the need for periodic maintenance testing and permanent supervision enables fast and precise actions in case of failures. IEC 61850 testing and simulation features enable fast and save isolation and testing of protection functions. Standard compliance enables efficient future retrofits of the secondary system (see Figure 5).

7. Conclusions and Recommendations

This work provides a review on digital substations importance and their working. This work emphasized on the importance of digital substations along with latest trends in modern smart grids. The provided comparison study proved the effectiveness of modern digital grids over conventional grids as well.

Apart of above-mentioned benefits, digital substations in Smart Grids provide a series of direct advantages to energy companies and end users that will help in the development of the energy market. The evolution of the electricity network provides the distribution companies with following benefits:

- Reduction of energy losses.
- Efficiency.
- Accomplish the optimization of the network infrastructure.
- Offering a better service to the client, with more commercial advantages (new rates, payment for use, etc.).

The users get the following benefits:

- Pay per use.
- Flexible rates.
- Remote management of the power supply.

Although at present, the home automation sector is in a complicated situation. It is promoting the instillation of control and automation systems in the home aimed at increasing energy efficiency, reducing energy losses, generated energy integration and accessibility to people with disabilities [46]. Currently, the number of agents involved in the electricity system has increased due to the liberalization of the energy market. The appearance of new participants is expected such as energy companies, system integrators, auditors, etc. It also predicts the development of renewable energies e.g., efficient solar cells, wind powered systems [47]-[48] and electric vehicles exponentially, in parallel with Smart Grids. So, it is necessary to keep these concepts in mind when carrying out any research or development related to the evolution of the electrical network.

R & D in the field of telecommunications for the development of new communication techniques and optimization of existing ones is still in progress. Technologies such as PLC, wireless connections or fiber optic, will provide a fundamental part of the Electrical Network of the future. Development in areas of power electronics, energy transport, electrical storage, energy efficiency, automated systems, new intelligent control techniques, etc. needs more investment and research [49]. Insofar, smart grids would play a great role in the development of smarter world having smart cities, smart homes, smart vehicles, etc. The ultimate requirement is to create systems of objects that are configured automatically and can operate autonomously [50]-[53].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like to thank Dr. Muhammad Nasir for his assistance on data organization and guidance for writing.

References

[1] Ramie, J. 2017 “Testing requirements for intelligent electronic devices in transmission & distribution facilities”in IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI), 7-11 Aug 2017: 1-36, Washington, DC: IEEE.

[2] Salvadori, F., Gehrke, C.S., Hartmann, L.V., de Freitas, I.S., Santos, T. D., and Texeira, T.A., "Design and implementation of a flexible intelligent electronic device for smart grid applications." in IEEE Industry Applications Society Annual Meeting, 1-5 Oct 2017: 1-6, Cincinnati, OH: IEEE.

[3] Chen, J., and Fu, Q., 2016 "Design of experiment platform for digital substation based on IEC 61850." In 5th International Conference on IEEE Computer Science and Network Technology (ICCSNT), 10-11 Dec 2016: 4-8, Changchun: IEEE.

[4] Hu, J., and Vasilakos, A., 2016 "Energy Big Data Analytics and Security: Challenges and Opportunities." IEEE Transactions on Smart Grid 7(5): 2423-2436, doi:10.1109/TSG.2016.2563461.

[5] Kong, P., 2016 "Wireless Neighborhood Area Networks with QoS Support for Demand Response in Smart Grid", IEEE Transactions on Smart Grid 7(4): 1913-1923, doi: 10.1109/TSG.2015.2421991.

[6] GVR "Substation Automation Market Size, Share | Global Industry Report 2024", (Accessed: 06- Jun- 2018) https://www.grandviewresearch.com/industry-analysis/substation-automation-market/.

[7] Faheem, M., Shah, S., Butt, R., Raza, B., Anwar, M., Ashraf, M., Ngadi, M. and Gungor, V., 2018. "Smart grid communication and information technologies in the perspective of Industry 4.0: Opportunities and challenges." Computer Science Review 30: 1-30, doi: doi.org/10.1016/j.cosrev.2018.08.001.

[8] Hong, T., Chen, C., Huang, J., Lu, N., Xie, L. and Zareipour, H. 2016. "Guest Editorial Big Data Analytics for Grid Modernization." IEEE Transactions on Smart Grid 7(5): 2395-2396, doi: 10.1109/TSG.2016.2593358.

[9] Siano, P., 2014. "Demand response and smart grids—A survey." Renewable and sustainable energy reviews 30: 461-478, doi: doi.org/10.1016/j.rser.2013.10.022.

[10] Mohtashami, S., Pudjianto, D., and Strbac, G., 2017. "Strategic Distribution Network Planning with Smart Grid Technologies" IEEE Transactions on Smart Grid 8(6): 2656-2664, doi: 10.1109/TSG.2016.2533421.

[11] Noriaki, K., 2013. "Electric network innovation by digital grid." IEEJ Transactions on Power and Energy 133(2): 141–144, doi: 10.1541/ieejpes.133.141.

[12] Hayashi, K., Kato, R., Torii, R., Taoka, H., and Abe, R., 2015. "Bi-directional power flow through a digital grid router" Journal of International Council on Electrical Engineering 5(1): 42-46, doi: https://doi.org/10.1080/22348972.2015.1068511.

[13] Rahim, M., Khalid, A., Javaid, N., Alhussein, M., Aurangzeb, K., and Khan, Z., 2018. "Energy efficient smart buildings using coordination among appliances generating large data", IEEE Access: 34670-34690, doi: 10.1109/ACCESS.2018.2805849.

[14] Meng, W., Ma, R. and Chen, H-H., 2014. "Smart Grid Neighborhood Area Networks: A Survey," IEEE Network 28(1): 24–32, doi: 10.1109/MNET.2014.6724103.

[15] Mortaji, H., Ow, S., Moghavvemi, M., and Almurib, H., 2017. "Load Shedding and Smart-Direct Load Control Using Internet of Things in Smart Grid Demand Response Management", IEEE Transactions on Industry Applications 53(6), 5155-5163, doi: 10.1109/TIA.2017.2740832.

[16] Chang, K.H., 2013. "Interoperable nan standards: a path to cost-effective smart grid solutions," IEEE Wireless Communications 20(3): 4–5, doi: 10.1109/MWC.2013.6549274.

[17] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., and Ayyash, M., 2015. "Internet of Things: A Survey on Enabling Technologies, Protocols and Applications," IEEE Communications Surveys & Tutorials 17(4): 2015 2347-2376, doi:10.1109/COMST.2015.2444095.

[18] Luo, F., Zhao, J., Dong, Z.Y., Chen, Y., Xu, Y., Zhang, X., and Wong, K.P., 2016. "Cloud-based information infrastructure for next-generation power grid: Conception, architecture, and applications," IEEE Transactions on Smart Grid 7(4): 1896-1912, doi: 10.1109/TSG.2015.2452293.

[19] Lim, J., Doh, I., and Chae, K., 2017. "Secure and structured IoT smart grid system management", International Journal of Web and Grid Services 13(2): 170, doi: https://doi.org/10.1504/IJWGS.2017.083383.

[20] Chiu, W., Sun, H., Thompson, J., Nakayama, K., and Zhang, S., 2016. "Integrated communications, control, and computing technologies for enabling autonomous smart grid", IEEE Communications Magazine 54 (12): 58-59, doi: 10.1109/MCOM.2016.7786111

[21] Yang, C. Vyatkin, V. and Pang, C., "Service-oriented extension of IEC 61850 for model-driven smart grid automation design." in IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, 29 Oct-11 Nov 2017: 5489-5496, Bejing: IEEE.

[22] Carvalheira, E., and Coronel, E., 2014. "Testing the protection system in IEC 61850 communication-based substations." in IEEE ANDESCON, 15-17 Oct 2014: pp. 1-1, Cochabamba:IEEE.

[23] Li, Y., Cheng, X., Cao, Y., Wang, D., and Yang, L., 2018. "Smart Choice for the Smart Grid: Narrowband Internet of Things (NB-IoT)", IEEE Internet of Things Journal, 5(3): pp. 1505-1515, doi: 10.1109/JIOT.2017.2781251.

[24] Krata, J. and Saha, T., 2018. "Real-Time Coordinated Voltage Support with Battery Energy Storage in a Distribution Grid Equipped with Medium-Scale PV Generation", IEEE Transactions on Smart Grid: 1-1, doi: 10.1109/TSG.2018.2828991.

[25] Veldman, E., and Verzijlbergh, R., 2015. "Distribution Grid Impacts of Smart Electric Vehicle Charging from Different Perspectives", IEEE Transactions on Smart Grid 6(1): 333-342, doi: 10.1109/TSG.2014.2355494.

[26] Aburukba, R.O., Al-Ali, A., Landolsi, T., Rashid, M., and Hassan, R., 2016. "IoT based Energy Management for Residential Area," in IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 27-29 May 2016: 1–2, Taiwan:IEEE.

[27] Baimel, D., Tapuchi, S. and Baimel, N., 2016. "Smart Grid Communication Technologies", Journal of Power and Energy Engineering 4(8): 1-8, doi:10.4236/jpee.2016.48001.

[28] Konstantelos, I., Giannelos, S., and Strbac, G., 2017. "Strategic valuation of smart grid technology options in distribution networks" IEEE Transactions on Power Systems 32(2): 1293-1303, doi: 10.1109/TPWRS.2016.2587999.

[29] Rossi, M., Viganò, G., Moneta, D., and Clerici, D., 2017. "Stochastic evaluation of distribution network hosting capacity: Evaluation of the benefits introduced by smart grid technology" in International Annual Conference, 20-22 Sep 2017:1-6, Cagiliari, Italy:IEEE.

[30] Wang, H., Qian, Y., and Sharif, H., 2013. "Multimedia communications over cognitive radio networks for smart grid applications," IEEE Wireless Communications 20 (4): 125–132, doi: 10.1109/MWC.2013.6590059.

[31] Cao, Z., Lin, J., Wan, C., Song, Y., Zhang, Y., and Wang, X., 2017. "Optimal cloud computing resource allocation for demand-side management," IEEE Transactions on Smart Grid, 8(4): 1943-1955, doi: 10.1109/TSG.2015.2512712.

[32] Cheng, X., Lee, W., and Pan, X., 2017. "Modernizing Substation Automation Systems: Adopting IEC Standard 61850 for Modeling and Communication". IEEE Industry Applications Magazine 23(1): 42-49, doi:10.1109/MIAS.2016.2600732.

[33] Miswan, N.S., Ridwan, M.I., Hayatudin, A., and Musa, I.A., 2015. "Interoperability testing for Digital Substation in Smart Grid domain: A power utility perspective" in International Symposium on Technology Management and Emerging Technologies (ISTMET), 25-27 Aug 2017: 154-158, Langkawai Island:IEEE.

[34] Abe, R., Taoka, H., and McQuilkin, D., 2011. "Digital grid: communicative electrical grids of the future". IEEE Trans. Smart Grid 2(2): 399-410, doi: 10.1109/TSG.2011.2132744.

[35] Deng, R., Yang, Z., Chow, M.-Y., and Chen, J., 2015. "A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches," IEEE Transactions on Industrial Informatics 11 (3): 570– 582, doi: 10.1109/TII.2015.2414719.

[36] Shen, B., Yuan, Y., Xie, L., Zeng, X., Liu, Y., and Luo, H., 2017. "Reasch on IED configurator automatic modeling based on IEC 61850" in IEEE Conference on Energy Internet and Energy System Integration (EI2), 26-28 Nov 2017 :1-5, Beijing:IEEE.

[37] Yang, Y., Xu, H., Gao, L., Yuan, Y., McLaughlin, K., and Sezer,K., 2017. "Multidimensional Intrusion Detection System for IEC 61850-Based SCADA Networks". IEEE Transactions on Power Delivery. Volume: 32 (2): 1068-1078, doi: 10.1109/TPWRD.2016.2603339.

[38] CRO "Renewable energy sources and climate change mtigation: special report of the Intergovernmental Panel on Climate Change", Choice Reviews Online, 49(11): 49-6309-49-6309, 2012, doi: https://www.ipcc.ch/report/renewable-energy-sources-and-climate-change-mitigation/

[39] Koko, S., Kusakana, K., and Vermaak, H., 2017. "Optimal energy management of a grid-connected micro-hydrokinetic with pumped hydro storage system", Journal of Energy Storage, 14: 8-15, doi: https://doi.org/10.1016/j.est.2017.09.007.

[40] Chena, W., Zhoua, K., Yanga, S., and Wuc, C., 2017. "Data quality of electricity consumption data in a smart grid environment", Renewable and Sustainable Energy Reviews, 75: 98–105, doi: https://doi.org/10.1016/j.rser.2016.10.054.

[41] Guerrero, J.L., García, A., Luque, J., and León, C., 2017. "Heterogeneous data source integration for smart grid ecosystems based on metadata mining", Expert Systems with Applications, 79: 254–268, doi: https://doi.org/10.1016/j.eswa.2017.03.007.

[42] Schett, G. 2017. "Digital substations for hydro power plants", in Proceedings of the Hydro Conference Series, Montreux, Switzerland 2017.

[43] Torbaghan, M.E., Burrow, M., Hunt, D., and Elcheikh, M., 2017. "Risk-Based Framework (RBF) for a UK Pan-European Supergrid" Energy, 124: 124-132, doi: https://doi.org/10.1016/j.energy.2017.02.058.

[44] Einfalt, A., Cejka, S., Diwold, K., Frischenschlager, A., Faschang, M., Stefan, M., and Kupzog, F., 2017. "Interaction of smart grid applications supporting Plug & Automate for intelligent secondary substations", CIRED - Open Access Proceedings Journal, 2017(1): 1257-1260, doi: 10.1049/oap-cired.2017.1080.

[45] Viswanatham, V., Chari, A., and Saritha, V., 2016. "Region-based group and hierarchical key management for secure smart grid communications." International Journal of Smart Grid and Green Communications, 1(1): 50, doi: 10.1504/IJSGGC.2016.077289.

[46] La, Q., Chan, Y., and Soong, B., 2016. "Power Management of Intelligent Buildings Facilitated by Smart Grid: A Market Approach", IEEE Transactions on Smart Grid, 7(3): 1389-1400, doi: 10.1109/TSG.2015.2477852.

[47] Waleed, A., Zhang, Q., Tavakoli, M.M., Leung, S.F., Gu, L., He, J., and Fan, Z., 2016."Performance improvement of solution-processed CdS/CdTe solar cells with a thin compact TiO2 buffer layer", Science bulletin, 61(1): 86-91, doi: https://doi.org/10.1007/s11434-015-0963-0.

[48] Waleed, A., Hassan, K.M., and Virk, U.S., 2014. "Designing a dual axis solar tracker for optimum power", Journal of Electrical Engineering 14(2): 168-173, doi: http://www.jee.ro/covers/art.php?issue=WH1375121304W51f6af987b6d1.

[49] Waleed, A., and Fan, Z., 2017. "Fabrication of stable organometallic halide perovskite NWs based optoelectronic devices" Science bulletin, 62(9): 645-647, doi: 10.1016/j.scib.2017.04.012.

[50] Tafreshi, S.M., and Lahiji, A.S., 2015. "Long-Term Market Equilibrium in Smart Grid Paradigm with Introducing Demand esponse Provider in Competition", IEEE Transactions on Smart Grid, 6(6): 2794-2806, doi: 10.1109/TSG.2015.2413951.

[51] Althaher, S., Mancarella, P., and Mutale, J., 2015. "Automated Demand Response from Home Energy Management System Under Dynamic Pricing and Power and Comfort Constraints", IEEE Transactions on Smart Grid, 6(4): 1874-1883, doi: 10.1109/TSG.2014.2388357.

[52] Kim, S-G., Hur, S-I, and Chae. Y-J., 2010. "Smart grid and its implications for electricity market design", Journal of Electrical Engineering and Technology, 5(1): 1-7, 2010.

[53] Sattarpour, T., Nazarpour, D., Golshannavaz, S., and Siano, P., 2015. "An optimal procedure for sizing and siting of DGs and smart meters in active distribution networks considering loss reduction", Journal of Electrical Engineering & Technology, 10(3): 804-811, doi:10.5370/JEET.2015.10.3.804.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 440-448 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Optimized Design of PM Halbach Array Linear Generator for Sea Wave Energy Converters Operate at Maximum Power Transfer

Ahmed Elsayed ELGebaly [*], Mohamed Kamal El-Nemr

Electrical power and machines department Faculty of engineering, Tanta university, Tanta, 31527, Egypt

A R T I C L E I N F O

Article history:
Received: 31 May, 2019
Accepted: 04 August, 2019
Online: 16 August, 2019

Keywords:
Marine Renewable Energy
Halbach Array
Linear Generators
Finite Element Method
Maximum Power Transfer

A B S T R A C T

This paper proposes Halbach array PM linear generators for direct conversion of ocean or sea wave energy. This generator is proposed to be directly coupled to a reciprocating wave energy system which may be a valued alternative for pneumatic and hydraulic systems. In this research, air-cored permanent magnet PM linear generator has been optimally designed according to optimal cost and performance indices. The design depends on Halbach array PM arrangement. The design this linear generator is appropriate for direct extraction of energy from sea waves specially at small wave amplitude. The finite element method and the space harmonic analysis are used to develop the electromagnetic design of this topology. In this article, the concept of the extraction of maximum permissible power from sea waves is established according to the direct wave energy converters hydrodynamics model. The dynamic analysis of the Halbach array linear generator is investigated during the state of maximum power extraction.

1. Introduction

This paper is an extension of work originally presented in 2018 Twentieth International Middle East Power Systems Conference (MEPCON) under title (Optimal Design of Slotless PM Halbach Array Linear Generator for Wave Energy Converters at Maximum Power Transfer Condition) [1]. Nowadays, the energy sources have great importance allover the world. Because of the expected deterioration of fossil fuels accompanied with hazardous environmental impacts, it is indispensable to establish clean renewable sources of energy. Ocean and sea waves energies are considered promising alternative renewable energies resources. On the Earth, it is considered as the largest renewable energy source. The running cost of this type of energy is low. Moreover, it is clean energy source. However, is not competently employed up till now. Preliminary reviews show that marine power has the ability to significantly contribute in energy market all over the world. Experts expect that the economical visibility of ocean or sea energies is very effective. It is expected that if wave-capturing technologies become totally developed, the predicted developed wave energy may equal 750 Trillion kWh/year [2].

Recently, numerous technologies are established to extract electric energy from seas or oceans waves [3]. These technologies are called the wave energy converters (WECs). They can transfer the mechanical wave energy to electrical energy. Such types of converters utilize the concept of a water column changing or buoy body movement [4]. Some of developed converters are more and more commercially established. There are some WECs directly convert the oscillating wave motion to reciprocal vertical movement. The small reciprocating motion with a speed about 2m/s may be obtained by these devices in some seas such as Mitterrandian Sea. Some researchers suggest the usage of induction generators to produce electric energy from sea waves. Nevertheless, this technology requires a hydraulic transmission system to convert the small velocity of the vertical motion to high rotational speed[5]. The wave energy converters which depend on the direct conversion of wave energy have great advantage related to the simple construction free of mechanical transmission system. The linear permanent magnet synchronous generators LPMSG have been developed and implemented as WECs which depend on the principle of Archimedes wave swing [1, 6]. The main shortage of such linear synchronous generators is the high cogging force. On the other hand, the air-cored tubular permanent magnet linear generators are suggested to overcome such force [7]. Therefore, the elimination of cogging force reduces the stress on the machine parts so there is no need for additional structure support. These machines can be built without the usage of back-iron. Therefore, the elimination of back-iron from stator and forcer leads to the

[*]Electrical power and machines department, faculty of engineering, Tanta university, Tanta, Egypt, +201153026587, ahmed.elgebaly@f-eng.tanta.edu.eg

elimination of their corresponding core loss. Additionally, such generators have significant lightweight compared with the conventional linear synchronous generators. The improvement of flux distribution in the air-cored PM generators improve their performance from the point of view of the developed voltage and power [1]. Halbach array arrangement is applied to provide low space harmonics magnetic field, which leads to further sinusoidal waveform of induced voltage and more developed power with the same forcer current. The field uniformity leads to extracting of higher power over small time. Therefore, the Halbach arrangement is proposed to be implemented in these generators to gain its performance improvements [1, 2, 8].

The optimized design of the proposed PM Halbach array linear generators requires the fulfilment of some indices; such as cost minimization and the production of appropriate voltage and power level. The recent optimization techniques are applied to obtain the optimal design of generator which is suitable for wave energy extraction [9]. In this paper, the design details are taken into consideration to obtain the optimal design. The design variables include the generator topology, the appropriate PM arrangement, the coils design and the generator overall dimensions. Moreover, this research proposes the concept of maximum power tracking MPT to make use of the most delivered wave energy. Preliminary, the MPT concept is explained in this paper according to the hydrodynamic performance of the WEC. The application of MPT mode requires the implementation specific control technique and power electronics converter [10].

2. Modelling of hydrodynamics of direct WEC

Figure1 demonstrates the structure of air-cored tubular PM linear synchronous generator for WEC based Halbach array configuration. According to this design, the floating buoy is directly connected to the movable translator of the PM linear generator where its base is strongly fixed to the sea bottom. The generator base may be connected to the sea bottom via wires. The changing of the height of the wave is the base of the operation of the WEC. The hydrodynamic model of the direct buoy WEC has been developed previously in numerous researches [5, 11]. The difference forces act on the moving part of the translator have been modeled. Without reaction force control, the moving part are affected by three forces; the input force produced by the wave, the extracted output power and the inertia force. It is recommended to integrate a suitable mechanical spring in the WEC to develop reaction force counteracts the force of inertia developed by the movable parts. The maximum power transfer condition in the WEC systems happens when both the inertia and the spring forces are equal in value and opposite in direction [1, 2].

The buoy may take a cylindrical shape with radius r and its motion is limited to be along the vertical axis (z). The cylinder is partly immersed. When the wave acts on the buoy in x direction as illustrated in Figure (1), a force is applied on the buoy cylinder due its tendency for floating. In this study, the incident wave is considered to have sinusoidal wave for simplicity.

The buoy motion equation can be formulated as the following:

$$m\frac{d^2z}{dt^2} = F_{stf} + F_{gen} + F_{hyd} \quad (1)$$

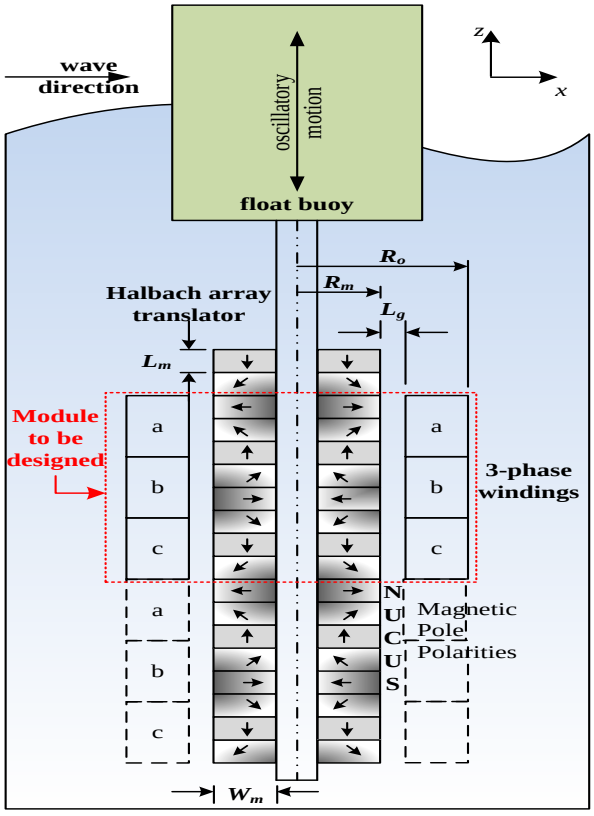


Figure 1: Proposed tubular Halbach array slotless linear generator for wave energy conversion

where m is movable parts mass i.e. buoy and Halbach array PMs. F_{stf} is the force produced by the spring, F_{gen} is the reacting generator force due to electrical power extraction, and F_{hyd} are hydrodynamic forces applied on the buoy exerted by the wave.

The stiffness or spring force is proportional to the translator displacement,

$$F_{stf} = -k_{st}\, z \quad (2)$$

where k_{st} is the constant of spring and z is the displacement on the vertical axis away from the balance point.

The force that causes the electrical power is modeled as a damping force where the produced force is proportional to the movable parts velocity. Therefore, the generated force is determined as:

$$F_{gen} = -B\frac{dz}{dt} \quad (3)$$

where B is the coefficient of damping which depends on the generated power. This coefficient should be continuously adjusted to enable the generator to extract the maximum allowable power from the wave.

The Archimedes' principle provides the equation that determine the hydrodynamic force correlated to the volume of water displaced:

$$F_{hyd} = -\rho g \pi r^2 z \quad (4)$$

To study the hydrodynamic model in steady state mode, all equation should be transfer to the frequency domain by applying Fourier transformation. At steady state, the cylinder is assumed to be exposed to sustained alternating force has a frequency ω. The output power from the generator can be maximized when the system is in the resonance where the stiffness force equals in magnitude the inertia force. Therefore, the natural frequency ω_r of the hydrodynamic system can be determined by [11]:

$$\omega_r = \sqrt{\frac{\rho g \pi r^2 + k_{st}}{m}} \quad (5)$$

The design of buoy cylinder relies on the hydrodynamic parameters which achieve the maximum power transfer condition. In this research, the proposed design is suitable for south shores of the Mediterranean Sea where the frequencies of the waves are within 3 radian per second [12]. The spring constant is determined depending on WEC rating and generator parameters [11]. Figure 2 illustrate the relation between the radius of buoy cylinder and the proper stiffness constant at different masses of movable parts. The required stiffness constant decreases with the increment of the radius. If the stiffness constant has negative values, the mass of the translator should be increased to achieve resonance condition [13]. So, the adjustment of spring constant is very vital to achieve the resonance condition; and therefore, to achieve the maximum power transfer condition. Nevertheless, the frequency of incident sea waves is continuously changed; so, the stiffness constant

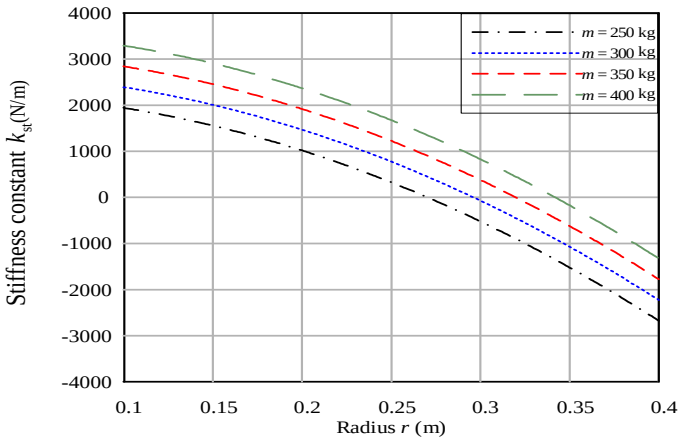


Figure 2: Variation of required stiffness constant with radius at different masses at ω_r = 3 rad/s

should be changed to change the natural frequency of the system to produce the maximum permissible power. The active adjustment of the stiffness constant is very complicated if it is mechanically implemented. Therefore, in this research, a adaptive electric spring is implemented to regulate the stiffness constant to run in resonance manner for different frequencies of the incident wave. The stiffness force or reaction force [10] can be created electrically due to certain current and voltage adjusted by power electronics converters. In this article, reaction force is applied during the operation of the generator to produce maximum power. Nevertheless, the movable parts mass should to be minimalized to reduce the reaction force and consequently, the produced generator current and its corresponding losses. Therefore, the mass of the generator movable parts is a vital index to obtain the optimized design.

3. Design of Halbach array WEC generator using FEM

Figure 1 illustrates that the proposed PM Halbach array linear generator has fixed coils surround a PM tubular Halbach array. The translator is composed of directed polarized permanent magnets for instance NdFeB. The magnets fluxes are directed with special sequence North (N) and South (S). But, the complete of the magnetic flux requires some magnets to be oriented in (C) direction. The PMs labeled (U) are oriented with 45 degrees or 135 degrees to produce smooth sinusoidal distributed flux. The resultant field tracks the translator motion, passing through the the concentrated coils. This paper studies the design of one module and the other modules have the same design parameters.

Low velocities WECs requires large reactive forces, therefore large dimension generators. In such iron-cored generators, the robust magnetic field leads to a noteworthy attraction force between the movable part and stator [14]. The structure can be significantly saved in the case when the cogging force can be decreased or even cancelled. The slotless air-cored linear generator is considered as a solution. The Halbach array has a vital advantage over the traditional air-cored related to the production of magnetic flux with less space harmonic content. The proposed generator module has the next ratings: 600 w rated power, 60 V rated phase voltage, the height of the wave about 1 m and the frequency of the wave about 3 rad/s. The ratings of the generator and wave is suitable to be applied on South Mediterranean Sea [12].

The proposed module may be repeated vertically to absorb larger amount of power from the incident wave. The overall length of generator depends on the depth of sea in the place of installation and on the stresses applied on the construction of the generator. In this research, the proposed generator has 3 kW; so, it consists of five modules. The following procedures explains the steps to obtain the optimal design of one module.

The investigation of the proposed model has been achieved by the application of finite element method by using FEMM package. Due to the slow motion of the translator in such type of generators, the analysis of the generator is considered magnetostatic problem. The problem of electromagnetic field is analyzed depending on FEM. The following equations is solved by FEM to analyze any electromagnetic problem. The vector of magnetic potential **A** is related to the vector of current density **J** by the following equation:

$$-\frac{1}{\mu}\nabla^2 \boldsymbol{A} = \boldsymbol{J} \quad (6)$$

where, (μ) is the permeability that depends the flux density for nonlinear materials. The advantage of solving FEM problem depending on the vector potential **A** is that all circumstances to be fulfilled are integrated within on equation. If **A** is determined for any geometry, the corresponding density of magnetic flux can be determined according:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \quad (7)$$

The proposed designs of the generators are analyzed by FEM to achieve the required ratings. Figure 1 illustrates the various dimensions of the module; R_o the outer radius of the module, τ is the pole pitch, L_g is air gap length, L_m is PM segment height, W_m is PM thickness and R_m is the PMs cylinder outer diameter.

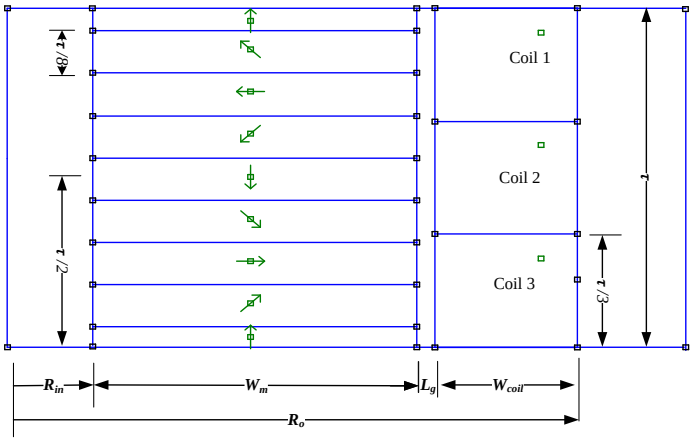


Figure 3: Full dimensions of one module of the proposed generator design

There are additional generator design parameters foi instance number of windings of each coil, the cross-section area of the copper conductors and the coil's filling factor. FEM analysis helps to determine the parameters of the design which achieve the suitable ratings. For any proposed dimensions and materials, FEM can calculate the different variables of the generator such as flux linkage, developed force overall weight. For each determined variable, the objective function is formulated to obtain the optimized design. Figure 3 illustrate the full dimensions applied in FEM for axisymmetric representation while R_{in} is the PMs cylinder inner radius, W_{coil} is the coil width and τ is the pole pitch. Along one pole pitch, the PM cylinder is divided into eight species. Along the full length of the translator, there are number of pole pitches equals N_{pit}. Over each pole pitch, the stator contains three coils where each coil presents one phase. The modularity of the proposed design allows to provide different generator ratings. Consequently, in case of 600 W module, generators of 1200, 1800 and 2400 W are developed. The control system applied for one module can be applied for the other modules without modifications.

Because of tubular shape of generator, the simulated model in FEMM is presented in axisymmetric dimensions. Several efforts have been achieved in [1], [2] and [5] which present a good guide to determine the probable ranges of design variables appropriate for the predetermined ratings demonstrated in Table.1.

The package FEMM is used to implement the finite element analysis of the proposed generator. The slow motion of the translator and the currents low frequencies which passes through coils lead to solving of the finite element problem as magnetostatic problem.

The optimization problem has some constrains; the first constrain is the filling factor of each coil should be less than 85 %; the second constrain is the losses which shouldn't exceed 50 % of the developed power. The dissipation of generator losses under water governs the description of constrains [2].

Some of post-process results of FEM present important terms which formulate the objective function. The obtained parametes are are as the following, the volume of PM segment Vol_{pm}, the volume of the coil Vol_{coil}, the produced loading force F_d, copper losses P_{cu}, and the maximum flux linkage at no-load λ_{max}. Some post-process results are essential to obtain some fitness function terms. The mass of both PMs and coils should be determined as the following:

$$M_{pm} = 8\ Vol_{pm}\ \rho_{pm} \tag{8}$$

$$M_{coil} = 3\ Vol_{coil}\ \rho_{cu}\ k_f \tag{9}$$

where M_{pm} is the mass PM cylinder for one module, ρ_{pm} is the PM material density, M_{coil} is the three-phase coils mass, ρ_{cu} is the density of copper and k_f is the filling factor. The maximum flux linkage per pole pitch λ_{max}/τ is an essential parameter should to be calculated to formulate the objective function.

The objective function is formulated to obtain the optimized design according to the next indices:

1- minimization of coils' mass
2- minimization of PM translator mass
3- achieving the rated power
4- achieving the rated voltage

Consequently, the objective function is formulated as the following:

$$O.F = k_1 M_{coil} + k_2 M_{pm} + k_3 \left| P_d - P_{dref} \right| + k_4 \left| \frac{\lambda_{\max}}{\tau} - \left(\frac{\lambda_{\max}}{\tau} \right)_{ref} \right| \tag{10}$$

and the fitness function to be maximized is:

$$fitness = -O.F \tag{11}$$

where P_d is the produced power by certain model, P_{dref} and $(\lambda_{\max}/\tau)_{ref}$ present the reference values and constants k_1 to k_4 are used as weighting factors to adjust the contribution of $O.F.$ terms.

Figure 3 illustrates the FEM presentation of one module located at the balance position where $z = 0$. With minimum number of calculations, this position has advantage related to the determination of required values of maximum flux linkage at no-load for coil 2 and developed force at loading if both coil 1 and coil 3 have maximum current. the proposed arrangement provides the benefit of obtaining the terms of objective function deprived of carrying out a complex dynamic modeling using FEM.

Table 1: the probable ranges of design variables appropriate for the proposed Halbach array generator

Variable	Symbol	Range	Unit
PM thickness	W_m	1: 9	Cm
Inner radius of PMs cylinder	R_{in}	0.5: 7	Cm
Air gap width	L_g	0.05: 0.5	Cm
Coil width	W_{coil}	2.5: 6	Cm
Pole pitch	τ	8: 15	Cm
Turns number of each phase	N	150: 300	Turn

In this research, the Genetic Algorithm (GA) is applied to obtain the optimized design with minimum weight and losses and provides voltage and power ratings.

Figure 4 demonstrates the flux produced by the Halbach array where the flux is directed out of the side area of the cylinder and partly deleted in the inner side of the cylinder. For the same generators volume, the characteristics related to providing higher power density of Halbach array linear generator is better than these of air-cored generator [2].

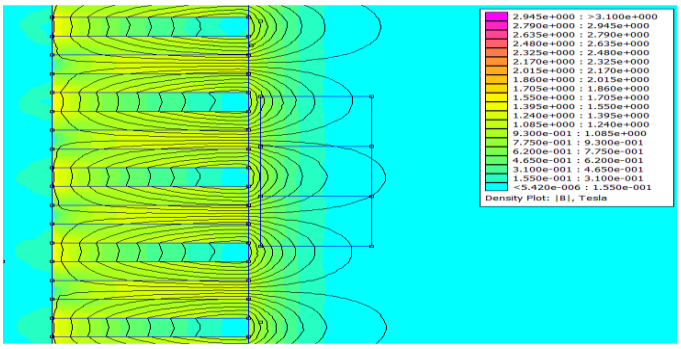


Figure 4: The Halbach array produced magnetic flux distribution and contour

To obtain the optimal values for the parameters, GA is applied to maximize the fitness function. In this study, the generation consists of 20 gene is applied. The six parameters are coded within one gene as binary string where these parameters are illustrated in Table.1. Therefore, the gene is composed of 120 bit. Then, the design parameters are passes to FEMM package to obtain their indices and design results. Then, the post-processing parameters are calculated. Therefore, the objective function and fitness value is determined for each gene. Then, GA applies its reproduction processes. The developed programs are carried out using programming language C++ and MFC library.

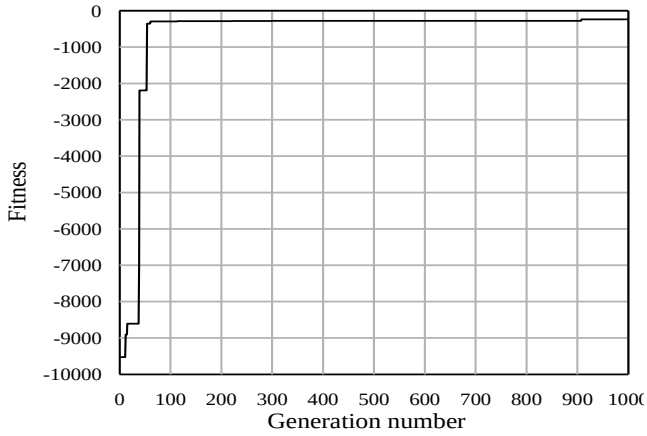


Figure 5: Fitness variation with GA generations for the optimization problem

The process of search is completed if the fitness value is stable at almost fixed value. Figure 5 demonstrates the changing of the fitness value versus the number of generation. After 1000 generation, the maximum fitness is stabilized at -270 as best vakue as shown in Figure 5. The stabilized fitness value is achieved approximately after 120 generations.

Table 2 illustrates the details of optimized design of the proposed generator topology for the specified ratings. The developed parameters of the optimized design are used in the dynamic analysis of the generator when it works in maximum power transfer state. The power density of the optimized design equals 41.6 W/kg which is more than the conventional air cored by 10% which equals 38 W/kg [15].

Table 2: The detailed parameters and dimensions of the optimized design

Dimensions		
Symbol	value	Unit
R_{in}	1.27	cm
W_m	4.37	cm
L_g	0.54	cm
τ	8.7	cm
W_{coil}	2.65	cm
Overall diameter	8.82	cm
Overall height	8.7	cm
The optimized parameters		
λ_{max}/τ	0.0676	wb/cm
M_{coil}	8.2	kg
M_{pmref}	6.2	kg
P_{dref}	600	W
Coil parameters		
N	393	turns
λ_{max}	0.588	wb
r_{in} Coil resistance	1.935	Ω
Inductance matrix $[M]$	$\begin{bmatrix} 0.033 & 0.0237 & 0.0236 \\ 0.0237 & 0.033 & 0.0237 \\ 0.0236 & 0.0237 & 0.033 \end{bmatrix}$	Henry

4. Performance assessment of Halbach array linear generator of WEC at maximum power transfer

According to the dimensions determined by the FEM, each coil of the there-phase has flux linkage over the displacement with one pole pitch illustrated in Figure 6. The flux linkage has sinusoidal waveform over τ.

Figure 7 demonstrations the equivalent circuit of the three-phase Halbach array linear generator. The there-phase equivalent

circuit model consists of three EMF supplies e_a, e_b and e_c. the phases assumed to be connected in star. Moreover, the model for each phase contains an internal winding resistance r_{in}, self-inductances L_a and the mutual induced voltage due to currents pass in the other phases is presented by two dependent sources. The load

$$\begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} = \frac{d}{dt}\begin{bmatrix} \lambda_a \\ \lambda_b \\ \lambda_c \end{bmatrix} = \frac{d}{dz}\begin{bmatrix} \lambda_a \\ \lambda_b \\ \lambda_c \end{bmatrix}.\frac{dz}{dt} = \frac{d}{dz}\begin{bmatrix} \lambda_a \\ \lambda_b \\ \lambda_c \end{bmatrix} v_z \quad (12)$$

$$\begin{bmatrix} v_a \\ v_b \\ v_c \end{bmatrix} = \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} - \begin{bmatrix} r_{in} & 0 & 0 \\ 0 & r_{in} & 0 \\ 0 & 0 & r_{in} \end{bmatrix}\begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} - \begin{bmatrix} L_a & M_{ab} & M_{ca} \\ M_{ab} & L_b & M_{bc} \\ M_{ca} & M_{bc} & L_c \end{bmatrix}\frac{d}{dt}\begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} - \begin{bmatrix} i_n & r_n \\ i_n & r_n \\ i_n & r_n \end{bmatrix} \quad (13)$$

$$\text{where } i_n = i_a + i_b + i_c \quad (14)$$

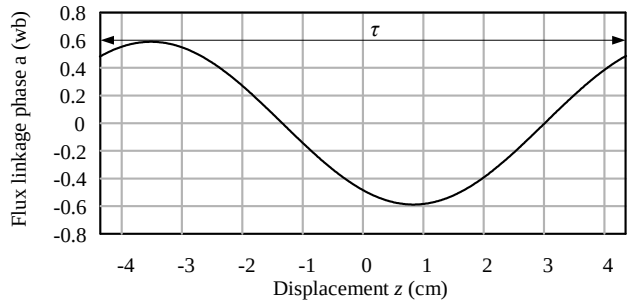


Figure 6: The changing of flux linkage with the vertical displacement over one pole pitch

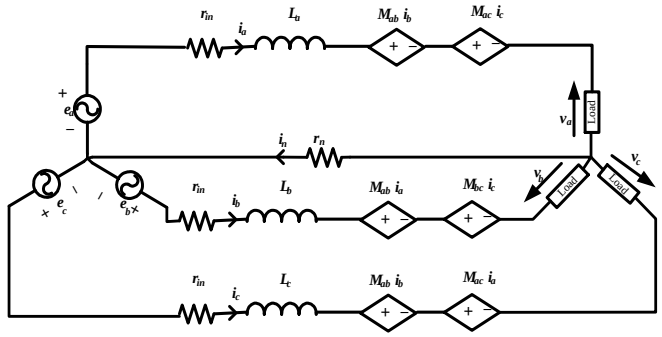


Figure 7: The equivalent circuit of the three-phase model of proposed Halbach array linear generator

To apply the maximum transfer condition from the mechanical point of view, the inertia force should be eliminated by stiffness force produced by applying current i_q* has phase shift of 90° with the induced phase voltage envelop [1]. From electrical point of view, the maximum power transfer happens when the circuit current is in phase with applied voltage. So, there is another current i_d* is in phase with voltage and therefore with the velocity. In (12), the voltage mainly relies on flux linkage and the velocity of the PM cylinder regarding the three-phase coils. In this generator, the energy is extracted from both rising and falling motions.

Now, the proposed Halbach array generator module is tested under two cases; the first when the maximum wave velocity equals 1.5 m/s and the second when it equals 1 m/s. Figure 8(a) shows the velocity applied on the floating buoy of WEC where the maximum velocity equals 1.5 m/s. This velocity is supposed to have sinusoidal waveform while its positive value means that the direction is up towards and the negative value means the reverse direction. As shown in Figure 8(b), the span of displacement ranges from 50 to -50 cm. Figure 8(c) illustrates the changing of the coils flux linkage during translator movement. As shown in Figure 8(d), the three phase voltages have 120° phase shift between each other although the voltages amplitude changes. Correspondingly, the velocity of the generator is directly related to the frequency of the induced voltages. The frequency related to speed should be considered during the voltage control of the generator output. The way of power extraction related to the value of currents passes through coils. Thus, the MPT is achieved by current regulation technique. Therefore, the MPT is achieved by allowing current i_q to pass through coils to produce stiffness force cancels the inertia force. Figure 8(e) illustrates the current i_q where its envelop has 90° phase shift with the envelop of voltage. Therefore, current i_q envelop is in phase with the translator displacement according to equation (2). The production of active power is achieved by allowing current i_d to pass through generator phases. The envelop of current id is in phase with the envelop of induced voltage as in Figure 8(e). Therefore, the maximum power transfer is achieved by allowing total current i_{tot} to be pass through each phase; where

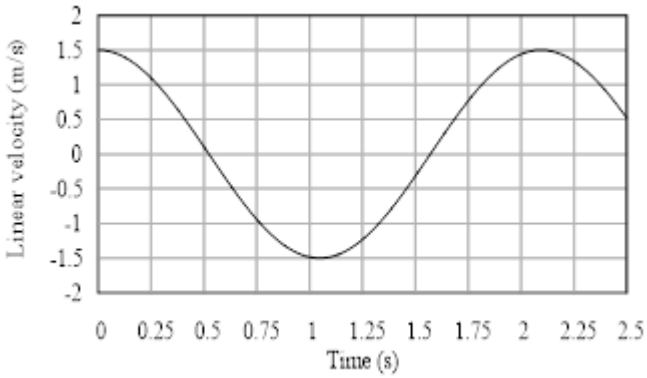
$$i_{tot} = i_d + i_q \quad (15)$$

These currents for phase a are illustrated in Figure 8(f). The total current with its two components can cancel mechanical inertia and electrical inductances. Figure 8(g) demonstrates both the stiffness and inertia powers which have the same amplitude and opposite sign. Figure 8(h) demonstrates the total output power P_{tot} produced by the generator at MPP and the output power P_{out}.

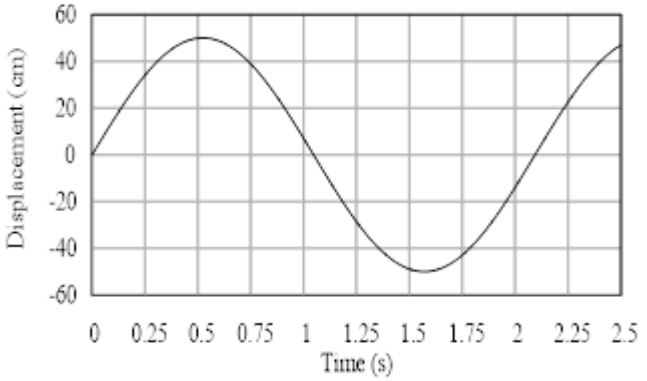
Figure 9 illustrates the generator performance at 1 m/s maximum velocity. It can be observed that the span of movement is reduced by 36 cm. Although the flux linkage, in this case, has the same amplitude, the induced phase voltages have less amplitude because they depend on the velocity. The currents are reduced due to the reduction of reaction forces and the total and output power. According to the results at different velocity, the concept of operation of MPT can applied to obtain the maximum power from the waves.

5. Conclusion

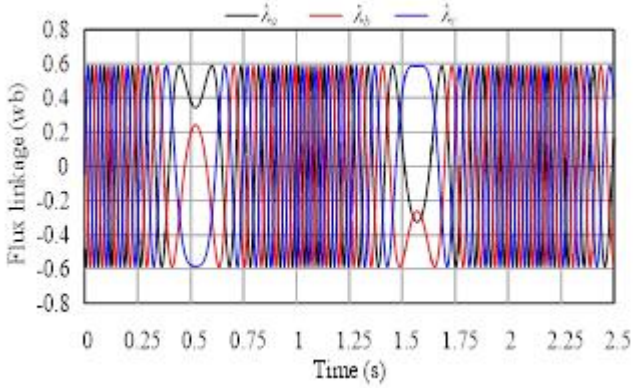
Slotless tubular Halbach array linear generator has been developed to be used within wave energy converter system at maximum power transfer condition. The simple hydrodynamic model of the WEC have been developed to describe the hydrodynamic circumstances to achieve MPT condition. The principle of operation of the proposed Halbach array linear generator has been introduced. Moreover, the parameters affects the design of the generator have been derived to formulate the proper objective function. The FEM has been used to get the design parameters for different dimensions of the generator. The GA in combined with FEM is used to get the optimal parameters of generator design achieves the maximum fitness value. The optimized design has been modeled and simulated to study its performance in the maximum power transfer condition. The WEC system has been simulated to work with low velocity sea waves as these in the Mediterranean Sea. The MPT has been successfully achieved by using the concept of current control to cancel all system inertias.



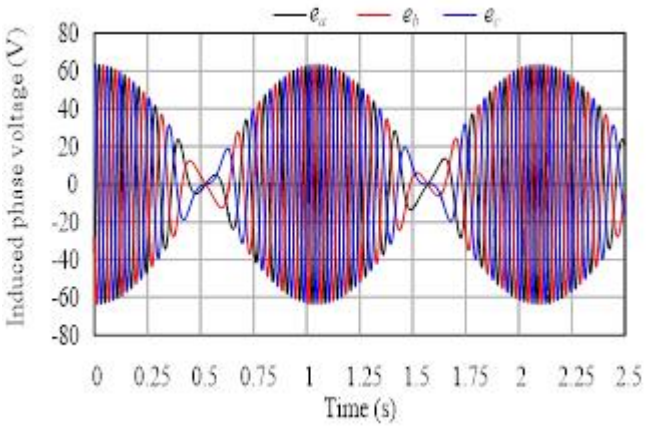
(a) Linear speed of WEC



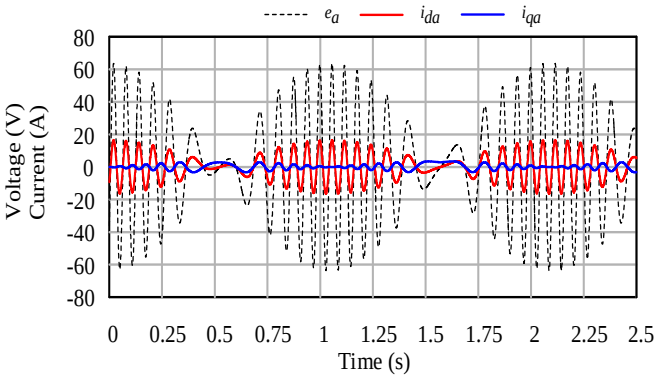
(b) Displacement of the buoy



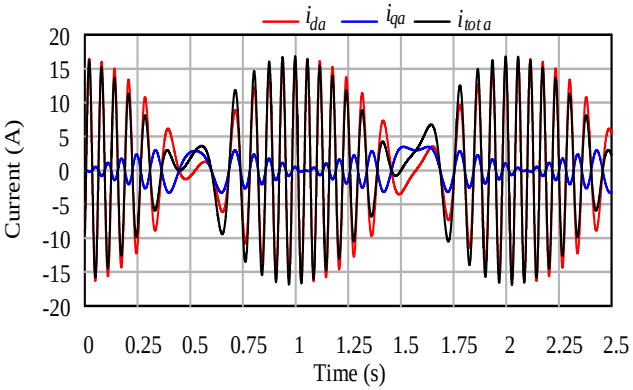
(c) Flux linkage of each phase



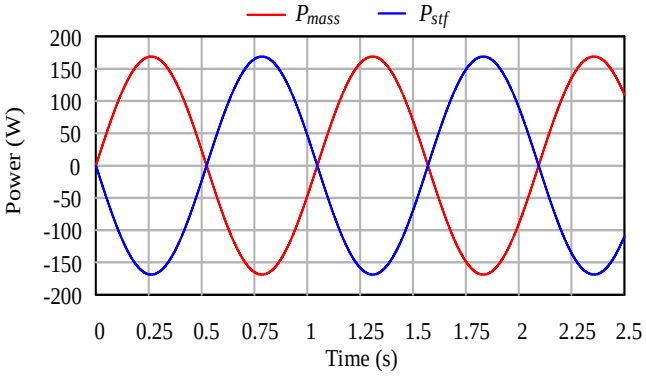
(d) Induced voltage of each phase



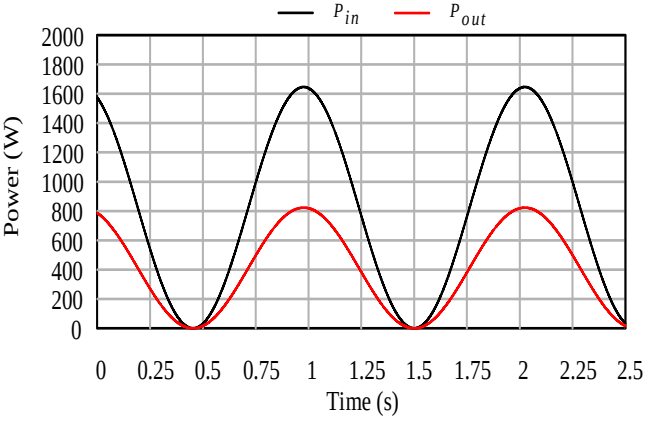
(e) Voltage and currents of phase a



(f) direct, quadrature and total currents of phase a

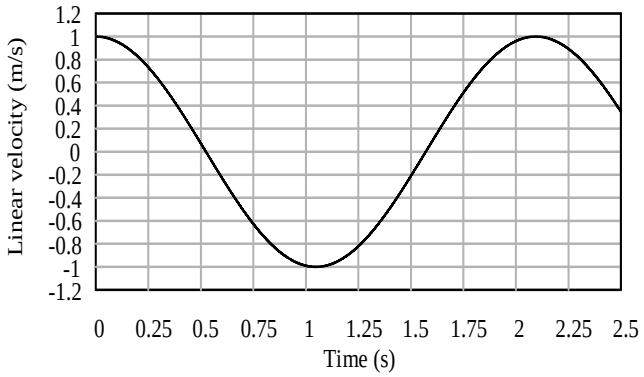


(g) Powers due to reactive forces (inertia and stiffness)

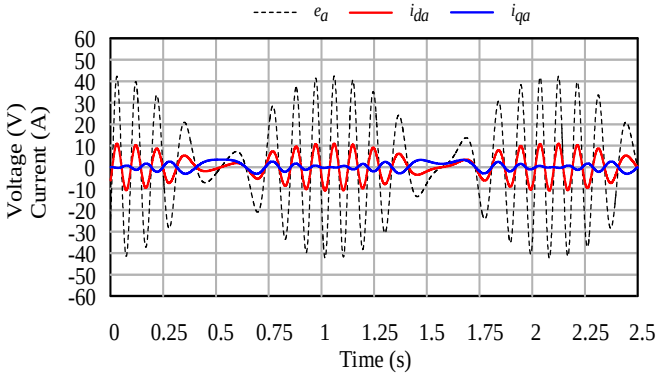


(h) Input and output power of WEC at MPT

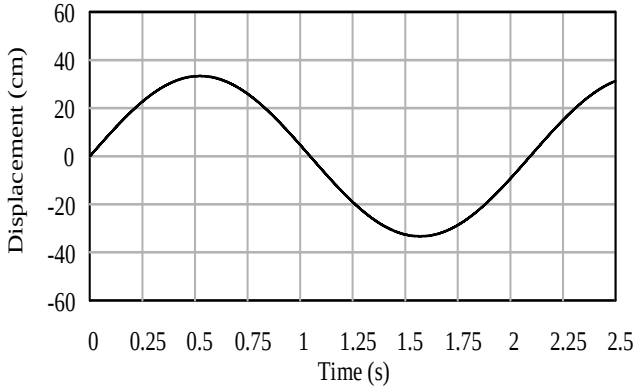
Figure 8: The performance of Halbach array linear generator as WEC at maximum linear speed with 1.5 m/s



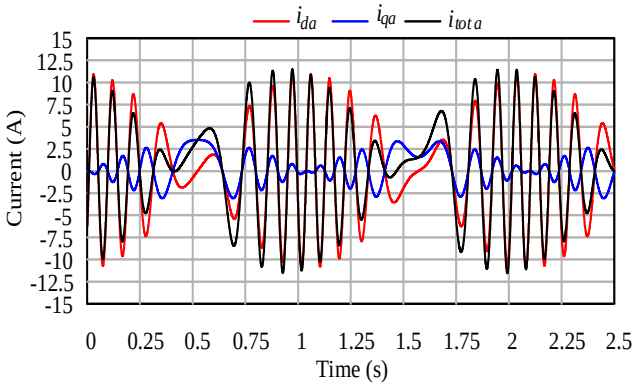
(e) Linear speed of WEC



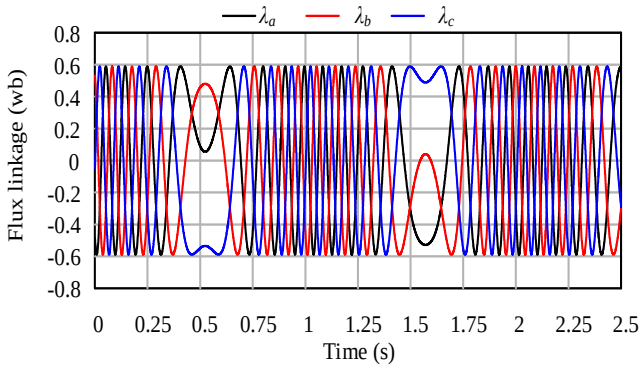
(e) Voltage and currents of phase a



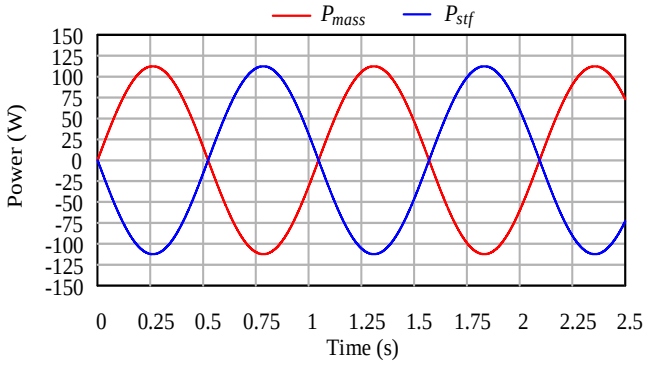
(f) Displacement of the buoy



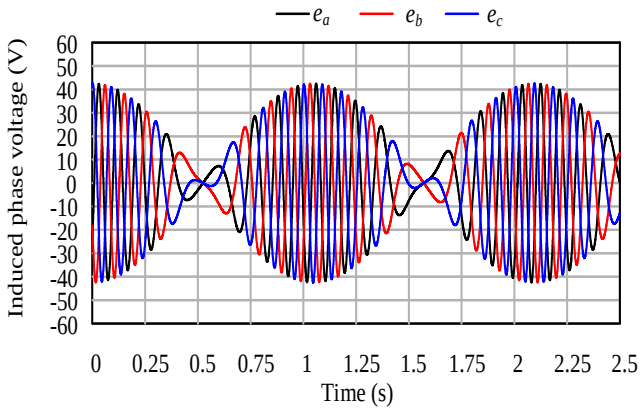
(f) direct, quadrature and total currents of phase a



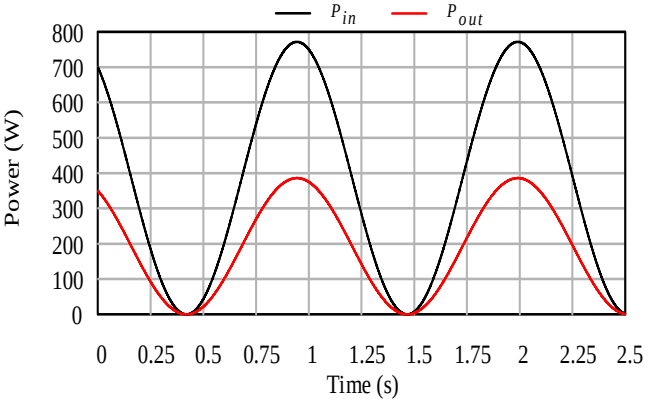
(g) Flux linkage of each phase



(g) Powers due to reactive forces (inertia and stiffness)



(h) Induced voltage of each phase



(h) Input and output power of WEC at MPT

Figure 9: The performance of Halbach array linear generator as WEC at maximum linear speed with 1 m/s

Acknowledgment

This study is performed at Energy Conversion Research Laboratory, Electrical power and Machines Department, Faculty of Engineering, Tanta University, Tanta, Egypt

References

[1] Ahmed E. ELGebaly, Mohamed K. El-Nemr, "Optimal Design of Slotless PM Halbach Array Linear Generator for Wave Energy Converters at Maximum Power Transfer Condition". in *2018* 20th International Middle East Power Systems Conference, MEPCON 2018, Cairo, Egypt, 2018. https://doi.org/10.1109/MEPCON.2018.8635113

[2] Ahmed E. ELGebaly, Mohamed K. El-Nemr, "Design and performance evaluation of halbach array linear generator for wave energy converters". in 2015 IEEE 8th GCC Conference and Exhibition, GCCCE 2015. Muscat, Oman, 2015. https://doi.org/10.1109/IEEEGCC.2015.7060035

[3] M. Mueller and H. Polinder, "Electrical Drives for Direct Drive Renewable Energy Systems. Electrical Drives for Direct Drive Renewable Energy Systems" 1–264 (Elsevier Ltd, 2013). doi:10.1533/9780857097491

[4] Bedard, R., Jacobson, P., Previsic, M., Musial, W. & Varley, R. An Overview of Ocean Renewable Energy Technologies. *Oceanography* 23, 22–31 (2011). doi: 10.5670/oceanog.2010.40.

[5] Vining, Jennifer, Tim Mundon, and Balky Nair. "Electromechanical Design and Experimental Evaluation of a Double-Sided, Dual Airgap Linear Vernier Generator for Wave Energy Conversion." *2017 IEEE Energy Conversion Congress and Exposition, ECCE 2017*. doi: 10.1109/ECCE.2017.8096926

[6] H. Polinder, B. C. Mecrow, A. G. Jack, P. G. Dickinson, and M. A. Mueller. "Conventional and TFPM Linear Generators for Direct-Drive Wave Energy Conversion" *IEEE TRANSACTIONS ON E NERGY CONVERSION*, VOL. 20, NO. 2, JUNE 2005

[7] Salauddin, M., M. A. Halim, and J. Y. Park. "A Low Frequency Vibration Energy Harvester Using Dual Halbach Array Suspended in Magnetic Springs." *Journal of Physics: Conference Series*. Vol. 660. Institute of Physics Publishing, 2015. doi: 10.1088/1742-6596/660/1/012011

[8] Zhu, Z.Q., and D. Howe. "Halbach Permanent Magnet Machines and Applications: A Review." *IEE Proceedings - Electric Power Applications* 148.4 (2002) doi: 10.1049/ip-epa:20010479

[9] Hodgins, Neil et al. "Design and Testing of a Linear Generator for Wave-Energy Applications." *IEEE Transactions on Industrial Electronics* 59.5 (2012): 2094–2103. *IEEE Transactions on Industrial Electronics* doi: 10.1109/TIE.2011.2141103

[10] B. Guo, R. Patton, M. Abdelrahman and J. Lan. " A continuous control approach to point absorber wave energy conversion 2016 UKACC 11th International Conference on Control (CONTROL), 31 Aug.-2 Sept. 2016, Belfast, UK. doi: 10.1109/CONTROL.2016.7737655

[11] Zheng, Xiong Bo, Jin Jiang, and Liang Zhang. "Power Characteristic Analysis and Optimization of Point Absorber Wave Energy Converter." *Applied Mechanics and Materials* 313–314 (2013): 837–842. https://dx.doi.org/10.4028/www.scientific.net/amm.313-314.837

[12] Kaiser, M. F.M., and O. E. Frihy. "Validity of the Equilibrium Beach Profiles: Nile Delta Coastal Zone, Egypt." *Geomorphology* 107.1–2 (2009): 25–31. doi: 10.1016/j.geomorph.2006.09.025

[13] Hazra, Samir et al. "Dynamic Emulation of Oscillating Wave Energy Converter." *2014 IEEE Energy Conversion Congress and Exposition, ECCE 2014*. Institute of Electrical and Electronics Engineers Inc., 2014. 1860–1865. *2014 IEEE Energy Conversion Congress and Exposition, ECCE 2014*. doi: 10.1109/ECCE.2014.6953645

[14] Dawande, M. S., Kanetkar, V. R. and Dubey, G. K. "Three-phase switch mode rectifier with hysteresis current control". *IEEE Transactions on Power Electronics* 11, 466–471 (1996). https://doi.org/10.1109/63.491640

[15] Farrok, Omar et al. "A New Optimization Methodology of the Linear Generator for Wave Energy Conversion Systems." *Proceedings of the IEEE International Conference on Industrial Technology*. Vol. 2016-May. DOI: 10.1109/ICIT.2016.7474965

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 449-457 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Fuzzy-Based Approach and Adaptive Genetic Algorithm in Multi-Criteria Recommender Systems

Mohamed Hamada[*,1], Abdulsalam Latifat Ometere[1], Odu Nkiruka Bridget[1], Mohammed Hassan[2], Saratu Yusuf Ilu[2]

[1]*African University of Science and Technology, Department of Computer Science & Engineering, 90001, Nigeria*

[2]*Bayero University of Nigeria, Department of Software Engineering, 700231, Nigeria.*

A R T I C L E I N F O

Article history:
Received: 25 April, 2019
Accepted: 29 July, 2019
Online: 16 August, 2019

Keywords:
Recommender System
Multi-criteria Recommendation,
Adaptive Genetic Algorithm
Fuzzy Logic

A B S T R A C T

Recommender Systems (RSs) are termed as web-based applications that make use of filtering methods and several machine learning algorithms to suggest relevant user objects. It can be said that some techniques are usually adopted or trained to develop these systems that generate lists of suitable recommendations. Conventionally, RS uses a single rating approach to preference user recommendation over an item. Recently, multi-criteria technique has been identified as a new approach of recommending user items based on several attributes or features of user items. This new technique of item recommendation has been adopted to solve several recommendation problems compared to the single rating approach. Furthermore, the predictive performance of the multi-criteria technique when tested proves to be further efficient as compared to the traditional single ratings approach. This paper gives a comparative study between two models that are based on the features and architecture of fuzzy sets system and adaptive genetic algorithm. Genetic Algorithms (GAs) are robust and stochastic search techniques centered on natural selection and evaluation that are often applied when encountering optimization problems. Fuzzy logic (FL) on the other hand, is known for its wide application in diverse fields in science. This study aims to evaluate, analyze, and compare the predictive performance of both methods and present their results. The study has been accomplished using Yahoo! Movies dataset, and the results of the performance of each model have been presented in this paper. The results proved that both techniques have significantly enhanced the system's accuracy.

1. Introduction

The advancement of Internet of things together with the rapid growth of e-commerce websites has caused uneasiness for customers to choose appropriately from the overwhelming number of items offered by these websites [1]. This fast development of web-based tools and the steady accessibility of a variety of information on the web have also given rise to the problem of information overload. As such, this has led online customers to make poor choices while purchasing items online [2]. To overcome this persistent problem, there is a need to introduce an intelligent decision support system that has the potential to scan through the available items using some computational and machine learning techniques to find and recommend appropriate items to users. Hence the need for RSs. Currently, RS is a significant tool that solves problems of information overload. It solves this problem by suggesting only items that are suitable to users. Increasing the number of items sold in e-commerce sites and an increase in customers' satisfaction from buying items online is one of the key benefits of RSs [3].

RS is a web-based application that supplies users with recommendations of items that might be of interest to them. The recommendations of items to users may be personalized or non-personalized [4]. The personalized recommendations are typically presented as orderly lists of items offered to the system user. It takes into consideration users' previous history for rating and predicting items. On the other hand, non-personalized recommendation systems recommend what is popular and relevant to all the users, which can be a list of top-5 items for every new user. Consequently, RS studies the user behaviour first, which could be expressed explicitly by the user through the rating of items or implicitly by just clicking on items. After learning the user's behaviour, it suggests items that might be useful to the user based on what it has learned from the user. It further suggests an ordered list of items in a way the user will like or rate them [3].

*Mohamed Hamada, University of Aizu, Software Engineering Lab, Japan, +81 242 372617, Email: mhamada2000@gmail.com

Rating of items by the user is one of the main forms of data transaction that is collected by the RS and uses it to classify RS into traditional RS and multi-criteria RS.

Traditionally, the commonly used recommendation approach adopted is either a hybrid-based, content-based or collaborative filtering approach. These approaches, together with subdivisions of collaborative filtering approach are shown in Fig. 1 [4]. These methods obtained an overall rating of a specific user on an item. Ratings of the users over items are used by recommendation algorithms to evaluate and predict users' preferences on new items. Single criterion rating tends to be inadequate in most cases because they have been proven to offer recommendations that are not quite efficient in achieving user desires, since users are unable to define their opinions based on various attributes of the item.

Multi-criteria Recommender Systems makes use of several items attributes to define the appropriateness of user items [5]. As an example, in a Movie RS, part of the attributes sighted may be the direction, action, visual, story. Multi-criteria RSs improves the Single criterion RS by putting into consideration varied items attributes in which users may like. Unlike Single Criterion RSs, users are capable of providing individual preference ratings on various item attributes. In most cases, additional relevant information offered by multi-criteria recommender systems helps to enhance recommendation value accurately. This enhancement depicts different features that a user may like about the item. In most RSs research, accuracy has been the most extensively examined point of emphasis [3]. However, more research is mandated on the effectiveness of new methods to integrate the multi-criteria rating information into the recommendation process efficiently. Therefore, this work aims to analyze and investigate how to use genetic algorithms and fuzzy technique to improve the precision in a multi-criteria RS. The study also provides a comparative analysis of the efficiencies of the two techniques.

The current paper is separated into five distinct sections, together with section 1. Section 2 presents a summary of traditional and multi-criteria RSs. Section 3 presents the methodology and frameworks used to implement both systems. Section 4 presents the results and comparison of both methods. Finally, section 5 presents the conclusion and discusses future work

2. Concept of Collaborative Filtering (CF)

CF is among the very commonly used approaches in generating valuable recommendations due to its human nature centred approach. It mainly "use the wisdom of the crowd or what seems to be common amongst my cycle" to recommend items to me. This is centred on the connection between users of the system and the item itself. It performs well in complex object recommendations (for example movies and music), which are seen to be totally autonomous from any machine-readable prediction of the items recommended.

Various companies have used RSs by implementing numerous algorithms that cut across different area. An example of such implementation is the statistical interactive learning techniques implemented by CleverSet, and also, the classical CF implemented in Net Perceptions [6] and Amazon [7].

CF approaches are mostly categorised into Memory-based and Model-based filtering method. The latter method makes use of earlier activities of the user to acquire a predictive model. This is done either by making use of some statistical analysis or machine learning techniques in making relevant recommendations [1]. Memory-based filtering approach makes use of recommendations grounded on similarities that exist in comparing previous activities or users' stored data. According to literature, there exist two ways commonly used to achieve memory-based techniques (item-based or user-based techniques) [8]. Fig. 1 shows a summary of the recommendation techniques. Model-based techniques can be developed using algorithms, some of such algorithms include Bayesian classifiers [9], Support vector machine, Neural networks [10], Boltzmann machine, Fuzzy patterns [11], Latent features, Genetic algorithms [12] and Matrix factorizations [13].

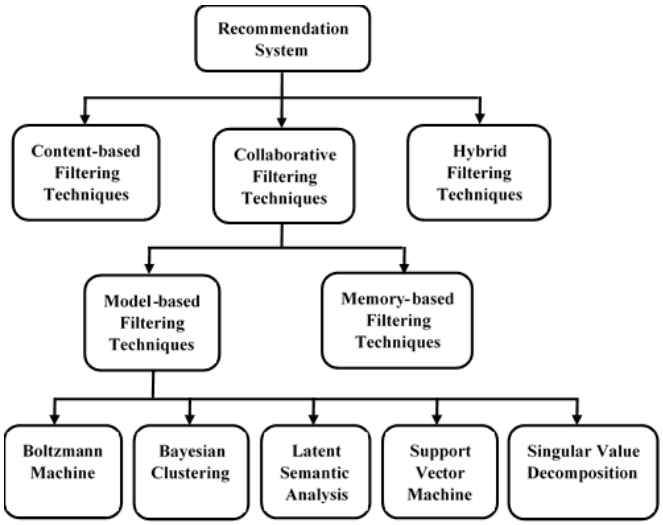


Figure. 1 Overview of existing recommendation methods

2.1. Asymmetric Singular Value Decomposition (ASVD)

ASVD is an extension of SVD. It relates to three vectors (c_i, d_i, $L_i \in \mathbb{R}^f$), also, users are signified by the objects they desire. User predictions to an item are represented in Eq. (1).

$$R'_{ui} = b_u + b_i + \mu + L_i^T(|R(u)|^{-0.5}\sum_{j\in R(u)}(R_{uj} - b_{uj})c_j + |N(u)|^{-0.5}\sum_{j\in N(u)} d_j) \quad (1)$$

by minimalizing the normalized squared error, we obtained Eq. (2).

$$\min_{b^*,L^*,c_i^*,d_i^*}\sum_{(u,i)\in K}(R_{ui} - \mu - b_u - b_i - L_i^T(|R(u)|^{-0.5}\sum_{j\in R(u)}(R_{uj} - b_{uj})c_j + |N(u)|^{-0.5}\sum_{j\in N(u)} d_j))^2 + \lambda(b_u^2 + b_i^2 + \| L_i\|^2 + \sum_{j\in R(u)}\|c_i\|^2 + \sum_{j\in R(u)}\|d_i\|^2) \quad (2)$$

where R_{ui} represent user ratings u to item i, μ represent the overall mean rating, b_u and b_i represent detected deviation of user u and observed deviation of item i correspondingly, $R(u)$ represent a set of the total items rated by the user u which are known ratings, $N(u)$ represent the set of all items which are rated by u, this can either be a known or an unknown rating and j represent an item.

The regularized squared error was minimised using stochastic gradient descent, and the boundaries λ and γ are

assigned positive real values of 0.002 and 0.005 respectively [14].

2.2. Multi-Criteria Recommender System (MCRS)

Majority of the RSs in use today are based on a single criterion rating which contains the overall user's satisfaction of an item [3]. Single criterion rating hides the users' choices and misleads the system when predicting items to the users [15]. It works in two-dimensional space of users and its utility function as expressed in Eq. (3)

$$R: Users \times Items \rightarrow R_0. \quad (3)$$

The utility function is gotten from the user inputs, e.g., transaction history or numeric ratings.

Single-criterion rating systems have shown successful recommendations in many areas. However, research in RS has identified the benefits of multi-criteria RS to enhance the prediction accuracy [16]. Multi-criteria RS does this by providing detailed information about the user ratings than a single-criterion RS and this enhances the recommendation process [5]. To determine the utility function of a user for a given item, we consider the overall rating R_0 and the user's ratings $R_1, \ldots, R_k$ for each criterion c (c = 1, 2, ..., k). Thus, the utility function R for multi-criteria RS is given as in Eq. (4)

$$R: Users \times Items \rightarrow R_0 \times R_1 \times R_2, \ldots, R_k. \quad (4)$$

2.3 Genetic Algorithm

Genetic Algorithms (GAs) are a robust and stochastic search technique centered on natural selection and evaluation that is frequently applied when encountering optimization problems. GA applications are frequently used when dealing with combinatorial optimization problems [17]. GA applies the same knowledge as any other evolutionary algorithm. It makes use of data collected from a set of individuals, represented as *G(i)*, where *G* represents the set (population), and *i* represents the number of the set (individuals). Each of the individuals is evaluated by assigning an appropriate fitness value, which is dependent on how respective individual is close to finding a solution to the problem. As its iteration continues, it finally approaches a local minima/maximum of the function.

A possible solution for a GA problem in GA is often called *chromosome or individuals*. Chromosomes are referred to as a set of genes; where each gene is represented by a distinct bit of nearby bit that encodes a part of the possible solution. Chromosomes are often scrambled as a sequence of bit string characterized abstractly in a genetic form. *Locus* is referred to as a location that has the encoding of some traits, while an *Allele* is a representation of each possible value of a locus (0 or 1) [18].

The *genotype* is referred to as a set of genes that exist in a genome, while the *Phenotype* is referred to as the physical representation of the genotype. After each encoding of the chromosome, the next step is to apply several reproduction procedures to the genotype. The genetic operators, such as the crossover, selection and mutation works on the genotype plot in other to return the value of the best chromosome.

Lastly, the stochastic nature of GA makes it challenging when trying to guarantee a solution that is optimal, and the algorithm may never stop running, which prompts for a termination condition. The algorithm is said to terminate only when the stopping criterion is met. The flowchart of a GA is shown in Fig. 2.

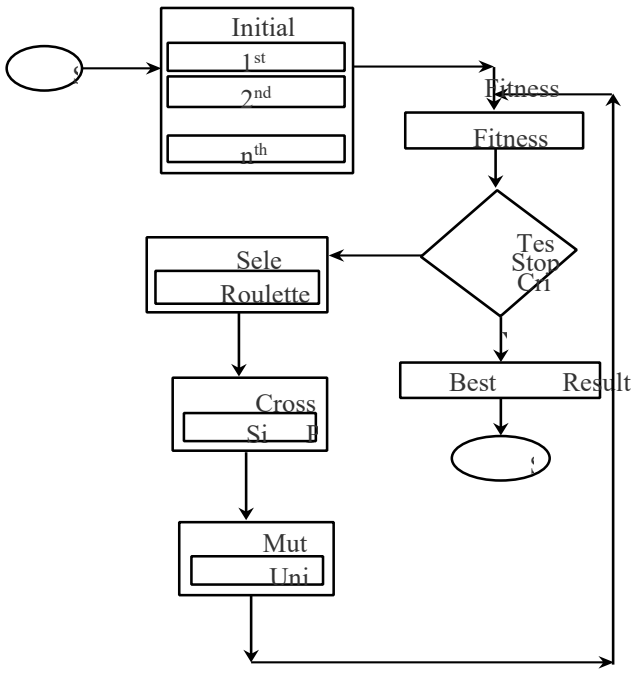


Figure 2 Flow chart of a standard genetic algorithm

2.3.1 Adaptive GA

Adaptive genetic algorithms are a subset of genetic algorithms. They provide substantial functionality improvements over the traditional GA implementations [19]. The configuration of genetic algorithms operators such as the probability of mutation and probability of crossover is a significant deterministic factor which affects the performance of a GA. The main function of GA is that, they help to modify the operators automatically, this is done by fine-tuning the operators centred on the algorithm's state [20]. These fine-tuning operations offered by GAs helps in maintaining convergence capacity and population diversity. The modification is centred on the values of the fitness solution which are, the best fitness value of the recent population and the best fitness value of the average population.

2.4 Fuzzy logic: An overview

Fuzzy Logic (FL) is a concept that is gotten from the fuzzy set theory which was developed by Lotfi Zadeh in 1975 [22]. It is a technique that can be used to mathematically express human reasoning or language by using a range of real numbers between 0 and 1. A membership function (MF) expressed as μA and defined as $\mu A: X \rightarrow [0, 1]$ is used to express the degree of membership of elements in a fuzzy set A.

2.4.1 Fuzzy logic architecture

Fuzzification

This step involves converting crispy set of inputs to a fuzzy set.

Linguistic variables

These are variables whose values are represented using ordinary language

Membership Function (MF)

It is used to represent graphically the amount of satisfaction a user got from an item.

Fuzzy Rule

It controls the output variable by applying some set of rules

Defuzzification

It is used to get a crispy value.

3. Methodology

This section presents the methods and frameworks used in the implementation of both systems.

3.1. Experimental dataset

Both systems were evaluated using Yahoo! Movie dataset collected from their website [5]. The dataset is a multi-criteria dataset that consists of user preferences ratings on movies based on four various movie features as follows: direction (c1), action (c2), story (c3), and visual (c4) effect of the movie. Each criterion ratings were considered on a 13-fold scale ranging from F through A^+ where F denotes the lowest preference, while A+ denotes the highest preference. An overall rating (c_0) denotes the total satisfaction derived by the user. The sample of the dataset is shown in Table 1.

Table 1: User Rating Obtained from *Yahoo*! Movies Rs

User ID	direction (c_1)	action (c_2)	Story (c_3)	visual (c_4)	Overall rating	Movie ID
1	10	10	10	8	9	100
	5	8	3	2	5	25
2	1	5	8	2	4	48
1	7	7	5	9	5	320
	10	13	4	13	10	360
3	7	8	10	10	9	224

For simplicity to model both systems, these categorical data were transformed to the numerical dataset as shown in Table 2.

The datasets were pre-processed to remove inconsistency that arose due to missing ratings. In the end, we achieved a total of 6078 users, 978 movies and 62,156 ratings with 0.0105 total sparsity estimate. The global median, global average, and standard deviation are 11.0000, 9.5221, and 3.5232, respectively. The summary of the records is shown in Table 3.

Table 2: User Rating Obtained from Yahoo Movies Rs (After Conversion)

Values	Frequency-rating	Percentage (%)	Cumulative-Percentage(%)
1	3395	5	5
2	1340	2	8
3	1522	2	10
4	1329	2	12
5	2051	3	16
6	2428	4	19
7	2489	4	23
8	3251	5	29
9	5586	9	38
10	7006	11	49
11	6702	11	60
12	12153	20	79
13	12904	21	100

Table 3: Rating Frequency of The $user \times item$ In Yahoo! Movie Datasets After Preprocessing.

User ID	direction (c_1)	action (c_2)	story (c_3)	visual (c_4)	Overall rating	Movie ID
1	B^+	B+	B+	B	B+	100
	D^-	B	D-	D	C	25
2	F	C	B	D	D+	48
1	C^+	C+	C	B-	C	320
	B^+	A+	D+	A+	B+	360
3	C^+	B	B+	B+	B-	224

3.2. Adaptive GA MCRS Methodology

The Adaptive GA method for rating approximation was divided into different stages:

1) Prediction of N multi-criteria Rating: This was done by decomposing the *k*-dimensional multi-criteria ratings into *k* single rating problem. To predict the unknown multi-criteria rating, an ASVD was used. For the next stage, the individual ratings were combined to provide an overall prediction of the ratings.

2) Acquiring the Function: This stage was aimed at approximating the connection between the original multi-criteria ratings and overall ratings of the items, such that $R_0=f(R_1,\ldots, R_k)$. An Adaptive GA was used to obtain the optimal weight of the individual benchmark for respective users over several criteria.

Feature weight is the priority a user gives to a feature, which may differ for each feature. In our method, Adaptive GA was applied to modify the weight to take user priorities for several features. User feature weight is denoted as a set, weight (u_a) = [w_i] (where i = 1, …, z, and z = number of features). Any feature with the weight zero is ignored for further calculation. Any double-valued vector strings are used to represent the genotype. The proposed adaptive GA parameters are reviewed below:

i. ***Initial Population*:** The original set of the population was created randomly with a size of 40 which consist of valid feature weights.

ii. ***Fitness Function*:** Fitness function is one of the most significant parts of GA. This function determines the success level of the individual chromosome after the population set is initialized. For a user who is active, a separate chromosome c is allocated a fitness function. The examined fitness function computes the accuracy of the predicted items using generated random weights. Eq. (5) represents the proposed fitness function.

$$fitness\ (c) = 1 - \left(\frac{\sum_{j=1}^{N_a} \left| R_{0j} - \left(\frac{w_1 R_{j1} + \cdots + w_z R_{jz}}{(w_1 + \cdots + w_z)} \right) \right|}{N_a} \right) \quad (5)$$

where N_a represent the aggregate number of times an individual chromosome c was tested, R_j represent the real ratings and R_{0j} represent the total ratings of u_a on an item, i_j.

iii. **Reproduction:** his step was solely dependent on the value of individual chromosome's fitness function. The fitness function defined for individual chromosome in the set of the population was computed and the chromosomes with the highest fitness rate were chosen as best chromosomes. The best chromosomes are reserved for the following generation.

iv. **Selection:** Roulette wheel selection technique was used, where parents are chosen according to their fitness value. Using the roulette wheel technique, each chromosome is allocated a slice of the wheel and the size of the slice is proportional to each chromosome's fitness. The wheel is then rotated and any chromosome the wheel lands each time was chosen.

v. **Crossover:** A single point crossover was selected to be implemented for the crossover. The selected crossover point was done randomly along the mated string length, then the next crossover point bits were swapped. The crossover probability was dependent on the fitness function, due to the fact that it's an adaptive GA.

vi. **Mutation:** Diversity was introduced by adopting uniform mutation. The value of the chosen gene was replaced with a uniform random value, selected between the specified upper and lower bounds for the gene by the mutation operator. The mutation probability was dependent on the fitness function, due to the fact that it's an adaptive GA.

vii. **Termination Condition:** Stopping criteria were centred on a well-defined number of initiations (100), and also on the convergence of best fit chromosomes.

3. Predicting Total rating: The calculation of the unknown total rating for test data was accomplished using the weighted sum of each predicted ratings. This was generated as depicted in Eq. (6).

$$R'_0 = \frac{\sum_{i=1}^{z} w'_i \times R'_{ji}}{\sum_{i=1}^{z} w'_1} \quad (6)$$

Where R'_j represent the predicted rating of u_a to i_j of the particular multi-criteria ratings. Developing Eq. (6) yields Eq. (7)

$$R'_0 = \frac{w'_1 R'_{j1} + \cdots + w'_z R'_{jz}}{(w'_1 + \cdots + w'_z)} \quad (7)$$

The aim of multi-criteria recommender systems is to generate a recommendation list for active users. The RS calculates the unknown total ratings by using feature weight function. Lastly, all unrated items and objects were sorted in non-increasing order with respect to the total ratings. Fig. 3 summarizes the working process of the Adaptive MCRS.

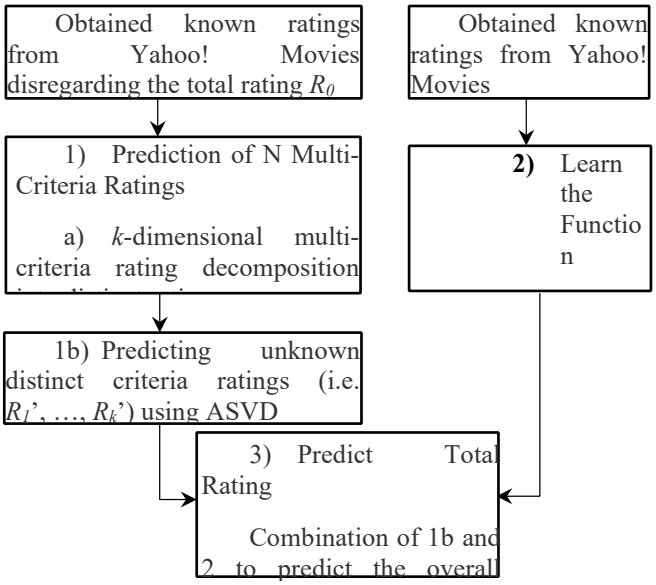


Figure 3 Structure of Adaptive Multi-Criteria Recommender System

3.3. Fuzzy-based MCRS Methodology

i. Predict N Multi-criteria rating

This step involves the decomposition of the n-dimensional multi-criteria rating space $R: Users \times Items \rightarrow Ro \times R_1 \times \ldots \times Rn$ where $(i = 1, \ldots, n)$ into n single-rating recommendation problems as $R: User \times Item \rightarrow R_c (c = 1, 2, \ldots n)$. The proposed model learns and predicts the user ratings for new items, however, we did not consider the overall rating at this stage.

$$(a, b, c) = \begin{cases} 0, & for\ x \leq a \\ \frac{x-a}{b-a} & for\ a \leq x \leq b \\ \frac{c-x}{c-b} & for\ b \leq x \leq x \\ 0 & for\ x > c \end{cases} \quad (8)$$

ii. Learn the Relationship f

This stage involves the computation of the relationship f relating to the overall rating r_0 and multi-criteria $r_1, \ldots, r_k$. Also, The aggregation function is considered as shown in the work of [21]. We integrated the FL technique to find the degree of satisfaction derived by a user from an item. Firstly, the user's ratings were represented using linguistic terms including (high, medium, low, etc.) and triangular membership function as shown

in TABLE 4. We considered a triangular membership function whose membership function is well-defined by the real numbers expressed as (a, b, c) as shown in Eq. (8).

We defined the membership function using 13 linguistic terms to represent interests of users and degrees to which each criterion was chosen. To compute the weight that determines the actual degree of membership of each criterion rating, this is calculated by dividing each user rating by n as shown in Eq. (9) below:

$$w_i = \frac{R_i}{n} \quad (9)$$

Where R_i represent user rating for each item and n represent the total number of fuzzy linguistic value.

We then derived the relationship f to learn how the user rates an item. Eq. (10) shows the aggregation function.

$$f(x_u^j) = \frac{\sum_{i=1}^{n} w_i\, c_i}{\sum_{i=1}^{n} c_i} \quad (10)$$

Where Ci is the users' criteria ratings and W_i is the weight associated with each criteria rating (for $i = 1, 2, \ldots, n$)

The relationship f for each user criteria rating for j items was computed, the function that resulted in the highest degree of membership for user u, for j ratings ($j = 1, 2, 3, \ldots n$) is chosen as the most preferred criterion for the user as shown in Eq. (11).

$$f(x_u) = max\{f(x_u^j)\}. \quad (11)$$

Table 4: Membership Function

User rating	TFN	Linguistic terms
1.0	(0,1,2)	V-V-V-V-low
2.0	(1,2,3)	V-V-V-low
3.0	(2,3,4)	V-V-low
4.0	(3,4,5)	V-low
5.0	(4,5,6)	Low
6.0	(5, 6,7)	Medium-low
7.0	(6,7,8)	Medium
8.0	(7,8,9)	Medium-high
9.0	(8,9,10)	High
10.0	(9,10,11)	V-high
11.0	(10,11,12)	V-V-high
12.0	(11,12,13)	V-V-V-high
13.0	(12,13,14)	V-V-V-V-high

i. Predict Overall rating

At this stage, the model predicts the overall rating $r'_0 = f(r'_1, r'_2, \ldots, r'_k)$ for an active user u_a. This is achieved through the integration of the trained fuzzy MCRS and the

single rating technique (Asymmetric SVD) so as to present the top-N recommendation to a user u_a. To calculate the overall rating for the user u_a, we use Eq. (12).

$$r'_0 = \frac{\sum_{i=1}^{n} w'_i \times P'_{ji}}{\sum_{i=1}^{n} w'_i} \quad (12)$$

Where

W_i represents the selected weight of the users to item i, and P_i represents the predicted rating of the users to the item i_j.

3.4. Implementation

Java programming language was used for implementing both systems. Java is a class-based, object-oriented programming language that is rich in FL and GA libraries that enhanced the process of both systems implementation,

3.4.1 Performance Metrics

To check the precision of both systems we explored the three categories of prediction precision measures [3]. As follows:

1. Rating predictions**:** Root Mean Squared Error (RMSE) and Mean Average Error (MAE) were used.

(a) RMSE: checks the precision of the predicted ratings as defined in Eq. (13) below:

$$RMSE = \sqrt{\frac{1}{|T_s|}\sum_{(u,i)}(\varepsilon T_s)\left(\widehat{P_{ui}} - r_{ui}\right)^2} \quad (13)$$

(b) MAE is defined in Eq. (14)

$$MAE = \frac{1}{T_s}\sum_{(u,i)}(\varepsilon T_s)\left|\widehat{P_{ui}} - \widehat{r_{ui}}\right| \quad (14)$$

Where: T_s = Test sets of user − item pairs (u, i) , $\widehat{P}_{ui}$ represent the predicted rating generated by the system and r_{ui} represent the actual rating of the user.

2. **Usage prediction**: This is used to measure the accuracy of the system based on how the system would predict the item a user would add to their content list. We considered Recall and Mean Average Precision (MAP) for the top-10 recommendation, which is defined in Eq. (15) and Eq. (17).

(a) $$recall = \frac{\#tp}{\#tn+\#fn} \quad (15)$$

Where tp is the number of positive useful items and fn is the number of useful predictions that are not in the top-n recommendation list.

(b) MAP: This calculates the value of Average Precision (AP) throughout the distinctive levels of recall. The value of MAP is expected to be less than or equal to 1 for a good algorithm and 0.5 for a bad algorithm. Eq. (16) and Eq. (17) was used to calculate these

$$AP = \frac{\sum_{i=1}^{N}(\text{precision(i)}\times\text{recall(i)})}{\text{number of relevant items}} \quad (16)$$

$$MAP = \frac{\sum_{u=1}^{M} AP_u}{M} \quad (17)$$

Where M represents the total number of items that are relevant in the catalogue of Top-N recommendation

3. **Ranking prediction**: This evaluates how accurate the

system would predict how a user would rank items according to their preference. Three approaches were considered, they are

(i) Area under the curve of Receiver Operating Characteristics (ROCAUC),
(ii) Fraction of Concordant Pairs (FCP),
(iii) Normalized Discounted Cumulative Gain (NDCG).

$$(a)\ ACU = \frac{1}{N}\left[\left(\sum_{i=1}^{\#tp_u} rank_{uj}^{+}\right)^{\#tp+1}_{2}\right] \quad (18)$$

where $rank_{uj}^{+}$ is the Position of the k^{th} relevant item on the top-N recommendation

$$(b)\ FCP = \frac{n_c}{n_c + n_d} \quad (19)$$

where n_c is the number of concordant pairs defined as:

$$n_c = \sum_u \varepsilon \cup |i,j| \left\{ \widehat{r_{ui}} > \widehat{r_{uj}} \Rightarrow \hat{r}_{ui} > \hat{r}_{uj} \right\} \quad (20a)$$

and n_d is the number of corresponding discordant defined as:

$$n_d = \sum_u \varepsilon \cup |(i,j)| \left\{ \hat{r}_{ui} > \hat{r}_{uj} \Rightarrow \hat{r}_{ui} \leq \hat{r}_{uj} \right\} \quad (20b)$$

4. Results and Discussion

This section discusses the results of the experiments done

4.1. Experiment

To verify the precision and efficiency of both systems, experiments were done using the Yahoo! movie dataset. An offline experiment was performed to simulate the actual system. Both systems were compared with their corresponding traditional collaborative filtering technique known as Asymmetric Singular Value Decomposition (AsySVD). We performed each test using a 10-fold cross-validation rule, which randomly divides the datasets into 10 separate subsets. The dataset was divided on a ratio of 90: 10 where 90% is used as training-sets and 10% as a test-set. The precision of both systems was checked by applying the metrics discussed in section 3.3.1. The results of evaluating the systems were compared with their corresponding traditional rating. TABLE 5 and TABLE 6 give the breakdown of the resulting performance evaluation of the system. We used Adaptive MCRS, Fuzzy-MCRS and AsySVD to represent Adaptive Genetic MCRS, Fuzzy logic model and Asymmetric SVD respectively.

Table 5: Performance Evaluation Results

Performance-measure	AsySVD	Adaptive-MCRS	%Accuracy Improvement
MAE	2.3980	1.5990	0.7990(49.9%)
RMSE	3.0900	2.1220	0.9680(45.6%)
MAP	0.0200	1.0290	1.0090(98.1%)
AUC of ROC	0.6880	0.9510	0.2630(27.7%)
FCP	0.7100	0.9460	0.236(24.9%)

It can be clearly seen from TABLE 5 above that the Adaptive MCRS outperforms the traditional AsySVD. The results proved that Adaptive MCRS achieved better predictive performance. The Adaptive MCRS approach achieved a decrease in prediction accuracy in terms of RMSE and MAE. RMSE and MAE had a value of 0.968 and 0.799 respectively, this decrease might seem minute, but it is able to generate a totally dissimilar answer and interpret into substantial enhancement on the accuracy and quality of recommendation. Adaptive MCRS achieved a MAP of 1.029, outperforming the traditional AsySVD by 50.5%. There is a necessity for recommendation algorithm to generate recommended ordering of objects or items that matches how users would have ranked the same item since users are frequently concerned of the items at the top-N recommendation list. The Adaptive MCRS achieved a more accurate and improved rank accuracy of 33.2% and 38.2% respectively for FCP and AUC when compared to the conventional AsySVD, which implies that Adaptive MCRS will provide a further precise top-N recommendation list to users. Fig. 4 is a line chart clearly demonstrating the high-performance result achieved by Adaptive MCRS as compared to AsySVD.

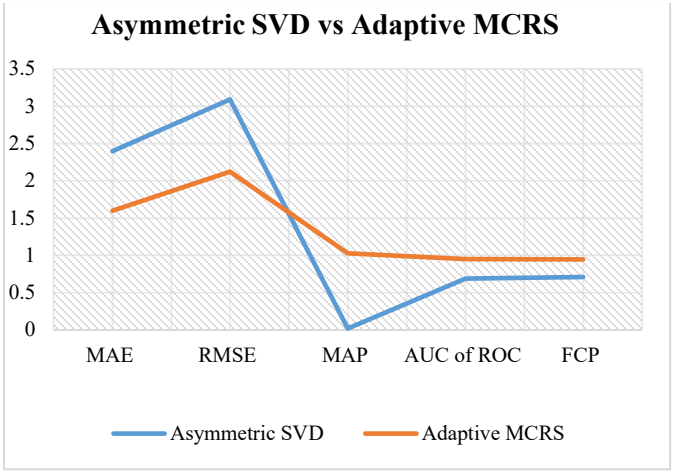


Figure. 4. The result of the evaluation of Adaptive MCRS

The result From Fig. 5 and TABLE 6, shows that Fuzzy-MCRS achieved a lower prediction error in RMSE and MAE, and similarly shows a higher-ranking accuracy in AUC and FCP. This indicates that modelling a multi-criteria RS using fuzzy logic would yield an efficient and more accurate system than the traditional AsySVD.

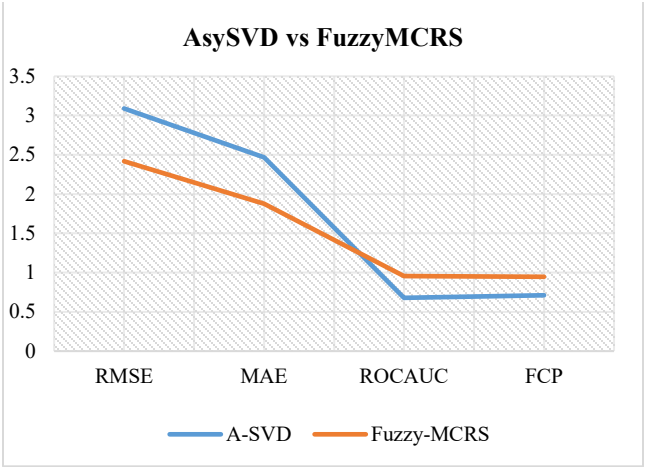


Figure 5. Comparing AsySVD and Fuzzy-MCRS

Table 6: Comparison of Result Of The Evaluation

Metrics	Fuzzy-MCRS	AsySVD	Accuracy Improvement
RMSE	2.4176	3.0895	0.6719(27.7%)
MAE	1.8753	2.4677	0.5924(31.5%)
ROCAUC	0.9558	0.6786	0.2572(26.9%)
FCP	0.9467	0.7118	0.2349(24.8%)

Table 7 and Fig. 6 clearly shows the comparison of the experimental result from both models. Adaptive MCRS resulted in lesser prediction inaccuracy in MAE and RMSE than the Fuzzy-MCRS, which implies that Adaptive MCRS yielded a better rating prediction than Fuzzy-MCRS. But both models seem to have equivalent usage accuracy, although Fuzzy-MCRS seem to have produced a higher usage accuracy level in ROCAUC and FCP than the Adaptive MCRS. This shows that Fuzzy-based model would be ideal for predicting the ranking behaviour of users for an item since it studies the exact degree to which a user will rank items to their content list and the order in which the items would be ranked. Similarly, Adaptive MCRS would be ideal to model a system that would be accurate in predicting and classifying the rate of satisfaction derived by a user from a purchased item.

Table 7: Comparison of Fuzzy-Mcrs and Adaptive-Mcrs

Metrics	Fuzzy-MCRS	Adaptive MCRS	Percentage Change
RMSE	2.4176	2.122	0.2956 (13.9%)
MAE	1.8753	1.599	0.2763 (17.3%)
ROCAUC	0.9558	0.951	0.0048 (0.5%)
FCP	0.9467	0.946	0.0007 (0.1%)

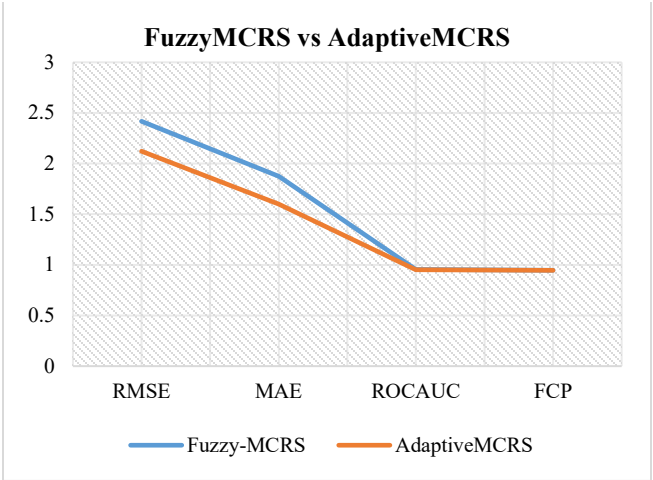


Figure. 6. Comparison of Fuzzy-MCRS and Adaptive MCRS

5. Conclusion

Providing efficient techniques for integrating the criteria ratings in multi-criteria RSs is of the utmost significance in predicting preferences of users based on several attributes of the items. Machine learning methods and other powerful techniques from the area of artificial intelligence should be the priorities, especially in product recommendations such as movies, song, and other items that could be purchased online. In this study, two important techniques (Adaptive genetic algorithm and fuzzy logic) have been proposed for developing recommendation models that could efficiently integrate the criteria ratings for making good recommendations. The experiments were conducted using a real-life dataset from Yahoo Movies, and the results have been analysed and compared. The study has shown that the two techniques realized the objectives of delivering better recommendation and predictive performance. This performance includes a reduction in error prediction and increased classification and rank accuracy. Lastly, the obtained experimental results from associating the predictive implementation of both multi-criteria recommender systems using the adaptive genetic algorithm and fuzzy logic, with existing conventional CF approach showed that the proposed system produced results that are better and also outperformed the conventional CF recommender system. Nonetheless, to achieve a better performance, it is recommended to design a hybrid model which could resolve some of the issues realized in the results of both systems.

Therefore, we propose to integrate both models to achieve higher performance. However, in the future, we plan to acquire a larger amount of data from a different domain to further certify the accuracy of the proposed system.

References

[1] L. Liu and D. Xu, "Multi-Criteria Service Recommendation Based on User Criteria Preferences," 2011.
[2] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems : Introduction and Challenges. 2015.
[3] D. R. Tobergte and S. Curtis, *Recommender Systems Handbook*, vol. 53, no. 9. 2013.
[4] R. Burke, "Hybrid Web Recommender Systems," *Adapt. Web*, pp. 377–408, 2007.
[5] K. Lakiotaki, N. F. Matsatsinis, and A. Tsoukiàs, "Multi-Criteria User Modeling in Recommender Systems."
[6] J. Mo, S. Sarkar, and S. Menon, "Know When to Run: Recommendations in Crowdsourcing Contests," *MIS Q.*, vol. 42, no. 3, pp. 919–944, 2018.
[7] G. Linden and B. Smith, "Amazon-Recommendations.pdf," no. February 2003.
[8] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egypt. Informatics J.*, vol. 16, no. 3, pp. 261–273, 2015.
[9] K. Miyahara and M. J. Pazzani, "Improvement of Collaborative Filtering with the Simple Bayesian Classifier 1," pp. 1–28, 2002.
[10] M. Hassan and M. Hamada, "A Neural Networks Approach for Improving the Accuracy of Multi-Criteria Recommender Systems," *Appl. Sci.*, vol. 7, no. 9, p. 868, 2017.
[11] R. Yera and L. Mart, "Fuzzy Tools in Recommender Systems : A Survey," vol. 10, pp. 776–803, 2017.
[12] U. Marung, N. Theera-Umpon, and S. Auephanwiriyakul, "Top-N recommender systems using genetic algorithm-based visual-clustering methods," *Symmetry (Basel).*, vol. 8, no. 7, pp. 1–19, 2016.
[13] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *Proc. 14th Annu. Conf. Uncertain. Artif. Intell.*, pp. 43--52, 1998.
[14] O. Alter and G. H. Golub, "Singular value decomposition of genome-scale

mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening," *Proc. Natl. Acad. Sci.*, vol. 103, no. 32, pp. 11828–11833, 2006.
[15] G. Adomavicius and Y. Kwon, "for Multicriteria."
[16] G. Adomavicius and A. Tuzhilin, "Towards the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions," pp. 1–43.
[17] R. P. Badoni, D. K. Gupta, and P. Mishra, "A new hybrid algorithm for university course timetabling problem using events based on groupings of students," *Comput. Ind. Eng.*, vol. 78, pp. 12–25, 2014.
[18] "Genetic Algorithms Thanks to : Much of this material is based on :," *Evol. Comput.*, 2009.
[19] L. Budin, M. Golub, and D. Jakobovic, "Parallel Adaptive Genetic Algorithm," pp. 157–163, 1998.
[20] L. Jacobson and B. Kanber, *Genetic Algorithms in Java Basics*. 2015.
[21] D. Warshawsky and D. Mavris, "Choosing aggregation functions for modeling system of systems performance," *Procedia Comput. Sci.*, vol. 16, pp. 236–244, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 462-467 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Marginalizing Last Mile Logistics Cost through 4th Party Milk Run

Robert de Souza[1], Linda William[2,*], Cher Kian Lee[3]

[1]*The Logistics Institute – Asia Pacific, National University of Singapore, 21 Heng Mui Keng Terrace #04-01, Singapore 119613*

[2]*School of Informatics and IT, Temasek Polytechnic, 21 Tampines Ave 1, Singapore 529757*

[3]*Independent Researcher*

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 28 July, 2019
Online: 21 August, 2019

Keywords:
Last mile logistics
Milk run
Excess capacity
Horizontal collaboration
Marginal cost

ABSTRACT

To reduce its high last mile logistics cost, a company may explore different options such as a horizontal collaboration with another company that has a similar logistics requirement. Traditionally, the collaboration can be conducted through an outsource contract mechanism where the company may need to guarantee a certain number of logistics demands or usage for a fixed period of time. This may incur a fixed transportation and logistics cost for the company. For a company who has more fluctuated logistics demands, it would be difficult to guarantee a certain number of logistics demands or usage for an outsourcing mechanism. The company may end up paying more than it should. Alternatively, the company may want to explore horizontal collaboration with a more flexible contract mechanism such as the "4th party milk run" (4PMR). The 4PMR model leverages on the last mile excess capacity of one company to fulfill the last mile logistics demands for another company based on a pay-per-use arrangement. Using the 4PMR model, the fixed transportation and logistics cost would be translated into a marginal cost. This paper describes the 4PMR model, including the optimization model and its computation experiment on two last mile logistics scenarios. The first scenario is a hypothetical scenario based on our field study in Jakarta, Indonesia with a small number of deliveries, while the second scenario is an actual scenario with a large number of deliveries based on existing routes of a Logistics Service Provider (LSP) in Surabaya, Indonesia. The experiment results show that 4PMR is able to provide a significant reduction in last mile logistics cost. To complement the experiment results, industry perspectives for implementing the 4PMR model is also reviewed.

1. Introduction

This paper is an extension of the work originally presented in the 2017 6th IEEE International Conference on Advanced Logistics and Transport (ICALT) in [1]. Loss of logistics efficiency becomes the highest across the supply chain at the nearest point to the aggregated or single demand point [2, 3]. This last (or sometimes first) mile logistics face significant fulfillment constraints, making it the most expensive part of the overall supply chain. For a company that moves millions of tons of materials every year, reducing the last mile logistics cost by a few cents can save tremendous amounts of money. As such, the company looks for new strategies to reduce the last mile logistics cost. Several strategies such as a horizontal collaboration with other similar companies or partners in the same level of the supply chain are now attracting the interest of those companies [4, 5, 6].

In supply chain, collaborations between companies can be grouped into two, namely: vertical and horizontal collaboration [7]. The vertical collaboration includes collaboration between companies in different level of supply chain [7], such as Vendor Managed Consignment Inventory (VMCI) model [8] and Collaborative, Planning, Forecasting and Replenishment (CPFR) model [9]. While the horizontal collaboration includes two or more

* Corresponding author: Linda William, School of Informatics and IT, Temasek Polytechnic, 21 Tampines Ave 1, Singapore 529757, Telp: +65-6780-5274, email: lwilliam@tp.edu.sg

unrelated or competing companies at the same level of the supply chain [7]. In horizontal collaboration, a company can work closely with another company that has similar logistic requirements by sharing resources and taking advantage of potential distribution synergies [10]. The benefits of horizontal collaboration are increasing vehicle utilization on the main tracking routes and vehicle efficiency as well as better improving resource management and delivery service level to the customers. Although vertical collaboration has been deeply studied and implemented [11], horizontal collaboration is still in its infancy [12, 13].

As illustrated in Figure 1, horizontal collaboration can be done in several forms. A company may explore the option of engaging one or more companies (e.g. Logistics Service Provider (LSP)) in a form of fixed outsourcing contracts to fulfill its last mile logistics demands. The company may need to guarantee a certain number of logistics demands or usage for a fixed period of time (e.g. monthly, quarterly or yearly) to the outsourcing companies. Using outsourcing, the company is able to save a substantial amount of money because the company does not need to invest in costly logistics-related assets such as transportation vehicles. Generally, this kind of outsourcing would incur a certain amount of fixed transportation and logistics costs.

For a company who has more fluctuated logistics demands, it would be difficult to guarantee a certain number of logistics demands or usage for outsourcing mechanism. The company may need to pay the fixed cost even though it has minimum logistics demands. For this kind of company, a model with a flexible (and often less binding) agreement, such as "uberization" and "4th party milk run" (4PMR) would be beneficial. These models use the pay-per-use mechanism that allows the company to pay only based on its logistics demands. It would reduce the logistics cost by translating the fixed costs to marginal costs. For example, the company does not need to pay fixed cost of 90,000/month for having only 50 last mile logistics demands.

These models use sharing economy concept to reduce the transportation costs. Uberization mainly focuses on conducting the deliveries using crowdsourcing, while the 4PMR model focuses on collaborating with one company (B2B) that has existing delivery routes/networks along the last mile delivery locations.

This paper focuses on analyzing the 4PMR model for reducing last mile logistics cost. It extends the model in [1, 14]. The 4PMR model is developed as an optimization model and solved using two metaheuristics solvers/algorithms. To investigate the cost reduction, we test the model on two scenarios. The first scenario is a hypothetical scenario based on our field study in Jakarta, Indonesia. It consists of 40 random logistics demand samples. While the second scenario is an actual scenario based on 9,622 logistics demands from an LSP in Surabaya, Indonesia. Our experiment results show that the 4PMR model is able to reduce 4.16% and 24.84% of the logistics cost in the first and the second scenario respectively. This provides a promising indication that the 4PMR model would be able to reduce the last mile logistics cost. To complement the experiment results, we also gather industry perspectives for implementing the 4PMR model in its day-to-day logistics operations.

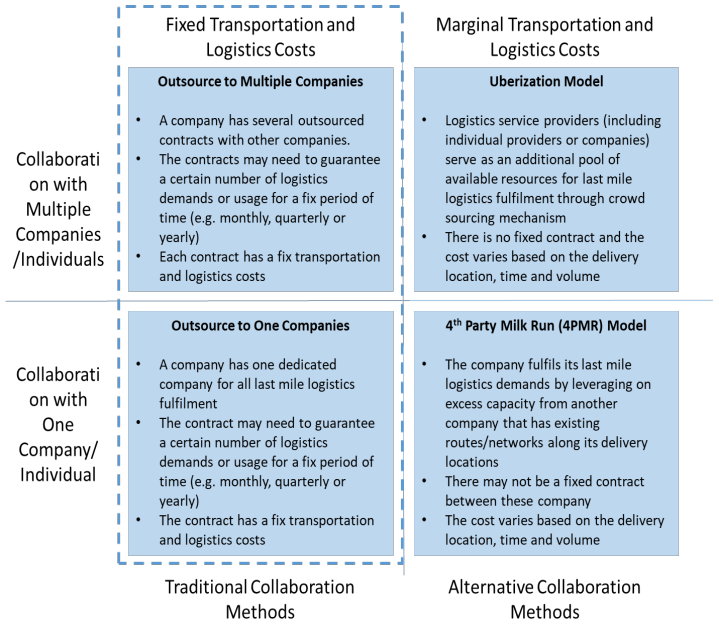


Figure 1. Horizontal Collaboration Strategies

The remainder of the paper is organized as follows. Section 2 presents the overview of 4PMR and its related works. Section 3 reviews the 4PMR optimization model. Section 4 describes the solvers/algorithms to solve the 4PMR and section 5 describes the computational experiments, including the scenarios and experiment results. Section 6 reviews the industry perspective on implementing 4PMR model. Lastly, section 7 presents the conclusions.

2. 4th Party Milk Run (4PMR) Model

2.1. Overview

The 4PMR model leverages on the excess capacities from one company that has existing designated routes along the last mile logistics demand locations of another company. The company with existing designated routes (henceforth as providing company) needs to serve those routes periodically regardless of the delivery volume. These routes may not be fully utilized and may have significant excess capacities. These unused capacities can be tapped and used by another company (henceforth as receiving company) to fulfill its last mile logistics demands.

In 4PMR model, the routing is based on the existing routes or networks from the providing company. It would stop at several locations to fulfill logistics demands from the receiving company.

2.2. Benefits of 4PMR model

The 4PMR model would bring benefits to the collaborating companies. These benefits can be categorized into three groups: costs and productivity, service and market position.

1. Cost and productivity benefits

One potential benefit of using the 4PMR model is to reduce logistics costs [15, 16]. Both collaborating companies, providing and receiving companies, would be able to achieve logistics cost reduction. By joining their delivery requirements, these companies can utilize and optimize their logistics-related assets. This can be

translated into a lower logistics cost and lower capital investment for logistics-related assets as it would need a lesser number of logistics-related assets.

One example of this cost reduction is on empty backhaul utilization. The providing company may have existing routes from area A to area B. If there is an imbalance of deliveries where all the loads come from area A to area B with no loads from area B to area A, the providing company must absorb the cost of the empty vehicle return (empty backhauls). However, if another company has last mile logistics demands in the opposite direction (from area B to area A), the company can make use of the empty backhauls. The providing company would be able to reduce its costs of absorbing empty backhaul while the receiving company would be able to negotiate a lower transportation cost.

2. Service benefits

For the receiving company, the 4PMR model would add the company capabilities to serve low demand areas. It allows the company to facilitate more frequent deliveries for wider geographic areas. The receiving company would be able to expand its area coverage by collaborating with the providing company that operates in different areas with minimum additional setup cost. It provides flexibility in providing more service and reaching more potential customers.

3. Market Position

For the providing company, the 4PMR model would enable the company to have "insinking" opportunities where the company would be able to replicate its 4PMR model to provide services to different companies and proactively select a group of companies with a strong synergy potential [10, 17]. It would create another business opportunity for the providing company. It would further increase vehicle utilization and reduce the last mile logistics cost.

2.3. Related Works

Two areas of works that is closely related to ours are the Vehicle Routing Problem (VRP) [18, 19] and Milk Run Logistics [20, 21]. VRP aims to provide optimal set of routes for a fleet of vehicles to fulfil logistics demands. It has many variants, including VRP with Time Windows [22, 23], Capacitated VRP [24, 25], green VRP [26, 27] and ride-sharing VRP [28, 29]. While Milk Run Logistics is essentially a logistics method that uses routing to consolidate the transportation of goods that has the following characteristics: small lot of transport, less than truckload, geographically sparse demands [20], as illustrated in Figure 2. As shown in Figure 2, milk run model reduces the excess capacity by consolidating the deliveries from different companies (i.e. Company X and Company Y).

3. Optimization Model

One potential concern with the 4PMR model is that the providing company may not have a direct route to fulfill the receiving company' logistics demands (direct from an origin location to a final destination location). It may need to go through several stops/hops/transit locations before it can finally deliver to the final delivery destination. There is a need to match the providing company existing routes with the receiving company logistics demand locations.

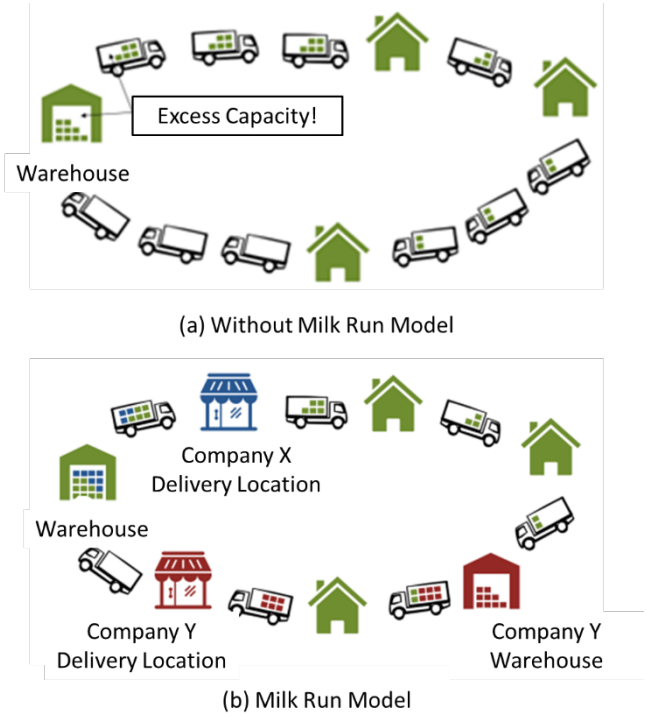


Figure 2. Milk Run Model Illustration

To match the logistics demands with the existing routes, an optimization model for the 4PMR is developed. The 4PMR model is defined as follows. Let $G = (v, e)$ be a directed graph representing the providing company existing routes. It consists of a set of locations, as vertexes or nodes (the term vertex and node are used synonymously throughout the paper), denoted by $V = \{v_1, v_2, \dots, v_n\}$, and a set of routes, as arcs E $\subseteq$ V x V, denoted by an ordered pair of nodes (u, v), where $u \in V$ and $v \in V$. e_{uv} represents a directed route from node u to node v. Each arc e_{uv} has an excess capacity Q_{uv} and time capacity T_{uv}. To simplify the problem, we consider each arc to have a fixed excess capacity Q_{uv}. For each arc e_{uv}, d_{uv}, t_{uv} and c_{uv} represent a non-negative travel distance, a non-negative travel time and a non-negative transportation cost for arc e_{uv}, respectively. A path is defined as a sequence of arc $p_{uv} = \{u = u_1, (u_1, u_2), (u_2, u_3), \dots, (u_{l-1}, u_l), u_l = v\}$.

Let $L = \{1, \dots, n\}$ be a set of logistics demands of the receiving company. Each delivery $l \in L$ has delivery load size q_l, origin node u and destination node v. $[w_u, z_u]$ denotes the time window in origin node u. While $[w_v, z_v]$ denotes the time window in destination node v. Each delivery service at origin node u and destination node v must take place between the time windows. For each logistics demand $l \in L$ from origin node u to destination node v fulfilled by 4PMR model, the delivery cost is defined as $c^l(p)$, where $c^l(p) = \sum_{(u,v)\in p} c^l_{uv}$. If the delivery fulfillment cannot be done using the 4PMR model (due to capacity or time constraints), it will use a direct delivery. The delivery cost for the direct delivery denotes as c^l_d, where in general $c^l(p) \ll c^l_d$.

There are a lot of constraints that can be included in the 4PMR model, such as capacity, transportation cost, overall cost, transportation distance, and time windows. These constraints depend on the nature of the collaboration between providing and

receiving companies and the nature of the products being delivered. In this paper, we only use one constraint, capacity. The capacity for logistics demands fulfilled by the 4PMR should not exceed the fixed weight capacity Q_{uv} for each arc. The model can be expended to include more constraints.

Using these notations, we define the 4PMR problem as:

$$Min\ Z = \sum_{l \epsilon L} \left(\min \left(\min \left(\sum_{(u,v) \epsilon E} c_{uv}^{l} \times x_{uv}^{l} \right), c_{d}^{l} \right) \right) \quad (1)$$

Subject to:

$$x_{uv}^{l} = \begin{cases} 1, if\ delivery\ request\ l\ can\ use\ arc\ (u,v) \\ 0, otherwise \end{cases} \quad (2)$$

$$\sum_{l \epsilon L} q_{l} \times x_{uv}^{l} \leq Q_{uv} \qquad \text{for every } (u,v) \in E \quad (3)$$

Constraint (2) and (3) ensures that the capacity for logistics demands fulfilled by the 4PMR model goes through a certain arc should not exceed the fixed weight capacity for that arc.

4. Solvers and Algorithms

The 4PMR optimization model is implemented in two solvers/algorithms. The first solver is designed for a small dataset with less than 50 logistics demands. It uses an evolutionary function in Microsoft Excel [30]. The function is based on genetic and evolutionary algorithm to seek "good" solutions for a limited dataset. This solver fits the purpose to provide a preliminary result for further study and exploration.

The second solver is used for large problem sizes. It is developed using a meta-heuristic algorithm that combines the Dijkstra algorithm with the Greedy algorithm [31]. The algorithm has two parts as illustrated in Figure 3. First, it uses the Dijkstra algorithm to produce all the shortest paths for each pair of delivery locations in the existing routes. It also calculates the logistics cost for each of the shortest path. Second, it runs the greedy algorithm to assign the logistics demands to the path with minimum cost. It will also re-configure the shortest paths and re-calculate the logistics cost, if necessary.

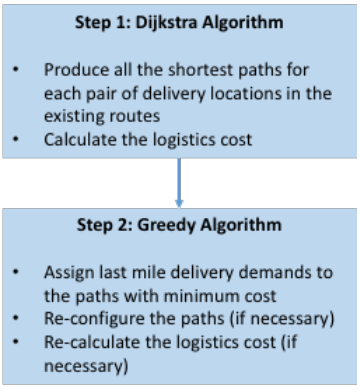


Figure 3. Meta-Heuristics Algorithm for Solving 4PMR

5. Computational Experiments

5.1. Small Problem Size

The 4PMR model was tested on a hypothetical scenario based on our field study in Jakarta, Indonesia in December 2016. The scenario comprises of a providing company with 5 existing routes. The routes connect 12 nodes. In each route, the delivery starts from and returns to the providing company' warehouse. The logistics demands are derived from a receiving company with 40 logistics demands in several locations along the providing company's routes. The receiving company produces goods and uses a hub-and-spoke model to deliver goods to all its customers within the same distribution network. The receiving company has a headquarter/main warehouse for its hub-and-spoke model. The providing company's warehouse and the receiving company's headquarter are located in different locations (represented by different nodes).

Given that the logistics demands originating from the receiving company, all the logistics demands share the same origin point – the receiving company's headquarter. For the hub and spoke distribution network, the utilization of the vehicles is 80% and that they return back to headquarter empty.

We run the first solver for this scenario. The experiment results are shown in Table 1. The results show that the 4PMR model is able to get a 4.16% reduction for the receiving company. For the no collaboration (own vehicle) model, the company will have additional costs to purchase the vehicles also (which is not included in the comparison). The detailed results for this experiment can be found in [1].

Table 1. Experiment Result of the 4PMR Model in Small Problem Size

Solution	Parameters		
	Total number of logistics demands	Number of logistics demands using 4PMR model	Logistics Cost
No collaboration (own vehicle)*	40	0	56.737
The 4PMR model	40	11	54.377

*: only logistics cost, without capital cost to purchase the vehicles

5.2. Experiment Result in Large Problem Size

To evaluate the performance of the 4PMR model in large problem size, we tested the model on scenario based on actual delivery routes from one Logistics Service Provider (LSP) in Surabaya, Indonesia. The LSP has an existing network comprising 27 nodes/locations with existing routes in Greater Surabaya, Indonesia. Each route has a certain excess capacity (around 20% of the overall capacity). The logistics demands were derived from real logistics demands from one FMCG company operated in Surabaya, Indonesia. We used 10 days of logistics demands in July

2015. It comprises of 9,622 logistics demands for different locations in Greater Surabaya.

Using the second solver, we run the 4PMR model for this scenario. In our experiment, we set the processing (i.e. execution) time to 10 second. To compare the 4PMR model with another horizontal collaboration, we also calculated the transportation cost for outsource mechanism. We assumed that the company needs to pay for 1 dollar for each delivery with minimum delivery of 1,000 deliveries/day.

The experiment results are shown in Table 2. The results show that the 4PMR model is able to get a 24.84% reduction from the no collaboration model and 29.54% reduction from the outsource model. For the no collaboration (own vehicle) model, the company will have additional costs to purchase the vehicles also (which is not included in the comparison).

6. Industry Perspective in Implementing 4PMR

While the benefits of the 4PMR model are clear, implementing it can be challenging. The challenges are varied depending on the company and also the industry. Based on our observation and close interaction with one e-commerce company in Singapore, we observe these two main challenges.

Table 2. Experiment Result of the 4PMR Model in Large Problem Size

Solution	Parameters		
	Total number of logistics demands	Number of logistics demands using 4PMR model	Logistics Cost
No collaboration (own vehicle)*	9,622	0	9,622.0
Horizontal collaboration with outsource	9,622	0	10,264
With 4PMR	9,622	2,656	7,231.6

*: only logistics cost, without capital cost to purchase the vehicles

Firstly, stakeholders of the 4PMR model (i.e. providing and receiving company) will require access to as much real-time data as possible. The providing company needs to provide information about its routes and excess capacities so that the receiving company can decide on whether to utilize those excess capacities or not. When excess capacities are eventually bought up, the providing company then has to update its information again to show the latest excess capacities and the routes to fulfill the receiving company's last mile logistics demands. Hence, software that leverages on secure and real-time data is needed. The Internet and the Internet of Things (IoT) can be used to facilitate information sharing (e.g. mobile applications to link providing and receiving companies).

For the providing company with existing the Transportation Management Software (TMS), another challenge will be to integrate the 4PMR solver/algorithm with its TMS. As TMS typically captures information about the vehicle, the cargo, the route and schedule, the logistics demands using the 4PMR model should be updated to the company's TMS.

Secondly, potential stakeholders of the 4PMR model (i.e. potential providing and receiving company) need to be convinced about the benefits of the 4PMR model. The potential providing company needs to evaluate whether the effort both in terms of IT infrastructure changes and manpower training to improve excess capacities worthwhile and whether there is an obvious return on investment. The potential receiving company needs to analyze its last mile logistics requirements and decides if the 4PMR model would be able to fulfill its business needs. For example, if a company requires trucking of large volumes (typically half to full truckload), it needs to understand that it risks its deliveries for not being fulfilled by the providing company because there are simply not enough excess capacities. Alternatively, if a company requires frequent trucking of easily forecasted volumes, it may need to evaluate the 4PMR model cost versus the traditional outsource model.

7. Summary and Key Take Away

This paper introduces the 4th Party Milk Run (4PMR) model for collaboration between similar companies in the same level of the supply chain. The 4PMR model leverages the excess capacities from one company (i.e. providing company) that has existing designated routes along the pick-up and delivery locations of another company (i.e. receiving company). The providing company needs to serve those routes periodically regardless of the delivery volume, and often these deliveries are not at full capacity. These unused capacities can be tapped and used by the receiving company to fulfill its last mile logistics needs. It would potentially bring benefits to the companies involved in terms of costs and productivity, service and market position.

Two solvers are designed to implement the 4PMR model. The solvers aim to match the providing company's existing routes with the receiving company's logistics demands. The first solver is used for small problem size scenario, while the second solver is used for large problem size scenario. We run these solvers for two scenarios – a small problem size scenario and a large problem size scenario. From the experimental result, it is also revealed that 4PMR is able to reduce logistics cost for up to 4.16% (in the small problem size scenario) and 24.84% (in the large problem size scenario).

Conflict of Interest

The authors declare no conflict of interest related to this paper.

References

[1] C. Lee, Lindawati and R. de Souza, "Utilizing Excess Capacity in Last Mile Using 4th Party Milk Run," in *6th IEEE International Conference on Advanced Logistics and Transport*, Bali, Indonesia, 2017.

[2] R. de Souza, M. Goh, H. Lau, W. Ng and P. Tan, "Collaborative urban logistics–synchronizing the last mile a Singapore research perspective," *Procedia-Social and Behavioral Sciences,* vol. 125, pp. 422-431, 2014.

[3] J. Lin, Q. Chen and K. Kawamura, "Sustainability SI: logistics cost and environmental impact analyses of urban delivery consolidation strategies," *Networks and Spatial Economics,* vol. 16, no. 1, pp. 227-253, 2016.

[4] I. Badraoui, J. G. Van der Vorst and Y. Boulaksil, "Horizontal logistics collaboration: an exploratory study in Morocco's agri-food supply chains," *International Journal of Logistics research and applications,* pp. 1-18, 2019.

[5] Y. Sheffi, M. J. Saenz, L. Rivera and D. Gligor, "ew forms of partnership: the role of logistics clusters in facilitating horizontal collaboration mechanisms," *European Planning Studies,* pp. 1-26, 2019.

[6] N. Martin, L. Verdonck, A. Caris and B. Depaire, "Horizontal collaboration in logistics: decision framework and typology," *Operations Management Research,* vol. 11, no. 1-2, pp. 1-19, 2018.

[7] L. Chen, X. Zhao, O. Tang, L. Price, S. Zhang and W. Zhu, "Supply chain collaboration for sustainability: A literature review and future research agenda," *nternational Journal of Production Economics,* vol. 194, pp. 73-87, 2017.

[8] M. Hariga, S. Babekian and Z. Bahroun, "Operational and environmental decisions for a two-stage supply chain under vendor managed consignment inventory partnership," *International Journal of Production Research,* vol. 57, no. 11, pp. 3642-3662, 2019.

[9] C. A. Hill, G. P. Zhang and K. E. Miller, "Collaborative planning, forecasting, and replenishment & firm performance: An empirical evaluation," *International journal of production economics,* vol. 196, pp. 12-23, 2018.

[10] F. Cruijssen, P. Borm, H. Fleuren and H. Hamers, "Supplier-initiated outsourcing: A methodology to exploit synergy in transportation," *European Journal of Operational Research,* vol. 207, no. 2, pp. 763-774, 2010.

[11] S.-L. D., P. Kaminsky and E. Simchi-Levi, Designing and Managing the Supply Chain: Concepts. Strategies, and Cases, McGraw-Hill, 2000.

[12] F. Cruijssen, W. Dullaert and H. Fleuren, "Horizontal cooperation in transport and logistics: a literature review," *Transportation journal,* pp. 22-39, 2007.

[13] R. Leitner, F. Meizer, M. Prochazka and W. Sihn, "Structural concepts for horizontal cooperation to increase efficiency in logistics," *CIRP Journal of Manufacturing Science and Technology,* vol. 4, no. 3, pp. 332-337, 2011.

[14] The Logistics Institute - Asia Pacific, "E-Commerce Trends and Challenges: A Logistics and Supply Chain Perspective," The Logistics Institute - Asia Pacific White Papers Series, Singapore, 2016.

[15] L. Wang, F. Kong, J. Cao and Y. Wang, "Research on the application and optimization countermeasures of auto parts milk-run logistics mode," in *IEEE 18th International Conference on Industrial Engineering and Engineering Management*, 2011.

[16] M. Jafari-Eskandari, S. Sadjadi, M. Jabalameli and A. Bozorgi-Amiri, "A robust optimization approach for the milk run problem (An auto industry Supply Chain Case Study)," in *IEEE International Conference on Computers & Industrial Engineering*, 2009.

[17] F. Cruijssen, B. Peter, H. Fleuren and H. Hamers, "Insinking: A Methodology to Exploit Synergy in Transportation," CentER Discussion Paper, 2005. [Online]. Available: https://ssrn.com/abstract=870403. [Accessed 26 03 2018].

[18] P. Toth and D. Vigo, The vehicle routing problem, Philadelphia: Society for Industrial and Applied Mathematics, 2002.

[19] B. L. Golden, S. Raghavan and E. A. Wasil, he vehicle routing problem: latest advances and new challenges, New York: Springer Science & Business Media, 2008.

[20] G. Brar and G. Saini, "Milk run logistics: literature review and directions," *Proceedings of the world congress on engineering,* vol. 1, pp. 6-8, 2011.

[21] G. S. Brar and G. Saini, "Milk run logistics: literature review and directions," in *In Proceedings of the world congress on engineering*, 2011.

[22] F. Errico, G. Desaulniers, M. Gendreau, W. Rei and L. M. Rousseau, "The vehicle routing problem with hard time windows and stochastic service times," *EURO Journal on Transportation and Logistics,* vol. 7, no. 3, pp. 223-251, 2018.

[23] Y. Niu, Z. Yang, P. Chen and J. Xiao, "Optimizing the green open vehicle routing problem with time windows by minimizing comprehensive routing cost," *Journal of cleaner production,* vol. 171, pp. 962-971, 2018.

[24] L. Wei, Z. Zhang, D. Zhang and S. C. Leung, "A simulated annealing algorithm for the capacitated vehicle routing problem with two-dimensional loading constraints," *European Journal of Operational Research,* vol. 265, no. 3, pp. 843-859, 2018.

[25] A. N. Letchford and J. J. Salazar-González, "The capacitated vehicle routing problem: stronger bounds in pseudo-polynomial time," *European Journal of Operational Research,* vol. 272, no. 1, pp. 24-31, 2019.

[26] G. Poonthalir and R. Nadarajan, "A fuel efficient green vehicle routing problem with varying speed constraint (F-GVRP)," *Expert Systems with Applications,* vol. 100, pp. 131-144, 2018.

[27] A. Shuib and N. Muhamad, "Mixed Integer Multi-Objective Goal Programming Model For Green Capacitated Vehicle Routing Problem," *Advances in Transportation and Logistics Research,* vol. 1, no. 1, pp. 356-368, 2018.

[28] L. Agussurja, S. F. Cheng and H. C. Lau, "A State Aggregation Approach for Stochastic Multiperiod Last-Mile Ride-Sharing Problems," *Transportation Science,* vol. 53, no. 1, pp. 148-166, 2019.

[29] V. F. Yu, S. S. Purwanti, A. P. Redi, C. C. Lu, S. Suprayogi and P. Jewpanya, "Simulated annealing heuristic for the general share-a-ride problem," *Engineering Optimization,* vol. 50, no. 7, pp. 1178-1197, 2018.

[30] FrontlineSolvers, "Excel Solver – Non-smooth Optimization," [Online]. Available: http://www.solver.com/excel-solver-non-smooth-optimization. [Accessed 01 05 2017].

[31] A. Benoit, R. Yves and F. Vivien, A guide to algorithm design: paradigms, methods, and complexity analysis, Boca Raton: CRC Press, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 468-477 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Estimation of Target Maneuvers from Tracked Behavior Using Fuzzy Evidence Accrual

Stephen Craig Stubberud[1], Kathleen Ann Kramer[2,*], Allen Roger Stubberud[3]

[1]*Advanced Programs, Oakridge Technology, San Diego, 92121, USA*

[2]*Department of Electrical Engineering, University of San Diego, San Diego, 92110, USA*

[3]*Department of Electrical and Computer Engineering, University of California, Irvine, 92697, USA*

ARTICLEINFO

Article history:
Received: 31 May, 2019
Accepted: 06 August, 2019
Online: 21 August, 2019

Keywords:
Evidence accrual
Fuzzy logic
Kalman gain monitoring
Maneuver detection
Pattern detection

ABSTRACT

While the Kalman filter, including its many variants, has been the staple of the tracking community, it also has been shown to have drawbacks, particularly when tracking through a maneuver. The most common issue is a lag in the position of the target track compared to the true target position as the target performs its maneuver. Another more problematic issue can occur where the filter covariance collapses, requiring the filter to be reinitialized. Techniques exist to compensate for maneuvers, but generating their response relies on detection of error between the estimated trajectory and the measured target position. In this effort, a maneuver detection routine is developed that can be used in conjunction with more standard maneuver compensation approaches. This routine is able to validate the existence of a maneuver more quickly than use of the inherent detection relied upon in the other methods. Maneuver detection is performed by an evidence accrual system that uses a fuzzy Kalman filter to incorporate new information and provide a level of evidence that maneuver is occurring. The input data uses behavior characteristics of the Kalman gain vector from the tracking algorithm.

1. Introduction

Feature objection extraction [1, 2] is an evidence accrual technique that can be applied to classification problems. An undirected tree [3] with various levels of information is used to connect evidence nodes. Each node has a level of evidence and an associated level of uncertainty similar to a state in a Kalman filter. Within the tree, nodes represent levels of evidence of what might be elementary information or a complex combination of information. This paper extends work originally presented in the 2018 Conference on Innovations in Intelligent Systems and Applications [1]. In theory, every node can be measured directly or indirectly. Unlike typical evidence accrual methods [4-6], the states are not probabilistic. When evidence points to more than one solution, multiple competing solutions can each have levels of evidence. Evidence can affect nodes at the same level, with the same evidence potentially increasing one or more than one node in levels of evidence, while others may have their levels decreased, and still other nodes at the same level may not be affected at all.

Feature objection extraction (FOX) propagates information within the tree using a variety of function or function-approximation relationships. The measured information is injected using a fuzzy Kalman filter [1, 7]. As will be described in detail, the FOX evidence accrual system decomposes high-level concepts into simpler concepts until it reaches root nodes which are comprised elementary information. Figure 1 describes a basic FOX tree with different connections and a variety of level of elementary information.

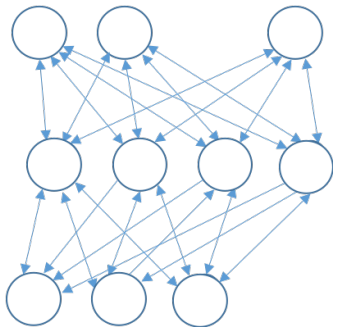


Figure 1: Feature-object extraction structure with various interconnections in a multi-level tree

* Kathleen Ann Kramer, Email: kramer@sandiego.edu

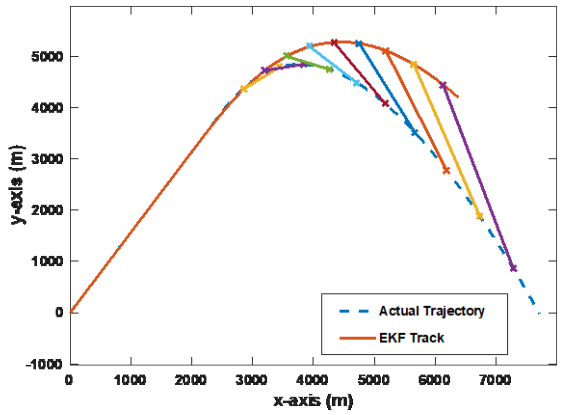


Figure 2: Ballistic target and standard tracker result in lagging of track through maneuver

FOX is incorporated into a tracking system to detect a target's maneuver and can use the computed values of the track to produce a measure of evidence that maneuver is occurring. Maneuver detection is important because, if the target dynamics are not modeled correctly for a maneuver, the tracking system can have an offset or a time-lag in its estimation of the target position and velocity. In Figure 2, an example of a maneuvering target trajectory is shown, alongside the target track from a Kalman-filter based tracking system, a more typical approach. The figure shows a track lagging the true target trajectory as the target proceeds through the maneuver. This comparative result illustrates the deleterious effects that can result when a maneuver occurs. Tracking problems also become more prevalent when the measurement lies in the unobservable space of the track kinematics, as seen in [8]. For example, if an angle-only tracker is used, the filter can become numerically unsound and require a reinitialization. In the tracking problem, losing a target or creating large target kinematic-errors can be life-or-death issue. If a maneuver can be detected, the tracking algorithm can be modified to improve performance and avoid catastrophic failures.

The Kalman filter, and its numerous variants, such as the extended Kalman filter (EKF) [9] and the unscented Kalman filter [10], have long provided the predominant core approaches for kinematic target-tracking systems [11]. To reduce the deleterious effects that maneuvers have on a tracking system, compensation approaches have been developed. The most widely used technique is the interacting multiple model (IMM) and its variants [12]. The IMM incorporates various maneuver models. The IMM system creates weighted combinations of models by comparing the residuals of each measurement to each model's prediction. The residual scores are used to interpolate between the models and create a weight for the state of each motion model to create a more accurate state estimate of the target position and velocity. Other techniques include adaptive Kalman filters such as a neural extended Kalman Filter (NEKF) [13] that also employ the Kalman filter residuals to adapt their maneuver parameters to more closely model that of the actual target dynamics.

All of these approaches rely on the residual measures to detect and make the adjustments. While the residual is an effective measure, variations in uncertainty can mask the maneuver until the residual value becomes significantly large. Fortunately, there exist other metrics as part of the Kalman filter that can detect a maneuver. Besides the residual, the behaviors of the Kalman gain can also indicate that the target is in a maneuver.

FOX can be employed to combine the disparate information and the associated uncertainty, providing for an effort at detection based upon more than one type of measure. In the problem demonstrated in this work, there is an injection of Kalman gain data provided by the aforementioned fuzzy Kalman filter, augmenting the residual measures, and the evidence is propagated using fuzzy, linear, and nonlinear relationships.

Four further sections provide an overview of the development of this approach and analyze its capability. Section 2 overviews the EKF, the most prominent variant of the Kalman filter used in tracking that provides the Kalman gains. Section 3 describes the Kalman gain behaviors that can provide information measures to indicate a target maneuver. Section 4 describes the FOX evidence accrual maneuver detection technique. In Section 5, examples of maneuvering targets are described and are used to exemplify the capabilities of FOX as a maneuver detection technique.

2. Target Tracking with the Extended Kalman Filter

For kinematic target tracking, the discrete-time dynamics of the are defined in (1).

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}\boldsymbol{x}_k \tag{1}$$

The subscript k indicates discrete time, and $\mathbf{x}$ is the standard state-vector representation of the target behavior represents position and velocity. In (2), the state vector represents three dimensions

$$\boldsymbol{x}^T = [\boldsymbol{x} \quad \dot{\boldsymbol{x}} \quad \boldsymbol{y} \quad \dot{\boldsymbol{y}} \quad \boldsymbol{z} \quad \dot{\boldsymbol{z}}]. \tag{2}$$

The target dynamics $\mathbf{F}$ are often described with the straight-line motion model

$$\mathbf{F} = \begin{bmatrix} \boldsymbol{F}_{2x2} & \mathbf{0}_{2x2} & \mathbf{0}_{2x2} \\ \mathbf{0}_{2x2} & \boldsymbol{F}_{2x2} & \mathbf{0}_{2x2} \\ \mathbf{0}_{2x2} & \mathbf{0}_{2x2} & \boldsymbol{F}_{2x2} \end{bmatrix}, \tag{3a}$$

where

$$\mathbf{F}_{2x2} = \begin{bmatrix} \mathbf{1} & \boldsymbol{dt} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{3b}$$

The term dt is the time difference between the last sensor report on the target and the latest sensor report.

While the dynamics are linear, the measurements provided by sensors are often nonlinear. Using active sensors, such as radar, a complete measurement space for the three-dimensional target-track would be a report ***h(x)*** of range, bearing, and elevation, shown in (4)

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} range \\ bearing \\ elevation \end{bmatrix} = \begin{bmatrix} \rho \\ \beta \\ \varepsilon \end{bmatrix}, \tag{4}$$

where

$$\rho = \sqrt{\left(x_{tgt} - x_{plt}\right)^2 + \left(y_{tgt} - y_{plt}\right)^2 + \left(z_{tgt} - z_{plt}\right)^2}$$

$$\beta = \arctan\left(\frac{x_{tgt} - x_{plt}}{y_{tgt} - y_{plt}}\right)$$

$$\varepsilon = \arctan\left(\frac{z_{tgt}-z_{plt}}{\sqrt{(x_{tgt}-x_{plt})^2+(y_{tgt}-y_{plt})^2}}\right),$$

and the subscript *tgt* denotes target component and the subscript *plt* denotes the platform.

Since these measurements are generated relative to the sensor platform, when they are reported to the tracking algorithm, they may be transformed to a universal coordinate system, such as earth-centered-earth-fixed (ECEF), be relative to the tracking system, or be relative to a localized flat earth [14]. To maintain the accuracy of the measurement, though, it is kept in a nonlinear-coordinate frame rather than being mapped into the track-coordinate frame which would linearize the tracking system. This is one reason the extended Kalman filter (EKF) is the most prevalent tracking algorithm [11].

These measurements are the driving inputs for the tracking algorithm. The EKF uses its estimate of the measurement and the residual between the estimated and reported measurements to correct its state-estimate of the target kinematics. The process of EKF to accomplish this is defined in (5a-e)

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{5a}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\left(\mathbf{z}_k - \mathbf{h}\left(\mathbf{x}_{k|k-1}\right)\right) \tag{5b}$$

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k|k-1} \tag{5c}$$

$$\mathbf{x}_{k+1|k} = \mathbf{f}\left(\mathbf{x}_{k|k}\right) \tag{5d}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k + \mathbf{Q}_k, \tag{5e}$$

where **H**, based on (4), provides the Jacobian of the output-coupling function:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\delta\rho}{\delta x} & 0 & \frac{\delta\rho}{\delta y} & 0 & \frac{\delta\rho}{\delta z} & 0 \\ \frac{\delta\beta}{\delta x} & 0 & \frac{\delta\beta}{\delta y} & 0 & 0 & 0 \\ \frac{\delta\varepsilon}{\delta x} & 0 & \frac{\delta\varepsilon}{\delta y} & 0 & \frac{\delta\varepsilon}{\delta z} & 0 \end{bmatrix}. \tag{6}$$

The function **f** is the modeled target dynamics, and the matrix **F** in (5d-e) is the associated Jacobian. As stated previously, the target dynamics are usually defined as (3a). The subscript indicates discrete time, with $k|k$ is the estimate at the time k given all the information up to that time and $k+1|k$ is the estimate for time $k+1$, given all the information up through time k.

The process noise, **Q**, indicates the accuracy of the system dynamics, and is usually modeled as integrated white noise [15]:

$$\boldsymbol{Q} = q^2\begin{bmatrix} \boldsymbol{Q}_{2x2} & \boldsymbol{0}_{2x2} & \boldsymbol{0}_{2x2} \\ \boldsymbol{0}_{2x2} & \boldsymbol{Q}_{2x2} & \boldsymbol{0}_{2x2} \\ \boldsymbol{0}_{2x2} & \boldsymbol{0}_{2x2} & \boldsymbol{Q}_{2x2} \end{bmatrix} \tag{7a}$$

$$\boldsymbol{Q}_{2x2} = \begin{bmatrix} \frac{dt^3}{3} & \frac{dt^2}{2} \\ \frac{dt^2}{2} & dt \end{bmatrix}. \tag{7b}$$

As the process noise is increased, the dynamic model, **F**, in (5d-e) is weighted less. The measurement becomes more dominant in the processing, allowing more of the measurement noise to be passed thought the filter to the track solution. As the process noise is decreased, in contrast, the reaction of the tracking algorithm becomes less responsive to the measurement and more smoothing occurs.

While not explicit in the (5a) and often overlooked, the EKF's Kalman gain is affected by the process noise, the error covariance **P**, and the state estimate **x**. The state estimate is injected into the Kalman gain though the output-coupling Jacobian **H**. The gain is also affected by the measurement noise **R**. The measurement noise relates to the quality of the sensor. A standard radar will have its angle accuracies based on its beamwidth, for example while the range accuracy will be based on the resolution of the pulse signal. The accuracy indicates that the target can be anywhere within the beam and so-called range bin, as with airborne radar.

Since a Kalman filter is used, the accuracy is represented as a Gaussian distribution. (If the accuracy is not Gaussian, techniques such as Gaussian sums could be used to represent the measurement accuracy [16], but that is beyond the needs of these developments.) In Kalman filtering, the ratio of **Q** to **R** represents the relative belief of the target-motion model versus that of the sensor reports. If **Q** is smaller relative to **R**, the measurements will have a reduced of effect on the target track and the results from measurements over time will be smoothed, with the effects of noise less pronounced. If this ratio is reversed, the measures will dominate the track behavior, with the effects of individual measurements more pronounced in the target track. It is of note that, since the EKF is a measurement driven method, measurements will always ultimately have some effect on the target track. When the measurements noise is smaller, the measurement effects will have a faster impact on the track as sensors are viewed as more accurate.

3. Kalman Gain Monitoring

The extended Kalman filter (EKF) behaves differently than the Kalman filter for linear time-invariant systems. While the Kalman filter for such systems is predictable in nature, the EKF will vary over time. The variations in the error covariance **P** are affected by the nonlinear behavior of the system and the local estimate in both the update equation of (5c) and the prediction equation of (5e).

The behavior of the Kalman gain within the EKF, the vector **K**, often precedes the noticeable changes of the error covariance. This arises as the Kalman gain is affected both by the nonlinearities directly in (5a) and by the variations in the error covariance matrix. As the gains change over time, the filter can inherently detect changes in the actual system that vary from the estimation model. In [1], for a target tracking application of the extended Kalman filter, the Kalman gains demonstrated that over time there existed features in their behaviors that coincided with target maneuvers. The analysis of [1] looked at the individual gains, tracking both position and velocity of the targets in two-dimensional space using a range and bearing measurement. The resulting behaviors of the eight element Kalman-gain vector were used as features to identify when various maneuvers, including turns and simple linear accelerations, occurred. Another example is provided here.

The details of the scenario are shown in Figure 3. The target simulates the behavior of a submarine as it is being tracked by another vessel. The submarine maintains a straight-line trajectory with a constant speed. The trailing vessel also maintains a constant

speed and course. After some time, the target maneuvers. This represents a concept in a submarine-tracking problem where, being tracked, the target maneuvers to clear baffles to "see" the trailing ship and possibly cause a break-in-track of the enemy's tracking system. The platform ship then maneuvers, followed quickly by the submarine maneuvering again. These successive maneuvers can result in the filter becoming numerically unsound.

Figure 4 shows the behavior of the four Kalman gains related to the range over this whole scenario and is marked where the target and the sensor platform perform maneuvers. The Kalman gains show different behaviors, including a sharp transient, a significant change in the slope of the gain, and a zero-crossing.

The other four gains, which relate to the bearing measurements, are shown in Figure 5, and are larger in magnitude, but similar in relative behavior. Other test cases [1, 17], have also shown the gains to high frequency behaviors during maneuvers.

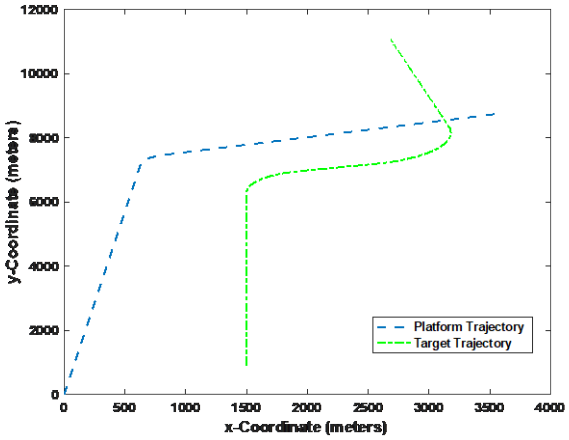


Figure 3: A manuevering target being tracked by a maneuvering sensor platform

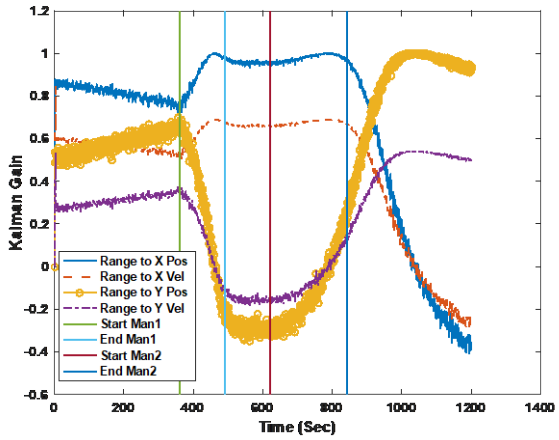


Figure 4: The behavior of the range-related Kalman gains over the course of the scenario with the target maneuvers marked

From these various test cases, the features of interest to detect a target maneuver have been determined to include monitoring the frequency behavior, the zero-crossings, the transient behavior, the slope performance, and the ownship (platform) maneuvers. It has also been determined that only a subset of the Kalman gains are necessary to observe. These include the gains related to the position states and range measurements K_{11} and K_{31} and all of the gains related to the bearing measurements K_{i2}, where those required vary by feature

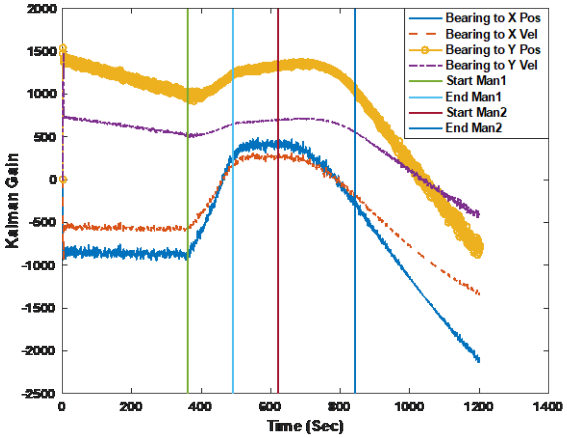


Figure 5: The behavior of the bearing-related Kalman gains over the course of the scenario with the target maneuvers marked

4. Evidence Accrual Using Feature Object Extraction

Many evidence accrual methods have been developed to combine information which, in this case, provide a level of evidence that a maneuver is occurring or has ceased to be. The primary evidence accrual techniques are the Dempster-Shafer method [5] and the Bayesian taxonomy approach [4], also referred to as the Pearl Tree. While these techniques have been utilized for decades, they have drawbacks, including that both represent the information as probabilities that update following Bayes rule. Also, uncertainty in the information that is the basis of the probabilities is not modeled easily, even with Dempster-Shafer.

FOX provides for distinct levels of evidence, such as of maneuver and for a non-maneuver separately. Since the measures of evidence are generated independently, the scores, unlike probabilities, are not related and an increasing evidence for one decision (i.e., a maneuver) need not decrease the other decision (i.e., no maneuver). Evidence can prove one or multiple decisions thus changing all or some of the decision.

FOX is also designed to provide a quality score with each level of evidence. The quality score is similar to a Kalman filter error covariance in that a high score indicates uncertainty while a low score indicates a high degree of certainty about the evidence score. The corroboration of evidence gathered from various behaviors can improve the certainty that an event occurred. Conversely, with multiple sources of evidence, maneuver events can be detected with one gain behavior being triggered while the others have not met the threshold.

For the maneuver detection problem, each of the elements of the Kalman gains' behaviors from the target tracking system can be used individually to indicate when a target maneuver is being initiated.

The FOX technique is designed to decompose a complex classification problem into a series of smaller and simpler problems. For the maneuver detection problem, a unidirectional tree is employed, as shown in Figure 6, illustrating use of FOX to accrue evidence based upon multiple behaviors of the Kalman

For maneuver detection, the top level is *Detected Maneuver.* The degree of evidence for this evidence node is limited to a range between 0 and 1. A corresponding tree for determining a level evidence for no maneuver occurring would be similar in concept with different linkages. The next level of this tree is comprised the individual-event detections of the tracking-system's Kalman gain vector: *Zero Crossing*, *Gain Frequency,* and *Slope Variation.* These three components are combined to create the *Detected Maneuver* scores. As the evidence levels in these states increase

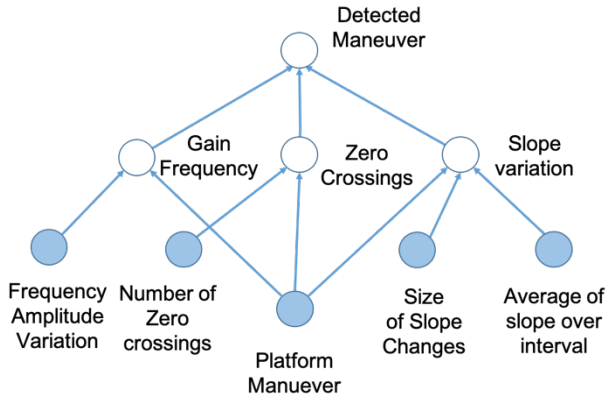


Figure 6: The proposed evidence accrual tree decomposition for Kalman gain maneuver detection approach of FOX

from 0 to 1, the overall detection-levels vary based not only on the lower-level scores but the quality level as well. Poor quality scores are weighted less than higher quality scores.

6, the shaded states indicate elementary-evidence input nodes. These represent the elemental measurements such as the measured frequencies or number of zero crossings. The unshaded nodes are referred to as states of interest. These are states of combined information.

approach is that, unlike the states of Markov chain [6], these states need not be disjoint nor do the states need be a complete representation of all the states of the system.

While the Kalman gain elements of the target tracking system provide evidence to the FOX evidence accrual system which employs a Kalman filter itself. To clarify, the Kalman filter of the FOX system in this case, actually a fuzzy Kalman filter, is different than that of the tracking system that is a source of evidence.

The level of evidence for a state of interest can be generated in two ways. The first is through direct observation, which implies a direct measurement is available, as shown with the shaded nodes. The evidence is then processed through a substate, depicted with a clear node. The tree of Figure 6 shows some substates have multiple injection nodes flowing into them. The FOX Kalman filter that injects the evidence into the tree nodes is a fuzzy Kalman filter (FKF). The FKF takes the multiple inputs and creates a single-output fuzzy measure. Substates that do not have direct evidence injection are generated using systems theory. The substates connect to state of interest through links that represent the functional relationship.

The FKF of the FOX system is based upon a development of Watkins [7] and modified in [2]. This approach was selected to allow a wide variety of measurement types and uncertainty models to be used with ease. The FKF implementation is a straightforward variant of the standard Kalman filter with a modification in the update equations. The FKF is defined in (8) with five equations, (8a) to (8e). Comparing the sets of equations, (5) and (8), two of these, the Kalman gain equation of (5a) and (8a), and the state update equation of (5b) and (8b), are the ones where counterparts differ. The fuzzy measures are incorporated by using the first moment, indicated as mom_1, of the consequent fuzzy membership function, referred to as m_{adj}, or the membership adjunct.

$$\mathbf{K} = \mathbf{P}_{k|k-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + mom_1(m_{adj}(\mathbf{R})))^{-1} \tag{8a}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}(mom_1(m_{adj}) - \mathbf{H}\mathbf{x}_{k|k-1}) \tag{8b}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_{k|k-1} \tag{8c}$$

$$\mathbf{x}_{k+1|k} = \mathbf{F}\mathbf{x}_{k|k}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^T + \mathbf{Q}_k \tag{8e}$$

The use of fuzzy logic provides for simplified linguistic based conversion from the measurement coordinate systems to the evidence space, which has been defined as a value between 0 and 1. If the measurement and uncertainty values are crisp, the FKF devolves into the standard Kalman filter.

The nature of the measurement data determines whether the fuzzy measure or a crisp measure is used. A true measure such as the number of zero crossings or a known even as the knowledge of a platform maneuver would be a crisp value that would map into crisp groupings. The mapping of data into groupings, such as a determination based upon the frequency behavior or degree of slope change, would be fuzzy.

Once elementary evidence is injected into the tree, it can propagate another FKF process or through the use system theory process as follows. A state of interest and its direct-substates are represented in vector form as

$$\mathbf{x}_S = [x_{s1} \quad x_1 \quad x_2 \quad \cdots \quad x_n \quad x_n] \tag{9}$$

The state of interest or of the node value is the first element of the state vector.

Using first-order observer decomposition [18], the evidence dynamics are given as

$$\mathbf{x}_{k+1} = \begin{bmatrix} f(xa, x_{s1}, \ldots, x_{sn}) \\ \begin{bmatrix} 0 & f_1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & f_n \end{bmatrix} \mathbf{x}_k \end{bmatrix}. \tag{10}$$

The state of interest is comprised of its substates and the previous value of the state of interest. The observer concept provides a forgetting factor. This forgetting factor and the observer updating equation are used to incorporate the quality factor of each sub-state. In (10), the uncertainty can be incorporated as a simple scaling in the state values or in the state-of-interest function, $f(\cdot)$.

The associated error covariance or uncertainty equation is propagated by (11).

$$P_k = f_a P_{a,k-1} f_a^T + f_1 P_{s1,k-1} f_1^T + \cdots + f_n P_{sn,k-1} f_n^T + q_x. \tag{11}$$

This decomposition of the problem into the individual substates reduces the complex model of the interactions of information into simplified operations. The decomposition also simplifies the incorporation of the uncertainty into the state vector.

4.1. Maneuver Detection Fuzzy Membership Functions

To generate the injection evidence, the following antecedent and consequent functions along with their associated inference engines were considered:

4.1.1. Frequency Amplitude Variation

When a maneuver occurs, some Kalman gain elements experience high frequency changes in amplitude. The high frequency behavior indicates that a maneuver could be occurring. As seen in [17], the high frequencies indicate potential issues with the Kalman filter that are a result of sharp maneuvers. The frequencies are mapped into a score using two input antecedent functions: *Max_frequency_value* and *Ratio_of_high_frequency_to_low_frequency_power.*

Figure 7 provides a four-element antecedent function that breaks the fast Fourier transform (FFT) spectrum of a time slice of the Kalman gain maps into the four trapezoidal-based functions. Figure 8 shows the five-element triangular antecedent membership functions that represents the ratio of the power of the highest-tenth of the spectrum to the lowest tenth of the spectrum. Table 1 provides the inference engine of the two sets of membership functions. These map to the consequence functions shown in Figure 9. These are defuzzified using the FKF.

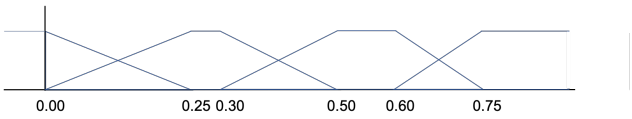


Figure 7: Kalman Gain Frequency amplitude variation FFT time slice antecedent function

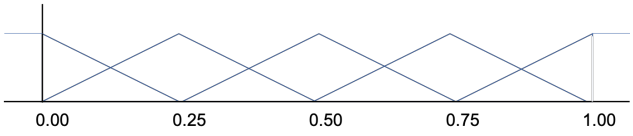


Figure 8: Kalman Gain Frequency amplitude power ratio membership function

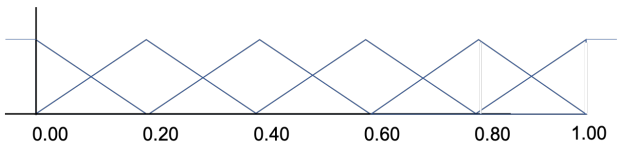


Figure 9: Kalman Gain Frequency Consquence Function

Table 1: Kalman Gain Frequency Inference Engine

	Frequency Maximum Membership Function			
Frequency Power Ratio Membership Function	1	2	3	4
1	1	2	3	4
2	1	2	4	4
3	2	3	4	5
4	2	4	5	5
5	3	4	5	5

4.1.2. Number of Zero Crossings

The number of zero crossings is a crisp value. The number of zero crossings num_{ZC} is mapped into a score ZC_{score}:

$$ZC_{score} = \begin{cases} 0 & num_{ZC} = 0 \\ 0.2 & num_{ZC} = 1 \; or \; 2 \\ 0.6 & num_{ZC} = 3, 4, or \; 5 \\ 1.0 & num_{ZC} > 5 \end{cases} \tag{12}$$

4.1.3. Slope Changes

The size of the slope changes in a given time interval is measured. The value is an absolute value of the difference between the initial slope and the ending slope compared to the maximum and minimum values. This crisp value is mapped similarly to the number of zero crossings but, are mapped evenly across four regions are mapped into the scores:

$$SC_{score} = \begin{cases} 0 & \frac{slope_{max}}{slope_{min}} \leq 0.5 \\ 0.33 & 0.5 < \frac{slope_{max}}{slope_{min}} \leq 2.5 \\ 0.67 & 2.5 < \frac{slope_{max}}{slope_{min}} \leq 4.0 \\ 1.00 & 4.0 < \frac{slope_{max}}{slope_{min}} \end{cases} \tag{13}$$

Unlike the other injection evidence mentioned here, the uncertainty is also fuzzy. The size in the slope changes creates the next score. Figures 10 depicts a five-element trapezoidal function that map the change in slopes. Table 2 provides the inference engine that maps into the consequence function represented by five triangular functions as seen in Figure 11. Again, the defuzzification is performed by the FKF.

Table 2: Kalman Gain Slope Inference Engine

	SC_{score}			
Slope Average Membership Function	0.0	0.33	0.67	1.0
1	1	1	2	3
2	1	2	3	3
3	3	3	4	4
4	5	5	5	5
5	5	5	5	5

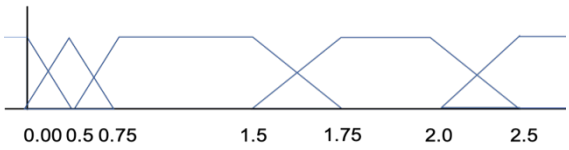


Figure 10: Changes in Kalman Gain Slope Membership Function

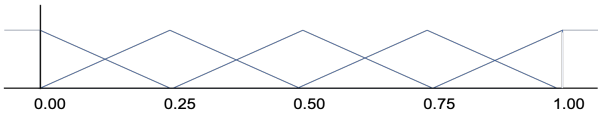


Figure 11: Changes in Kalman Gain Slope Consquence Function

Now that the direct evidence has been defined, the higher level nodes, depicted in Figure 4, can also be defined. The function relating the three sets of substate evidence to the *Detected Maneuver* state is defined with the limited linear combination in (14).

$$\boldsymbol{x_{MD}}(k) = \min\big(1, \mu(k)\big),$$

$$\mu(k) = [0.7 \quad 0.47 \quad 0.7]\, \boldsymbol{x_{MD}}(k),$$

with

$$\boldsymbol{x_{MD}} = \begin{bmatrix} x_{MDfreq} \\ x_{MDZC} \\ x_{MDslope} \end{bmatrix}, \tag{14}$$

where $\mathbf{x_{MD}}(k)$ is the estimated level of evidence at time k for detection of a maneuver state. The states that comprise $\mathbf{x_{MD}}$ are $\boldsymbol{x_{MDfreq}}$, the frequency behavior of the Kalman gain used to detect a maneuver, $\boldsymbol{x_{MDZC}}$, which is the incidence of zero-crossings for a Kalman gain value, and $\boldsymbol{x_{MDslope}}$, which is the change in slope of the Kalman gain value that indicates a manuever. For $\boldsymbol{x_{MD}}$, the score was not directly weighted by the measurement uncertainty and is solely based on the current values of the components.

The subnode x_{MDZC} is defined at time k as

$$x_{MDZC}(k) = [0.2 \quad 0.95] \cdot \begin{bmatrix} x_{MDNZC}(k-1) \\ \max\big(1, x_{lc}(k-1)\big) \cdot \big(1 - x_{pltmnv}(k-1)\big) \end{bmatrix}, \tag{15}$$

where $x_{pltmnv}(k)$ is either 1 or 0, with a 1 indicating the platform is, and a 0 indicating that it is not, in a maneuver at time k. x_{lc} is defined based upon whether the measurement is range-bearing in (16) or bearing only in (17), using elements of the Kalman gain. For range-bearing,

$$x_{lc}(k) = 0.4\, x_{K11}^{ZC}(k) + 0.4\, x_{K31}^{ZC}(k) + 0.4\, x_{K12}^{ZC}(k) + 0.4\, x_{K22}^{ZC}(k) + 0.4\, x_{K32}^{ZC}(k) + 0.4\, x_{K42}^{ZC}(k), \tag{16}$$

or, if the measurement is bearing-only

$$x_{lc}(k) = 0.67\, x_{K11}^{ZC}(k) + 0.67\, x_{K21}^{ZC}(k) + 0.67\, x_{K31}^{ZC}(k) + 0.67\, x_{K41}^{ZC}(k). \tag{17}$$

Here, the subscript *Kij* indicates the ith, jth element of the Kalman gain and $x_{Kij}^{ZC}(k)$ indicates the Kalman gain number of zero-crossings upto time *k*.

The subnode x_{MDfreq} is defined in (18) and (19) for the range-bearing measurement and in (20) and (21) for the bearing measurements.

For the range-bearing subnode $\boldsymbol{x_{MDfreq}}$,

$$x_{MDfreq}(k) = \Big[\min\Big(1, \mu_{freq}(k)\Big)\Big] \cdot \big(1 - x_{pltmnv}(k-1)\big), \tag{18}$$

where

$$\mu_{freq}(k) = [0.9 \quad 0.2 \quad 0.3 \quad 0.3 \quad 0.3 \quad 0.2]\, \mathbf{x}_{\mathrm{Krb}}^{freq}(k),$$

and

$$\mathbf{x}_{\mathrm{Krb}}^{freq}(k) = \begin{bmatrix} x_{MDfreq}(k-1) \\ x_{K11}^{freq}(k) \\ x_{K31}^{freq}(k) \\ x_{K22}^{freq}(k) \\ x_{K32}^{freq}(k) \\ x_{K42}^{freq}(k) \end{bmatrix}. \tag{19}$$

For the bearing-only measurement, the subnode x_{MDfreq} is defined similarly, but with different Kalman gains, as

$$x_{MDfreq}(k) = \Big[\min\Big(1, \mu_{freq}(k)\Big)\Big] \cdot \big(1 - x_{pltmnv}(k-1)\big), \tag{20}$$

where

$$\mu_{freq}(k) = [0.9 \quad 0.6 \quad 0.6 \quad 0.5]\, \mathbf{x}_{\mathrm{Kb}}^{freq}(k),$$

and

$$\mathbf{x}_{\mathrm{Kb}}^{freq}(k) = \begin{bmatrix} x_{MDfreq}(k-1) \\ x_{K21}^{freq}(k) \\ x_{K31}^{freq}(k) \\ x_{K41}^{freq}(k) \end{bmatrix}. \tag{21}$$

Finally, the subnode $x_{MDslope}$ is defined in (22) and (23) for range-bearing measurements and in (24) and (25) for bearing-only measurements. For range-bearing,

$$x_{MDslope}(k) = \Big[\min\Big(1, \mu_{slope}(k)\Big)\Big] \cdot \big(1 - x_{pltmnv}(k-1)\big), \tag{22}$$

where

$$\mu_{\mathrm{slope}}(k) = [0.7 \quad 0.4 \quad 0.4]\, \mathbf{x}_{\mathrm{Krb}}^{slope}(k),$$

and

$$\mathbf{x}_{\mathrm{Krb}}^{slope}(k) = \begin{bmatrix} x_{MDslope}(k-1) \\ x_{K12}^{freq}(k) \\ x_{K32}^{freq}(k) \end{bmatrix}. \tag{23}$$

For bearing only,

$$x_{MDslope}(k) = \left[min\left(1, \mu_{slope}(k)\right)\right] \cdot \left(1 - x_{pltmnv}(k-1)\right), \quad (24)$$

where

$$\mu_{slope}(k) = [0.7 \quad 0.4 \quad 0.4]\, \mathbf{x}_{\mathrm{Kb}}^{slope}(k),$$

and

$$\mathbf{x}_{\mathrm{Kb}}^{slope}(k) = \begin{bmatrix} x_{MDslope}(k-1) \\ x_{K11}^{freq}(k) \\ x_{K31}^{freq}(k) \end{bmatrix}. \quad (25)$$

5. Scenario, Example, and Analysis

In each of two examples, an EKF target tracking system is employed and the FOX system as previously described is applied. The scenario used to demonstrate the effectiveness of FOX technique as applied to the maneuver detection problem was summarized in Section 2 and is illustrated in Figure 3. The simulator was developed in MATLAB by the authors. The scenario lasts 1200 seconds. A platform is heading 5 degrees west of north for 480 seconds at 15 kts. Then, the platform changes heading and speed over 60 seconds to 65 degrees west of north and to 9 kts. The platform remains at this heading and speed for the rest of the scenario. The target heads due north for 360 seconds at 15 kts. Then the target changes course and speed over 130 seconds. The course is changed to 70 degrees west of north, and the speed is slowed to 10 kts. At 620 seconds into the scenario, the heading and speed is changed again. This time the acceleration of the target is changed over 224 seconds. The speed is changed to 15 kts while the heading is changed to 10 degrees east of north.

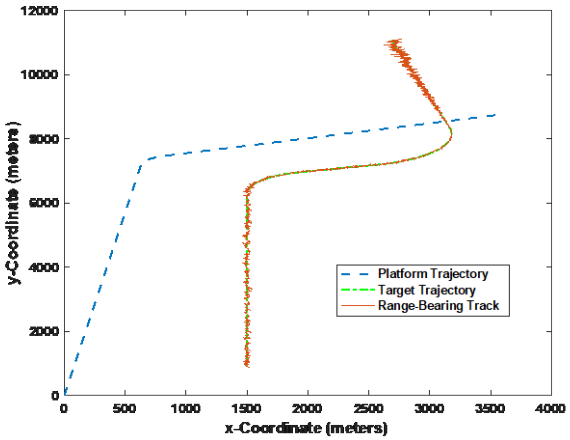


Figure 12: The test scenario with the tracked target using range and bearing meaasurements.

The platform sensor in the first example is a range-bearing sensor with a reporting time of 1 second. The range accuracy is 0.01 m while the bearing accuracy is 0.0003 radians. In the second example, the range sensor is turned off, while the bearing sensor has the same sample time and accuracy in the first example.

5.1. Example 1

The target track of the scenario when using the range-bearing measurement is shown in Figure 12. The associated Kalman filter gains associated to the range measurements are seen in Figure 4 and the gains associated to the bearing measurements were provided in Figure 5. The gains are windowed over a 20 second segment. The window continually slides until the end of the scenario. Figure 13 shows the resulting maneuver detection score.

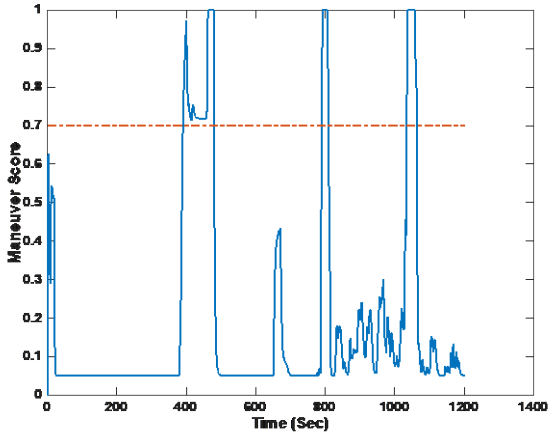


Figure 13: FOX maneuver-detection system results for the range-bearing example

The maneuver detection results indicate that the FOX maneuver detection does detect the target maneuvers when the measurements driving the tracking system contain both a range and a bearing. Using a detection threshold of 0.7, the first target maneuver is detected 31 seconds after it begins (near 400 seconds into scenario) and FOX continues to detect it until the platform begins its maneuver. The second maneuver is detected 170 seconds into its 224 second increase-in-speed and turn. When the target switches from east to west in absolute bearing, the event is detected as a maneuver for 30 seconds. The sharper maneuver (the first maneuver) is detected earlier and for a longer percentage of time of the maneuver. The maneuver has a significant effect on the range and bearing change and impacts the Kalman gain relatively more than a measurement that has little relative impact as with the second maneuver.

In the second maneuver, the effect on the Kalman gain vector values is smaller; the change in bearing and range is smaller at first, and then builds. This compares similarly to the results in [1, 17] in that the closer a target is to the platform and the sharper the maneuvers, the greater the Kalman gain behavior. When the target moves from east to west of the platform, the bearing changes from positive to negative. While the tracker handles the transition smoothly, the sign-changes affect the Jacobian of the measurement-coupling function **H**, as in Eq. (5a). This would be the same if the platform were to maneuver. The results indicate that using the Kalman gains are able to provide indication of maneuvers. The underlying scores indicate that slope behavior is important in target turns. The frequency behavior complements the zero-crossing in the last two detections.

5.2. Example 2

Figure 14 shows the target track overlaid the truth track for the scenario. The target track is, as expected, terribly inaccurate as bearing-only measurement makes the target location and velocity partially unobservable [19]. The target also maneuvers, and this exacerbates the issue [20]. The associated Kalman gains are shown in Figure 15. The Kalman gains are similar to the bearing gains of

range-bearing measurement of Example 1. The gains are windowed over a 20 second segment. The window continually slides until the end of the scenario. Figure 16 shows the maneuver detection score.

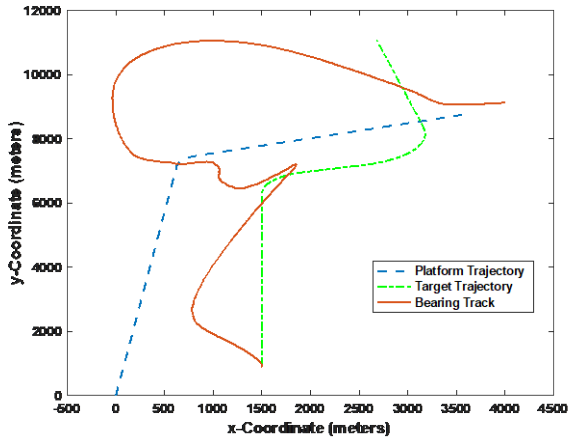


Figure 14. The test scenario with the tracked target using bearing-only measurements where the target position and velocity become unobservable

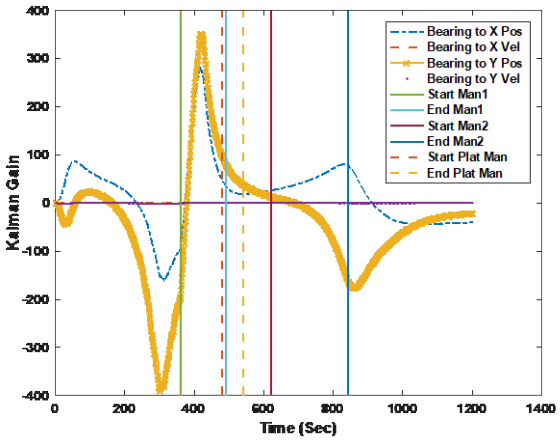


Figure 15: The bearing-only related Kalman gains as mapped over the entire scenario with target and platform maneuvers indicated

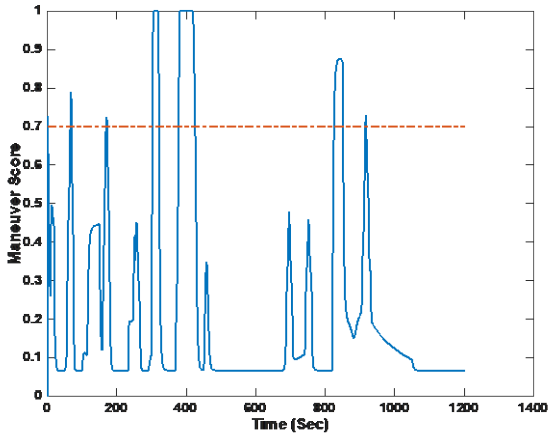


Figure 16: FOX maneuver-detection system results for the bearing-only tracker

The maneuver detection algorithm shows six detections of maneuvers. As discussed in [8], the estimations from a bearings-only track have significant observability issues. The fact that the target is moving originally parallel to the sensor platform results in the range information not being observable. This often collapses the range uncertainty covariance. Three times the target track is detected as a maneuver. In watching the tracker behavior, the Kalman gain is actually providing indications of the failure of the tracking system. When the target maneuvers, the algorithm detects this behavior as the EKF becomes more inaccurate. The platform maneuver provides a pseudo cross-fix which temporarily corrects the track but does not provide enough information to stabilize the filter. The second maneuver is detected later than the maneuver when the range-bearing tracking. For this unobservable measurement type, the FOX algorithm will detect issues with the tracker performance besides maneuvers.

5.3. Summary of Results

The two test scenarios indicate that the FOX maneuver-detection algorithm was able to provide triggers when the target is in or concluding a maneuver. The algorithm also detects other events, particularly the failure of the EKF. This would indicate that at times a reset in the tracker could be useful. It also indicates that behaviors, such as the change from east-to-west of the target, needs to be incorporated into the algorithm similar to the platform maneuver.

6. Conclusions

In this paper, the FOX evidence accrual technique was applied to the problem of detecting target maneuvers using measures from the Kalman gain behaviors from an EKF tracking routine. The algorithm was demonstrated to a variety of events most notably, when the trackign system begins to numerically collapse. Kalman gains of an EKF are important element in detecting tracking failures or changes. The results indicate that other highest-level nodes,i.e., target cross-overs and Kalman collapse, should be incorporated into the FOX system. Maneuver detection can be exploited to work with EKF tracking systems. Incorporation of the method into an interacting multiple-model (IMM) tracking system is planned to advance the applicability of the technique.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] S.C. Stubberud, K.A. Kramer, and A.R. Stubberud, “Fuzzy-Based Evidence Accrual for Target Maneuver Detection,” in Proc. of Innovations in Intelligent Systems and Aplications, INISTA ’18. 2018, pp. 1-7. https://doi.org/10.1109/INISTA.2018.8466310

[2] S. Stubberud and R. Pudwill, in "Feature object extraction - a fuzzy logic approach for evidence accrual in the Level 1 Fusion classification problem," in Computational Intelligence for Measurement Systems and Applications, 2003. CIMSA '03. 2003 IEEE International Symposium on, 2003, pp. 181-185. https://doi.org/10.1109/CIMSA.2003.1227224

[3] R. Sedgewick and K. Wayne, *Algorithms*, *4th Edition*, Addison-Wesley, 2011.

[4] J. Pearl, Causality: Models, Reasoning and Inference, 2nd Edition ed.: Cambridge University Press, 2009.

[5] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," Annals of Mathematical Statistics, vol. 38, pp. 325-339, 1967. https://doi.org/10.1214/aoms/ 1177698950

[6] J. R. Norris, *Markov Chains*. New York: Cambridge Press; 1997.

[7] F.A. Watkins. *Fuzzy Engineering*. MS thesis, Electrical and Computer Engineering, University of California, Irvine; 1994.
[8] R. Lobbia and M. Kent, "Data fusion of decentralized local tracker outputs," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30 (3), July 1994. pp. 787-799. https://doi.org/10.1109/7.303748
[9] R. G. Brown, P. Y. C. Hwang. Introduction to Random Signal Analysis and Kalman Filtering, 3rd Ed. New York: Wiley; 1996.
[10] S. Haykin, Kalman Filtering and Neural Networks, Prentice-Hall, Englewood Cliffs, NJ, 2002.
[11] S. Blackman and R. Popoli, Design and Analysis of Modern Tracking Systems, Artech House, Norwood, MA, 1999.
[12] H. A. P. Blom and Y. Bar-Shalom. "The interacting multiple model algorithm for systems with Markovian switching coefficients." In IEEE Trans. On Automatic Control, 33 (8), 1988, pp. 484-488. https://doi.org/10.1109/9.1299
[13] M. W. Owen and A. R. Stubberud, "A neural extended Kalman filter multiple model tracker," Proc. of OCEANS 2003, Vol 4, Sept. 2003, pp. 2111–2119. https://doi.org/10.1109/OCEANS.2003.178229
[14] P.J. Shea, T. Zadra, D. Klamer, E. Frangione, R. Brouillard, and K. Kastella, "Precision tracking of ground targets," *Proceedings of the IEEE Aerospace Big Sky Conference,* vol. 3, pp. 473-482, 2000. https://doi.org/10.1109/AERO.2000.879873
[15] Y. Bar-Shalom, X.-R.K.T. Li, Estimation with applications to tracking and navigation, Wiley, 2001.
[16] D. Alspach and H. Sorenson, (1972). "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, Vol. 17 , No. 4 , pp. 439 – 448, Aug., 1972.
[17] S. C. Stubberud and KA Kramer, "Monitoring the Kalman Gain Behavior for Maneuver Detection," 25th International Conference on Systems Engineering (ICSEng), 2017, pp. 39-44. https://doi.org/10.1109/ICSEng.2017.71
[18] M.S. Santina, A.R. Stubberud, G.H. Hostetter, Digital control system design, Saunders College Pub., 1994.
[19] W.S. Levine (ed.), *The Control Handbook,* CRC Press, Boca Raton, Florida, 1996.
[20] SC Stubberud, KA Kramer, "Navigation using angle measurements," Mathematics in Engineering, Science & Aerospace (MESA) 8 (2), 2017, 1-15

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 478-484 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Conceptual Architecture for the Continuity if Workflow in Activities on Multi-Devices: Case Study Co-Kitchen

Mirian Janeth Avalos-Viveros [1,*], Luis Gerardo Montané-Jiménez[1], Gabriela Sánchez Morales[2], Carmen Mezura-Godoy[1], Edgard Benítez-Guerrero[1]

[1]*Facultad de Estadística e Informática, Universidad Veracruzana, 91020, Mexico*

[2]*Department of instruction and knowledge management, Universidad de las Américas Puebla, 72810, Mexico*

ARTICLEINFO

Article history:
Received: 31 May, 2019
Accepted: 02 August, 2019
Online: 21 August, 2019

Keywords:
Continuity
Workflow
Multi-device

ABSTRACT

Nowadays, the realization and follow-up of a team activity takes place in different execution environments (e.g. mobile phone, PC) through the use of computer applications. This implies context changes that generate interruptions in the activity. To improve the continuity of an activity with the reduction of interruptions during the development of this activity, would help to finish a task in a shorter time. An example of interruption occurs when a user initiates an activity on a device and decides to continue working on it on a different device, thus facing additional steps that must be performed to obtain the updated information on the desired device. The interruptions can be found in several domains as in the previous example that refers to the work developed through multi-devices or collaborative systems (CS). In this article we present a proposal for a model that incorporates the continuity attribute in the CS development process, to support its users to carry out Collaborative Activities (CA) through different devices with the least possible number of interruptions, so that that way the user does not require additional steps that extend the time of the task and negatively affect the process of transferring his CA to another device.

1. Introduction

This article is an extension of a previous work, in which a Conceptual model to improve the continuity of the workflow of collaborative activities in multi-devices is presented [1]. Collaborative activities can be developed within the Collaborative Systems (CS). CS allow a team of people to work on a common task, through a user interface in a shared environment, virtually [2]. Currently, due to the increase in the capabilities of mobile devices, these are used to perform different activities, gradually replacing the need for the use of a personal computer. This allows to conceptualize that the users of a CS can interact from any of their devices. An example of this scenario is a group of people which for different reasons of time and space decide to use a tool that supports distance collaboration, such as Google Drive®, with which it is possible to give continuity to an activity from a laptop, a tablet or a mobile phone.

The flow of continuity of an activity is limited due to the additional steps that are required to perform activities in multi-devices [3, 4]. These additional steps are considered interruptions, which are presented once the user wishes to continue performing an activity in a new device, such as: saving progress made in the source device, to later log in to the target device, open the application and position to the last state or download the advances in the new device; to name a few. It is important to mention the existence of tools that offer the user a flow of activity that involves fewer steps such as Apple AirPlay®, which in conjunction with Apple TV® wirelessly transmits movies, music and photos from any IOS® device; or Google Drive® offering the creation of content and editing flow, documents, photos and videos, allowing users to access them from any device.

Due to the problem described above, in this work we propose a model that offers continuity to the workflow of CA with support to multiple devices, seeking to reduce the number of interruptions that the user faces each time he wishes to resume a task in a different device. That is, for any user, the process of transferring the last state of information from one device to another is transparent.

* Mirian Janeth Avalos-Viveros, Email: zs17000719@estudiantes.uv.mx

The structure of this article is as follows: Section 2 addresses the aspects of Collaborative Systems, Collaborative Activities, Workflow and Continuity. Section 3 presents papers that support continuity in the workflow. Section 4 shows the challenges and opportunities identified from the analysis of the state of the art. Section 5 presents the proposal of a conceptual model that offers continuity in the CA workflow in multi-devices. Section 6 shows the proposal of a conceptual architecture for the construction of a CS under the concept of continuity. In section 7 a case study is presented to be implemented under the proposed model and architecture, as well as the design of the prototype. Finally, section 8 presents the conclusions and introduces future work.

2. Collaborative activities and Workflow

2.1. Collaborative systems

CS are computer systems that support collaborative work among groups of people who share common goals [2]. These CS are classified based on space, where collaborators can be found in the same place or in different geographical spaces; also, they are classified based on time, where people work at the same time (synchronous), or at different times (asynchronous) [5].

The CS offer functionalities for communication, coordination and collaboration, also known as the 3C's. The communication is through the sending and receiving of messages between team members, as well as the way in which these users observe what the rest of their team is doing. The coordination provides the way to organize the activities to be carried out. The collaboration considers functionalities necessary for users to achieve a collaborative activity [6, 7].

2.2. Collaborative activity

In [8] the activity theory (AT) is presented, representing it in a triangular figure with the elements that intervene in a collaborative activity and its relationship. The elements that make up the triangular system of activity correspond to: subjects, tools, community, division of labor, rules, objects and results; in [9] it is described as the relationship given by the link that exists between two elements of the system with another of the same. During the development of a CA, the people who are involved are immersed in a little-explored element so far: the workflow, by means of which one can know the interruptions that are generated during the life cycle of a task; for example, an interruption to the continuity of the workflow could occur when the user is writing a text on his cell phone and receives a phone call. At this moment, the flow is interrupted and the continuity suffers pauses due to the change of context.

2.3. Workflow

The workflow of an activity is a set of actions that need to be carried out to achieve a specific purpose. The [3] workflow considers two types of activities: individual activity flow and flow of sequenced activities. In the flow of individual activity, the user carries out only one activity, for example: watching a video, reading a book, editing a text, among others, making use of different devices, where each of them takes up the activity from the same point where the previous device left it. While in the stream of sequenced activities, the task flow is composed of different activities that must be performed by users to achieve the final goal; generally, each activity is carried out in a different device according to the needs of the activity, in different contexts of space and time [10]. The flow of activity suffers interruptions in its process due to changes in context, as previously mentioned; that is, there is no continuity in the process of completing a task.

2.4. Continuity

Continuity can be defined as the execution of steps without interruptions. The duration of an activity is an important factor that allows to determine the continuity within a task that is being carried out through a device. For that purpose, it [3] shows three ways in which the flow of continuity is carried out in accordance with the duration of the activity: multiple sessions, context changes and their tasks. With multiple sessions, when an activity is extensive, the user manages to complete it in multiple steps and different devices, depending on the context. While, in context changes of an activity, the user ends session and later takes it back, perhaps in a new device depending on the context where the user is. However, in the case of subtasks, when an activity is relatively extensive, it can be divided into subtasks; which can be different from each other and require a different place of context (time, location, among others.), therefore they can be made in different devices. Figure 1 shows the example of the user John who is in different context changes during the performance of a particular activity. He begins his work at his school using a desktop computer but changes the context to work using his laptop; this generates interruptions (represented by spikes, see Figure 1) that the user faces when he decides to transfer the information to a new device. Later, the user John decides to stop using the desktop computer to continue working from his tablet or cell phone, because he needs to be in motion (in this case he is experiencing interruptions to make possible the transfer of workflow). Finally, when he returns to his work area (school kitchen), he projects the content of the activity on a screen.

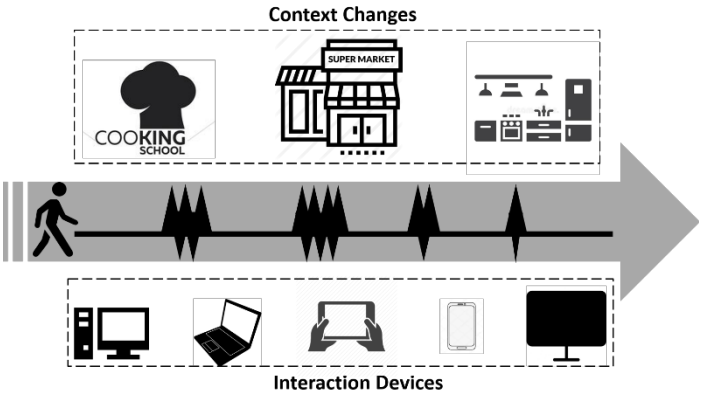


Figure 1: Continuity of a task with changes of context and devices.

3. Works on the continuity in the workflow

Currently, most people have multiple devices and interact with each other to perform their daily tasks [4, 10, 11], this because mobile devices (cell phones or tablets) have capabilities similar to those of a desktop computer or laptop, as people seek to try to keep their work at their disposal at all times. But this has brought unattractive events for users, for example: [12] discovered that when there was no storage in the cloud, people e-mailed the URLs of the websites they visited, to access them from a different device; in [11], a study finds that there are systems with an unintuitive design so that users have continuity in the interaction with multiple devices. In scenarios like this, where the user wants to work with

the devices, the transfer of the task between each device becomes a negative charge.

As a consequence of the lack of continuity in the process of carrying out the activity, there is a disconnection between the actions required of the user when he is using multiple devices [7].

As time passes, proposals have been sought to solve the lack of continuity with the aim of providing the user with the way to transfer information from one device to another without the intervention of unfavorable solutions such as: email, desktop software remote, USB drives, among others. Koren and others, in [13] propose an architecture for the mobility of the session of applications between different mobile devices, within the contributions include support for the discovery of nearby devices and the transfer of sessions with support for interactions based on the exchange of messages and files. However, they present problems in the transfer of the status of the activity.

Cabarcos and others, in [14] propose the design of an architecture for single session initiation in multi-devices (SuSSo) that includes the definition of storage formats of flow of operation session for the transfer of information in an abstract level, so that it can be easily implemented in any electronic device and guarantee interoperability of the cloud service in all devices; in this way sessions initiated from a device transfer their content to a second device without any problem.

Salminen and others, in [15], propose the implementation and evaluation of a middleware to migrate the user interface. The middleware is responsible for having control in the handling of events of the user interface and, in this way, the aforementioned component is transparent for the application. According to the tests carried out by the authors, the studies show that the preference for doing a task in a mobile device depends on the degree of familiarity that the user has with the keyboard of his cell phone, and they consider not having the need to incorporate a new method in order to finish their task before.

Cheng in [16]proposes a middleware capable of handling communication through different devices. The virtual browser facilitates operations between devices according to a predefined file. This research divides a DOM tree into multiple sub_DOM trees, to dynamically manage subviews on multiple devices and link them into a single virtual device. The author presents problems in the development of his proposal, for example: the difficulty to share events and the synchronization of elements and statuses.

Ghiani and others, in [17] present a process of migration for the permanence of status of the tasks that the user is doing in the device called source to the target device, that is, to which he wants to move the user interface to continue his interaction. The authors seek to provide: i) flexible support in several aspects taking into account the variety of possible devices that users may want to use, and ii) an accessible solution for the migration of any application developed with standard Web language, without using any specific complement. According to the tests carried out in this work, they obtained two variants as a result of user evaluations: ease of use and utility.

Pyla and others, in [18] designed a continuous user interface prototype that guarantees a seamless migration of tasks for users who try to perform a specification of requirements and collect tasks, using a tablet and a desktop computer. This system provides support for automatic migration of the task context (eg, applications that were in use) between two devices. In an evaluation carried out, the participants informed that the proposal helped to mitigate the effects associated with the disconnections of tasks, also informed that the automatic availability of the necessary data in the computers and modules contributed directly to a higher perceived reliability and lower error probability.

Currently, there are proprietary applications that offer continuity during the flow of activity, these are available for public use where some have cost and others are freely available, such as: Google Drive®, Netflix®, Office package ®, among others. In Google Drive®, the user has information availability from any device, as long as the device is linked to the same email account; the changes that are made are automatically saved by the tool. In the case of Netflix®, the user can access the exact minute in which he was watching a movie or series in his last session. In desktop systems such as the Office® parcel, if a document that has already been edited is opened, it asks the user if he wishes to continue working from the place where he was previously doing it. In [18] and [19], studies are presented to users where they interact with at least two devices to perform a task. Among the results obtained, an aspect that stands out from the others is the need to have a continuous interface design, for the process of information transfer with fewer steps towards a new device. Additionally, in [19], the authors consider that this aspect could improve the user´s experience when interacting with multi-devices in the development of tasks.

Table 1: Summary of related works

Work	Collaborative	Problems
Koren el at. (2013)	Yes	When the transfer of activity states is made to a new device, there is loss of information.
Cabarcos (2012)	No	The entire state of the task is not preserved.
Cheng (2012)	No	Conflict when several devices share information.
Ghiani et al. (2012)	Yes	Registration to the platform is required to obtain continuity in the devices.
Salmien et al. (2007)	No	Additional steps that users seek to avoid.
Pyla et al. (2009)	No	It does not allow working in parallel in a task with two devices.

In our review of the literature, limited efforts were found that seek to provide continuity in the workflow through multi-devices. Table I summarizes the problems of the papers presented in this section; it is also indicated that only two out of the six works have a collaborative approach as proposed in this research work.

4. Challenges and opportunities to support the continuity of the workflow.

In an exploratory way, we conducted a preliminary study with two teams of five members, each of which held a Collaborative

Activity that included two tasks; the Google Drive® tool was used to perform the tasks. In the first task, they used their computer; once finished this, they had to change device. The second task was developed using their cell phone. After performing the activity by teams, the users were questioned about their experience when doing the activity in multi-devices.

As a result of the exploratory study, we observed that the majority of users usually do tasks through their different devices, but they express displeasure when they need to execute a series of steps to change the device, such as opening the application in the cell phone or downloading the advance in the new device; reason why some users requested to continue the activity from their computer, to avoid extending the time of completion of the task due to interruptions.

Therefore, within the identified challenges and opportunities, is to build a mechanism for the development of Collaborative Systems that transfers information between multi-devices with the least number of interruptions, achieving so that users can work on their tasks not only from their office, school or area where they normally do it, but from any context in which they find themselves and wish to continue working through the device that best adapts to the context. In this way, the user would be provided with a way to carry out Collaborative Activities from any of their devices whenever they wish, according to their needs and interaction preferences, and without being limited to working only in specific spaces.

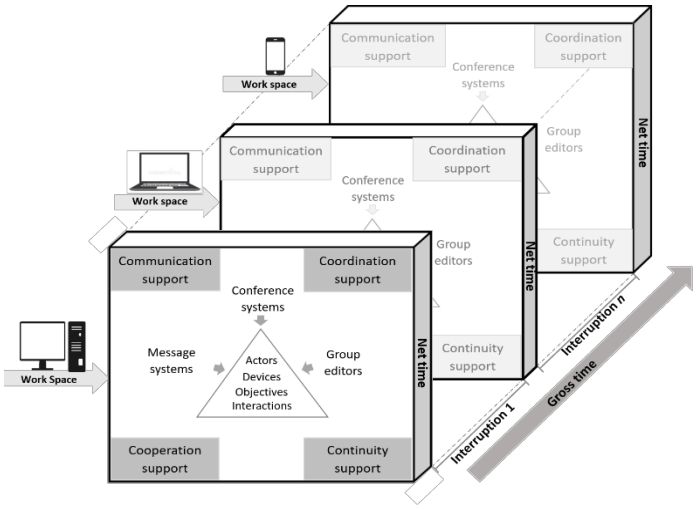
5. Conceptual model for the continuity of the workflow

As a consequence of the analysis of the works related in this research, this article proposes the Model 4C's, as a frame of reference to provide the Collaborative Systems developers with the continuity element.

With the Model of the 4C's, the number of interruptions can be reduced through continuity support, to minimize additional steps when the user resumes work on a new device. Another contribution of the Model of the 4C's, is to allow users to carry out collaborative activities in multi-devices without affecting the workflow of the CAs by the operating system or the user's context change.

The Model of the 4C's is an instance of the Model 3C's, which arises from the models: 3C's [2] , [19] , Mars [20] and CAMCOS [21], taking components such as: communication, cooperation, coordination, workspaces, actors, interactions; which in turn are linked to the continuity element proposed in this work. The 4C's model is shown in Figure 2.

In a Collaborative Activity, groups of actors (users) participate in a group work space to carry out joint activities; These actors also have an individual work space, a space in which each user carries out their activities independently. Within both spaces, users perform different interactions that represent actions to complete a particular activity, which are executed within the three properties of the CS: communication, coordination and cooperation; but, in addition, continuity is required to offer the user the possibility of interacting through different devices during the performance of a task.



For this reason, it is necessary to include in the 3C's model of [4] the continuity element, which provides the user with permanence of the last status of the task when a context change is presented, that is, when the user for external reasons or preference of interaction is in the need of leaving the device with which he initiated the CA and requires to retake it in a different device. In order to provide continuity to the CA workflow through multi-devices, the 4C's model considers 8 elements: 1) Family of actors, 2) Actors, 3) Roles, 4) Objects, 5) Tasks, 6) Interactions, 7) Artifacts and 8) Work spaces; which are described in Table II. These elements are interrelated to propitiate an environment that minimizes interruptions in the sense that the statuses of each of them can be transferred to another device.

Table 2: Elements of the 4C′s model for collaborative applications.

Elements	Description
Family of actors	Set of names of actors where each actor has a different role.
Actors	People or software that carry out CA, through the interaction of different artifacts
Roles	Role that each user plays within the realization of an AC.
Objects	Everything that the actor can manipulate through interactions to achieve goals.
Actions	Actions that the user carries out to fulfill the collaborative activity.
Interactions	Realization of an action by a user through different artifacts within the system.
Mechanism	Hardware that the user uses to carry out his AC, within the SC, such as cell phone, tablet, laptop, desktop computer, among others.
Workspaces	Place where actors perform their actions in a group or individual way through an mechanism.

6. Conceptual architecture

Based on the 4C's conceptual model, we designed a conceptual architecture that supports the continuity of the CA workflow, through which the user can work in a collaborative activity from different mobile devices with a reduced number of interruptions in the process of information transfer and status of the task. The proposed architecture is shown in Fig. 3.

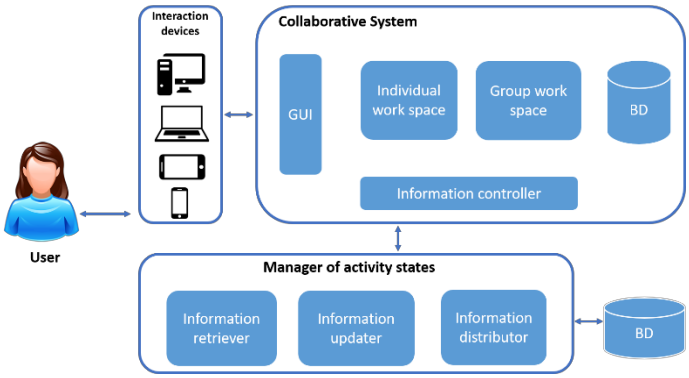


Figure 3: Conceptual architecture of the 4C's model

The continuity support starts once the user selects the device to start the CA. The interaction device is the means of communication between the application and the user. It is responsible for sending requests to the system in order to obtain updated information, and the system receives these requests through the activity's status manager.

Within the activity's status manager, three components are considered: an information retriever, an information updater and an information distributor. The information retriever obtains the last modifications of the users, in addition to the last workspace where they stopped.

The information updater is responsible for unifying the information of the individual and group workspace of the users and generates a new version of the progress of the team's work. When the user wants to resume the activity in a new device, the information distributor obtains the latest version of the work's progress and shows it to the user from the workspace in which he/she is working, supporting in this way the continuity of the workflow of the collaborative activity in multi-devices.

7. Case study

Based on the elements of the 4C's model and the conceptual architecture proposed above, a case study is being addressed through a scenario where, during the process of developing an CA, it is required to transfer the information to different mobile devices, in order to test whether the elements identified in the conceptual architecture are sufficient to provide the user with a continuous workflow, in the performance of collaborative activities in multi-devices.

The case study was modeled so that a team of actors worked on different activities for a common goal (to prepare a dish and generate a recipe). There are three types of actors: the administrator (teacher of the cooking school), the team representative and the members of the work team. Each actor plays a role within the collaborative activity; the administrator is the evaluator of the final product that each team elaborates, the team representative is the organizer of activities within the team and the member of the team is the buyer of ingredients. The actors interact with different artifacts during the collaborative work process such as: a desktop computer or a laptop to send the list of ingredients to the work team, as well as to make the recipe for the final product; a cell phone to perform tasks that require user mobility such as going to a supermarket to buy the ingredients of the dish; a tablet where users can count on a broader work space compared to the telephone, but also with mobility in the different workspaces where the actors must move to meet the objective of the activity.

The design of the Co-Kitchen system was made using the user-centered design methodology, an approach that allows iterations with the design throughout the life cycle of the product and in this way to redesign the system in order to meet the expectations and needs of end users [22]. Within the work framework for the design of human-centered interactive systems [23], the design process starts from the understanding of the context of use to specify the requirements of users, make a proposal of design solution and finally evaluate the design.

Phase one of the design process has been specified at the beginning of this section, where the modeled study case is detailed. Once analyzed and understood the context of use, we identified the requirements that the actors have in accordance with what they do within the system; in this way we knew the functionalities that the system should provide. In Table III, the functionalities of the system are described.

Table 3: Functions of the prototype Team Cooking

Function	Description
Create team	The administrator can create a work team.
Create list of ingredients	The administrator generates a list of the ingredients that the team must use to prepare a dish.
Assign tasks	The team representative is responsible for assigning tasks to each member of the team. Tasks such as buying ingredients, looking for kitchen utensils, lighting the oven, among others.
Create recipe	All the team at the end of the proposed dish, should generate a recipe for cooking it; for this each one must describe within the kitchen format the procedure for the realization of it, resulting in the recipe of the dish.

Once the requirements and functionalities that the system should provide the user in order to carry out his tasks were identified, the next step was to create the proposed design solution, which is shown and described below in Figures 4, 5, 6 and 7.

For the design of the proposal we use standard components such as: buttons, menus, text fields, etc., as well as: pointers, chat, and lists of connected users, according to [24] [25], who propose to simplify the development of groupware interfaces and to reduce the time of information search by grouping related elements in such a way that it is easy for the user to identify in which section is the element that he requires in order to perform a task.

In Figure 4, the start screen that the team representative actor displays when logging into the system is displayed. On the left side are the actions that can be executed within the system, in the center part of the screen you can see the work area and on the right

side there is the collaboration area where, through a list of connected users, the actor is kept informed of the presence of users who are collaborating during the realization of the CA and, at the top of that same side, we can also see the presence of each member, as well as the progress that each one of them has.

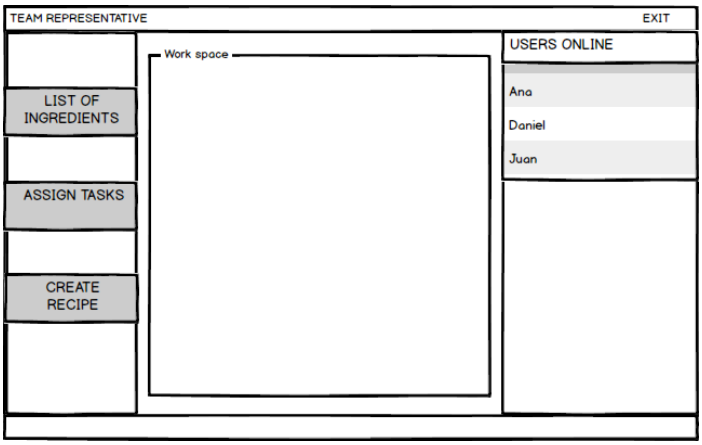


Figure 4: Team representative window.



Figure 5: Window to visualize list of ingredients.

In Figures 5, 6 and 7, what happens when the user clicks on one of the buttons on the left side is shown. When selecting "List of ingredients" (see Figure 5), in the work area the list of ingredients that was sent by the teacher of the cooking class is loaded. In case of selecting the option "Assign tasks" (see Figure 6), within the work area the team representative can enter the tasks assigned to each member of their team, which once assigned arrive as notification to the members to begin to work on them.

Finally, for the option to create a recipe (see Figure 7) within the work area, the actor works together with the rest of the team members in making the previously prepared recipe. Since it is a shared work area, each user has access to specific areas of the recipe according to the tasks previously assigned, which can be reported and described in the recipe.

The design of the solution proposal corresponds to a Web version loaded from a laptop or a tablet, but in the case of using a mobile device such as a cell phone, the design of the interface must be redesigned to fit a smaller screen, so that the work area is not affected and so the actor can work through it without complications.

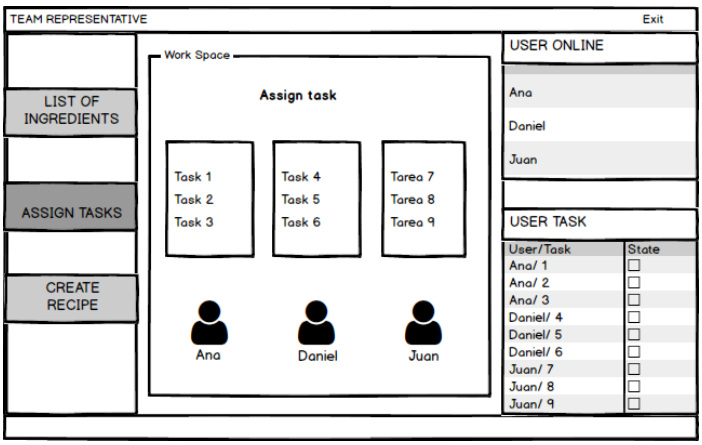


Figure 6: Window to assign tasks

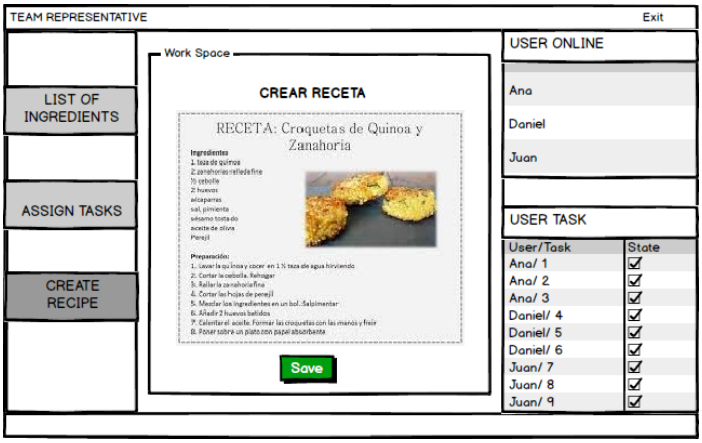


Figure 7: Window to create recipe.

8. Conclusions and future work

The advance of technology generates systems for collaboration between groups of people that interact with each other through the computer. Nowadays, it is possible to carry out an activity from different mobile devices such as: a cell phone, a tablet or a laptop, providing the user mobility to perform their activities. However, there is a gap in the conservation of the task status when moving to a new device. Therefore, this problem needs to be addressed in order to offer users systems that provide continuity to their workflow by requiring changes in the context of use during the performance of a task. That is why in this article we propose the model of the 4C's and the conceptual architecture to support users to carry out a collaborative activity in multiple devices in a shorter time. This proposal includes the continuity element of the model 3C's of [2], and the manager of states in the conceptual architecture, seeking to minimize the number of interruptions during the transfer of information from the collaborative activity through different mobile devices

As future work, we plan to continue with phase four of the framework for the design of interactive human-centered systems [23], corresponding to the evaluation of the design proposal

presented in this article, to subsequently work on a redesign under the observations and suggestions obtained from its evaluators.

Once the iterations of the design and evaluation process of the system design proposal have been finalized, we will give the appropriate follow-up to the construction of the Co-Kitchen system, to subsequently test the system and test whether our proposal effectively improves the continuity of the workflow of a collaborative activity when it is carried out through multi-devices, reducing the time in the interruptions that the user experiences when transferring and resuming the activity in a new device.

References

[1] L. G. M. -J. G. S. M. C. M.-G. a. E. B.-G. M. J. Avalos-Viveros, ""Conceptual Model to Improve Continuity of the Workflow of Collaborative Activities in Multi-Devices" International Conference on Information Systems and Computer Science (INCISCOS). IEEE, pp. 233-238, 2018.

[2] C. A. ELLIS, S. J. GIBBS y G. REIN, "Groupware: some issues and experiences" Communications of the ACM, vol. 34, n 1, pp. 39-58, 1991.

[3] M. Levin, Designing multi-device experiences : an ecosystem approach to creating user experiences across devices, O'Reilly Media, 2014.

[4] K. MAJRASHI, M. HAMILTON y A. UITDENBOGERD, "Evaluating Cross-Device Transitioning Experience in Seated and Moving Contexts" de Proceedings of the 22nd Pacific Asia Conference on Information Systems (PACIS 2018)(Yokohama, Japan2018), AISEL, Yokohama, Japan, 2018.

[5] G. GERONIMO, V. CANSECO y H. DE LEÓN, "Breve Introducción a los Sistemas Colaborativos: Groupware & Workflow" Temas-Universidad Tecnológica de la Mixteca, pp. 49-54, 2002.

[6] L. N. Yann Laurillau, "Clover architecture for groupware" In Proc. Conference on Computer supported cooperative work, ACM, pp. 236-245, 2002.

[7] C. G. a. S. G. D. Pinelle, "Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration" ACM Transactions on Computer-Human Interaction (TOCHI), vol. 10, n 4, pp. 281-311, 2003.

[8] Y. Engeström, "Expansive learning at work: Toward an activity theoretical reconceptualization" Journal of education and work, vol. 14, n 1, pp. 133-156, 2001.

[9] L. G. M. Jiménez, Presencia social en sistemas groupware, Tesis doctoral, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, México, 2016.

[10] D. R. J. K. M. B. S. E. V. R. a. B. M. L. J. Paay, "Investigating Croos-Device Interaction between a Handheld Device and a Large Display" In Proc. of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, pp. 6608-6619, 2017.

[11] J. K. a. M. B. S. D. Raptis, "Continuity in multi-device interaction: an online study" In proc. of the 9th Nordic Conference on Human-Computer Interaction. ACM, p. 29, 2016.

[12] H. B. a. S. D. W. Jones, "Keeping found things found on the web" Tenth international conference on Information and knowledge management. ACM, pp. 119-126, 2001.

[13] D. S. a. T. S. I. Koren, "Session mobility for collaborative pervasive apps using XMPP" Pervasive Computing and Communications Workshops (PERCOM Workshop), IEEE Internacional Conference on. IEEE, pp. 169-174, 2013.

[14] F. A. M. R. S. G. A. M. L. a. D. D.-S. P. A. Cabarcos, "Susso: seamless and ubiquitous single sign-on for cloud service continuity across devices" IEEE Transactions on Consumer Electronics, vol. 58, n 4, pp. 1425-1433, 2012.

[15] S. H. a. J. R. T. Salminen, "Middleware based user interface migration: implementation and evaluation" In Proc. of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in bobile technology, ACM, pp. 358-363, 2007.

[16] B. Cheng, "Virtual browser for enabling multi-device web applications" In Proc. of the Workshop on Multi-device App Middleware. ACM, pp. 3-7, 2012 .

[17] F. P. a. C. S. G. Ghiani, "Push and pull of web user interfaces in multidevice environments" International Working Conference on Advanced Visual Interfaces. ACM, pp. 10-17, 2012.

[18] P. S. T. M. H. J. &. P.-Q. M. A. Pyla, "Continuous user interfaces for seamless task migration" International Conference on Human-Computer Interaction, Springer, pp. 77-58, 2009.

[19] E. I. B. M. D. C. M. a. E. M. L. G. Montané, "Studying Social Interactions in Groupware Systems" IEEE Latin America Transactions, vol. 13, n 10, pp. 3488-3497, 2015.

[20] M. R. a. S. T. C. Mezura-Godoy, "Mars: Modelling arenas to regulate collaborative spaces" In International Conference on Collaboration and Technology, Springe, pp. 10-25, 2003.

[21] C. M.-G. a. L. G. M.-J. E. Benítez-Guerrero, "Context-aware mobile collaborative systems: Conceptual modeling and case study" Sensors, vol. 12, n 10, pp. 13491-13507, 2012.

[22] J. R. a. D. Chisnell, Handbook of usability testing : how to plan, design, and conduct effective tests, Indianapolis: Wiley Publishing, Inc., 2008.

[23] ISO, E. N. 9241-210: Ergonomics of human-system interaction. Human-centred design for interactive systems, 2010.

[24] V. M. R. M. L. L. M. D. G. J. L. G. J. A. N. M. .. &. H. M. V. Penichet, " Una aproximación al proceso de diseño e implementación de interfaces de usuario para aplicaciones groupware" IX Congreso Internacional Interacción , pp. 19-28, 2008.

[25] C. Stephen, " A cognitive approach to the design of information graphics" CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF PSYCHOLOGY, pp. AIP-143, 1991.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 485-502 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Material, Structural Optimization and Analysis of Visible-Range Back-Illuminated OPFET photodetector

Jaya V. Gaitonde*, Rajesh B. Lohani

Electronics and Telecommunication Department, Goa Engineering College, Farmagudi-Ponda-Goa, 403401, India

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 02 August, 2019
Online: 21 August, 2019

Keywords:
Material optimization
Structural optimization
Back-illuminated OPFET
Photodetector
Visible range
Visible/UV contrast ratio

ABSTRACT

High gain-bandwidth product and visible/UV contrast photodetectors are vital in Visible Light Communication (VLC) and Ultraviolet (UV) reflectance imaging applications respectively. We adopt material and structural optimization to perceive such photodetectors with back-illuminated Optical Field Effect Transistor (OPFET) wherein any potential difference in absorption coefficient of the semiconductor material between the visible and the UV range (higher in the UV region) can be explored at its full potential. The results have been analyzed using the photoconductive and the photovoltaic effects, the series resistance effects, scaling rules- induced effects, and channel length-variation effects. We consider the three most prominent and functional materials in the visible range (Si, GaAs, and InP) for material-based optimization. Structural optimization is performed employing a range of medium gate lengths. The gate electrodes utilized are Indium-Tin-Oxide (ITO) for Si and GaAs with high Schottky barrier heights of ~0.71 eV and ~0.98 eV respectively whereas the Schottky contact on InP is gold (Au) with a high barrier height of ~0.8 eV. The operating visible and UV wavelengths are 600 nm and 350 nm respectively. The results suggest that GaAs OPFET has wide bandwidth potential in the gigahertz range apart from its high sensitivity and visible/UV contrast features. The InP-based OPFET exhibits high sensitivity and sub-gigahertz frequency response; and can compete or surpass the GaAs OPFET in terms of the visible/UV contrast ratio. The Si OPFET shows bandwidth in the megahertz range along with high sensitivity but exhibits low contrast ratio. The structural parameters have a significant effect on the detector response. The results are in-line with the experiments. This paper reflects the performance of the investigated detectors towards the said applications through optimization and the associated analysis represents the dependence of the obtained response on the device material and structural parameters, thus, opening the door for further research.

1. Introduction

This paper is an extension of work originally presented in *The IEEE Conference on Emerging Devices and Smart Systems* (*ICEDSS'18*) [1].

Visible-Light Communication (VLC) is an emerging alternative to RF-based communication to suffice for the increasing demand for bandwidth. Its potential applications include Li-Fi (Light-Fidelity) Networks, vehicle to vehicle communication, robots in hospitals, underwater communication, and information displayed on sign boards [2]. The photodetectors contributing to VLC should feature high gain-bandwidth product apart from low dark current. Presently, the state-of-art photodetectors being employed in VLC include the p-i-n and the Avalanche photodiodes (APDs). p-i-n detectors provide high speed detection but without amplification whereas the APDs offer high gain but as the frequency increases, the gain reduces significantly. On the other hand, Ultraviolet (UV) reflectance imaging is a well-known application in biomedical engineering. It requires photodetectors with high visible/UV contrast ratios alongwith high visible sensitivity and low dark current. There is no much information available on the visible/UV contrast features of any photodetector. The contrast ratio should be high in order to appropriately control the ambient visible light in the presence of background UV radiation during imaging and enabling erroneous

*Jaya V. Gaitonde, Goa Engineering College, Farmagudi-Ponda-Goa-India, 403401, Mobile: (+91) 9420687029, Email: jayagaitonde46@gmail.com

operation. Low dark current improves the device performance. Both the applications have one thing in common: they need photodetectors operating in the visible range. Three most featured materials capable of operating in the visible region include Si, GaAs, and InP. These possess distinct electrical and optical properties which motivate one to conduct research using these materials. One potential candidate to cater to high gain-bandwidth product applications is a phototransistor or a photoFET. Optical Field Effect Transistor (OPFET) or optically controlled Metal-Semiconductor Field Effect Transistor (MESFET) has been a widely studied device over the past few decades for its applications as photodetectors, optically-controlled amplifiers, oscillators, and switches [3]-[13]. However, the research lacks the assessment or the impact of the OPFET device incorporating these materials from the optimization point of views i.e. material and structural optimization. Further, analysis of the attained results and the correlation of the detector responses with the material and the structural parameters can substantiate further research. Thus, in this paper, we call for material and structural optimization of the visible-range back-illuminated OPFET detector for communication and visible/UV contrast applications. The back-illuminated OPFET model (without substrate effect) is purposely chosen so that any significantly high difference in absorption coefficient at the visible and UV wavelengths is potentially used for contrast applications. The transparent gate materials considered for simulation are Indium-Tin-Oxide (ITO) for Si and GaAs whereas gold (Au) for InP. The corresponding barrier heights are ~0.71 eV, ~0.98 eV, and ~0.8 eV respectively. Possibly large barrier heights have been chosen since these can have considerable positive effect on the photovoltage and the photoconductive effect through the modulation of the depletion width.

A theoretical semi-analytical model accounting for photoconductive and photovoltaic effect is presented. The continuity equations have been solved analytically to yield the carrier densities whereas the charges are calculated numerically using the Trapezoidal method. The total drain-to-source current is estimated using the model given in [14]. The trap effects have been neglected in this work.

We begin with the theory followed by the results, discussion, and the relevant analysis. We conclude with a brief summary.

2. Theory

The back-illuminated device structure is schematically shown in Fig. 1. The gate is shorted to the source for a reduction in the number of power supplies. The channel is n-type uniformly moderately doped whereas the substrate is p-type semi-insulating. The illumination is guided by a fiber inserted from the rear side of the device through the substrate upto the active layer-substrate interface. The radiation is absorbed in the channel and the substrate regions creating electron-hole pairs. The holes contribute to photovoltage after crossing the Schottky junction which increases the drain-to-source current through a reduction in the depletion width (photovoltaic effect). On the other hand, the electrons are directed toward the channel which increases its conductivity upon the application of the drain-to-source voltage (photoconductive effect).

The device characteristics are depicted by the three regions of operation: the linear, the pinch-off, and the saturation regions. The drain-to-source saturation current depends upon the total charge, the gate width, and the saturation velocity and is limited by the voltage drop across the source series resistance. The drain-to-source conductance also depends upon the total charge (however, at zero drain-to-source voltage), the carrier mobility, and the gate width; is inversely proportional to the gate length, and is limited by the source and drain series resistances. The total drain-to-source current is given by [14]:

$$I_{ds} = I_{dss}(1+\lambda V_{DS})\tanh(\eta V_{DS})$$

where the symbols have the same meaning as stated in [14]. The parameter η is defined as the ratio of the drain-to-source conductance to the saturation current as the drain-to-source voltage tends to zero. Thus, on one hand, the drain-to-source current is directly proportional to the saturation current; on the other hand it is inversely proportional to the saturation current through the parameter η under the hyperbolic tangent expression. This shows that there is limitation on the total current that can be achieved. The parameter η also signifies that the total current is directly proportional to the conductance as V_{DS} tends to zero under the hyperbolic tangent expression. Hence, the conductance as V_{DS} tends to zero and the saturation current are the prominent factors determining the total drain-to-source current apart from the voltage, V_{DS}, and the total current is limited by the source and drain series resistances.

Under illumination, the total charge is boosted due to the photovoltaic and the photoconductive effects which enhances the saturation current and the drain-to-source conductance limited by the series resistances. In addition, with the increase in the optical power, the photovoltage increases reducing the series resistances and augmenting the current. The source and drain series resistances are calculated based on the method provided in [15]. The reader is advised to follow [1] for the modeling equations and the calculation of the carrier densities, photovoltage, and the total charge.

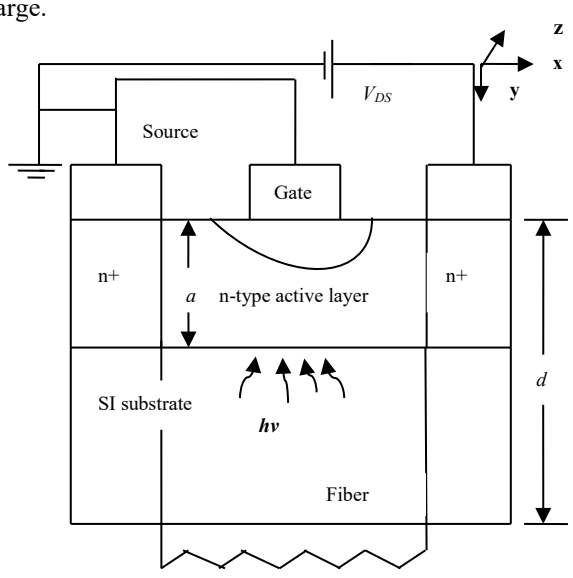


Figure 1. The schematic structure of the back-illuminated model of OPFET.

The detector parameters such as responsivity, dark current, and visible/UV contrast ratio are defined in [1]. Here we define the 3-

dB bandwidth and the unity-gain cut-off frequency (f_T) of the device. The bandwidth is the highest modulation frequency which the photodetector can detect without errors. Mathematically, it is given by the frequency at which the photocurrent response falls to 0.707 times its DC response. The unity-gain cut-off frequency is the operating frequency at which overall gain of the device falls to unity. The f_T is given by the ratio of the transconductance to 2π times the gate-to-source capacitance. The transconductance is the variation in the drain-to-source current with the variation in the gate-to-source voltage and is the amplification factor. The gate-to-source capacitance is the change in the space charge below the gate and its sidewalls with the change in the gate-to-source voltage and is a switching parameter. Thus, fT represents the amplification bandwidth of the device. The simultaneous detection-cum-amplification bandwidth is given by either of the two (3-dB bandwidth or the f_T value), whichever is smaller.

3. Results and Discussion

The simulations have been carried out in MATLAB with MEX coding feature. The results have been validated with experimental work published elsewhere [3] (Figure 2). The results have been compared with the commercially available software (Visual TCAD)-based simulations and have shown close resemblance (Figure 3) (Gate length of 4 μm, gate width of 4 μm, active layer thickness of 0.27 μm, and doping concentration of 4×10^{22} /m^3). The drain-to-source bias voltage is variable whereas the gate-to-source voltage is set to zero volts. The wavelengths used are 600 nm for the visible light and 350 nm for the UV radiation. The photon flux densities of 10^{16}, 10^{19}, and 10^{22} /m^2-s used in these simulations correspond to optical power densities of 0.33 μW/cm^2, 0.33 mW/cm^2, and 0.33 W/cm^2, respectively at 600 nm and 0.57 μW/cm^2, 0.57 mW/cm^2, and 0.57 W/cm^2 respectively at 350 nm. The structural optimization is performed over a wide range of medium gate lengths from 3 μm to 5 μm. However, only the four most relevant sets are provided here for comparison. The design variables are chosen based on the scaling rules stated in [16]. The materials being employed are Si, GaAs, and InP for the back-illuminated OPFET device with the gate electrodes being Indium-Tin-Oxide (ITO) for Si and GaAs, and gold (Au) for InP. The simulated optimization results are presented in Table I. The parameters used for calculation are given in Table II. The comparison with the state-of-art photodetectors and amplifiers is provided in Table III showing significantly enhanced response. Figure 4 depicts the drain-to-source current responses under dark and illumination for all the three devices at a gate length of 3 μm and an active layer thickness of 0.27 μm at an intensity of 10^{19} /m^2-s clearly showing the sensitivity to visible light. Figure 5 portrays the frequency responses of the three devices at a gate length of 3 μm and an active layer thickness of 0.27 μm at an intensity of 10^{19} /m^2-s demarcating that the GaAs, InP, and Si OPFET devices operate at high, moderate and low data frequency ranges respectively.

3.1 Series resistance and photovoltage analysis in Si, InP, and GaAs OPFETs

Beginning the discussion with the Si back-illuminated OPFET, this device exhibits the largest series resistances among all the material devices with similar dimensions in the visible range. With the smallest Schottky barrier height of ITO-Si junction (~0.71 eV), the series resistance ought to be the smallest due to the larger undepleted channel thickness in the spacing between the gate and the source, and the gate and the drain (depletion width is dependent upon the barrier height). This should have drawn larger current through the channel charge. However, due to the low electron mobility (~0.09 cm^2/(V.s)) and the said barrier height, upon which the current flowing in the spacing between the gate and the source, and the gate and the drain depends upon (barrier height directly affects the voltage drop across the channel), the current significantly falls. This raises the series resistance value to compensate for the fall with the increment in the depletion width through the voltage drop across the series resistance. Other materials, GaAs and InP exhibit lower and almost equal series resistances. This is attributed to the higher mobility of GaAs (~0.5 cm^2/(V.s)) and higher barrier height of ITO-GaAs junction (~0.98 eV) (through the voltage drop) which tend to increase the current in the spacing and lower the resistance. The same barrier height also tends to decrease the current (through the increment in the depletion width) and raise the resistance, however, the factors supporting the lowering of the resistance supersede those raising it, resulting in net reduction of resistance. For InP, the mobility is (~0.43 cm^2/(V.s)) and the barrier height of Au-InP junction is (~0.8 eV). In this case, the mobility is significantly high whereas the barrier height is moderate. With the moderate barrier height, the series resistance ought to be between that of Si and GaAs, but due to the high mobility the series resistance in InP features the same as that in GaAs. The series resistances play an important role in limiting the photocurrent. The series resistances decrease with illumination.

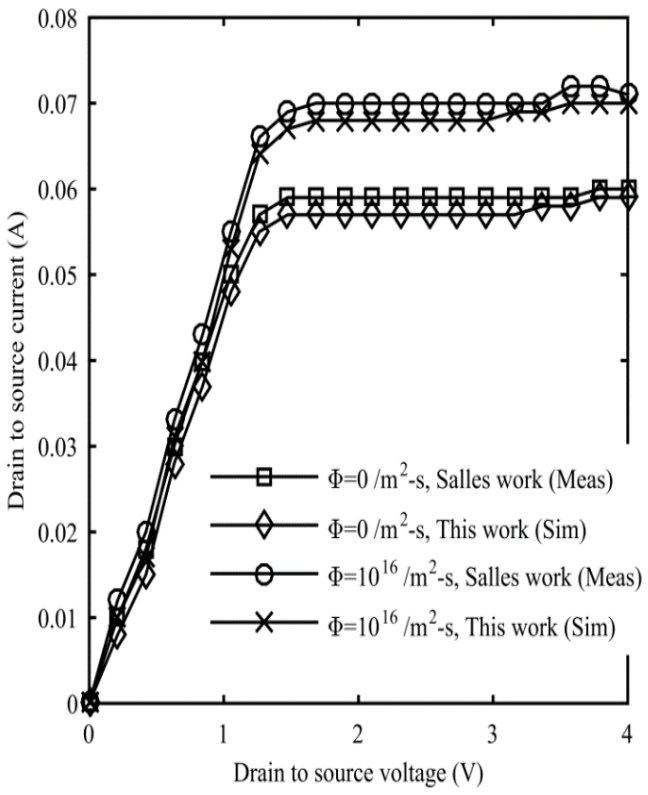


Figure 2 Comparison of the present model results with the experiments in [3].

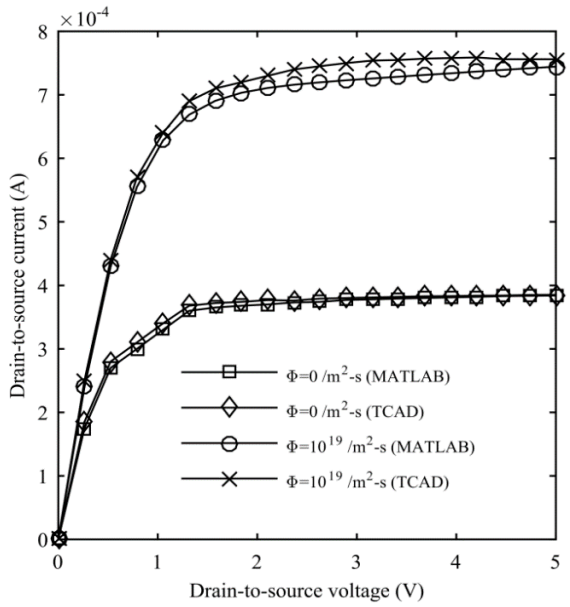


Figure 3 Comparison of the present model results with the simulations using Visual TCAD Software.

The photovoltage generated is almost the same in both Si and InP devices (~0.27 V, ~0.46 V, and 0.61 V) in the visible range considering most of the sets. This can be explained as follows: The photovoltage depends upon the photogenerated hole density crossing the Schottky junction which in turn, depends upon the absorption coefficient, the photon flux density, the hole lifetime, and the saturated hole velocity. The absorption coefficient of Si at 600 nm is 7.5×10^5 /m, the hole lifetime is 75 µs, and the saturated hole velocity is ($\sim6.3\times10^4$ m/s). On the other hand, the corresponding parameters for InP are 1×10^7 /m, 0.3 µs, and ($\sim5.6\times10^4$ m/s) respectively. The hole density is directly proportional to the absorption coefficient, the photon flux density, and the hole lifetime ($\alpha\Phi\tau_p$), whereas it is inversely related to the saturated hole velocity through ($1+\alpha v_y\tau_p$).The values suggest that the terms directly as well as inversely related to the hole density are almost the same in both the cases since any difference in individual parameters between the two material systems is cancelled out when their product is evaluated. One more parameter upon which the photovoltage is dependent upon is the reverse saturation current density through the Schottky barrier height. The photovoltage is inversely related to the saturation current density, and the saturation current density is inversely related to barrier height. Hence, the photovoltage is directly related to barrier height. Since, the barrier heights of ITO-Si andAu-InP junctions are (~0.71 eV) and (~0.8 eV) respectively, the photovoltage tends to be more in the InP device. However, since InP possesses a high absorption coefficient of 1×10^7 /m corresponding to an absorption depth of 0.1 µm, the hole density experiences an exponential decrease with distance and fewer holes cross the Schottky junction as compared to the Si device. The moderate absorption coefficient of 7.5×10^5 /m results in a large absorption depth thus, causing a lesser degradation of hole density with distance. As a result, the photovoltage developed is almost the same in both the cases. On the other hand, the photovoltage generated in GaAs OPFET is enhanced (~0.398 V, ~0.576 V, and ~0.755 V). This is due to the moderate absorption coefficient of GaAs in the visible region (4×10^6 /m), the higher saturated hole velocity ($\sim9\times10^4$ m/s), and the significantly higher Schottky barrier height of the ITO-GaAs junction (~0.98 eV) superseding the effect of lower hole lifetime (10^{-8} s). The saturated hole velocity positively affects the photovoltage through the photogenerated hole current density in the equation for photo voltage.

At the lowest gate length under consideration of 3 µm and the highest doping concentration of 5×10^{22} /m^3 (gate length-doping concentration product is constant from scaling rules [16]) with an active layer thickness of 0.15 µm, the OPFET devices exhibit the lowest dark and photocurrents among all the sets. This is ascribed to the very less active-layer thickness, as well as the low gate length and the gate width along with the reduced sensitivity at the higher doping concentrations. This proportionally decreases the current although the high doping concentration induces significant channel charge.

3.2 Analysis of Si OPFET at 3 µm gate length and 0.15 µm channel thickness

The drain-to-source currents obtained using Si OPFET are (0.36 mA, 0.9 mA, 1.7 mA, and 15.6 mA). The large series resistances owing to the small active layer thickness (0.15 µm) and significantly high barrier height (~0.71 eV) limit the said currents. At the flux density of 10^{16} /m^2-s, the photovoltaic effect is significant (0.28 V) whereas the photoconductive effect is negligible as compared to the photovoltaic effect (due to the lower power level) resulting in a current of 0.9 mA. The photoconductive effect will add to the response only when the photoconductive charge is comparable to the doping-induced charge. As the flux density is increased to 10^{19} /m^2-s, the photoconductive effect also contributes in addition to the photovoltaic effect (0.46 V) owing to the higher power level and the long electron lifetime (75 µs). Most of the photoconductive contribution emanates from the depletion region since the major portion of the channel is depleted owing to high barrier height, small channel thickness, and the large series resistance which incurs significant voltage drop across it. This results in a quite high current of 1.7 mA. At the flux density of 10^{22} /m^2-s, the current is significantly enhanced (15.6 mA) owing to the considerably high contribution from the photoconductive effect in the depletion region. There is no photovoltaic contribution; instead the device shows negative sensitivity i.e. increase in depletion width with illumination. This arises due to the substantially high current from the photoconductivity which incurs large voltage drop across the series resistance. Thus, the depletion width increments with illumination surpassing the effect of photovoltage-induced reduction of the depletion width. As such, the responsivities attained are (5.3×10^8 A/W, 1.32×10^6 A/W, and 1.49×10^4 A/W) which are significantly high owing to the above discussed phenomena and the low dark current. The 3-dB bandwidths exhibited by this device are (0.16 MHz, 0.136 MHz, and 0.57 MHz). These megahertz range bandwidths are due to the long electron and the hole lifetimes (75 µs) upon which the bandwidth is inversely related through the modulation of carrier lifetime with frequency. At the lowest flux density, the bandwidth is 0.16 MHz due to the contribution from the photovoltaic effect alone. At the higher flux density, the bandwidth falls to 0.136 MHz due to both photovoltaic and photoconductive effects adding to the response.

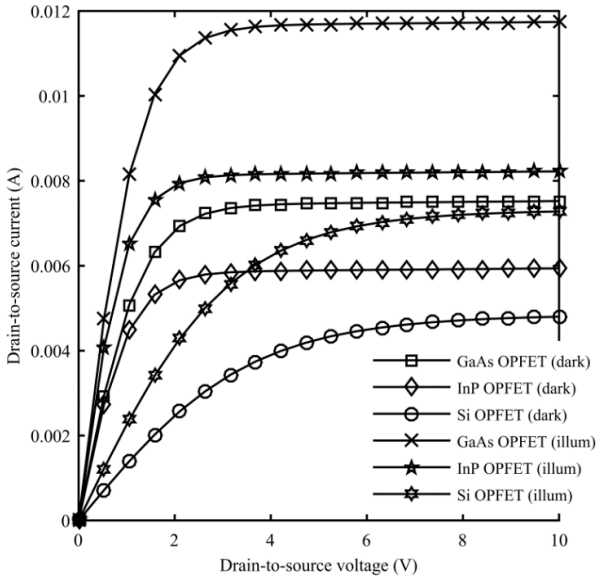


Figure 4 Comparison of the Si, InP, GaAs OPFET devices under dark and illumination.

This can be explained by the fact that when the identical frequency responses of two independent comparable effects are added, the net frequency response exhibits reduced bandwidth as compared to the individual ones. As the flux density is increased further, the bandwidth rises to 0.57 MHz. This is attributed to the sole and substantially high contribution from the photoconductive charge as compared to the previous responses. The transconductances obtained are the lowest among all the sets (1.4 mS, 2.4 mS, 3.9 mS, and 29.1 mS) due to the reduced dimensions at this gate length which decreases the overall current and the reduced sensitivity at high doping concentrations. Note that at constant photovoltage, the depletion width sensitivity is higher at lower doping concentrations. The transconductance increases with the increase in the optical power [6], [7], [9], [10] since at a constant doping concentration, the depletion width sensitivity is higher at larger photovoltages. However, at the highest flux density, this does not hold good where the device experiences negative sensitivity. In this case, it is the high photoconductive charge which results in high saturation current and is independent with the change in the gate-to-source voltage due to the complete depletion of the channel. The drain-to-source conductance evaluated at zero drain-to-source voltage also is high due to the induced charge and is sensitive to the variation in the gate-to-source voltage since the channel is undepleted (zero drain-to-source voltage). Further, the conductance is affected by the dependence of the series resistances on the gate-to-source voltage. The above reasons suffice for the observed high transconductance. The gate-to-source capacitances are also the lowest among all the sets (0.0546 pF, 0.084 pF, 0.105 pF, and 0.0076 pF) owing to the same reasons as applicable to the transconductances, however, here the gate region space charge replaces the drain-to-source current. The reasons hold good for the lower flux densities. However, at the higher flux density, the space charge is considerably reduced due to the photoconductive electrons in the depletion region de-ionizing the charged donor ions. Also, the gate-to-source capacitance is independent of the change in the photoconductive charge (which is a major contributor to the response) since it is a measure of the space charge variation with the gate-to-source voltage. These reasons explain the significant drop in the capacitance at the higher flux density. As such, the unity-gain cut-off frequency attained is (4.08 GHz, 4.53 GHz, 5.89 GHz, and 606.7 GHz). Under UV illumination of 350 nm, the photovoltage is slightly enhanced (0.31 V, 0.49 V, and 0.61 V) as compared to that under visible light since Si possesses a larger absorption coefficient of 1×10^7 /m at the UV wavelength. The smaller active layer thickness (0.15 μm) enables significant number of holes to cross the junction before the generated carriers experience degradation due to the dependence on the distance. The series resistances experienced by the device are almost the same due to the slight difference between the photovoltages. The drain-to-source current is almost the same at the lower intensity (1 mA) whereas there is a considerable boosting of the current at the higher intensities. This is attributed to the large absorption coefficient of Si under UV light which augments the photoconductive charge in the depletion region. At these intensities, the large currents develop large voltage drops across the series resistances thus inducing negative sensitivity and eliminating the photovoltaic contribution. The currents so obtained are 7.9 mA and 16.1 mA. The lesser enhancement of current at the flux density of 10^{22} /m^2-s is due to the limitation of the drain-to-source conductance (at zero drain-to-source voltage) by the series resistances and the large saturation current due to the induced charge also limiting the factor η as discussed earlier thus, restricting the total drain-to-source current. As such, the responsivities attained are (3.53×10^8 A/W, 4.33×10^6 A/W, and 9×10^3 A/W). The responsivities at the flux densities of 10^{16} and 10^{19} /m^2-s are slightly lower than the visible range responsivities. This is because the device uses more power to detect UV light than the visible light since the power is directly proportional to the photon energy. Hence the visible/UV rejection or the contrast ratios are very low (1.5, 0.3, and 1.65).

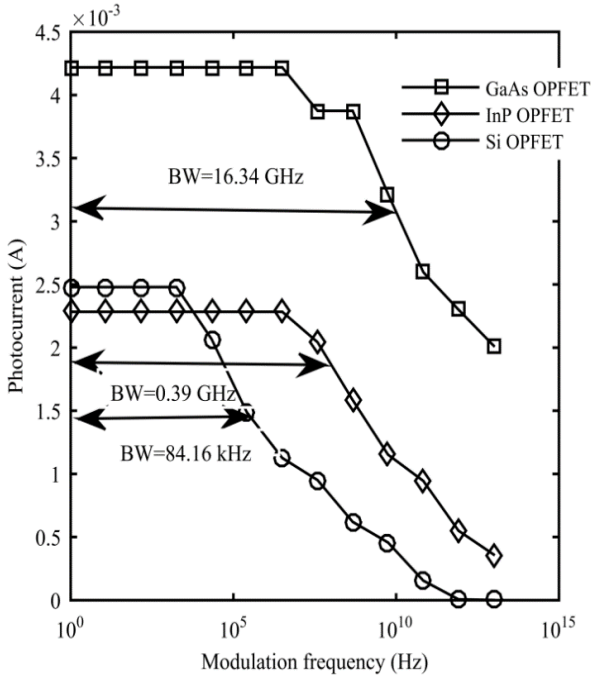


Figure 5 Frequency responses of the Si, InP, GaAs OPFET devices.

3.3 Analysis of InP OPFET at 3 μm gate length and 0.15 μm channel thickness

Discussing the InP OPFET with a gate length of 3 μm and an active layer thickness of 0.15 μm, the photovoltage is slightly enhanced (0.297 V, 0.476 V, and 0.65 V) as compared to the Si OPFET due to the small active layer thickness which eliminates the degradation of the generated carriers with distance as would occur in other cases. The series resistances are significantly lower than the Si OPFET but comparatively higher than that at higher gate length devices due to the reasons discussed earlier. The conducting photovoltaic charge is higher in the case of the InP OPFET in the presence of series resistances. This is because, the Si device possesses larger series resistances which incurs considerable voltage drop across these resistances thus, limiting the photovoltaic effect. Further, the photovoltage is slightly higher in the InP detector. This charge determines the saturation current. Hence, the saturation current is higher in the InP device. The conducting photovoltaic charge in the absence of series resistances i.e. at V_{DS}=0 V, is lower in the case of InP OPFET. The above emanates from the fact that at a fixed doping concentration and at a given gate-to-source voltage, the depletion width sensitivity to illumination is higher in both the lower barrier height and the lower dielectric constant device i.e. Si device. This charge determines the drain-to-source conductance at V_{DS}=0 which is higher in the Si OPFET however, is limited by the presence of series resistances which are larger in the case of Si detector. One more factor which differentiates the two devices is the electron mobility upon which the conductance depends upon and is substantially higher in the case of InP. The dark current is lower in the InP device due to the higher barrier height. The above factors result into slightly higher photovoltage-induced currents (1.3 mA, and 2.2 mA) in the InP OPFET than the Si OPFET. At the flux density of 10^{22} /m^2-s, the photoconductive charge from the depletion region alone contributes since the device exhibits negative sensitivity as mentioned earlier and the channel is totally depleted. The InP detector generates a current of 27.8 mA as opposed to 15.6 mA in the Si OPFET. This can be explained as follows: Since the absorption coefficient of Si is moderate (7.5×10^5 /m), there is almost uniform photogeneration throughout the fully depleted channel of 0.15 μm thickness. The absorption coefficient of InP is high (1×10^7 /m) due to which only a part of the depletion region carries significant number of carriers. Thus, the contributing charge is one order higher in the case of Si OPFET than the InP OPFET. This results in a large saturation current in the Si device. However, the conductance is lower in the Si device owing to lower electron mobility and larger series resistances. Since the saturation current appears in the numerator and as well as the denominator (through the hyperbolic tangent expression (see (1))), the large saturation current also limits the total drain-to-source current in the Si device. As a cumulative effect, the current is larger in the InP OPFET. These currents are the lowest among all the sets due to the reasons discussed earlier in the case of the Si OPFET. The 3-dB bandwidths achieved are (67.35 MHz, 0.26 GHz, and 3.84 MHz). These bandwidths are higher than that in Si OPFET owing to the shorter hole and electron lifetimes (0.3 μs) in InP as compared to 75 μs in Si. The boosting of the photovoltage as the intensity is increased from 10^{16} to 10^{19} /m^2-s also boosts the bandwidth from 67.35 MHz to 0.26 GHz. At the flux density is raised to 10^{22} /m^2-s, the bandwidth drops to 3.84 MHz due to the following: Under dc conditions or at lower frequencies, the photoconductive effect is significantly high and is the sole contributor. However, as the frequency is increased, the photoconductive contribution starts decreasing due to the dependence of the electron lifetime on frequency. This reduces the voltage drop across the series resistance thus, widening the effective channel width and increasing the photovoltaic contribution. Hence, at the juncture of the 3-dB point, both the effects effectively contribute causing the drop in the bandwidth. The responsivities attained are (9.97×10^8 A/W, 1.9×10^6 A/W, and 2.7×10^4 A/W) which are higher than that in the Si device due to larger photocurrents. The InP detector exhibits transconductances of (2.3 mS, 4.3 mS, 5.6 mS, and 6.7 mS). These values are higher than that for Si OPFET except at the flux density of 10^{22} /m^2-s. As explained earlier, the saturation currents, the drain-to-source conductances, and the mobility involved are higher in the InP device whereas the series resistances are lower in the InP device at the lower intensities. The depletion width sensitivity to the applied gate-to-source voltage with the other factors kept constant is higher in the lower barrier height and the lower dielectric constant device (Si OPFET). But, in practice, due to larger series resistances and the lower mobility of Si, the net effect is the larger sensitivity in the InP device. Further, the photovoltage is slightly larger in the InP device. Due to the above factors, the transconductances are higher in the InP OPFET at the lower flux densities. But at the intensity of 10^{22} /m^2-s, the channel is fully depleted. The depletion state does not affect the saturation current in terms of magnitude (it grows without bounds in spite of the presence of the series resistances). This is because, when the channel gets depleted (due to the large voltage drop across the series resistances), it is the photoconductivity in the depletion region which takes over role of the otherwise contributing photoconducting charge from the neutral channel region. Since the photoconductive charge in Si OPFET is very high and one order higher than that of the InP OPFET, the saturation current is considerably higher in the Si device. This current is insensitive to the applied gate-to-source voltage owing to the complete depletion of the channel. However, the drain-to-source conductance changes with the gate-to-source voltage since it is evaluated at V_{DS}=0 which means a finite undepleted channel width. This conductance is lower in magnitude as well as less sensitive to the change in the gate-to-source voltage in the Si detector due to the lower mobility and limiting effects of the series resistances. But, the effect of very large saturation current in the Si device supersedes the effect of lower conductance and its sensitivity to generate higher transconductance. The transconductance increases with the increase in the optical intensity due to the enhancement of the photovoltaic effect upto the flux density of 10^{19} /m^2-s and high contribution of photoconductive effect at the intensity of 10^{22} /m^2-s. On the whole, the transconductances obtained at this gate length are the lowest among all the sets due to the reasons stated earlier for the Si OPFET. The gate-to-source capacitances attained in the InP device are (0.034 pF, 0.0695 pF, 0.092 pF, and 0.016 pF). These values are lower than that of Si device. This is explained by the fact that at a constant or slightly varying photovoltage and a fixed doping concentration, the depletion width sensitivity to the applied gate-to-source voltage is larger for the lower barrier height (and dielectric constant) device i.e. Si OPFET. The capacitances increase with optical power and fall at the highest flux density and are lowest among all the sets due to the same reasons as described earlier in the case of Si device. The unity-gain cut-off frequencies

thus obtained are (10.76 GHz, 9.74 GHz, 9.65 GHz, and 65.4 GHz). Under UV illumination, the photovoltages developed are substantially lower (0.0943 V, 0.27 V, and 0.44 V) than that in the visible region owing to the very high absorption coefficient of InP in the UV region (7.6×10^7 /m).This generates significantly large number of carriers only upto a small distance from the absorption surface (i.e. the active layer-substrate interface). The generated carriers crossing the Schottky junction are considerably low due to the significant degradation of the carriers with distance (the active-layer thickness (0.15 μm) is much larger than the absorption depth (0.013 μm)). The drain-to-source currents obtained are (0.272 mA, 0.54 mA, 1.2 mA, and 30.1 mA). Due to the lesser photovoltages, the currents obtained are lower in the UV region upto the intensity of 10^{19} /m^2-s. At the flux density of 10^{22} /m^2-s, the current is higher (30.1 mA) owing to the substantially high contribution from the photoconductive effect in the depletion region emanating from the larger absorption coefficient of InP in the UV region. The photovoltaic effect plays no role at this intensity since the device exhibits negative sensitivity. The responsivities obtained are (1.53×10^8 A/W, 5.2×10^5 A/W, and 1.71×10^4 A/W) based on the above photocurrents and further assisted by the fact that the device uses more optical power in the UV region than the visible region. The visible/UV contrast ratios obtained are (6.516, 3.65, and 1.58).

3.4 Analysis of GaAs OPFET at 3 μm gate length and 0.15 μm channel thickness

Discussing the GaAs OPFET device possessing a gate length of 3 μm and an active-layer thickness of 0.15 μm, this device develops photovoltages of (0.41 V, 0.58 V, and 0.76 V) in accordance with [7] which are substantially higher than the Si and the InP devices. Under dark, the channel is totally depleted on account of a large Schottky barrier height of ITO-GaAs junction (0.98 eV). Thus, the current from the device channel ceases to exist but there is a small current flowing in the device due to the presence of the shunt resistance. The higher photovoltages ought to have produced larger photovoltaic charge responses as compared to the other detectors (InP OPFET). However, since the channel is deeply depleted (beyond the active-layer thickness), the developed photovoltages do not open the channel wide enough (of course, limited by the series resistances) as opposed to the InP device. The InP device is partially depleted under dark and the generated photovoltages (0.297 V, 0.476 V, and 0.65 V) are able to significantly reduce the depletion width thus yielding greater photovoltaic charges than the GaAs device. In spite of this, the drain-to-source currents (97.1 μA, 1.4 mA, 2.5 mA, and 54.9 mA) and the transconductances (0, 5.5 mS, 7.5 mS, and 67.2 mS) achieved in GaAs OPFET under illumination are higher than the InP device. This is attributed to the higher electron saturated velocity (1.2×10^5 m/s) of GaAs as compared to InP (7.6×10^4 m/s) which considerably increases the saturation current further assisted by slightly lower source series resistance in GaAs. Also, the slightly higher electron mobility of GaAs (0.5 m^2/(V.s)) as compared to InP (0.43m^2/(V.s)) and the lower source and drain series resistances in the GaAs device enhances its drain-to-source conductance even if the conductance in the absence of series resistances is almost the same in both the devices. As a result, the higher currents are attained. The larger transconductances emanate from the amplification of the charge due to the depletion width sensitivity to the applied gate-to-source voltage by the higher electron saturation velocity in GaAs producing larger change in the saturation current. Further, the amplification of the said charge under the absence of series resistance by the higher mobility of GaAs so also the lower series resistances results in larger variation in the drain-to-source conductance. This, on the whole, produces a larger change in the drain-to-source current with the applied gate-to-source voltage. Also, the photocurrents are slightly enhanced due to the lower dark current. At the flux density of 10^{22} /m^2-s, the photocurrent as well as the transconductance are considerably enhanced. This is owing to the boosting of the saturation current and the conductance due to the substantially high contribution from the photoconductive effect in the depletion region as a result of moderate absorption coefficient of GaAs and one order of magnitude higher lifetime (1 μs). The transconductance is much higher than the other two devices because of high saturation current and greater magnitude of the drain-to-source conductance as well as its higher sensitivity to the applied gate-to-source voltage. The gate-to-source capacitances under illumination (0.064 pF, 0.086 pF, and 0.018 pF), on the whole, are lower than the other devices due to the earlier stated fact that the sensitivity of depletion width to the change in the gate-to-source voltage is higher with devices bearing lower barrier height and lower dielectric constant when other factors are almost constant. The capacitance dependence upon the optical power is the same as analyzed with other devices. As such, the f_T values attained are (13.68 GHz, 13.79 GHz, and 596.5 GHz). The 3-dB bandwidths exhibited by the device are (1.5 GHz, 6.3 GHz, and 1.1 MHz) which are considerably higher than the other two detectors at the lower intensities owing to the significant photovoltaic effect and lower hole lifetime (10^{-8} s) with the photovoltage increasing with the optical power. However, at the higher intensity, the bandwidth falls due to the significant sole contribution from the photoconductive effect with a large electron lifetime of 1μs. The responsivities obtained are (1.24×10^9 A/W, 2.4×10^6 A/W, and 5.37×10^4 A/W) in line with [27] which are larger than the other two devices due to larger photocurrents. Under UV light, the response falls significantly (1.57×10^8 A/W, 6.4×10^5 A/W, and 2.56×10^4 A/W), due to the lower photovoltages (0.19 V, 0.37 V, and 0.55 V) owing to the larger absorption coefficient of GaAs (8×10^7 /m) which enables lesser hole density to traverse the Schottky junction. In this case, the significantly high generation takes place only at or near the absorption surface. Also, the larger photoconductive charge contribution at the high intensity of 10^{22} /m^2-s, emanating from more than one order of magnitude larger absorption coefficient of GaAs than that in the visible region produces large saturation current and large conductance wherein the large saturation current limits the total current. The currents obtained are as (97.1 μA, 0.37 mA, 1.2 mA, and 44.9 mA). The contrast ratios obtained are (7.9, 3.75, and 2.1).

3.5 Comparative Analysis of Si, InP, and GaAs OPFETs at 3 μm gate length and 0.27 μm channel thickness

As the active layer thickness is widened to 0.27 μm, keeping the gate length constant at 3 μm, the series resistances are significantly reduced owing to the wider thickness. However, they are still higher in the Si device as compared to the other two devices. The photovoltage remains almost the same in the Si and the GaAs OPFETs whereas it is slightly decreased in the InP OPFET explained by the possession of moderate absorption

Table 1: Comparative Studies of Si, InP, and GaAs OPFET detectors (1st set)

Si back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **1st set** (L=3 μm, Z=100 μm, t_{sm}=0.15 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**	**Trans-conductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	0.36 mA	------	1.4 mS	0.0546 pF	4.08 GHz	-------	0	0.36 mA	------
10^{16}	0.28 V	0.9 mA	5.3×10^8	2.4 mS	0.084 pF	4.53 GHz	0.16 MHz	0.31 V	1 mA	3.53×10^8
10^{19}	0.46 V	1.7 mA	1.32×10^6	3.9 mS	0.105 pF	5.89 GHz	0.136 MHz	0.49 V	7.9 mA	4.33×10^6
10^{22}	0.606 V	15.6 mA	1.49×10^4	29.1 mS	0.0076 pF	606.7 GHz	0.57 MHz	0.61 V	16.1 mA	9×10^3
InP back-illuminated OPFET with Au gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **1st set** (L=3 μm, Z=100 μm, t_{sm}=0.15 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**	**Trans-conductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	0.27 mA	------	2.3 mS	0.034 pF	10.7 GHz	-------	0	0.27 mA	------
10^{16}	0.297 V	1.3 mA	9.97×10^8	4.3 mS	0.0695 pF	9.74 GHz	67.35 MHz	0.094 V	0.54 mA	1.53×10^8
10^{19}	0.476 V	2.2 mA	1.9×10^6	5.6 mS	0.092 pF	9.65 GHz	0.26 GHz	0.27 V	1.2 mA	5.2×10^5
10^{22}	0.65 V	27.8 mA	2.7×10^4	6.7 mS	0.016 pF	65.4 GHz	3.84 MHz	0.44 V	30.1 mA	1.71×10^4
GaAs back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **1st set** (L=3 μm, Z=100 μm, t_{sm}=0.15 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**	**Trans-conductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	97.1 μA	------	0 mS	0 pF	0 GHz	-------	0	97.1 μA	------
10^{16}	0.41 V	1.4 mA	1.24×10^9	5.5 mS	0.064 pF	13.7 GHz	1.5 GHz	0.19 V	0.37 mA	1.57×10^8
10^{19}	0.58 V	2.5 mA	2.4×10^6	7.5 mS	0.086 pF	13.8 GHz	6.3 GHz	0.37 V	1.2 mA	6.4×10^5
10^{22}	0.76 V	54.9 mA	5.37×10^4	67.2 mS	0.018 pF	596 GHz	1.1 MHz	0.55 V	44.9 mA	2.56×10^4

coefficient by Si and GaAs which enables significantly high hole density to cross the Schottky junction even if the channel thickness is increased.

On the other hand, the high absorption coefficient of InP results in lesser hole density to traverse the junction when the channel is widened. With the increase in active-layer thickness, the drain-to-source currents and the photocurrents are substantially enhanced due to the increment in the conducting charge as well as the reduction in the series resistances, however, with some exceptions at higher intensity of 10^{22} /m^2-s. At this flux density, in InP OPFET, due to the low series resistance, and the moderate Schottky barrier height, the increase in the channel thickness leads to the partial depletion of the channel. Hence, both the neutral channel and depletion regions contribute to the photoconductive effect significantly. However, owing to the large absorption

coefficient of InP, the contribution from the depletion region is lesser as compared to that at the lower active-layer thickness since the absorption depth is small and the most of the generated carriers in the far end of the depletion region are negligible as compared to that in the near end. But, it is still larger than the neutral channel contribution attributed to the factor in the denominator of the electron density equation in the depletion region which is absent in the equation for electron density in the neutral channel region. This results in one order higher density in the case of depletion region considering the same thicknesses for their contribution (by inspection of the two equations). At this intensity, the photovoltaic effect also contributes to the response, due to partial depletion of the channel and the device does not exhibit negative sensitivity. However, in overall, the response falls as compared to that at the lower channel thickness resulting in a lower current of 16.8 mA. In the other two detectors, these phenomena do not occur as the series resistance is sufficiently large in Si OPFET and the barrier height is considerably high in the GaAs OPFET thus maintaining complete depletion of the channel thickness at this flux density. Further, due to the moderate absorption coefficient of Si and GaAs (absorption depth is large), the photoconductive effect from the whole of the depletion region contributes significantly, thus, boosting the photocurrent. The devices exhibit negative sensitivity as usual at this intensity, thus, eliminating the photovoltaic response. The transconductances are boosted as the channel thickness is increased owing to the larger currents involved and the lesser limitation of the depletion width sensitivity due to the lower series resistances. This produces greater change in the saturation current and the drain-to-source conductance, however, again with some exceptions. In the Si OPFET, at the flux density of 10^{22} /m^2-s, the transconductance falls to 27.1 mS from 29.1 mS which is attributed to the larger photoconductive effect contribution from the depletion region induced by the greater depletion area leading to substantially high enhancement of the saturation current which is independent of the gate-to-source voltage. In this case, the channel is totally depleted sowing to the large voltage drop across the series resistance. The large saturation current limits the total current, and the only significant variation with the gate-to-source voltage is the considerable drain-to-source conductance (being evaluated at V_{DS}=0), on the whole, causing a fall in the transconductance. Similarly, in the GaAs OPFET, analogous phenomena occur, but in this case the saturation current involved is much lower than in the case of Si OPFET. This is due to the lesser contribution from photoconductivity in the depletion region since the electron lifetime in Si is more than one order of magnitude higher than that in GaAs. Also, the significant conductance variation with gate-to-source voltage is limited by the series resistances causing a substantial drop in the transconductance. These phenomena do not occur in InP OPFET, since the channel is partially depleted with a significant contribution from the photovoltaic effect, and the photoconductive effect from the depletion and the neutral channel regions with lower series resistances (as discussed earlier). However, this contribution is comparatively lower than the photoconductivity alone in the case of Si and GaAs OPFETs. This results in the rise in the transconductance from its previous value at the lower active layer thickness but exhibiting smaller values compared to that in Si and GaAs OPFETs. One anomalous behavior is observed at the flux density of 10^{19} /m^2-s in the InP OPFET, wherein the transconductance falls to 4.9 mS from its previous value of 5.6 mS at the lower active layer thickness and is also lower than its value (5.3 mS) at the flux density of 10^{16} /m^2-s. The reason behind this is still under our own investigation. The gate-to-source capacitance values increase with the increase in the channel thickness owing to the fact that at the lower channel thickness, with the applied drain-to-source voltage of 10 V, significant channel area is depleted as compared to that with the higher channel thickness. This is because the applied voltage drops itself across the channel in the form of the channel voltage which varies nearly linearly from zero at the source end to V_{DS} at the drain end. Thus, the change in the space charge with the change in the gate-to-source voltage is greater in the structure possessing larger channel thickness leading to the enhancement of the gate-to-source capacitance. As such, with the obtained tranconductance and capacitance values, the Si OPFET experiences an overall rise in f_T whereas the InP and the GaAs devices perceive a drop in the f_T with the widening of the channel thickness. With the enhanced photocurrents, on the whole, all the detectors show increased responsivities with the increment in the channel thickness. The 3-dB bandwidths obtained using Si device with the active layer thickness of 0.27 μm are as (0.66 MHz, 84.16 kHz, and 0.57 MHz) as compared to its previous values at the lower channel thickness (0.16 MHz, 0.136 MHz, and 0.57 MHz). The boosting of the bandwidth at the lower flux density is owing to the increase in the photovoltaic effect whereas the fall at the flux density of 10^{19} /m^2-s is due to the rise in the photovoltaic effect but simultaneous augmentation of the photoconductive effect due to widening of the channel thickness, thus, causing the drop. The reason for the constant bandwidth at the intensity of 10^{22} /m^2-s is still under our own study. The frequency responses of the InP device are (32.9 MHz, 0.39 GHz, and 4.156 MHz) at the higher active layer thickness as opposed to (67.35 MHz, 0.26 GHz, and 3.84 MHz) at the lower channel thickness. At the intensity of 10^{19} /m^2-s, the rise in the bandwidth is due to the increase in the photovoltaic effect. The slight increase in the bandwidth at the flux density of 10^{22} /m^2-s can be discussed as follows: At the lower channel thickness, at dc or lower frequencies, the photoconductive effect alone contributes but as stated earlier at the juncture of the 3-dB point, both the photovoltaic and the photoconductive effects are equally contributing. This results in a bandwidth of 3.84 MHz. On the contrary, at the higher channel thickness, under dc or lower frequencies, both the effects contribute significantly but the photoconductive charge is higher. But as the frequency increases, the photovoltaic charge starts incrementing whereas the photoconductive charge starts decreasing. Thus, at the juncture of the 3-dB point, the photovoltaic charge has rose to significantly high value but the photoconductive charge has decreased significantly. Hence, the non-equally contributing charges results in a wider bandwidth (4.156 MHz).The reason behind the drop in the bandwidth at the optical intensity of 10^{16} /m^2-s is still being investigated (shows anomalous behavior being contrary to the increase in the photovoltaic effect). On the other hand, the bandwidths attained using GaAs OPFET are (12.87 GHz, 16.34 GHz, and 0.72 MHz) as compared to that at the lower channel thickness (1.5 GHz, 6.3 GHz, and 1.1 MHz). At the lower flux densities, the enhancement of the photovoltaic effect results in the rise in the bandwidth whereas at the higher intensity, the device shows anomalous behavior being contrary to the increase in the photoconductive effect alone. The UV responses of the Si OPFET

device (4.8 mA, 5.9 mA, 9.9 mA, and 28.1 mA) are almost the same as that in the visible region except at the flux density of 10^{19} /m^2-s (higher in the UV region) owing to the two orders higher absorption coefficient of Si in the UV region. This significantly boosts the photoconductive effect in the depletion region at this flux density in addition to the contribution from the photovoltaic effect thus increasing the response. The photovoltages developed are almost the same as that in the visible region and slightly lower than that at the lower active layer thickness attributed to the larger absorption coefficient of Si in the UV region. This signifies that the generated carriers undergo significant degradation before these traverse the Schottky junction after travelling through the active layer thickness of 0.27 µm starting from the absorption surface. Due to the almost same photovoltages as that in the visible region, the series resistances almost remain unaltered. At the flux density of 10^{16} /m^2-s, the response remains unchanged due to similar photovoltaic effect. At the intensity of 10^{22} /m^2-s, the insensitive response to the wavelengths under consideration is because although the photoconductive effect in the depletion region is considerably enhanced in the UV region owing to the larger absorption coefficient, this leads to a larger saturation current and the conductance; the total current limited by the saturation current. As such, the responsivities attained are lower at the flux densities of 10^{16} and 10^{22} /m^2-s due to the insensitive response with the use of higher photon energy in the UV region. The slightly higher responsivity at the intensity of 10^{19} /m^2-s is due to the enhanced photocurrent. The contrast ratios obtained are (1.45, 0.826, and 1.73) which are very low signifying that the Si detector can function well in both the visible and UV regions. In the InP OPFET, there is significant degradation of photovoltage at the UV wavelength (0.105 mV, 42.1 mV, and 0.2107 V) owing to the very large absorption coefficient of InP (7.6×10^7 /m) and the higher channel thickness (0.27 µm). Due to the low photovoltages, the series resistances are higher. The photoconductive responses are negligible compared to the photovoltaic currents due to the small absorption depth and generation area. The currents involved are (5.9 mA, 5.9 mA, 6.1 mA, and 6.9 mA) leading to responsivities of (0.0764 A/W, 1.02×10^5 A/W, and 5.4×10^2 A/W). The contrast ratios thus obtained are (1.75×10^{10}, 22 and 19.7). This shows that the InP device with the structure under consideration can detect visible light with high responsivity under background UV radiation with rejection ratio of 1.75×10^{10}at the intensity of 10^{16} /m^2-s. At the higher intensities, the ratios are not high but can suffice in certain instances of applications. In the GaAs OPFET, similar situation occurs under UV illumination since the absorption coefficient of GaAs (8×10^7 /m) is identical to that of InP (7.6×10^7 /m) signifying small absorption depth. With this, the photovoltages produced are (2.6 mV, 0.12 V, and 0.3 V), the currents induced are (7.5 mA, 7.5 mA, 8.2 mA, and 9.5 mA) and the responsivities registered are (48.88 A/W, 4×10^5 A/W, and 1.15×10^3 A/W) rejecting UV wavelength with the factors of (5.21×10^7, 10.25, and 68.7).Thus, the rejection ratio is inferior to the InP OPFET at the lower intensities and vice versa at the higher intensity.

3.6 Comparative Analysis of Si, InP, and GaAs OPFETs at 4 µm gate length and 0.3 µm channel thickness

When the gate length is elongated to 4 µm from 3 µm with a proportionate increase in the active layer thickness to 0.3 µm from 0.27 µm and a corresponding increment in gate width from 100 µm to 150 µm, the doping concentration is reduced to 4×10^{22} /m^3 from 5×10^{22} /m^3 considering constant gate length-doping concentration product from scaling rules [16]. These changes in the structural parameters have a significant change in the device parameters and some of them can be correlated based upon certain basic facts and phenomena. At first, the series resistances are reduced due to the larger channel thickness. Secondly, the photovoltages remain almost unaltered owing to the materials' absorption depths which are sufficient enough to cause the hole density crossing the junction to experience very less carrier decay with distance. The drain-to-source currents proportionally increase with the gate length ascribed to the corresponding increase in the gate width, the active layer thickness, the electron saturated velocity (velocity decreases with the doping concentration) and the lower series resistance which raises the value of the saturation current although the lower doping concentration induces less conducting channel charge under dark. The increase in electron mobility (mobility also decreases with the doping concentration), the gate width to the gate length ratio, and the lower drain and source series resistances boosts the drain-to-source conductance in spite of the lower doping concentration under dark. The longitudinal increment of gate length induces more conducting charge along the gate length which increases the current under dark. Under illumination, the photocurrents introduced also increase with the gate length: Since the photovoltages are maintained at almost the same values, it's a fact that at a fixed photovoltage and a constant gate-to-source voltage, the depletion width sensitivity to the applied illumination is higher at the lower doping concentrations i.e. the structure possessing a higher gate length. However, the reduced doping concentration induces less conducting charge which supersedes the depletion width sensitivity to give lesser photovoltaic charge. Nevertheless, due to the parameters and the factors mentioned under the dark condition in addition to the above-mentioned facts, the photocurrents are enhanced as compared to their values at the lower gate length.

At the flux density of 10^{22} /m^2-s, at the most at the intensity of 10^{19} /m^2-s in Si OPFET (due to high electron lifetime), the photoconductive effect significantly contributes to the response owing to the larger channel thickness in addition to the above stated factors to produce very large response. The transconductances also rise with gate length due to the fact that at a fixed photovoltage, the depletion width sensitivity to the applied gate-to-source voltage is larger at the lower doping concentrations (higher gate length). This is assisted and limited by the other factors as mentioned earlier for the drain-to-source current and the photocurrent enhancement. The significant boost of transconductance at the intensity of 10^{22} /m^2-s is due to the significant photoconductive effect (in Si and InP OPFET) wherein the saturation current is sufficiently high but at its optimum to produce significant change in the drain-to-source current with the applied gate-to-source voltage through the drain-to-source conductance. In GaAs OPFET, there is a fall in the transconductance owing to the large saturation current limiting the conductance change with the gate-to-source voltage. The gate-to-source capacitances are also incremented with the gate length on account of the fact that at a given photovoltage, the depletion width sensitivity to applied gate-to-source voltage is higher at the lower doping concentrations.

Table 2: Comparative Studies of Si, InP, and GaAs OPFET detectors (2nd set)

Si back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **2st set** (L=3 μm, Z=100 μm, t_{sm}=0.27 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	4.8 mA	------	3.7 mS	0.1 pF	5.72 GHz	-------	0	4.8 mA	------
10^{16}	0.28 V	5.8 mA	9.35×10^8	4.3 mS	0.12 pF	5.57 GHz	0.66 MHz	0.28 V	5.9 mA	6.43×10^8
10^{19}	0.46 V	7.3 mA	2.43×10^6	5.5 mS	0.14 pF	6.39 GHz	84.16 kHz	0.46 V	9.9 mA	2.94×10^6
10^{22}	0.606 V	28.3 mA	2.3×10^4	27.1 mS	0.0078 pF	554.3 GHz	0.57 MHz	0.59 V	28.1 mA	1.33×10^4
InP back-illuminated OPFET with Au gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **2nd set** (L=3 μm, Z=100 μm, t_{sm}=0.27 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	5.9 mA	------	4.5 mS	0.098 pF	7.23 GHz	-------	0	5.9 mA	------
10^{16}	0.266 V	7.3 mA	1.34×10^9	5.3 mS	0.11 pF	7.6 GHz	32.9 MHz	0.1 mV	5.9 mA	0.0764
10^{19}	0.44 V	8.2 mA	2.2×10^6	4.9 mS	0.13 pF	5.9 GHz	0.39 GHz	42.1 mV	6.1 mA	1.02×10^5
10^{22}	0.62 V	16.8 mA	1.064×10^4	7.6 mS	0.034 pF	35.7 GHz	4.15 MHz	0.21 V	6.9 mA	5.4×10^2
GaAs back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=10 V, λ=600 nm (visible) and 350 nm (UV) **2nd set** (L=3 μm, Z=100 μm, t_{sm}=0.27 μm, N_{dr}=5×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	7.5 mA	------	6.7 mS	0.088 pF	12.1 GHz	-------	0	7.5 mA	------
10^{16}	0.39 V	10.1 mA	2.55×10^9	7.6 mS	0.11 pF	10.9 GHz	12.87 GHz	2.6 mV	7.5 mA	48.9
10^{19}	0.57 V	11.7 mA	4.1×10^6	8.6 mS	0.13 pF	10.8 GHz	16.34 GHz	0.12 V	8.2 mA	4×10^5
10^{22}	0.75 V	88.1 mA	7.9×10^4	38.5 mS	0.018 pF	342 GHz	0.72 MHz	0.3 V	9.5 mA	1.15×10^3

This is further assisted by the greater gate width, gate length and active layer thickness, which offer larger transverse, longitudinal, and lateral space charges superseding the effect of lower doping concentration on the total charge. Owing to the enhancement of the transconductances and the gate-to-source capacitances, the unity-gain cut-off frequencies experience a rise with the gate length since the rate of transconductance increase is greater than that of the gate-to-source capacitance except in GaAs OPFET at the flux density of 10^{22} /m^2-s due to the fall in the transconductance. The responsivities decrease on the whole with gate length although the photocurrents are enhanced since the devices use greater optical power to produce a given photocurrent, optical power being directly proportional to the gate length and the gate width of the devices. However, in the case of Si and GaAs OPFETs, at the

intensity of 10^{22} /m^2-s the responsivity increases due to large photoconductive effects. The bandwidths exhibited by the devices are: In Si OPFET, with the elongation of the gate length, the bandwidths attained are (0.72 MHz, 44.5 kHz, and 0.41 MHz) as opposed to (0.66 MHz, 84.16 kHz, and 0.57 MHz) at the lower gate length. The slight rise in the bandwidth at the intensity of 10^{16} /m^2-s is attributed to the fact that the depletion width sensitivity increases at lower doping concentration (or higher gate length) which creates the roll-off slope of the photovoltaic charge or current frequency response curve to be gentler thus, broadening the response. The fall in the bandwidth at the flux density of 10^{19} /m^2-s in spite of the roll-off slope of the photovoltaic effect being gentler is owing to the increase in the photoconductive contribution due to larger channel thickness. The reason behind the drop in the bandwidth at the intensity of 10^{22} /m^2-s is still under study. On the other hand, in the InP OPFET, the bandwidths obtained are (92.6 MHz, 0.36 GHz, and 3.84 MHz) as compared to (32.9 MHz, 0.39 GHz, and 4.156 MHz) at the lower gate length. The same principles apply for the attained bandwidths at the flux densities of 10^{16} and 10^{19} /m^2-s, however, the slight fall in the bandwidth at the intensity of 10^{19}/m^2-s is due to the limitation by the series resistance. Whereas at the flux density of 10^{22} /m^2-s, there is slight drop in bandwidth owing to the fall in the photovoltaic charge response as compared to that at the lower gate length due to lower doping concentration but maintaining almost the same photoconductive contributions. This is owing to the contribution adjustment between the neutral channel and depletion regions signifying that at the juncture of the 3-dB point, both the effects contribute significantly resulting in the fall of bandwidth (3.84 MHz). The 3-dB bandwidths attained with the GaAs OPFET device are (4.57 GHz, 42.4 GHz, and 0.615 MHz) as compared to (12.87 GHz, 16.34 GHz, and 0.72 MHz) at the lower gate length. The reduction in the bandwidth at the lower intensity is contrary to the fact of gentler roll-off slope of photovoltaic charge frequency response (reason behind it is under study) whereas it is in line with the said fact at the intensity of 10^{19} /m^2-s. The slight drop in the bandwidth at the flux density of 10^{22}/m^2-s is contrary to the increment in photoconductive response with gate length and is under investigation. Under UV illumination, in the Si OPFET, the photovoltages are almost the same as that in the visible region since the increase in the absorption coefficient in the UV region by two orders of magnitude is neutralized by the lesser absorption depth maintaining the same values. The photocurrent is almost the same at the flux density of 10^{16} /m^2-s due to the same photovoltage. At the intensity of 10^{19} /m^2-s, the photocurrent is enhanced as compared to its value in the visible region owing to the substantial enhancement of the photoconductivity on account of the two orders of magnitude higher absorption coefficient in the UV region. At the flux density of 10^{22} /m^2-s, almost the same photocurrent is maintained although there is boost in the photoconductive effect due to the large saturation current which limits the total current. The responsivities obtained are (3.5×10^8 A/W, 2.34×10^6 A/W, and 2.4×10^4 A/W) with the visible/UV contrast ratios as (1.71, 0.8, and 1.74). This shows again that at this gate length, the Si device can detect both the visible and UV wavelengths with high sensitivities. The InP OPFET can reject very well the UV radiation and detect visible light with the rejection ratios as (Very High, Very High, and 28.556) at this gate length ascribed to the very high absorption coefficient of InP in the UV region which limits its absorption depth and the generation area. Also, due to the larger active layer thickness, the photovoltages generated are significantly degraded which produces zero response at the lower intensities whereas considerable response at the higher intensity due to somewhat moderate photovoltage owing to the higher optical power level. On the other hand, the GaAs OPFET exhibits rejection ratios of (Very High, 16.05, and 137.54). This can be explained as follows: Although the absorption coefficient of GaAs is almost the same as that of InP, the photovoltages generated in GaAs are slightly higher than that in InP owing to the larger barrier height. Hence, the GaAs device produces zero response (due to very low photovoltage) at the lower intensity whereas substantial responses (due to reasonable photovoltages) at the higher intensities which account for the obtained ratios.

3.7 Comparative Analysis of Si, InP, and GaAs OPFETs at 5 μm gate length and 0.4 μm channel thickness

As the gate is elongated to 5 μm from 4 μm, with a corresponding increase in the active layer thickness to 0.4 μm from 0.3 μm, and the gate width to 200 μm from 150 μm,thedoping concentration is reduced to 3.2×10^{22} /m^3. This tends to increase the depletion width sensitivity but is significantly limited by the voltage drop across the series resistances since the large channel thickness induces high conducting charge under dark which considerably raises the dark current. The series resistances are substantially lowered due to large active layer thickness. In the Si OPFET, the photovoltages remain almost unaltered due to the moderate absorption coefficient of Si. The currents induced in the Si OPFET are (11.9 mA, 13.8 mA, 17.5 mA, and 150.2 mA). The current increase with gate length is due to the reasons described earlier. It is the photovoltaic effect enhancement at the lower intensities whereas the photoconductive effect is boosted at the higher intensity due to larger channel thickness resulting in a larger current of 150.2 mA. The transconductances are considerably boosted (7 mS, 8.1 mS, 7.6 mS, and 100 mS) as compared to that at the lower gate length due to the principles stated earlier. The transconductance increases with illumination initially as usual, but falls at the flux density of 10^{19} /m^2-s due to the limitation by the voltage drop incurred across the series resistance owing to larger current involved (17.5 mA). At the intensity of 10^{22} /m^2-s, although the current is very large, it is due to the photoconductive effect which is insensitive to the voltage drop across the series resistance and can gain sensitivity through the change in the conductance with the applied gate-to-source voltage. The gate-to-source capacitance increases further with the gate length as usual due to the reasons explained earlier (0.18 pF, 0.216 pF, 0.23 pF, and 0.013 pF). These result in the f_T values of (6 GHz, 6 GHz, 5.2 GHz, and 1.42 THz). The Si detector shows bandwidths of (0.72 MHz, 30 kHz, and 0.41 MHz) at this gate length. The unchanged value at the lower flux density is owing to the limitation posed by the series resistance effect. The drop in the bandwidth at the intensity of 10^{19} /m^2-s is attributed to the fall in the photovoltaic charge response with gate length (due to the lower doping concentration) and the increase in photoconductivity from larger channel thickness. The reason behind the unaltered bandwidth of 0.41 MHz at the intensity of 10^{22} /m^2-s is still unknown. The responsivities attained at this gate length are (5.54×10^8 A/W, 1.63×10^6 A/W, and 4.07×10^4 A/W) which are lower than that at the previous gate length as the

Table 3: Comparative Studies of Si, InP, and GaAs OPFET detectors (3rd set)

Si back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=25 V, λ=600 nm (visible) and 350 nm (UV) 3rd set (L=4 μm, Z=150 μm, t_{sm}=0.3 μm, N_{dr}=4×10^{22}/m^3)										
	Visible							UV		
Photon Flux Density (/m^2-s)	Photo-voltage (V)	Drain-to-source current (A)	Responsi-vity (A/W)	Trans-con-ductance (S)	Gate-to-source capacitance (F)	f_T(Hz)	Band-width (Hz)	Photo-voltage (V)	Drain-to-source current (A)	Responsivity (A/W)
0	0	6.4 mA	------	4.6 mS	0.12 pF	6 GHz	-------	0	6.4 mA	------
10^{16}	0.286 V	7.7 mA	6×10^8	5.7 mS	0.145 pF	6.2 GHz	0.72 MHz	0.28 V	7.7 mA	3.5×10^8
10^{19}	0.46 V	10.3 mA	1.87×10^6	7.4 mS	0.16 pF	7.3 GHz	44.5 kHz	0.46 V	14.6 mA	2.34×10^6
10^{22}	0.607 V	91.6 mA	4.17×10^4	84.7 mS	0.011 pF	1.25 THz	0.41 MHz	0.58 V	90.2 mA	2.4×10^4
InP back-illuminated OPFET with Au gate (v_{gs}=0 V, V_{ds}=25 V, λ=600 nm (visible) and 350 nm (UV) 3rd set (L=4 μm, Z=150 μm, t_{sm}=0.3 μm, N_{dr}=4×10^{22}/m^3)										
	Visible							UV		
Photon Flux Density (/m^2-s)	Photo-voltage (V)	Drain-to-source current (A)	Responsi-vity (A/W)	Trans-con-ductance (S)	Gate-to-source capacitance (F)	f_T(Hz)	Band-width (Hz)	Photo-voltage (V)	Drain-to-source current (A)	Responsivity (A/W)
0	0	7.9 mA	------	5.7 mS	0.116 pF	7.73 GHz	-------	0	7.9 mA	------
10^{16}	0.26 V	9.4 mA	7.4×10^8	7 mS	0.13 pF	8.4 GHz	92.6 MHz	Very Less	7.9 mA	0
10^{19}	0.44 V	10.9 mA	1.47×10^6	7.9 mS	0.15 pF	8.2 GHz	0.36 GHz	10.7 mV	7.9 mA	0
10^{22}	0.61 V	24.2 mA	7.97×10^3	18.2 mS	0.044 pF	65.6 GHz	3.84 MHz	0.15 V	8.9 mA	279.1
GaAs back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=25 V, λ=600 nm (visible) and 350 nm (UV) 3rd set (L=4 μm, Z=150 μm, t_{sm}=0.3 μm, N_{dr}=4×10^{22}/m^3)										
	Visible							UV		
Photon Flux Density (/m^2-s)	Photo-voltage (V)	Drain-to-source current (A)	Responsi-vity (A/W)	Trans-con-ductance (S)	Gate-to-source capacitance (F)	f_T(Hz)	Band-width (Hz)	Photo-voltage (V)	Drain-to-source current (A)	Responsivity (A/W)
0	0	9.4 mA	------	8.3 mS	0.106 pF	12.3 GHz	-------	0	9.4 mA	------
10^{16}	0.398 V	13 mA	1.75×10^9	9.4 mS	0.13 pF	11.4 GHz	4.57 GHz	0.3 mV	9.4 mA	0
10^{19}	0.576 V	14.6 mA	2.6×10^6	10.9 mS	0.15 pF	11.5 GHz	42.4 GHz	66.2 mV	10 mA	1.62×10^5
10^{22}	0.755 V	180.5 mA	8.39×10^4	28.9 mS	0.027 pF	171 GHz	0.615 MHz	0.24 V	11.5 mA	6.1×10^2

device uses more illuminated area for achieving the said photoresponse. In the UV region, the Si OPFET device generates slightly lower photovoltages owing to the higher absorption coefficient which means a smaller absorption depth and due to the large channel thickness, slightly fewer holes make it to the Schottky junction. Due to slightly lower photovoltage, the drain-to-source current is slightly reduced as compared to that in the visible region (13.4 mA) at the flux density of 10^{16} /m^2-s. However, at the flux density of 10^{19} /m^2-s, the response is raised to 18.6 mA due to the significant photoconductive contribution from the depletion region emanating from larger absorption coefficient of Si. At the intensity of 10^{22} /m^2-s, the response falls to 143.7 mA since the photoconductivity is enhanced whereby a large saturation current limits the total current. The responsivities achieved are

(2.55×10^8 A/W, 1.14×10^6 A/W, and 2.26×10^4 A/W). As such, the contrast ratios obtained are (2.17, 1.43, and 1.8).

In the InP OPFET with a gate length of 5 μm, the currents are further enhanced at the lower intensities whereas at the higher intensity, the photocurrent drops owing to the limited sensitivity caused by the presence of large dark current. This, in addition to the photocurrent creates large voltage drop across the series resistance. Thus, with a finite channel width due to moderate barrier height of Au-InP junction and the high absorption coefficient of InP, the photoconductive effect contributes non-significantly as compared to the lower gate length. In this case, the contributions are divided between the neutral channel region and the depletion region, with depletion region still the major contributor. This induces lesser change in the photovoltaic charge as compared to that at the lower flux density. Hence, both the photovoltaic and photoconductive effects fall at this flux density at this gate length causing a drop in the photocurrent. The transconductance rises as opposed to its value at the lower gate length under dark explained by the reasons discussed earlier since the current is comparatively lower which does not significantly limit the sensitivity through the series resistance. However, as the flux density increases, significantly high drain-to-source currents are introduced which considerably limit the depletion width sensitivity due to the voltage drop across the series resistance. At the intensity of 10^{22} /m^2-s, the transconductance is nearly zero since the reduction in photovoltaic sensitivity as explained earlier is almost totally compensated by the rise in the photoconductive sensitivity with the applied gate-to-source voltage resulting in the zero transconductance. The capacitances as usual are raised above their values at the lower gate length. The limitation of the transconductance and the rise in the capacitance leads to the fall in the unity-gain cut-off frequencies (5.5 GHz, 5.6 GHz, 5.5 GHz, and 0.653 mHz). The bandwidths drop at the lower flux densities (62.2 MHz, and 0.33 GHz) owing to the series resistance limiting effects. At the higher flux density, the bandwidth is raised (6.2 MHz). This can be discussed as follows: At the lower gate length, at dc or lower frequencies, both the photoconductive and the photovoltaic effects contribute with significantly high contribution from the photoconductive effect in the depletion region (owing to the lower channel thickness). As the frequency is increased, the photovoltaic charge starts increasing since the otherwise limited sensitivity under large current is converted into non-limited sensitivity. In this case, the increment in frequency lessens the photoconductive charge in the depletion region due to the dependence of electron lifetime on frequency. This reduces the overall current and hence, the voltage drop across the series resistance. This signifies that at the juncture of the 3-dB point, both the photovoltaic and the photoconductive effects are equally contributing. Further, it is known that when two similar effects are contributing to the response with equal operating lifetimes, the net frequency response has a bandwidth which is lower than that of the individual ones. This results in a bandwidth of 3.84 MHz. On the contrary, at the gate length of 5 μm, under dc or lower frequencies, both the effects contribute significantly but here the photovoltaic effect contribution has increased and the photoconductive contribution has decreased (still photoconductive charge is higher). But as the frequency increases, as explained in the previous case, the photovoltaic charge starts incrementing whereas the photoconductive charge starts decreasing. Thus, at the juncture of the 3-dB point, the photovoltaic charge has rose to significantly high value but the photoconductive charge has decreased significantly. Hence, the non-equally contributing charges results in a wider bandwidth (6.2 MHz). The responsivities attained are (5.1×10^8 A/W, 9.3×10^5 A/W, and 3.5×10^3 A/W) which account for the generated photocurrents and the optical power used under visible light. Under UV illumination, the photovoltages are severely degraded and the photoconductive responses are negligible owing to the very large absorption coefficient of InP and the larger active layer thickness which produce zero responsivities. Thus, the visible/UV contrast ratios are very high. This shows that the InP OPFET with a gate length of 5 μm is a high sensitivity detector with a very high contrast between visible and UV wavelengths but at the expense of a larger dark current. Discussing the GaAs OPFET device with a gate length of 5 μm, this device exhibits the same photovoltages as that with other gate lengths owing to the moderate absorption coefficient of GaAs even if the channel thickness is increased to 0.4 μm. Due to the larger active layer thickness and the high photovoltages, the series resistances are substantially reduced. As usual with the elongation of the gate length, the drain-to-source currents and the photocurrents are significantly enhanced (17.5 mA, 21.5 mA, 24.2 mA, 275.4 mA), however, this time with the limitation by the series resistances owing to the larger currents involved. The transconductance drops under dark (7.9 mS) since the dark current is very high (17.5 mA) as compared to its value at the previous gate length (9.4 mA) thus, limiting the sensitivity. At the flux density of 10^{16} /m^2-s, the transconductance is enhanced (12.7 mS) due to significant photovoltaic effect. At the flux density of 10^{19} /m^2-s, the transconductance falls (10.6 mS) due to larger current involved (24.2 mA) thus inducing series resistance-limited effect. At the intensity of 10^{22} /m^2-s, the transconductance slightly rises (30 mS) due to the substantial photoconductive effect in the depletion region insensitive to the voltage drop across the series resistance, thus, inducing a large saturation current independent of the gate-to-source voltage also limiting the total current and the device gaining the sensitivity through the significant conductance change with the applied gate-to-source voltage. The capacitances are increased as usual (0.16 pF, 0.2 pF, 0.22 pF, and 0.036 pF). As such, the f_T values obtained are (7.6 GHz, 10.11 GHz, 7.65 GHz, and 133.1 GHz) which are lower than the values at the lower gate length. The 3-dB bandwidths attained are (19.15 GHz, 19.15 GHz, and 0.615 MHz). At the intensity of 10^{16} /m^2-s, the bandwidth enhancement is due to the larger sensitivity. At the flux density of 10^{19} /m^2-s, the drop in the bandwidth is due to the limitation by the series resistance. The unaltered bandwidth at the intensity of 10^{22} /m^2-s is owing to no significant change in the photoconductive charge. The responsivities achieved are (1.18×10^9 A/W, 2×10^6 A/W, and 7.59×10^4 A/W) which account for the generated photocurrents utilizing larger illuminated areas. Under UV light, the photovoltages are significantly degraded owing to the very large absorption coefficient of GaAs and wider channel thickness (0.4 μm). This produces zero photoresponses at the lower flux densities but at the higher flux density since the optical power level is high and on account of the very high absorption coefficient of GaAs, the photoconductive effect from the depletion region is large. This is because the photovoltage is very low and the high photoconductive current develops significant voltage drop across the series resistance thus, making the channel totally depleted and producing high response (4.42×10^4 A/W). Hence, the visible/UV

Table 4: Comparative Studies of Si, InP, and GaAs OPFET detectors (4thset)

Si back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=30 V, λ=600 nm (visible) and 350 nm (UV) **4th set** (L=5 μm, Z=200 μm, t_{sm}=0.4 μm, N_{dr}=3.2×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	11.9 mA	------	7 mS	0.18 pF	6 GHz	-------	0	11.9 mA	------
10^{16}	0.289 V	13.8 mA	5.54×10^8	8.1 mS	0.216 pF	6 GHz	0.72 MHz	0.26 V	13.4 mA	2.55×10^8
10^{19}	0.47 V	17.5 mA	1.63×10^6	7.6 mS	0.23 pF	5.2 GHz	30 kHz	0.449 V	18.6 mA	1.14×10^6
10^{22}	0.61 V	150.2 mA	4.07×10^4	100 mS	0.013 pF	1.42 THz	0.42 MHz	0.56 V	143.7 mA	2.26×10^4
InP back-illuminated OPFET with Au gate (v_{gs}=0 V, V_{ds}=30 V, λ=600 nm (visible) and 350 nm (UV) **4th set** (L=5 μm, Z=200 μm, t_{sm}=0.4 μm, N_{dr}=3.2×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	13.8 mA	------	6.2 mS	0.178 pF	5.5 GHz	-------	0	13.8 mA	------
10^{16}	0.24 V	15.5 mA	5.1×10^8	7 mS	0.197 pF	5.6 GHz	62.2 MHz	Very Less	13.8 mA	0
10^{19}	0.42 V	17 mA	9.3×10^5	7.6 mS	0.22 pF	5.5 GHz	0.33 GHz	Very Less	13.8 mA	0
10^{22}	0.596 V	25.7 mA	3.5×10^3	0 mS	0.106 pF	0.65 mHz	6.2 MHz	6 mV	13.8 mA	0
GaAs back-illuminated OPFET with ITO gate (v_{gs}=0 V, V_{ds}=30 V, λ=600 nm (visible) and 350 nm (UV) **4th set** (L=5 μm, Z=200 μm, t_{sm}=0.4 μm, N_{dr}=3.2×10^{22}/m^3)										
	Visible							**UV**		
Photon Flux Density (/m^2-s)	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsi-vity (A/W)**	**Trans-con-ductance (S)**	**Gate-to-source capacitance (F)**	**f_T(Hz)**	**Band-width (Hz)**	**Photo-voltage (V)**	**Drain-to-source current (A)**	**Responsivity (A/W)**
0	0	17.5 mA	------	7.9 mS	0.16 pF	7.6 GHz	-------	0	17.5 mA	------
10^{16}	0.39 V	21.5 mA	1.18×10^9	12.7 mS	0.2 pF	10.1 GHz	19.1 GHz	Very Less	17.5 mA	0
10^{19}	0.57 V	24.2 mA	2×10^6	10.6 mS	0.22 pF	7.65 GHz	19.1 GHz	0.13 mV	17.5 mA	0
10^{22}	0.75 V	275.4 mA	7.59×10^4	30 mS	0.036 pF	133 GHz	0.615 MHz	0.046 V	274.8 mA	4.42×10^4

rejection ratios attained are (Very High, Very High, and 1.72). Thus, the GaAs OPFET is comparable to the InP OPFET at the lower intensities in terms of contrast ratio whereas it is inferior to the InP OPFET at the higher intensity.

Conclusion

We studied in detail the visible light and the UV photoresponses of the three efficient materials-based (Si, InP, and GaAs) back-illuminated OPFET photodetectors for VLC and UV reflectance imaging applications i.e. high gain-bandwidth product and high visible/UV contrast applications respectively. The results were analyzed based on the photoconductive and the photovoltaic effects, the series resistance effects, scaling rules- induced effects, and channel length-variation effects. The GaAs OPFET emerges

as a high-performance photodetector with higher responsivities, higher 3-dB bandwidth in the gigahertz range, and high visible/UV rejection ratios. On the other hand, the InP OPFET device can compete or even surpass the GaAs OPFET device in terms of contrast ratio with high photoresponse but inferior to the GaAs device and bandwidth in the sub-gigahertz range. Whereas the Si OPFET device exhibits high responsivities inferior to the InP device at the lower intensity and ranging between that of the InP and the GaAs devices at the higher intensities but the bandwidth is in the sub-megahertz range and the contrast ratios are very low. All the devices show significant unity-gain cut-off frequencies. On the whole, the drain-to-source currents, the photocurrents, the transconductances, and the gate-to-source capacitances increase with the elongation of the gate length and the widening of the active layer thickness. The responsivities decrease with the gate length. The photovoltages remain almost unaltered. The series resistances reduce with the gate length and the active layer thickness. Very High visible/UV rejection ratios are observed at the higher gate lengths in the case of the InP and the GaAs OPFETs. The 3-dB bandwidth under photovoltaic conditions in normal circumstances increase with the gate length but in some instances is limited by the series resistance effects. Also, with the enhancement of the photoconductive effect with gate length, the bandwidth increases. However, in other circumstances, wherein the bandwidth depends upon both the photoconductive and the photovoltaic effects, and their relative contributions, the bandwidth can either increase or decrease with gate length. The unity-gain cut-off frequency has an arbitrary dependence upon the gate length based upon the relative contributions of the transconductance and the gate-to-source capacitance. The GaAs OPFET detector exhibits the highest detection-cum-amplification bandwidth of around 11 GHz using a gate length of 3 μm and an active layer thickness of 0.27 μm at the flux densities of 10^{16} and 10^{19} /m^2-s whereas it shows a bandwidth of 11.5 GHz with a gate length of 4 μm at the intensity of 10^{19} /m^2-s, thus, showing its potential as detector-cum-amplifier in Opto-Electronic Integrated Circuits (OEICs)-based high bandwidth VLC applications such as Li-Fi (Light-Fidelity Networks). At 3 μm gate length, it exhibits the highest responsivities of 2.55×10^9 A/W and 4.1×10^6 A/W at the intensities of 10^{16} and 10^{19} /m^2-s respectively. Further, the corresponding visible/UV contrast ratios are 5.21×10^7 and 10.25. The InP OPFET attains a modest detection-cum-amplification bandwidths of 92 MHz, 0.36 GHz and 3.84 MHz at the 4 μm gate length and at the flux densities of 10^{16}, 10^{19}, and 10^{22}/m^2-s respectively with the corresponding high responsivities as 7.4×10^8 A/W, 1.47×10^6 A/W, and 7.97×10^3 A/W and contrast ratios as (Very High, Very High, and 28.5). These values suffice for moderate bandwidth VLC applications viz. hospitals and healthcare, under-water communication, defence, and security. The Si OPFET device can operate in low data rate VLC applications such as general positioning, vehicle and transportation, and smart lighting, due to its low detection-cum-amplification bandwidths of 0.66 MHz, 84.16 kHz, and 0.57 MHz at the gate length of 3 μm and an active layer thickness of 0.27 μm at the flux densities of 10^{16}, 10^{19} and 10^{22} /m^2-srespectively, with the corresponding responsivities of 9.35×10^8 A/W, 2.43×10^6 A/W, and 2.3×10^4 A/W. The Si OPFET is not suitable for visible/UV contrast applications. These devices show good prospects for communication and contrast applications. The comprehensive analyses presented here which relate the effect of the structural and the material parameters on the device parameters through physics-based discussion will aid in conducting further research.

Table 5: Parameters employed in calculation.

Parameters used in calculation (Si)			
Parameter	**Name**	**Value**	**Unit**
μ	Low field electron mobility	(~ 0.09)	(m^2/V.s
Φ_B	Schottky Barrier Height (ITO-Si)	(~ 0.71)	(eV)
v_{vn}	Saturated electron velocity	~9.4×10^4	(m/s)
v_{vp}	Saturated hole velocity	~6.3×10^4	(m/s)
τ_p	Lifetime of holes	75×10^{-6}	(s)
τ_n	Lifetime of electrons	75×10^{-6}	(s)
ε	Permittivity	1.04×10^{-10}	(F/m)
α	Absorption Coefficient @ 600 nm	7.5×10^5	(/m)
α	Absorption Coefficient @ 350 nm	1×10^7	(/m)
Parameters used in calculation (InP)			
μ	Low field electron mobility	(~ 0.43)	(m^2/V.s
Φ_B	Schottky Barrier Height (Au-InP)	(~ 0.8)	(eV)
v_{vn}	Saturated electron velocity	~7.6×10^4	(m/s)
v_{vp}	Saturated hole velocity	~5.6×10^4	(m/s)
τ_p	Lifetime of holes	0.3×10^{-6}	(s)
τ_n	Lifetime of electrons	0.3×10^{-6}	(s)
ε	Permittivity	1.11×10^{-10}	(F/m)
α	Absorption Coefficient @ 600 nm	1×10^7	(/m)
α	Absorption Coefficient @ 350 nm	7.6×10^7	(/m)
Parameters used in calculation (GaAs)			
μ	Low field electron mobility	(~ 0.5)	(m^2/V.s
Φ_B	Schottky Barrier Height (ITO-GaAs)	(~ 0.98)	(eV)
v_{vn}	Saturated electron velocity	~1.2×10^5	(m/s)
v_{vp}	Saturated hole velocity	~9×10^4	(m/s)
τ_p	Lifetime of holes	1×10^{-8}	(s)
τ_n	Lifetime of electrons	1×10^{-6}	(s)

ε	Permittivity	1.14×10^{-10}	(F/m)
α	Absorption Coefficient @ 600 nm	4×10^{6}	(/m)
α	Absorption Coefficient @ 350 nm	8×10^{7}	(/m)

Table 6: Performance comparison with state-of-art photodetectors and amplifiers.

Ref. No.	Year	Detector	Optical Power Density	Responsivity (A/W)	Bandwidth	Amplifier Type and f_T (Hz)
[17]	2004	PIN	NP (Not Provided)	0.36 A/W @ 660 nm	1.35 Hz	TIA (25 GHz)
[18]	2010	Photodiode (PD)	NP	0.21 A/W @ 660 nm	70 MHz	TIA (NP)
[19]	2012	PD	NP	0.36 A/W @ 670 nm	60 MHz	TIA (4.5 GHz)
[20]	2012	PIN	NP	0.52 A/W @ 660 nm	720 MHz	TIA (0.622 GHz)
[21]	2013	PIN	NP	0.54 A/W @ 675 nm	995 MHz	TIA (NP)
[22]	2014	APD	0.147 W/cm^2	27.9 A/W @ 670 nm	700 MHz	NP
[23]	2016	APD	0.016 W/cm^2	30 A/W @ 670nm	850 MHz	TIA (0.731 GHz)
[24]	2016	APD	NP	0.28 A/W @ 450 nm	650 MHz	NP
[25]	2017	APD	0.75 mW/cm^2	14.2 A/W @ 675 nm	960 MHz	TIA (NP)
[26]	2017	APD	NP	2.61 A/W @ 450 nm	>90 MHz	TIA (NP)
This Work	2019	**OPFET**	**0.33 mW/cm^2**	**4.1×10^{6} A/W @ 600 nm**	**16.34 GHz**	**OPFET (10.8 GHz)**

References

[1] J. V. Gaitonde, and R. B. Lohani, "Back-illuminated GaAs OPFET: A High Visible/UV Contrast Photodetector," In 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), IEEE, Namakkal, Tamil Nadu, India, 2018.https://doi.org/10.1109/ICEDSS.2018.8544284

[2] L. U. Khan, "Visible light communication: Applications, architecture, standardization and research challenges," Digital Communications and Networks 3(2), 78-88, 2017. https://doi.org/10.1016/j.dcan.2016.07.004

[3] C. Baack, G. Elze, and G. Walf, "GaAs MESFET: A high-speed optical detector," Electronics Letters, 13(7), 193-193, 1977.https://doi.org/10.1049/el:19770141

[4] A. A. A. De Salles, "Optical control of GaAs MESFET's," IEEE Trans. on Microwave Theory and Techniques, 31(10), 812-820, 1983.https://doi.org/10.1109/TMTT.1983.1131611

[5] H. Mizuno, "Microwave characteristics of an optically controlled GaAs MESFET," IEEE Trans.on Microwave Theory and Techniques, 31(7), 596-600, 1983.https://doi.org/10.1109/TMTT.1983.1131551

[6] S. R. Saxena, R. B. Lohani, R. U. Khan, and B. B. Pal, "Generalized dc model of GaAs optical field effect transistor considering ion-implanted profile," Optical Engineering, 37(4), 1343-1352, 1998..https://doi.org/10.1117/1.601895

[7] P. Chakrabarti, S. K. Shrestha, A. Srivastava, and D. Saxena, "Switching characteristics of an optically controlled GaAs-MESFET," IEEE Trans. on Microwave Theory and Techniques, 42(3), 365-375, 1994.https://doi.org/10.1109/22.277428

[8] N. S. Roy and B. B. Pal, "Frequency-dependent OPFET characteristics with improved absorption under back illumination," Journal of lightwave technology, 18(4), 604-613, 2000.https://doi.org/10.1109/50.838136

[9] J. Gaitonde and R. B. Lohani, "GaAs Optical Field Effect Transistor (OPFET): A High Performance Photodetector for Automotive Applications," SAE International Journal of Passenger Cars-Electronic and Electrical Systems, 9(2016-01-0094), 204-211, 2016.https://doi.org/10.4271/2016-01-0094.

[10] N. S. Roy, B. B. Pal, and R. U. Khan, "Frequency-dependent characteristics of an ion-implanted GaAs MESFET with opaque gate under illumination," Journal of Lightwave Technology, 18(2), 221-229, 2000. https://doi.org/10.1109/50.822796

[11] M. K. Verma and B. B. Pal, "Analysis of buried gate MESFET under dark and illumination," IEEE Trans. on Electron Devices, 48(9), 2138-2142, 2001.https://doi.org/10.1109/16.944207

[12] R. Gautam, M. Saxena, R. S. Gupta, and M. Gupta, "High Sensitivity Photodetector Using Si/Ge/GaAs Metal Semiconductor Field Effect Transistor (MESFET)," in AIP Conference Proceedings, Optics: Pheneomenon, Materials, Devices, and Characetrization, 1391(1), 232, 2011.https://doi.org/10.1063/1.3646835

[13] A. Singh, A. Kumar, M. Gupta, and R. S. Gupta, "Model for dc and rf characteristics of optically biased GaN metal semiconductor field effect transistor for electronic/optoelectronic microwave applications," Optical Engineering, 41(11), 2915-2922, 2002.https://doi.org/10.1117/1.1512909

[14] T. Chen and M. S. Shur, "Analytical models of ion-implanted GaAs FET's," IEEE Trans. on Electron Devices, 32(1), 18-28, 1985.https://doi.org/10.1109/T-ED.1985.21903

[15] Y. H. Byun, M. S. Shur, A. Peczalski, and F. L. Schuermeyer, "Gate-voltage dependence of source and drain series resistances and effective gate length in GaAs MESFETs," IEEE Transactions on Electron Devices, 35(8), 1241-1246, 1988.https://doi.org/10.1109/16.2543

[16] J. M. Golio, "Ultimate scaling limits for high-frequency GaAs MESFETs," IEEE Transactions on Electron Devices, 35(7), 839-848, 1988. https://doi.org/10.1109/16.3334

[17] R. Swoboda, and H. Zimmermann, "2.5 Gbit/s silicon receiver OEIC with large diameter photodiode," Electronics Letters, 40(8), 505-507, 2004. https://doi.org/ 10.1049/el:20040346

[18] F. Tavernier, M. Steyaert, "A high-speed POF receiver with 1 mm integrated photodiode in 180 nm CMOS," In Optical Communication (ECOC), 2010 36th European Conference and Exhibition on, IEEE, 2010, Torino, Italy. https://doi.org/10.1109/ECOC.2010.5621473

[19] Y. Dong, and K. W. Martin, "A high-speed fully-integrated POF receiver with large-area photo detectors in 65 nm CMOS," IEEE Journal of Solid-State Circuits, 47(9), 2080-2092, 2012.https://doi.org/10.1109/JSSC.2012.2200529

[20] M. Atef, R. Swoboda, and H. Zimmermann, "1.25 Gbit/s over 50 m step-index plastic optical fiber using a fully integrated optical receiver with an integrated equalizer," Journal of lightwave technology, 30(1), 188-122, 2012. https://doi.org/10.1109/JLT.2011.2179520

[21] P. Brandl, and H. Zimmermann, "3 Gbit/s optical receiver IC with high sensitivity and large integrated pin photodiode," Electronics Letters, 49(8), 552-554, 2013. https://doi.org/10.1049/el.2013.0558

[22] W. Gaberl, K. Hornstein, R. Enne, and B. Steindl, "Avalanche photodiode with high responsivity in 0.35 μm CMOS," Optical Engineering, 53(4), 043105, 2014. https://doi.org/10.1117/1.OE.53.4.043105

[23] P. Brandl, T. Jukić, R. Enne, K. Hornstein, H. Zimmermann, "Optical wireless APD receiver with high background-light immunity for increased communication distances," IEEE Journal of Solid-State Circuits, 51(7), 1663-1673, 2016. https://doi.org/10.1109/JSSC.2016.2557815

[24] X. Li, N. Bamiedakis, X. Guo, J. McKendry, E. Xie, R. Ferreira, E. Gu,M. D. Dawson, R. V. Penty, and I. White, "Wireless visible light communications employing feed-forward pre-equalization and PAM-4 modulation," Journal of Lightwave Technology, 34(8), 2049-2055, 2016.https://doi.org/10.1109/JLT.2016.2520503

[25] D. Milovancev, T. Jukić, B. Steindl, P. Brandl, and H. Zimmermann, "Optical wireless communication using a fully integrated 400 um diameter APD receiver," The Journal of Engineering, 2017(8), 506-511, 2017. https://doi.org/10.1049/joe.2017.0247

[26] S. Rajbhandari, A. Jalajakumari, H. Chun, G. Faulkner, K. Cameron,R.Henderson, D. Tsonev, H. Haas, E. Xie, J. J. McKendry, and J. Herrnsdorf, "Multigigabit per Second Integrated Multiple-Input Multiple-Output VLC Demonstrator," Journal of Lightwave Technology, 35(20), 4358-4365, 2017. https://doi.org/10.1109/JLT.2017.2694486

[27] J. Luo, A. Grot and D. Psaltis, "Optical FET receivers for neural network applications," In Proc. Optoelectronic Interconnects III (Vol. 2400, pp. 355-362). International Society for Optics and Photonics, Photonics West '95, 1995, San Jose, CA, United States. https://doi.org/10.1117/12.206325

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 522-528 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

The Fusing Framework Between Lifting Carrier and Tractor-Trailer for Modern Transportation

Ha Quang Thinh Ngo[1], Thanh Phuong Nguyen[2,*], Hung Nguyen[2]

[1]*Department of Mechatronics Engineering, HCMC University of Technology, Vietnam National University-Ho Chi Minh City (VNU-HCM), Vietnam*

[2]*Institute of Engineering, HCMC University of Technology (HUTECH), Vietnam*

ARTICLE INFO

Article history:
Received: 12 June, 2019
Accepted: 06 August, 2019
Online: 25 August, 2019

Keywords:
Robotics
Motion Control
Industrial Automation

ABSTRACT

The integrating model becomes one of key trends in engineering design. The more the product has functional specifications, the more the industrial consumer apply in many fields. In this paper, a fusion design of architecture platform for autonomous system is demonstrated. It is greater due to combining the specifications of various types in commercial market such as driving structure, multi-functional model or autonomous level. Therefore, plentiful advantages of systems are converged into one platform. From customer's requirements, the system parameters and specifications are offered to consult during the process of design. To compare with previous works, the enhancement of tractor-trailer, lifting-carrier and auto-feeder using differential driving mechanism is merged successfully. The detailed design of prototype involves the development of hardware components and infrastructure. Later, the modeling of autonomous system is simulated on computer to meet the employed conditions. The force analyzing scheme helps to predict the working ability to complete their mission. To verify the proposed design, an experimental version of this system is tested in different cases. In mode of lifting carrier, the vehicle and cargo that become a rigid body, follow the reference trajectory in practical map. In another circumstance, vehicle plays a role as leader which take along with cargo, as follower. From these results, it can be seen that the proposed design is feasible, effective and capable in real world. In future works, a fleet of autonomous vehicles mixing multi-modes in the same map should be considered.

1. Introduction

In the worldwide era, a large-scale system becomes more popular due to its huge advantages. Especially, in online shopping business, it needs a thousand of employers who work in distribution center or warehouse every day. Reducing the labor cost means that enormous profits are earnt. Therefore, the producer would always discover the methods to lessen the operating payment. One of common ideas is to use autonomous vehicle that carry goods to human operators [1, 2]. Various challenges had to be overcome in order to make this viable system, from design of robust autonomous vehicles, real-time communication, the coordination of vehicle and various control algorithms which permit the system to adapt and reconfigure itself leaning on the environment and working conditions. Generally speaking, grounded automated vehicle (GAV) is digested into several sub-categories. Authors in [3, 4] posed the lifting type of vehicle to elevate cargo in production site. The veridical product has three wheels including one driven wheel and two driving wheels. Two driving wheels of GAV have independent brush motors while the other is passive wheel. This model is very widespread nowadays since its movement is free and high loading capability.

The second one is forklift type with steering mechanism [5, 6]. This kind of vehicle often appears in port transportation, harbor storehouse or container yard. The main benefits are to provide heavy loading ability and powerful lifting mechanism. In most of its applications, it requires human to sit on. The diesel engine version is cheaper than one using electric engine. However, because its dimension is large, it needs more space to operate than others. In the other style of GAV, i.e. tractor-trailer transporter [7, 8], the long combination vehicle is gradually becoming more public in certain circumstances because of vantages related to decrease costs for material transportation and diminish fuel

* Thanh Phuong Nguyen, Email: nt.phuong@hutech.edu.vn

consumption. This research implements active trailer steering control to improve the maneuverability of long truck-trailer combinations. The problem of swept path width during cornering could be solved by nonlinear control strategies. The effectiveness of proposed approach is evidenced on experimental platform.

In reality, there is a need to integrate these above vehicles into a compact unit. The emerging trend helps to overcome the existing limitations such as particular control problems of model type, enhancing carrier, reducing shared workspace, improving low speed maneuverability and so on. Besides, the saving cost on buying an integrated platform is rather than separated ones. In this paper, we present a fusion design of automated vehicle in multi-applications. After the problem statement is given in Section II, the novel idea based on combining between lifting carrier and truck-trailer model is to provide many expandable developments. The study of proposed pattern that is demonstrated in Section III enables the modern mechatronics system design. Section IV performs the actualizing works in each type whenever vehicle adapts to different scenarios. Later, the verification of proposed design is proved through experimental tests of Section V. Finally, some conclusions about this approach are commented for further development in Section VI.

2. Problem Statement

GAV or autonomous robot which is energized by electric power, play the role as a potential means of transportation in future. Due to great rewards in logistics, e-commerce, factory automation, social service and in some situations, public carriage, it has attracted the attention of industry and academic. There are numerous forms of classifications in GAV, such as wheeled structure, driving mechanism or their purposes. Referring to four wheels and driving by front-side, authors [11] investigated the control scheme to realize the yaw moment in presence of unknown mismatched disturbances. An integral sliding mode control was embedded into differential drive assistance steering (DDAS) to guarantee the transient control performance. In the case of complete failure of active front-wheel steering system, a robust composite nonlinear feedback strategy [12] to achieve path following was studied in detail. Besides, by utilizing fault-tolerant control with DDAS, an innovative multiple-disturbances observer-based method was applied in considering the tire force saturations. With a variable speed integral PID controller and a coordinated control system based on adaptive weighting dependent vehicle speed [13], both the ideal steering wheel torque and expected yaw rate could be obtained. However, they are hard to employ in commercial market owing to unreachable real-time performance, especially in emergency stop case. Via driving mechanism in rear-side, some car prototypes have been commercialized effectively. The control problem seems to be more difficult since the leading angle is adjusted lately. Furthermore, the head of car is light weight. It tends to be unstable when running at high velocity. To be more flexible, four wheels of electric vehicle have been driven. In [14], the wheel driving torque on each motor could be modulated precisely and continuously to attain not only maneuverability but also energy-saving mode simultaneously. The distribution strategy of torque control using multi-objective optimization was categorized into high layer and low layer. Although the developed method was simulated, the effectiveness of this approach has not verified in experimental platform yet.

In the other configuration, the three wheels framework becomes one of key topics to develop. In [15], to decline the error dynamic stabilization problem, a nonlinear sliding mode control law is employed to surpass. The control organization is uniformly asymptotically stable if unknown disturbances and modeling uncertainties are bounded. It utilized the nonholonomic constraints and transitional trajectory to acquire feasible state trajectories. Though, in practical context, the mass changes incessantly and it is hard to measure inertial moment of whole system (vehicle and load) precisely. The most concern of two-wheeled configuration is balancing control algorithm. The reporters [16] analyzed a passive inverted pendulum model as an investigational device. The sampling effect of the digital control is simulated as a zero-order hold. The stabilization of the upright position is possible by choosing proper control parameters as function of the sampling period of the controller. There are many limitations such that nonlinear characteristics (friction, behavior of electro-mechanical) are not mentioned while they still occur in model. Also, the system coefficients are not prearranged exactly. Sometimes, the omni-directional mobile robot including four mecanum wheels was installed in manufacturing fields. The authors [17] introduced the design and development of omni-based robot orienting to intelligent factory. The upper sections are named as control level and application level while the lower one is chassis frame. Each modularized wheel was located on a vertical suspension mechanism to ensure the moving stability and keep the distances of four wheels invariable. In the unknown semi-structured indoor environment, the data from a Kinect visual sensor and four-wheel encoders were fused to localize the mobile robot using an extended Kalman filter with specific processing. The existing difficulties in the model are to eliminate the slipping phenomenon on mecanum wheel and restrict loading competence.

The other types of GAV are forklift truck and tractor-trailer vehicle. The prototype forklift is regularly applied in container yard, pallet-based storehouse or cold storage. The structure of forklift [18-20] entails of a single active rear wheel and two non-driven front wheels balanced in electric stacker. Around the automatic forklift, several sensing modules, for instance, positioning encoder, laser scanner or GPS-indoor are often attached to collect environmental information. The drawbacks of forklift robot are too dangerous to actuate in medium or small-scale area, high maintenance cost and inflexible motion. In the coordinated control approach [21, 22], a tractor-trailer vehicle could be ensured to track trajectory and maintain vehicle kinematics restriction and dynamics maneuvers simultaneously. In the level of kinematics, linear quadratic regulator and model predictive control were used to assess the posture controller separately. For dynamics analysis, sliding mode control and global terminal sliding mode control were inspected to design the dynamic controller for the tracking of the desired velocities generated online. Nevertheless, the viability and application of this method are still questioned.

As a result, the incorporated requirements become an inevitable trend in our era. The contributions in this paper are, (a) analyzing the fusing hardware of models between tractor-trailer form and lifting carrier form, (b) build up an experimental platform of GAV based on the analysis design positively and enforce it safely on working map.

3. Analysis of Fusing Model

Based on the new trend of fusion, multi-function should be merged various roles into one platform. For instance, the autonomous vehicle is able to elevate cargo in vertical direction and pull freight horizontally. The vertical lifting vehicle modeling is illustrated as Fig. 1. It is assumed that vehicle and freight become a rigid body which has total weight m_{vl} .

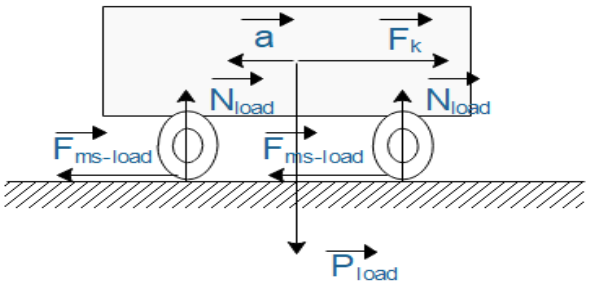


Figure 1: Forces act on lifting type model.

$\vec{a}$: acceleration of rigid body

$\vec{P}_{load}$: gravity force of rigid body

$\vec{N}_{load}$: reaction force on plane due to Newton's law

$\vec{F}_{ms-load}$: friction force between wheel and plane

$\vec{F}_k$: driving force from motors

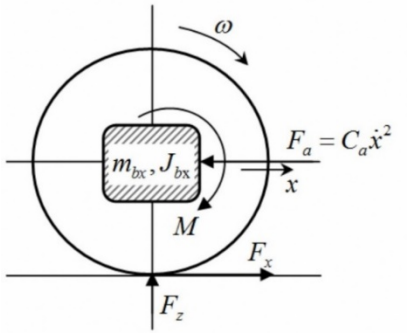


Figure 2: Analysis of driving wheel actuator.

According to second law of Newton, we have

$$\sum \vec{F} = m\vec{a} \Leftrightarrow m_{vl}\vec{a} = \vec{F}_k + \vec{F}_{ms-load} \quad (1)$$

$$F_k = m_{vl}a + F_{ms-load} \quad (2)$$

$$F_k = m_{vl}a + P_{load}\mu_{ms-load} \quad (3)$$

The force F_k is estimated as 17,405 N approximately. It is considered that vehicle moves on high stiffness plane as Fig. 2. If there is no slipping phenomenon, then, the constraints of design parameters to avoid this problem must be met below condition.

$$P_c \le P_k \le P_\varphi \quad (4)$$

where $P_c = F_k + P_f + P_i + P_\omega + P_j$

The force P_f helps maintaining driving wheel and plane in contact.

$$P_f = \mu' N = \mu' m_{vl} g \quad (5)$$

The inertial moment of driving wheel J_{wheel} is computed as.

$$J_{wheel} = \frac{1}{2} m_{wheel} 2r_{wheel}^2 \quad (6)$$

The symbol r_{wheel} is radius of driving wheel. Therefore,

$$P_j = \left(m_{vl} + \frac{J_{wheel}}{r_{wheel}^2} \right) a \quad (7)$$

Due to the movement on flat, the drag force P_i could be cancelled. The volume of vehicle is small, hence, the air resistance force P_ω is ignored.

By choosing adhesion coefficient $\varphi = 0.7$, the force P_φ is determined as.

$$P_\varphi = \varphi m_{vl} g \quad (8)$$

A relationship between driving force and moment of driving wheel is presented as.

$$P_k = \frac{M_1}{r_{wheel}} \quad (9)$$

As a result, the moment of driving motor M_1 , which is essential to control the system, should be complied with following constraint.

$$2,71 \le M_1 \le 24,89 \quad (10)$$

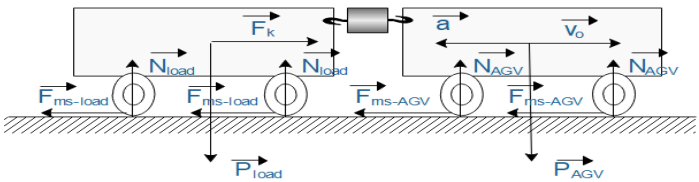


Figure 3: Forces act on tractor-trailer type model.

To certify the operation of truck-trailer type, an analysis of this modeling is shown as Fig. 3. The leader and follower are connected via a hook. The first vehicle is active drive while the other is passive one. In this case, we concern on interacting forces among them and angular driving wheels.

Theoretically speaking, the leader vehicle obligates to provide larger traction force to drive a chain of followers. This additional force is deliberated as $F_k^{'}$. Thus, the new force F_k^{new} that is necessary to drive whole system is denoted.

$$F_k^{new} = F_k + F_k^{'} \tag{11}$$

Consider that it pulls the followers not greater than its weight. Adding the safe ratio as our experiences, the new force F_k^{new} could be rewritten as.

$$F_k^{new} = 2,2F_k \tag{12}$$

Besides, the factors (material, structure or inside components) of driving wheel would affect on adhesion coefficient ϕ. The estimated moment of driving motor M_2 is renewed as.

$$4,32 \le M_2 \le 38,73 \tag{13}$$

4. Fusion Design in Proposed Hardware

In this section, the detailed design of proposed model is mentioned. Fig. 4 shows the inside frame of mechanical parts in vehicle. Generally speaking, the architecture of system is classified into three main layers (base, middle and lifting layer) and one intermediate layer. The base layer is the lowest one and stiff enough to bear the entire mass of whole system. The driving mechanism connoting DC motors with gears, drivers, soft couplings, battery, electric cylinder and so on, is located on base layer. Four rolling actuators that are hang on middle layer play a role as feeding actuator automatically. One electric cylinder is situated at center of the middle layer to upraise goods. Eventually, the lifting layer is at top. The electromagnetic lock, dwelling in this layer, is exploited to pull load horizontally. The intermediate one is a framework for holding and suspending among layers.

1-Driving wheels

2-Bearing

3-Soft coupling

4-Jig of motor

5-DC motor

6-Caster wheel

7-Loadcell

8-Shaft of intermediate layer

9-Battery tray

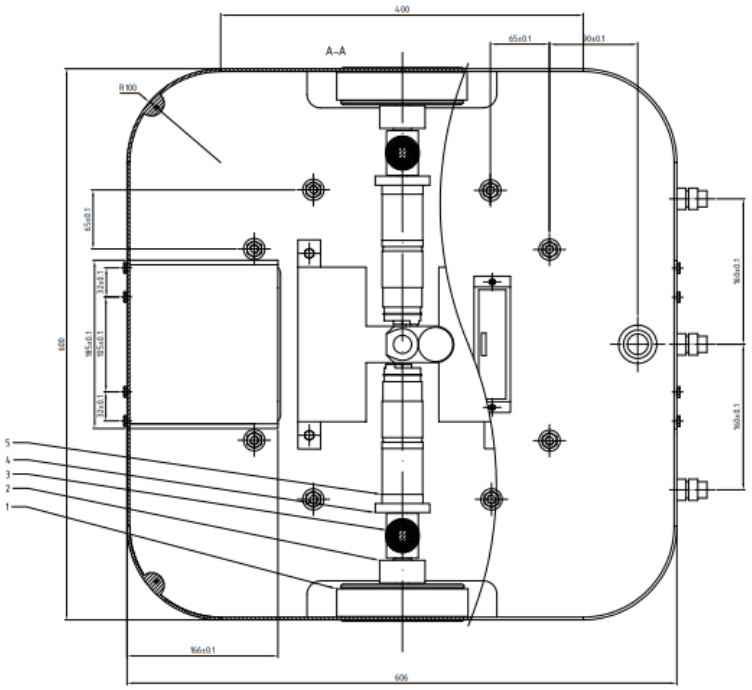
10-Electric piston

11-Inductive proximity sensor

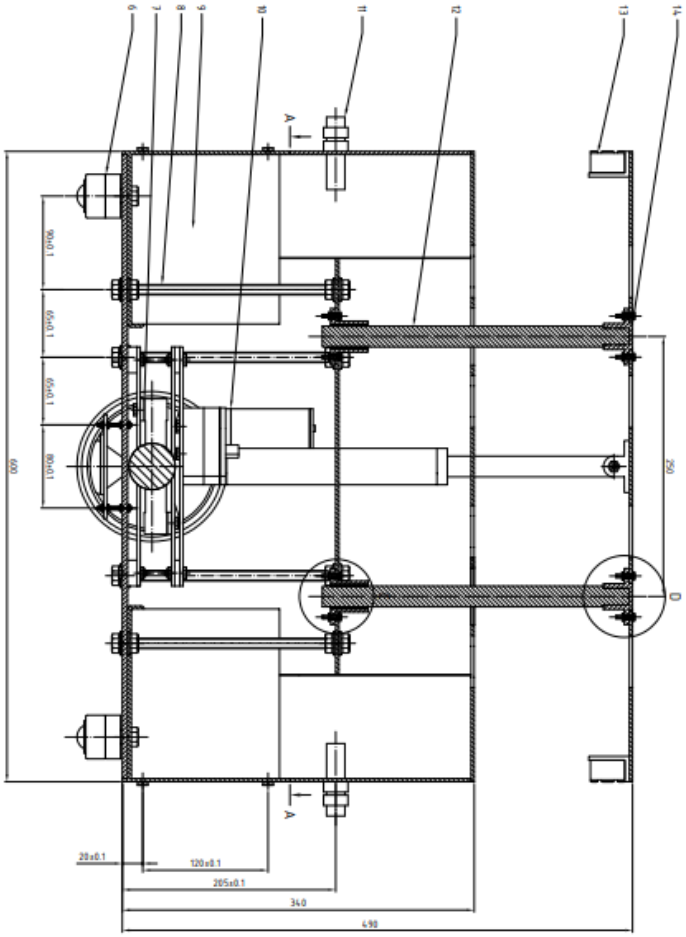
12-Linear actuator

13- Electromagnetic lock

14- Lifting layer



(a)



(b)

Figure 4: Vertical view (a) and horizontal view (b) of hardware structure design.

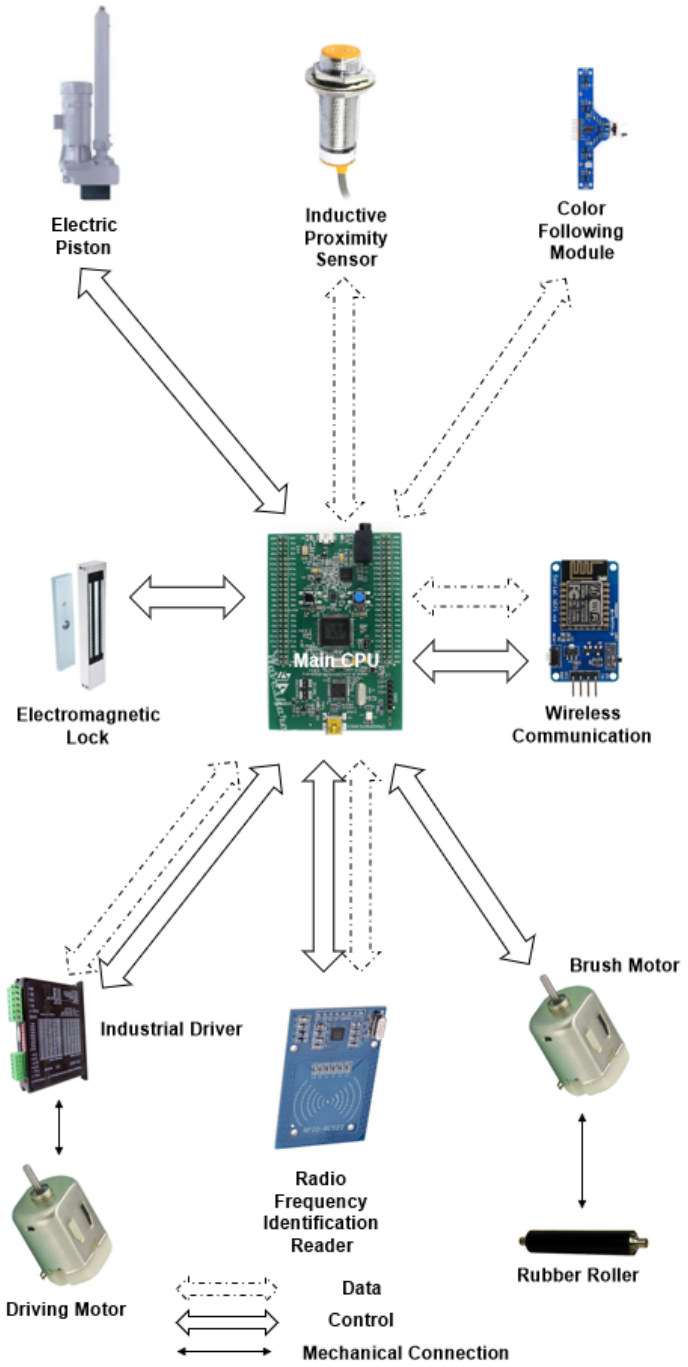


Figure 5: Diagram of data and control bus in proposed system.

Fig. 5 demonstrates the threads of data transmission and elements of control scheme in proposed system. It is powered by 32-bit MCU up to 168 MHz with floating point unit, 1 Mbyte flash memory, 192 Kbyte RAM, various I/Os, three channels of 12-bit ADC, two channels of DAC providing high performance and wide range of applications. For the internal signals, the data bus accommodates exchanging information (sensors or network protocol) whilst the control bus encompasses synchronizing interface (actuators, drivers). Due to centralized control method, the system is effective, reliable highly and simple to debug. However, to enlarge the real time performance, the coding needs to be optimized in firmware level.

The navigation problem of vehicle is also an attractive issue in control of system. The Radio Frequency Identification (RFID) technology is involved to lessen the burden data transmission. In each crossroad or corner, RFID card is occupied under the line guide. In the bottom side of vehicle, the card reader is stayed at central point to read and notify the unique code to host computer. The operator, who sits in center control room, could determine where the system locates in working map.

5. Research Results

It has been known that the combining approach must deal with various challenges and the gap between computational design and real world should be as small as possible. Henceforth, the system is validated experimentally in real scenario shown in Fig. 6. The architecture of control topology consists of local controllers and central one. The main microprocessor in local controller is powerful with high computing ability, fast processing and expandable peripherals. The host personal computer is with i7 core, high speed clock and visual monitoring capability. The supervisor would stay in control area to supervise the operation of system. The local controllers and central manager are contacted via wireless communication. The control messages are unceasingly transferred among them to prevent unfortunate incident.



Figure 6: Experimental model of proposed vehicle.

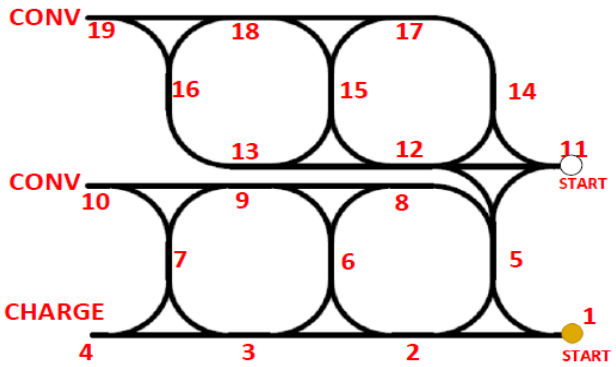


Figure 7: Overview of mapping area in workspace.

To imitate the practical situations in factory or distribution center, an overview of mapping working sector is drawn in Fig. 7. For multi-agent system, there are two reference regions are in parallel to enhance industrial productivity. The start points, charging section and conversing node are formed to offer various circumstances. At initial stage, vehicle moves to cargo's location

and bring to target position. In the case that the cargo must be out of storage, vehicle takes to conversing node. The battery management is gathered at charging station. Whenever level of energy is low, autonomous vehicle sends a notification to host server and move to charging node.



Figure 8: Experimental test scenario of tractor-trailer model.



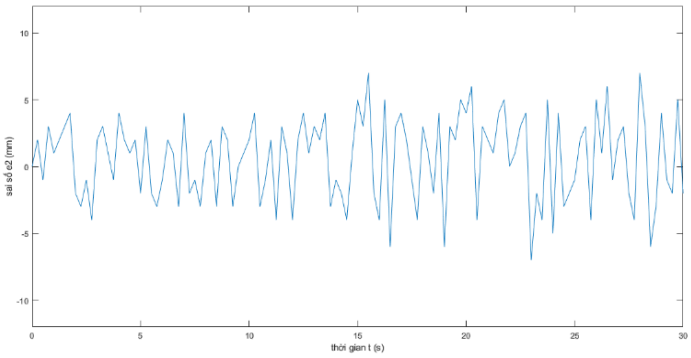
Figure 9: Experimental test scenario of lifting model.

The shelf of cargo is made of steel and its height comply with industrial standard. There are only two floors, but its floors can be extendable. Four caster wheels are attached under the shelf to assist the free motion. The experimental validations include two-fold: first, vehicle begins at start point, it passes serial points to target position. Then, the lifting layer is gone up as height as the bottom floor of shelf. At that moment, the local controller activates electromagnetic lock to maintain force. It drives the shelf according to reference path as Fig. 8. The reverse process is actuated in target position: inactivate the electric lock, lift down and gradually move to next position.

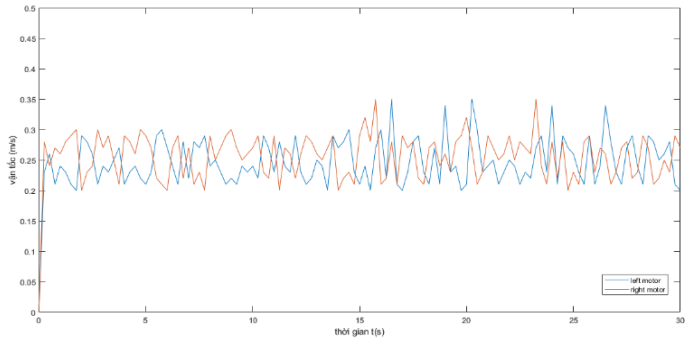
Sequently, the lifting mode is experimented as Fig. 9. In this case, the autonomous vehicle is compelled to accomplish the mission: brings the rack of payload from initial position to final location. Launching at start point, it comes to site of shelf by tracking linear and circular path. Later, vehicle orders lifting part to move up elevating the shelf. At this time, vehicle and cargo become a rigid body. Both of them are conveyed to final place and released there. The grounded robot could come back start point and ready for next task.

Table 1: Comparative Tracking Error Results in Experimental Tests

Test case	RMS	Average
Lifting model	0.157	0.268
Truck-trailer model	0.082	0.194



(a)

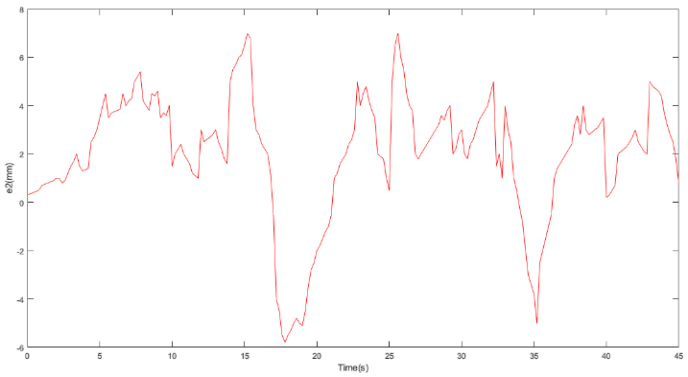


(b)

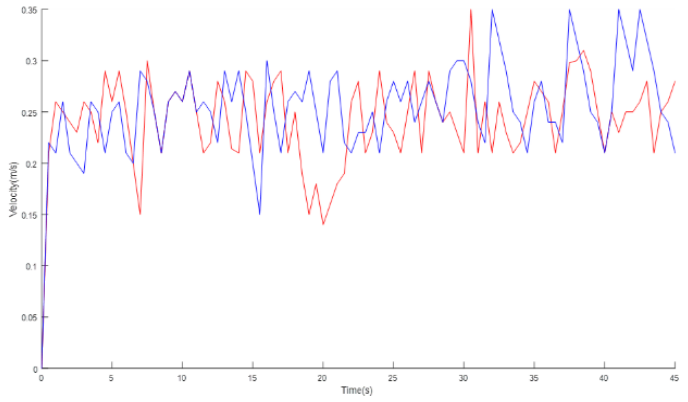
Figure 10: Experimental performance of truck-trailer model, (a) tracking error, (b) velocity of left wheel and right wheel.

The monitoring parameters in this study are tracking errors and control speeds. In differential drive, the variations in velocity of left and right wheels have an effect on directional movement. The tracking error governs the output performance in whole system. Fig. 10 and Fig. 11 show experimental results in two case studies respectively. In truck-trailer test, the leader must take a follower behind carefully. The motion of trailer affects on tracking error of front-runner. If leader moves so fast, it tends to deviate from reference path at corner. Otherwise, the force is too small to steer a follower or time-consuming motion costs too much.

In lifting test, the weight of whole system including total mass on shelf plus itself is heavier. Consequently, the motion seems to be more difficult. The value of tracking error, in this scene, is larger than others while the velocities of wheels are adjusted continuously. To visualize the differences in two cases, Table 1 reports the tracking error (Root-Mean-Square value and average value). From these results, it could be seen obviously that tracking error in lifting mode is larger than one in truck-trailer mode. The reason is that the control scheme drives harder when vehicle and cargo become a rigid body.



(a)



(b)

Figure 11: Experimental performance of lifting model, (a) tracking error, (b) velocity of left wheel and right wheel.

6. Conclusions

In this paper, a multi-purposes implementation for autonomous grounded vehicle is fused into an unique framework in practical factory. The proposed design is feasible, low cost and complies with industrial solicitations whilst it still conserves flexible motion. Furthermore, the integrating approach ensures that it is not only useful in various applications, but also it provides wide development trends, such as smaller, smarter or more convenient. From the successful results of research, it could be achievable in logistics, manufacturing industry, material transportation and training.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. D'Andrea and P. Wurman, "Future challenges of coordinating hundreds of autonomous vehicles in distribution facilities" in IEEE International Conference on Technologies for Practical Robot Applications, 2008, https://doi.org/10.1109/TEPRA.2008.4686677.

[2] H. Q. T. Ngo, T. P. Nguyen and H. Nguyen, Research and develop of AGV platform for logistics warehouse environment; Vancouver: Springer-Verlag, 2018, https://doi.org/10.1007/978-3-030-02683-7_32.

[3] D. Chen, Z. Shi, P. Yuan and Z. Li, "Trajectory tracking control method and experiment of AGV" in IEEE Advanced Motion Control, 2016, . https://doi.org/10.1109/AMC.2016.7496323.

[4] V. Q. Nguyen, H. M. Eum, J. S. Lee and C. H. Huyn, "Vision Sensor-based Driving Algorithm for Indoor Automatic Guided Vehicles," International Journal of Fuzzy Logic and Intelligent Systems, **13**(2), pp. 140-146, 2013.

[5] T. A. Tamba, Q. T. T. Bui and K. S. Hong, "Trajectory generation of an unmanned forklift for autonomous operation in material handling system" in SICE Annual Conference, 2008, https://doi.org/10.1109/SICE.2008.4654955.

[6] J. X. Xiao and F. L. Lu, "An improvement of the shortest path algorithm based on Dijkstra algorithm" in International Conference on Computer and Automation Engineering, Singapore, 2010, https://doi.org/10.1109/ICCAE.2010.5451564.

[7] P. Ritzen, E. Roebroek, N. van de Wouw, Z. P. Jiang and H. Nijmeijer, "Trailer steering control of a tractor-trailer robot," IEEE Transactions on Control Systems Technology, **24**(4), 1240-1252, 2016, https://doi.org/10.1109/TCST.2015.2499699.

[8] M. A. Guney and I. A. Raptis, "A robotic experimental platform for testing and validating warehouse automation algorithms" in International Conference on Technologies for Practical Robot Applications, MA, USA, 2015, https://doi.org/10.1109/TePRA.2015.7219693.

[9] C. T. V. Kelen, P. M. G. Jorge, S. B. Thales, C. A. Roberto, M. S. Rafael, B. Marcelo and A. P. C. Glauco, "Robotic forklifts for intelligent warehouses: routing, path planning and auto-localization" in International Conference on Technologies for Practical Robot Applications, Vina del Mar, Chile, 2010, https://doi.org/10.1109/ICIT.2010.5472487.

[10] H. Q. T. Ngo, T. P. Nguyen, V. N. S. Huynh, T. S. Le, C. T. Nguyen, "Experimental comparison of complementatry filter and Kalman filter design for low-cost sensor in quadcopter," International Conference on System Science and Engineering, pp. 488-493, 2017, https://doi.org/ 10.1109/ICSSE.2017.8030922.

[11] C. Hu, R. Wang, F. Yan, Y. Huang, H. Wang and C. Wei,"Differential steering based yaw stabilization using ISMC for independently actuated electric vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 2, pp. 627-638, 2018.

[12] C. Hu, R. Wang, F. Yan and H. R. Karimi,"Robust composite nonlinear feeback path following control for independently actuated autonomous vehicles with differential steering," IEEE Transactions on Transportation Electrification, vol. 2, no. 3, 2016.

[13] T. Chen, X. Xu, Y. Li, W. Wang and L. Chen, "Speed-dependent coordinated control of differential and assisted steering for in-wheel motor driven electric vehicles," Proc. of the Institution of Mechanical Engineers Part D: Journal of Automobile Engineering, vol. 232, no. 9, pp. 1206-1220, 2018.

[14] C. Lin, Z. Xu, "Wheel torque distribution of four-wheel-drive electric vehicles based on multi-objective optimization," Energies, vol. 8, no. 5, pp. 3815-3831, 2015.

[15] H. Ashrafiuon, S. Nersesov, G. Clayton, "Trajectory tracking control of planar underactuated vehicles," IEEE Transactions on Automatic Control, vol. 62, no. 4, 2016.

[16] B. A. Kovacs, G. Stepan, Z. Wang, T. Insperger, "On the stability of two-wheeled vehicle balancing passive human subjects," 12th IFAC Symposium on Robot Control, vol. 51, no. 22, pp. 337-342, 2018.

[17] J. Qian, B. Zi, D. Wang, Y. Ma, D. Zhang, "The design and development of an omni-directional mobile robot oriented to an intelligent manufacturing system," Sensors, vol. 17, pp. 1-15, 2017.

[18] D. Ilangasinghe, M. Parnichkun, "Navigation control of an automatic guided forklift," 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence and Robotics, pp. 123-126, 2019.

[19] H. Q. T. Ngo, Q. C. Nguyen, T. P. Nguyen, "Design and implementation of high performance motion controller for 2-D delta robot," Seventh International Conference on Information Science and Technology, pp. 129-134, 2017, https://doi.org/10.1109/ICIST.2017.7926505.

[20] J. L. Syu, H. T. Li, J. S. Chiang, C. H. Hsia, P. H. Wu, C. F. Hsieh, S. A. Li, "A computer vision assissted system for autonomous forklift vehicles in real factory environment," Multimedia Tools Applications, vol. 76, no. 18, pp. 18387-18407, 2017.

[21] Ngo, Ha Q.T.; Phan, Mai-Ha. 2019. "Design of an Open Platform for Multi-Disciplinary Approach in Project-Based Learning of an EPICS Class." Electronics 8, no. 2: 200.

[22] M. Yue, X. Hou, R. Gao, J. Chen, "Trajectory tracking control for tractor-trailer vehicles: a coordinated control approach," Nonlinear Dynamics, vol. 91, no. 2, pp. 1061-1074, 2018.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 545-560 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Study and Implementation of Various Image De-Noising Methods for Traffic Sign Board Recognition

M. Monica Subashini[*,1], Abhinav Deshpande[1], Ramani Kannan[2]

[1]School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore, 632014, India

[2]Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS (UTP), 32610, Malaysia

ARTICLE INFO

Article history:
Received: 04 March, 2019
Accepted: 29 July, 2019
Online: 25 August, 2019

Keywords:
Color
Shape
Image De-noising
Gaussian Filter
Deep Neural Network (DNN)
Harmonic Filter
Arithmetic Mean Filter
Circular Averaging Filter (Pillbox)
Anisotropic Diffusion Filtering
Non-Local Means

ABSTRACT

The problem of recognizing traffic sign boards in a correct fashion is one of the major challenges since there is an alarming rate of increase in the number of road accidents happening because of incorrect interpretation of traffic sign boards in bad weather conditions. In this paper, a comparative analysis of various noise removal techniques based on calculating different parameters which decide the quality of input roadway symbol like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) as well as Structural Similarity for measuring Image Quality (SSIM) is being performed and the best technique will be chosen among them which gives minimum Mean Squared Error (MSE) value and maximum Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) values. This technique will be quite useful for de-noising a given image which is present in both the testing and the training image databases.

1. Introduction

There is a huge demand in the market as there is a rapid development in improving the existing system which ensures in giving security to the driver during the course of driving a vehicle. This leads to the development of a system that provides a helping hand thereby incorporating new techniques used for the purpose of detecting as well as recognizing the traffic signs. Nowadays, many automobile industries are extending in the market due to an increase in the demand for smart cars as a result of the rising competition among its fellow industries [1-5].

In the areas where there is a large amount of travelling, the regulation of the speed of the vehicles is done by the use of sign boards. They are used for limiting the current velocity of moving vehicles and are considered to be a part that comprises of some major categories of traffic signs [6-10].

A rise in the number of roadway accidents is observed because of ignoring the sign boards which try to limit the velocity of a moving vehicle for the persons who are driving a vehicle. Generally, it is observed that the persons who are not aware of the potential danger during the course of driving consequently meet a serious accident thereby even leading to their deaths [11-15]. There is an urgent need for development of a system which can be used to recognize the Speed Limit Signs (SLS) in an automatic fashion and may be used for alerting the driver [16-20]. There are two major steps which are involved in the processes of detecting and recognizing the Speed Limit Signs (SLS). The first step is to detect the potential candidates which are known as the Regions of Interest (ROI) usually comprising sign board images and the second step is to recognize the ROI in order to extract the correct category of sign board images [21-25].

During the course of driving a vehicle [26-35], the driver tends to ignore the speed limit sign boards that convey a message to restrict the velocity of a moving vehicle thereby suffering from a

[*]M.Monica Subashini, School of Electrical Engineering, VIT, Vellore.
monicasubashini.m@vit.ac.in

collision which may ultimately lead to the death of the driver consequently becoming a reason for increasing number of roadway accidents [36-45]. The development of a system that is capable of recognizing the Speed Limit Signs (SLS) in an automatic fashion as well as alerting the driver at the same time is the need of the hour [46-50].

It is generally observed that various automobile industries are increasing in the market because of the demand of smart cars. This is the result of the rising competition among its fellow industries thereby facilitating the growth of systems. It would provide a helping hand thereby incorporating some new techniques used for the purpose of detecting as well as recognizing the traffic signs. In recent years, since there is a large requirement in the market because of rapid growth in the improvement of the existing system, a secure driving environment is provided [1-5]. The maximum speed limit sign boards are generally employed to restrict the velocity of a moving vehicle and are used for regulating the velocity of a vehicle in the case of areas where there is a large amount of travelling thereby comprising of a major category of traffic signs [6-10]. During the course of driving a vehicle, the driver tends to ignore the speed limit sign boards that convey a message to restrict the velocity of a moving vehicle thereby suffering from a collision which may ultimately lead to the death of the driver consequently becoming a reason for increasing number of roadway accidents. The development of a system that is capable of recognizing the Speed Limit Signs (SLS) in an automatic fashion as well as alerting the driver at the same time is the need of the hour [11-15]. The process of detecting as well as recognizing the SLS normally comprises of two major steps namely, detecting the potential candidates known as the Regions of Interest (ROI's) consisting of the sign board images and recognizing the ROI so that the category of sign board images can be extracted in a correct fashion [16-20]. The main objective of detecting a sign board image can be accomplished with the help of extracting the primary visual features like the color and the shape which are present inside a given input image [21-25] such as the color feature is employed for detecting the red SLS while the shape feature is employed for detecting circular or rectangular sign board images [26-30]. If we study the research work which is done in the past, we will come to know that the rectangular speed limit signs [31-35] are usually detected with the help of shape-based methods [36-40], for example, if we study the research work which is carried out by [41-45] we will find that the rate of detecting the sign board images was increased in the rough conditions, especially, during the night time [46-50].

The circular sign board images are detected by making use of the color-based and/or the shape-based methods, for example, if we consider the research work which is carried out by [1-5]. Next step which is involved in the process of recognizing the SLS comprises of two important techniques, namely, the processes of recognizing sign boards that are available in a broad range as well as recognizing each and every character on a given sign board. If we take an example of strategy that incorporates recognizing a given sign board which is made available in a wide range, one may find a complete candidate sign to be considered for recognizing process [6-10]. If we take the example of digit-based methods, the process of extracting multiple characters for recognizing numerical portrait of velocity which is exhibited by a moving vehicle rather than classifying a complete sign board image treating it as a single entity is performed [11-15]. The process of recognizing sign board images is treated as one of the important factors for driver assistance system because it plays an important part in controlling the road traffic thereby issuing a warning signal and giving proper guidance to the drivers [16-20]. It is observed that ignoring the sign boards because of distraction or being in a psychological state by the drivers is the main cause of roadway accidents which are occurring in last few years [21-25]. There is an urgent need for incorporating an automatic system for the purpose of recognizing the sign boards and is considered to be a main factor for building an autonomous navigation system [26-30]. In order to detect and recognize the sign boards [31-35] with a faster speed and higher efficiency, particularly, in a real time scenario, the proposed system [36-40] should have a higher precision so that the process identification of sign board images can be performed in a correct manner [41-45]. The detection and the recognition efficiencies can be affected if the system is busy in managing certain complicated issues [46-50].

The various problems that are associated with these systems include the variations in the illumination such as light levels, twilight, fog, rain and shadow, motion blur and sign occlusion [1-10]. The effectiveness can be considered to be one among the major factors because the complete navigation system is damaged by the improper classification and lack of detection of sign board images. But, later it was revealed that the systems that are currently present in the literature part are unable to provide 100% accuracy. The research scholars were motivated by the results which are given in the section above and there was an improvement in the performance of system that is capable of detecting as well as recognizing the sign board images even in the complicated situations and thus highlights the importance of our research work in building this method that is being presented in this review paper. A novel technique which is used for the process of detecting, tracking and classifying the sign boards at a faster speed from a moving vehicle even in the difficult conditions was proposed by [11-20]. In order to extract the candidate regions of the sign boards, a color-based segmentation technique was being proposed in the course of detecting the sign board images. The HOG features are applied in order to encode the sign board images which are detected and the feature vectors are also computed. The process of detecting and recognizing the sign board images in an automatic manner remains as one of the daunting task that comprises of a variety of areas of application, for example, if we consider reinforcements resulting from a rapid growth in the arena of providing a safe driving environment to a driver who is driving a vehicle, taking a survey of existing road conditions that are prevalent in our surroundings so that potential dangers can be judged and proper steps can be implemented in order to prevent them from happening [21-30].

1.1. Traffic signs interpretation

When we compare the research work which is done on automatically detecting and recognizing symbol-based traffic sign board images and recognizing the textual information which is present inside the sign board images, we will find that a small amount of research work is done on recognizing the textual information which is present within a sign board image in comparison with the former [31-40]. One of the main difficulties that are associated with this procedure such as presence of variation in lighting conditions as well as shadowing effect that is caused due to improper lighting conditions, the effect of causing a blur, causing an occluding effect due to improper and less amount of light as well as deteriorating of sign board images can be considered responsible for causing this scenario to a smaller extent [41-50]. During the course of driving, we cannot distinguish in between sign board images as well as images that belong to some another class other than sign boards merely on grounds that provides the information which is provided by their primary visual features like the color and the shape but some extra temporal or contextual information should also be provided supplementing this information, since the information that is given by color and the shape of traffic signs as well as non-traffic signs proves to be inadequate [1-10]. Based on this hypothesis, it is observed that there are large numbers of images that does not fall in the class of sign boards which are classified from images falling in a particular class of sign boards although detected successfully from the original images which is found most frequently within a sign board image that tends to appear in somewhat cluttering fashion, as we normally observe in the case of a road scene [11-20]. It is generally observed that when there is lack of some supplementary information such as the structural and temporal information, there are large numbers of false positives (FPs) that are detected by the system especially when we observe the top scene, thus facilitating the process of detecting the traffic signs which is present in both the images in a rather successful manner [21-30]. The possible solution to this problem was provided by the use of primary visual features like the color and shape of traffic sign board images that are employed in the process of detection of a large number of text-based traffic sign board images. One should also take proper care that no True Positives (TPs) are ignored thus reflecting the importance of the process of over segmentation [31-40]. In order to eliminate unlikely candidates, an enormous amount of candidate regions which are detected with the help of structure of the scene and the temporal information that is conveyed by them were diminished to a greater extent [41-50]. One of the main objectives of an intelligent transportation system is to provide a safe and secure driving environment thereby ensuring the safety of the surrounding traffic. We can devise a new technique ensuring the safety of the surrounding traffic by implementing an on-board camera-based driver alert system which helps in recognizing sign board images such as stop signs and speed limit signs [1-10]. Traffic Sign board images convey important information regarding the present condition about the roadway as well as tend to provide other information for the purpose of navigation. They can be thought of as planar rigid objects which are having different colors and shapes. The primary visual characteristics of an image constitute the color, shape and pictogram which help us to provide relevant information which is embedded in it. The area of Intelligent Transportation Systems is gaining popularity nowadays and many car manufacturing companies are focusing their eyes on the area of Advanced Driver Assistance Systems since it contains a broad spectrum available for carrying out research and development especially in the domain of Traffic Sign Recognition [11-20]. In the year 2008, Mobil eye entered into a joint venture ship with Continental AG and three additional new updates were installed in the BMW 7 series that comprised of lane departure warning, speed limit information based on the method of detecting sign board images and intelligent headlight control. One of the major challenging tasks which is confronted by modern car manufacturing companies is to recognize the sign boards correctly particularly in an uncontrolled environment [21-30].

There are three steps in a typical traffic sign board image recognition system that mainly consists of detecting a particular traffic sign from other signs by using some detection method followed immediately by the process of elimination of noise from it with the help of some pre-processing techniques and finally recognizing that traffic sign by using some pattern recognition and machine learning approaches. We can augment the existing system with the help of tracking algorithm so that the speed of recognition can be increased, so that we can focus on a tiny region of the object which is detected for the purpose of recognition. In this research paper, more emphasis is laid on the process of recognizing traffic sign board images; we can leave the remaining three methods, viz., detection, rectification and tracking. One can find the research work done on the problem of recognizing traffic signs dating from the last century, and the proof of this work is found in an ancient survey which was lastly updated in May 1999, given in [31-40]. Since, there are plenty of techniques available in the field of recognizing traffic sign boards over a past few decades, we do not try to give a comprehensive literature survey, and instead concentrate on several background work most relevant to the present research. Due to the growth of different intelligent methods pertaining to driver assistance systems (DAS), it is trying to prove its superiority over the currently existing approaches which are present in Intelligent Transportation Systems. In the DAS, the task of detecting and recognizing traffic signs can be considered as one of the major approaches for acquiring information regarding a safe driving environment and we can also take precautionary measures in order to overcome the potential hazards faced by driver particularly while driving a vehicle along the road. The process of recognizing the sign board images, that too in a real environment, proves to be a herculean task if we want the technique to be done with greater accuracy and within a stipulated time, since the visibility of sign boards can be affected due to some undesirable factors to a larger extent. So far, many algorithms have been proposed for traffic sign recognition [41-50]. In [1-10], an entirely different approach of the currently prevailing AdaBoost algorithm is utilized in order to detect a

traffic sign by deployment of batteries of classifier algorithms that are given proper training to split the classes in an error correcting output code framework. A robust sign similarity measure which is built by using the SimBoost or a fuzzy regression tree framework, a technique which is used for recognizing traffic signs is proposed in [11-20]. We all are familiar with the Support Vector Machine (SVM) algorithm which happens to be a good classifier that is very popular in the domain of computer vision. In order to facilitate the process of detecting and recognizing road signs in an automatic fashion, an entirely different method is presented in [21-30] in which a SVM is used for traffic sign detection while a Gaussian kernel SVM is adopted for traffic sign recognition [31-40]. In order to classify an unknown traffic sign, an Eigen-based traffic sign recognition system was proposed in [41-50] taking the help of Principal Component Analysis (PCA algorithm) by selecting the most effective components of traffic sign board images. Boosted by the successful applications on handwritten digits recognition, convolutional neural network (CNN) has also been employed on traffic sign classification [1-10]. In [11-20], instead of various features, a CNN is trained directly with the raw pixel values of traffic sign images. Moreover, a better result is obtained by integrating the results obtained by a CNN and a multilayer perceptron (MLP). Gradient orientation is one kind of useful information in various object recognition methods, including the traffic signs. Traffic scene analysis is a very important topic in computer vision and intelligent systems [21-30]. The traffic signs are manufactured in such a fashion that they convey important information regarding the present condition and other information of the road. They are rigid and simple shapes with eye catching colors and the information they carry is easy to understand. However, accidents may still occur when drivers do not pay attention to a traffic sign in time [31-40]. So, in order to overcome these difficulties, there is an urgent need for implementing an automatic real time traffic sign recognition system which is capable of detecting and recognizing traffic sign board images. The above factors make the problem of detecting and recognizing traffic sign boards very complicated and still remains a daunting task for most of the research scholars [41-50].

1.2. Signs interpretation methods

There has been a lot of improvement in the field of detecting and recognizing the textual information which is present inside an image [1-10]. The different techniques which has led to an improvement in this area can be classified into two major types, one is the Region-Based Methods and another one is the Connected Component (CC) based methods [11-20]. The local features for example, texture are used by the region based methods for the purpose of locating the regions which contain the textual information in an image and the connected component-based methods are used for the purpose of segmentation of the characters that are present in the textual information within a particular traffic sign board image on an individual basis with the help of information like the intensity, color distribution and the edges of an image [21-30]. The Connected Component based methods normally comprises of three stages, the first stage is used for the purpose of detecting CCs which are present inside an image followed by a second stage in which the CCs that are not very similar are eliminated based on their features which is subsequently followed by a final stage in which grouping of the remaining CCs is performed into words or lines [31-40]. It was revealed that the amount of research work which targeted specifically on the task of detecting the textual information which is present inside a particular traffic sign board image is restricted to a narrow spectrum because of the difficulty in performing the given task. The methods that are currently available in the literature consist of two steps, the first step is the process of detecting a given sign board image and the second step is the process of recognizing a particular image [41-50]. The research work which is carried by founding the candidate regions by combining the characteristics that are obtained by combination of Shi characteristics as well as Tomasi characteristics, the prototypes that are obtained through the mixing of characteristics that are derived from Gaussian functions as well as that which are obtained by the process of analyzing a given image on the basis of its geometrical properties. It was postulated on the basis of the textual information which is present inside a given sign board image that is visible if we take some plane surface that is oriented in the upward direction thereby giving due regards towards the movement as well as the axis of the camera that exhibits the photonic characteristics and properties. If we consider the real world, it may be possible that the sign boards which contain the textual information within it may tend to appear if one looks it from its own perspective to be somewhat in synchronism with the original image of the sign board [1-10]. So, there is an urgent need for incorporating a perspective transform which gives the OCR a better possibility of performing the task of recognizing the textual information that is contained inside a sign board image. The process of matching of the candidate regions with the help of consecutive frames was done and the process of interpretation was then performed with the help of an OCR system when their size was considered to be adequate. The detection rate was found to be 88.9% and the false detection rate was found to be 9.2%, which was based on database containing 22 video sequences with video sequence having a length of 30 seconds duration. The work which was done by [11-20] comprised of segmentation of the areas that are present inside an image and the extraction of these regions seem to attract the research scholar for the purpose of carrying out his own research which is propounded on the grounds of the information that is conveyed by the primary visual characteristics of an image thereby taking help of some limiting criterion that is applied to channels that contain the chromaticity as well as the luminosity components if we consider the most popular technique which is deployed for the purpose of segmentation of sign board images [21-30]. If we compare the tracing of outer boundaries that are present inside every single blob with the help of Fourier transformation by adopting an approach that increases the speed

of computation of the process thereby reducing the duration that is required for every single computation with those obtained by adopting the same method from an arbitrary uniform plane lamina which can be considered as an example for the purpose of implementing our technique was being employed in order to find the rectangular regions [31-40]. The quadruple corners which can be thought of as an area that resembles a rectangle were being searched by deploying the process by taking peaks of the area that is traced from an entire image with the help of transformation of complete image using Fourier method thereby increasing the speed for carrying out the process of computation thus reducing the duration required for every single computation and by taking help of aforementioned corners, thereby carrying out the task of rotation of corresponding areas that are chosen for the purpose of extraction from complete image trying for the process of alignment of the textual information consisting of mainly the characters that is present inside a given sign board image particularly in upward direction. If we consider the point of view of a vehicle, it is revealed that the above method is not sufficient if we try to consider the process of recovering above area that is present on the chassis of the vehicle if the point of view of the observer is taken for the process of consideration which consists of textual information inside a sign board image and the perspective correction which is given in [41-50] tries to provide a more robust solution to this problem. There was a lack of output that represents the content which is illustrated in a more sophisticated fashion thereby giving the amount of content which is extracted from a given sign board image which is contributed through the research work that is illustrated in the existing literature. If we consider some existing techniques that are given by [1-10], we will come to know that they exhibit some peculiar characteristics that mainly focuses on taking help from MSERs that mainly exerts a thrust on detecting both the images that represent the significant information related to existing road conditions as well as if we consider some characters that are present in the textual information which is contained within a given sign board image. If we try to combine the processes of segregating the original image as well as the characters that are present in an image if we try to see the image particularly in sunlight thereby constituting a container, we come to know of the interesting fact that the areas which are present on the chassis of a vehicle having a whitish as well as bluish appearance are subjected to the process of detecting each of the components that are present inside a given video sequence [11-20]. By taking help from both the mechanical systems that tend to classify the basis functions thereby supporting the conventional technique which can be deployed for the process of classifying a given image as well as the classification method that is presented by some research scholars, these regions were then again subjected to the process of classification. More emphasis was placed on the geo-localization of sign board images with the help of information which is supplied by a Global Positioning System, and the technique was then applied to single images, without the use of temporal information [21-30]. If we try to make an approximation about size of our complete world, we will come to know that height of textual content which is present inside an image was used along with the spacing in between our camera and images that carry some message about the prevalent road conditions. The rates that are required for the processes of detecting, recognizing as well as classifying the words, numerical characters as well as symbolic information that is represented by the alphabets present on a particular sign board image that are taken from shorter, longer as well as medium distances in between the observer and the sign board images were being given by the existing authors. If we consider the ranges for the process of detecting and recognizing the traffic sign board images, it was found that the range for detecting the sign boards was lying in between 13.09% and 90.18%, while the range for the process of recognizing the sign boards was found lying in between 8.51% and 87.50%. Our proposed method overcomes the numerous limitations which are faced by the method that were proposed by [31-40] in which postulates were propounded for the method of detecting traffic sign board images that comprised of textual information within it and are sufficient for providing a high recall rate. The additional FPs was eliminated by making use of structural and temporal information that is conveyed by a particular image by our technique which is being propounded in the above section. One can find the output of system which we derived in the course of performing experiments that get highlighted from our research work are sufficient enough for giving a proof to the proposed method which is described in the section above. The research work which was done by [41-50] used a rough OCR technique in detecting and recognizing the textual information contained in the sign boards was superceded by our proposed method that uses the concept of recovering original information contained in sign board images comprising point of view from an observer as well as techniques that involve the usage of merging two portions of any kind of objects that is present along lateral portions of an object. One may find manuscripts from earlier and past research works which reflect our propounded system performing with better outcomes alone is sufficient enough for giving a proof to our proposed method which is given the section above. If we consider the research work which was done [1-10] takes the help of temporal information used for detecting symbol-based traffic sign board images which is further contributed by the research work that is described in this review paper by embedding the structural information that is present in the scene. We can find a huge amount of research work that focuses on the problem of detecting the textual information which is present in sign board images by taking help from a mobile sensor and mainly depends on the devices that can be worn by the driver, for example, the techniques that are mentioned in [11-20] and the methods which are mainly applicable to mobile robotics can be found in the research work that is presented by [21-30]. We can find a number

of databases which are used for the purpose of validation of systems used for detecting and recognizing the sign board images which includes the famous dataset which is included in the research work that is done by [31-40] as well as the database that is included in the research work which is done by [41-50]. The above databases find no use in order to validate our proposed method which is entirely based upon detecting and recognizing the sign boards containing textual information within them and are more concentrated upon the process of detecting the symbol-based sign boards. A camera with knowledge of all the parameters used in the process of calibration was used for extracting the data [1-10]. One may consider the process of recognizing a particular sign board image amongst all other members, a Herculean task when it comes to recognizing sign board image from a group of other sign boards. It is found that a variety of techniques are available in the arena of classifying known sign board images ranging from ordinary methods like the matching of templates to sophisticated machine learning techniques. One of the very significant and most important algorithms which are employed to perform the task of classifying multiple sign board images can be attributed to well-known Support Vector Machine (SVM) algorithm. If the transcripts authored by [11-20] are considered, one can find the significance attached to automatically detecting and recognizing sign board images taking help of Support Vector Machines (SVMs) in combination with Gaussian kernels. However, the system was required to classify candidate blobs into a shape class before recognition. As a sequence, only the pixels that were part of the sign were used to construct the feature vector. In [21-30] different methods other than those which are currently available in the context of detecting and recognizing sign boards in an automatic fashion were brought to the notice of people. It can be very well understood from this study that a comparatively greater emphasis was laid in enhancing the accuracy of existing methodologies prevalent in detecting and recognizing sign boards in an automatic fashion thereby resulting in the reduction of number of support vectors which are required in due course of the complete process leading to a sudden fall in the necessity of memory and time for testing new samples [31-40]. An SVM segmentation approach for traffic sign recognition was given in [41-50] while in [1-10], an effective strategy helping in the process of recognizing slanted speed limit signs by extracting the rotation invariant features with the help of Fourier based wavelet descriptor was introduced. The different categories of sign board images were classified with the help of Support Vector Machines (SVMs) consisting of binary tree architecture. In [11-20], a shape-based classification was developed using an SVM. In order to represent the features, two types of features including a binary image and Zernike moments were used. The main objective behind carrying out this research work was recognizing seven categories of traffic sign shapes as well as five categories of speed limit signs. In addition to SVMs, AdaBoost is also a popular classification approach. A color insensitive Haar wavelets feature combined with the AdaBoost algorithm was introduced in to develop a country-independent recognition module [21-30]. By using the concept of a class similarity measure that is learnt from image pairs, realization was achieved with the help of a new newer version of the existing AdaBoost algorithm which is known as SimBoost algorithm [31-40]. The solution to the problem of classification between multiple classes can be obtained if we compare the similarities between an unknown example and prototype from each class, subjected to the condition that an estimate of the similarities between any two images is made available. The stop and give way signs in different and complex surroundings were successfully classified by taking help from an array of weak classifiers in a reliable fashion. [41-50] took the help of classifiers which are trained to refine the type of traffic sign like neural networks and AdaBoost classifiers was also being taken. With the help of AdaBoost feature selection procedure, suitable features were chosen from a very large pool of features consisting of Haar, moment, symmetry, frequency and other features. The results which were obtained from the various classifier algorithms were combined over time and merging of the signs that are quite similar on both the sides of the roadway was accomplished taking help of fusion module that incorporated combining a Bayesian network and a decision tree [1-10]. An Error Correcting Output Codes (ECOC) which takes help from a forest of optimal tree structures that are embedded in the ECOC matrix were also being formulated along with the existing methods. An entirely different approach that is suitable for classifying sign boards belonging to different classes by using the ECOC technique was introduced [11-20]. The ensemble of binary classifiers that were given training on categories of sign board images, having two partitions was taken and the ECOC was being implemented with the help of above scheme [21-30]. The decomposition of the original problem into a set of complementary problem having two classes that are encoded in ECOC matrix which tend to share the classifiers across the original classes was performed. In [31-40], a comparison was performed considering the class specific sets of discriminative local regions in between a discrete color image of the viewed sign with other images. By using the principal of one versus all dissimilarity maximization, they were learnt in offline mode from the standard templates of sign board images [41-50]. The comparison in between the robust discrete color image was initiated by using a color distance transform which served as a basis for defining this dissimilarity. It becomes a challenging task when it comes to the method of comparison of the actual performance since there is no such kind of ideal database or procedure for performance evaluation of several machine learning approaches that try to find a solution to the problem of recognizing traffic signs, even though they are present in the literature part [1-10]. There has been a rapid growth in research and development sector especially in the area of detecting and recognizing traffic sign board images over the last few decades.

Several novel ideas and effective methods have been proposed [11-20]. Usually, the detection part hunts [21-30] potential regions of traffic signs whereas the category to which a traffic sign belongs is determined by the recognition part [31-40]. The traditional methods which are used for the purpose of detecting sign boards can be grouped into three main classes, viz., color-based methods, shape-based methods and the sliding window-based methods [41-50].

1.3. Definitions of various Performance Metrics used for Image Quality Evaluation

The various performance metrics such as the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) as well as the Structural Similarity for measuring the Image Quality (SSIM) which are used for evaluating the quality of a given traffic sign board image are defined as follows:

1.3.1. Mean Squared Error (MSE)

The Mean Squared Error (MSE) or the Mean Squared Deviation (MSD) of an estimator or of a procedure for estimating an unobserved quantity is defined as the average of the squares of the errors that is the average squared difference in between the estimated values and what is estimated. The mathematical equation which governs the Mean Squared Error (MSE) is given as below:

$$\text{Mean Squared Error (MSE)} = \frac{1}{n}\sum \qquad I = n + 1^{n+q}\,(Y_i - Y_i') \quad (1)$$

where n is the total number of predictions generated from a sample of n data points on all the variables, Y is the vector of observed values of the variable being predicted.

where M × N represents the size of an image.

1.3.2. Peak Signal to Noise Ratio (PSNR):

The Peak Signal to Noise Ratio (PSNR) is defined as the ratio in between the maximum possible power of a given signal and the power of corrupting noise that affects the fidelity of its representation. The mathematical equation which governs the Peak Signal to Noise Ratio (PSNR) is given as below:

$$\text{Peak Signal to Noise Ratio (PSNR)} = 20\log_{10}(MAX_I) - 10\log_{10}(\text{MSE}) \quad (2)$$

where MAX_I is the maximum possible pixel value of the image and MSE is the Mean Squared Error, I is the m × n noise free monochrome image and its noisy approximation is denoted by K.

1.3.3. Structural Similarity for measuring the Image Quality (SSIM)

The Structural Similarity for measuring the Image Quality (SSIM) is defined as the technique which is used for measuring the similarity in between two images and it is also employed to predict the perceived quality of digital television and cinematic pictures as well as other kinds of digital images and videos. The mathematical equation which governs the Structural Similarity for measuring Image Quality (SSIM) is given as below:

$$\text{Structural Similarity for measuring Image Quality (SSIM)} = (2\mu_x\mu_y + c_1)(2\sigma xy + c_2)/(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2) \quad (3)$$

where

μ_x is the average of x
μ_y is the average of y
σ_x^2 is the variance of x
σ_y^2 is the variance of y
σ_{xy} is the covariance of x and y

c_1 and c_2 are the two variables to stabilize the division with weak denominator.

2. Software Tool used

The MATLAB R2019a software is utilized in the synthesis of various roadway symbols from input test dataset during the course of performing experimental procedure.

3. Experimental Results and Discussion

A set of various roadway symbol images belonging to numerous categories is taken into consideration which comprising a total number of 679 roadway symbols constituting training dataset belonging to 18 different classes of roadway symbols and there are a total number of 48 roadway symbols incorporating testing dataset.

The training dataset comprises of symbols containing some printed details embedded on their outer surface and symbols that do not have any kind of printed details embedded on their outer surface. This manuscript depicts the numerous roadway symbols incorporated in dataset used for implementing this research project.

Some examples of traffic sign board images comprising of various colors and shapes which are present in the training image database as well as testing image database are depicted in the section of this manuscript (Figure 1 and Figure 2) shown as below:



Figure 1 Examples of various roadway symbols from the Training image database

A short summary of numerous methods incorporated during the course of investigation is given in section 3.1.



Figure 2 Examples of various roadway symbols from the Testing image database

3.1. Anisotropic Diffusion Filtering Type 1

The technique of Anisotropic Diffusion Filtering which is also popular by another name Perona Malik Diffusion is primarily used to reduce the unwanted disturbance present in roadway symbols. The technique of Anisotropic Diffusion Filtering Type 1 can be represented with the help of a mathematical equation which is given as follows:

$$c\,(\|\otimes I\|) = 1/1 + (\|\otimes\ I\|/K)^2 \qquad (4)$$

where the constant K controls the sensitivity to edges and is usually chosen experimentally or as a function of the noise in the image, □ denotes the Laplacian operator and it also denotes the gradient and c (x, y, t) is the diffusion coefficient and it controls the rate of diffusion and is usually chosen as a function of the image gradient so as to preserve the edges in the image, I (:, t) : Ω → R be a family of gray scale images.

The results which are obtained after carrying out the process of Anisotropic Diffusion Filtering Type 1 of the roadway symbols comprising input testing image database (Figure 3) are depicted as below:

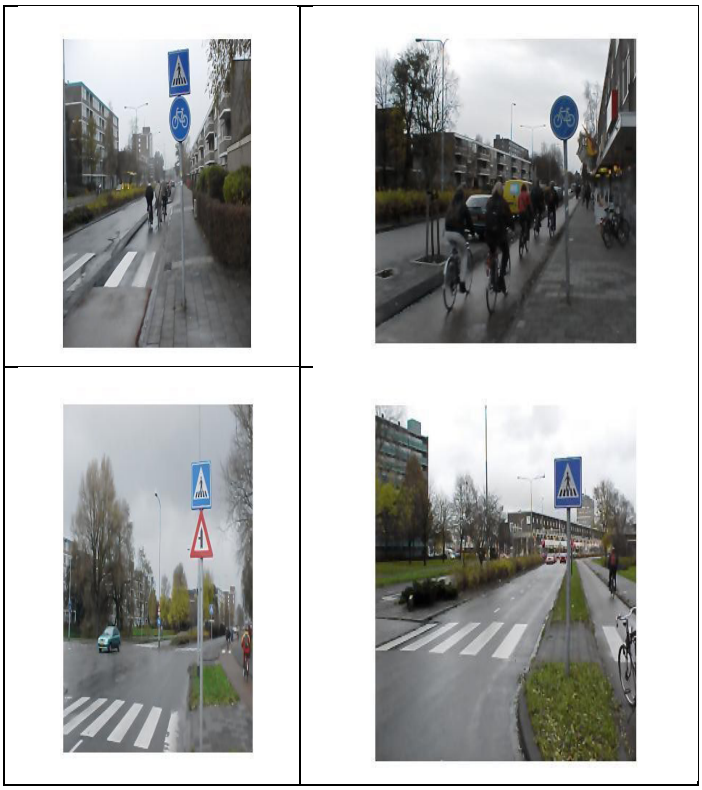


Figure 3 Testing image database Anisotropic Diffusion Filtering Type 1 Results

The results which are obtained after carrying out the process of Anisotropic Diffusion Filtering Type 1 of the roadway symbols comprising input training image database (Figure 4) are depicted as below:



Figure 4 Training image database Anisotropic Diffusion Filtering Type 1 Results

3.2. Non-Local Means Type 2

In this technique of Non-Local Means Filtering, a given color image is converted into the L*a*b color space thereby facilitating the Non-Local Means Filter to smooth the colors which are having similarity in visual perception. For the purpose of computing the numerical value of standard deviation and the Euclidean distance from the origin (e_{dist}), a homogenous L*a*b patch from the image is being extracted so that the standard deviation of e_{dist} is calculated in order to estimate the amount of noise contained in it. The mathematical equation which governs the process of Non-Local Means Type 2 is given as follows:

$$u(p) = 1/C(p) \int_{\Omega} v(q) f(p, q) dq \quad (5)$$

where u (p) is the filtered value of the image at the point p, v (q) is the unfiltered value of the image at the point q, f (p, q) is the weighting function and the integral is evaluated over $\forall q \in \Omega$ and the normalizing factor C (p) is given by the formula which is given as follows:

$$C(p) = \int_{\Omega} f(p, q) dq \quad (6)$$

The results which are obtained after carrying out the process of Non-Local Means Type 2 of the roadway symbols comprising input testing image database (Figure 5) are depicted as below:



Figure 5 Testing image database Non-Local Means Type 2 Results

The results which are obtained after carrying out the process of Non-Local Means Type 2 of the roadway symbols comprising input training image database (Figure 6) are depicted as below:

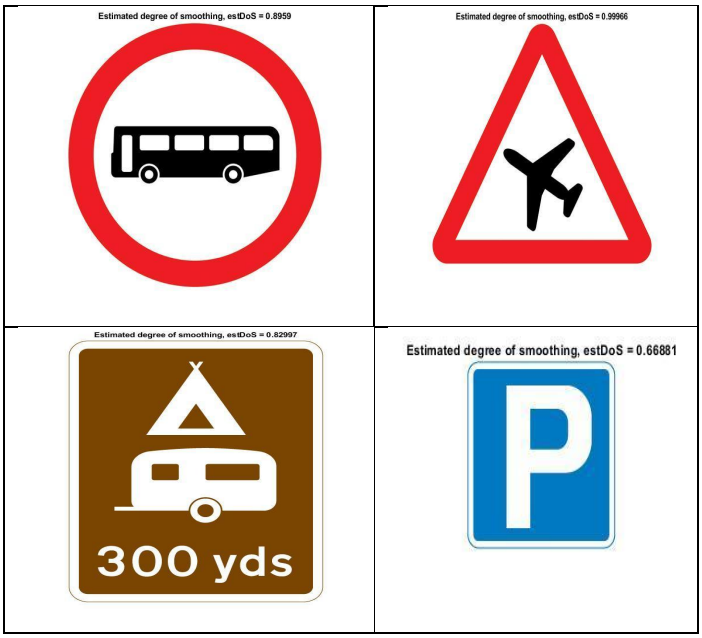


Figure 6 Training image database Non-Local Means Type 2 Results

3.3. Image Denoising using Deep Neural Network (DNN)

In this technique of denoising a given gray scale image by using a Deep Neural Network (DNN), the pretrained denoising convolutional neural network named as 'DnCNN' is being retrieved and the amount of noise which is present in the noisy version of the original gray scale image is subjected to the process of removal.

The results which are obtained after carrying out the process of Image Denoising using Deep Neural Network (DNN) of the roadway symbols comprising input testing image database (Figure 7) are depicted as below:



Figure 7 Testing image database Image Denoising using Deep Neural Network (DNN)

The results which are obtained after carrying out the process of Image Denoising using Deep Neural Network (DNN) of the roadway symbols comprising input training image database (Figure 8) are depicted as below:

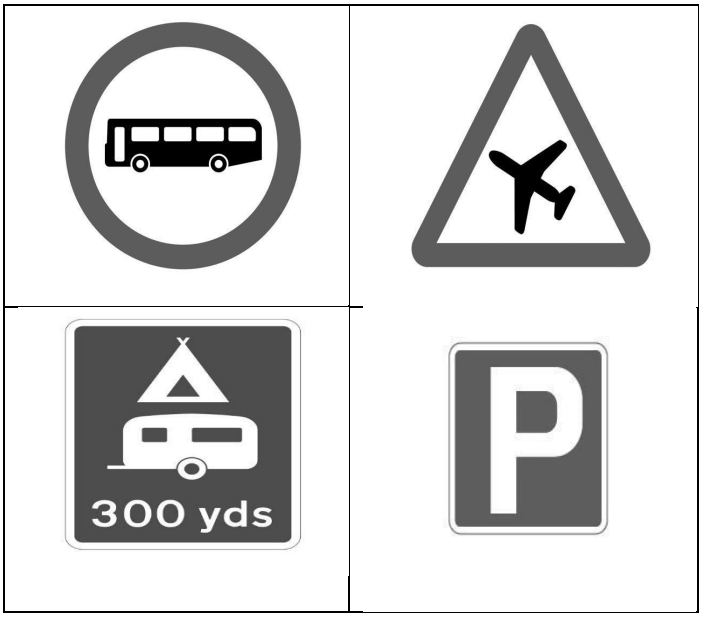


Figure 8 Training image database Image Denoising using Deep Neural Network (DNN) Results

The comparative analysis of various image denoising techniques based on various image quality metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the image quality (SSIM) from the past 2-3 years comprising input testing image database is given in the tabulated fashion which is shown (Table 1) as below:

Table 1: Comparative Analysis of Image Denoising Techniques based on various Image Quality Metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) for Testing image database from the past 2-3 years

S. No.	Year	Techniques used	Mean Squared Error (MSE) Values	Peak Signal to Noise Ratio (PSNR) Values	Structural Similarity for measuring image Quality (SSIM) Values
1.	2018	Anisotropic Diffusion Filtering Type 1	40.282	32.121	0.892
2.	2018	Non-Local Means Type 2	0.676	51.479	0.995
3.	2017	Image Denoising using Deep Neural Network (DNN)	1.640	47.1174	0.997

The comparative analysis of various image denoising techniques based on various image quality metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the image quality (SSIM) from the past 2-3 years comprising input training image database is given in the tabulated fashion which is shown (Table 2) as below:

Table 2: Comparative Analysis of various combinations of Image Denoising Techniques based on various Image Quality Metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) for Training image database from the past 2-3 years

S. No.	Year	Techniques used	Mean Squared Error (MSE) Values	Peak Signal to Noise Ratio (PSNR) Values	Structural Similarity for measuring image Quality (SSIM) Values
1.	2018	Anisotropic Diffusion Filtering Type 1	110.887	27.897	0.877
2.	2018	Non-Local Means Type 2	0.080	61.175	0.999
3.	2017	Image Denoising using Deep Neural Network (DNN)	0.970	51.267	0.998

3.4. Arithmetic Mean Filter

The Arithmetic Mean can be defined as being equal to the sum of the numerical values of each and every observation divided by the total number of observations. The arithmetic means of a given set of numbers each one is having a numerical value a_1, a_2, a_3, ……..., a_n can be calculated by using the formula which is given as below:

$$\text{Arithmetic Mean (A.M.)} = a_1 + a_2 + a_3 + \ldots\ldots\ldots + a_n/n \quad (7)$$

where a_1, a_2, a_3,……., a_n are n number of physical quantities, n is the total number of physical quantities and A.M. is the Arithmetic Mean of the n number of physical quantities which are taken into consideration.

The results which are obtained after carrying out experiments with the help of an Arithmetic Mean Filter (Figure 9) are presented as below:



Figure 9 Testing image database Arithmetic Mean Filtering Results

3.5. Circular Averaging Filter (Pillbox)

The technique of removing salt and pepper noise from an image can be accomplished with the help of smoothing filter which also helps in removing the sharp edges thereby refining the small contours as well as ridges from it. It is useful in removing minute information before the operation of extracting an object having greater size from the image resulting in the amalgamation of tiny regions in between the two portions that are present in lines or curves, thereby telling its significance about the reason for including in the pre-processing step. The results that are obtained after applying Averaging Filter to the roadway symbols that are present in our image dataset are presented in the below section of this manuscript (Figure 10) as follows:



Figure 10 Testing image database Circular Averaging Filter (Pillbox) Results

3.6. Harmonic Mean Filter

A Harmonic Mean Filter can be defined as the reciprocal of the arithmetic mean of the reciprocals of a given set of observations. The Harmonic Mean can be calculated by using the mathematical expression which is given as below:

$$\text{Harmonic Mean (H.M.)} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots\ldots\ldots + \frac{1}{x_n} \quad (8)$$

The results which are obtained after carrying out experiments with the help of a Harmonic Mean Filter (Figure 11) are presented as below:

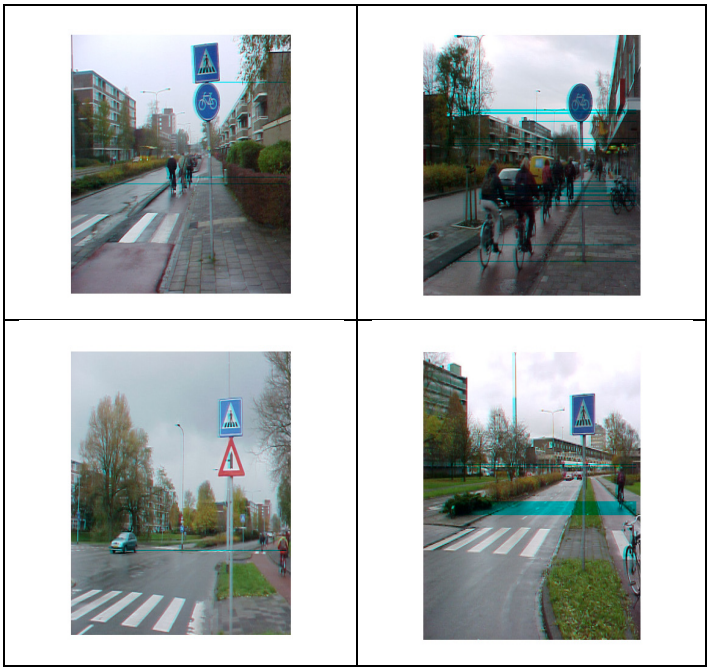


Figure 11 Testing image database Harmonic Filter Results

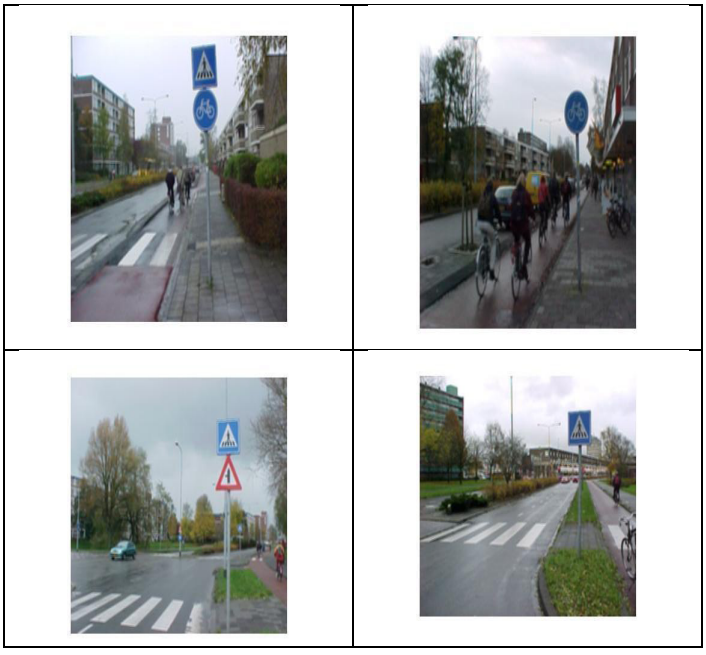


Figure 12 Testing image database Box Filter Results

3.7. Box Filter

The Box Blur or sometimes called as Box Linear filter can be defined as a spatial domain linear filter in which each pixel in the resulting image possess a value which is equal to the average value of its neighboring pixels present in the original input image. It is a type of low pass or a blurring filter.

The results obtained after carrying out experiments with the help of a Box Filter are presented in Figure 12.

3.8. Gaussian Filter

The Gaussian function is defined as follows:

$$g(x, y) = 1/2\pi\sigma^2 e^{-(x2 + y2)/2\sigma2} \quad (9)$$

where a, b and c are the arbitrary constants, x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis and σ is the standard deviation of the Gaussian distribution. The Gaussian function is derived from the name of a well-known scientist named Carl Friedrich Gauss. The results which are obtained after carrying out experiments with the help of a Gaussian Filter (Figure 13) are presented as below:



Figure 13 Testing image database Gaussian Filter Results

The comparative analysis of various image denoising techniques based on various image quality metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) comprising the input testing image database is given in a tabulated fashion which is shown (Table 3) as below:

Table 3 Comparative Analysis of various Image Denoising Techniques based on various Image Quality Metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) for Testing image database from the research paper presented in IEEE ESTCON-ICIAS-2018 International Conference held at UTP, Malaysia from 13th-15th August 2018

S. No.	Year	Techniques used	Mean Squared Error (MSE) Values	Peak Signal to Noise Ratio (PSNR) Values	Structural Similarity for measuring image Quality (SSIM) Values
1.	2006	Gaussian Filter	22.820	35.054	0.985
2.	2006	Box Filter	529.370	21.329	0.634
3.	-	Arithmetic Mean Filter	381.839	22.656	0.729
4.	2006	Circular Averaging Filter (Pillbox)	459.299	21.961	0.666
5.	-	Harmonic Mean Filter	321.160	24.146	0.862

4. Novelty of the Proposed Research Work

The main objective of our proposed research work is to identify the best technique which helps in removing the noise which is naturally present in an image among various noise removal techniques based on different image quality metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring Image Quality (SSIM).

The best technique is identified from its counterparts by assuming that it is having the least numerical value of Mean Squared Error (MSE) as well as it is also having the greatest numerical value of Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring Image Quality (SSIM).

The Non-Local Means Type 2 and the Image Denoising using Deep Neural Network (DNN) techniques were combined and applied to the input traffic sign board images which are present in the testing image database and the results so obtained are shown in (Figure 14) as below:

After the process of calculating the numerical values of Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) was completed, it was observed that the Mean Squared Error (MSE) value was found out to be 113.83027446809 and the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) values were found out to be 31.578017021277 and 0.90815744680851 respectively.

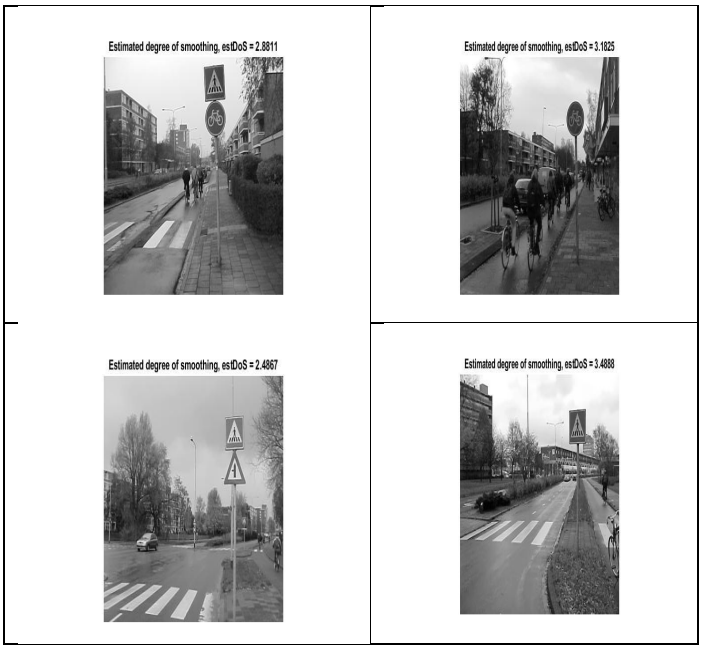


Figure 14 Testing image database Results obtained by combining the Non-Local Means Type 2 + Image Denoising using Deep Neural Network (DNN) Techniques

The technique of Anisotropic Diffusion Filtering Type 1 was combined with Non-Local Means Type 2 technique and applied to the input traffic sign board images which are present in the testing image database and the results so obtained are shown in (Figure 15) as below:

After the process of calculating the numerical values of Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) was completed, it was observed that the Mean Squared Error (MSE) value was found out to be 1.5655462333826 and the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) values were found out to be 47.416585376662 and 0.99865745937961 respectively.

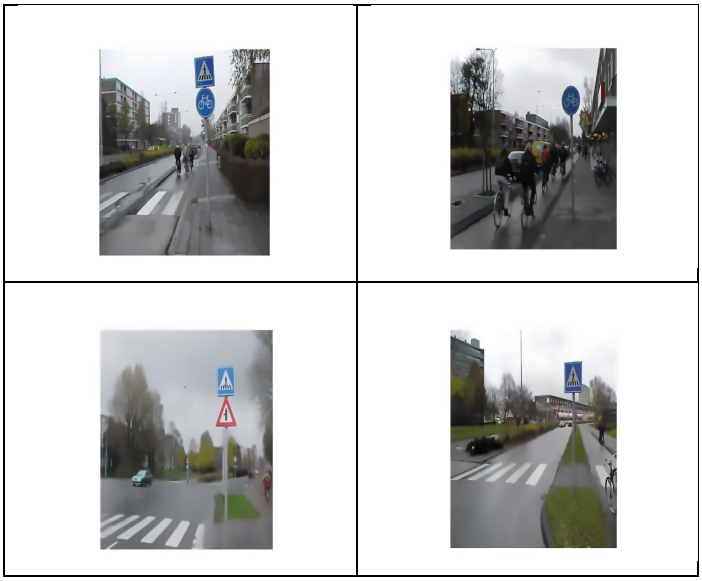


Figure 15 Testing image database Results obtained by combining the Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2 Techniques

The techniques of Non-Local Means Type 2 and Anisotropic Diffusion Filtering Type 1 were combined and applied to the input traffic sign board images which are present in the training image database and the results were obtained after the execution of this experimental procedure which are shown in the figure that is given (Figure 16) as below and the numerical values of the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) were found out to be 255.94895760709, 24.440101624815 and 0.95028419497784 respectively.

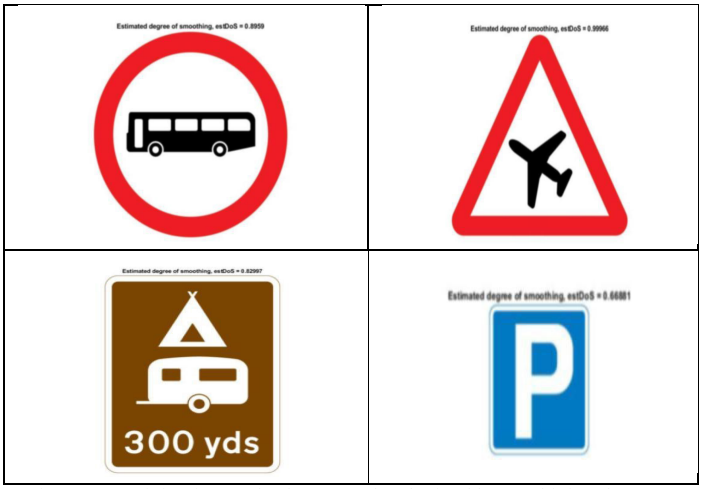


Figure 16 Training image database Results obtained by combining the Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1 Techniques

The Non-Local Means Type 2, Anisotropic Diffusion Filtering Type 1 and Non-Local Means Type 2 techniques were combined and applied to the input traffic sign board images which are present in the training image database and the results which are obtained after the execution of the experimental procedure are shown in the figure which is given (Figure 17) as below and the numerical values of the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) were found out to be 76.349860118168, 31.671063072378 and 0.97056218611521 respectively.

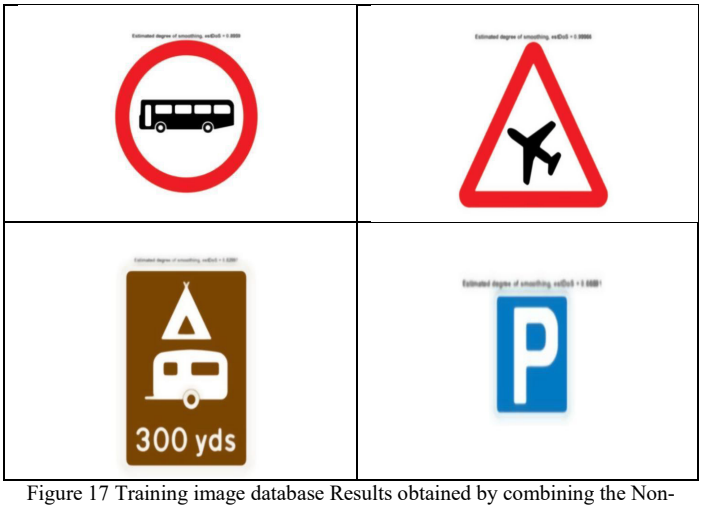


Figure 17 Training image database Results obtained by combining the Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2 Techniques

The techniques of Non-Local Means Type 2, Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and Image Denoising using Deep Neural Network (DNN) were combined and applied to the input traffic sign board images which are present in the training image database and the results which are obtained after carrying out the experimental procedure are shown in the figure which is given (Figure 18) as below and the numerical values of the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) were found out to be 0.022547858197932, 66.103574150665 and 0.99989054652881 respectively.

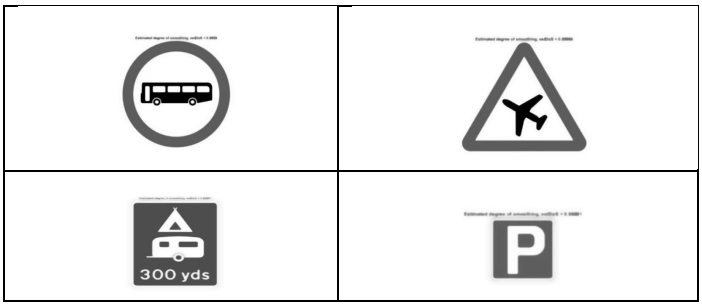


Figure 18 Training image database Results obtained by combining the Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2 + Image Denoising using Deep Neural Network (DNN) Techniques

The comparison of these image quality parameter numerical values for different combinations of Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and the Image Denoising using Deep Neural Network (DNN) techniques which are employed for removal of noise present in the testing image database is given in the tabulated fashion shown (Table 4) as below:

The comparison of these image quality parameter numerical values for different combinations of Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and the Image Denoising using Deep Neural Network (DNN) techniques which are employed for removal of noise present in the training image database is given in the tabulated fashion shown in Table 5:

Table 4 Comparative Analysis of various combinations of Image Denoising Techniques based on various Image Quality Metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) for Testing image database

S. No.	Year	Techniques used	Mean Squared Error (MSE) Values	Peak Signal to Noise Ratio (PSNR) Values	Structural Similarity for measuring Image Quality (SSIM) Values
1.	2018	Anisotropic Diffusion Filtering Type 1	40.282	32.121	0.892
2.	2018	Non-Local Means Type 2	0.676	51.479	0.995
3.	2017	Image Denoising using Deep Neural Network (DNN)	1.640	47.117	0.997
4.	2018 + 2018	Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2	113.830	31.578	0.908
5.	**2018 + 2018 + 2017**	**Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2 + Image Denoising using Deep Neural Network (DNN)**	**0.137**	**61.466**	**0.999**

Table 5 Comparative Analysis of various combinations of Image Denoising Techniques based on various Image Quality Metrics like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity for measuring the Image Quality (SSIM) for Training image database

S. No.	Year	Techniques used	Mean Squared Error (MSE) Values	Peak Signal to Noise Ratio (PSNR) Values	Structural Similarity for measuring Image Quality (SSIM) Values
1.	2018	Anisotropic Diffusion Filtering Type 1	110.887	27.897	0.877
2.	2018	Non-Local Means Type 2	0.080	61.175	0.999
3.	2017	Image Denoising using Deep Neural Network (DNN)	0.970	51.267	0.998
4.	2018 + 2017	Non-Local Means Type 2 + Image Denoising using Deep Neural Network (DNN)	1.565	47.416	0.998
5.	2018 + 2018	Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1	255.948	24.440	0.950
6.	2018 + 2018 + 2018	Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2	76.349	31.671	0.970
7.	**2018 + 2018 + 2018 + 2017**	**Non-Local Means Type 2 + Anisotropic Diffusion Filtering Type 1 + Non-Local Means Type 2 + Image Denoising using Deep Neural Network (DNN)**	**0.022**	**66.103**	**0.999**

If the testing image database is considered, it is observed that if the techniques of Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and Image Denoising using Deep Neural Network (DNN) are combined, the numerical values of different image quality metrics like the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) were found out to be 0.13772127659574, 61.466536170213 and 0.99981489361702 respectively. Therefore, it is clear that the above technique which is mentioned in this research paper proves to be the best technique which can be used for the purpose of noise removal from input traffic sign board images present in the testing image database since the numerical values of Mean Squared Error (MSE) is the least as compared with its other counterparts and the numerical values of the Peak Signal to Noise Ratio (PSNR) as well as the Structural Similarity for measuring the Image Quality (SSIM) are the greatest as compared with its other counterparts.

If the training image database is considered, it is observed that if the techniques of Non-Local Means Type 2, Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and Image Denoising using Deep Neural Network (DNN) are combined, the numerical values of different image quality metrics like the Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) and the Structural Similarity for measuring the Image Quality (SSIM) were found out to be 0.022547858197932, 66.103574150665 and 0.99989054652881 respectively. Therefore, it is clear that the above technique which is mentioned in this research paper proves to be the best technique which can be used for the purpose of noise removal from input traffic sign board images present in the training image database since the numerical values of different image quality metrics like the Mean Squared Error (MSE) is the least as compared with its other counterparts and the numerical values of the Peak Signal to Noise Ratio (PSNR) as well as the Structural Similarity for measuring the Image Quality (SSIM) are the greatest as compared with its other counterparts.

Conclusion and Future Scope

In this paper, a comparison of various image denoising techniques was done in order to choose the best technique among a set of three different techniques and their three different combinations which was being applied to the traffic sign board images present in the testing image database and four different techniques and their seven different combinations to the traffic sign board images present in the training image database based on calculating different parameters which decide the quality of input roadway symbols like Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR) as well as Structural Similarity for measuring Image Quality (SSIM) and after a thorough comparison, it was concluded that the combination of Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and Image Denoising using Deep Neural Network (DNN) techniques which was being applied to the traffic sign board images present in the testing image database and the combination of Non-Local Means Type 2, Anisotropic Diffusion Filtering Type 1, Non-Local Means Type 2 and Image Denoising using Deep Neural Network (DNN) techniques which was being applied to the images which are present in the training image database has the maximum Peak Signal to Noise Ratio (PSNR) as well as Structural Similarity for measuring Image Quality (SSIM) values and a minimum Mean Squared Error (MSE) value as compared to other techniques. These techniques will be applied to the traffic sign board images which are present in both the testing as well as training image databases in future research work for the purpose of noise removal from input traffic sign board images.

Conflict of Interest

The authors mention and acknowledge the funding sources which were provided during the course of performing the research work and there are no significant conflicts of interest.

Acknowledgments

A sincere vote of thanksgiving must be extended to Vellore Institute of Technology, Vellore administration and every one among the stakeholders whose contribution towards the completion of this work in a successful fashion paved the way for implementing it on a broad level.

References

[1] Yu-Mei Huang, Hui-Yin Yan, You-Wei Wen, Xue Yang, “Rank minimization with applications to image noise removal”, Elsevier- Information Sciences, Volume 429, March 2018, pp. 147-163.

[2] Anissa Selmani, Hassene Seddik, “Anisotropic smart shape-adapted image smoothing without conductance function efficient for impulse noise removal”, Elsevier- Digital Signal Processing, Volume 75, April 2018, pp. 83-95.

[3] Tingting Yang, Xiang Long, Arun Kumar Sangaiah, Zhigao Zheng, Chao Tong, “Deep detection network for real-life traffic sign in vehicular networks”, Elsevier- Computer Networks, Volume 136, March 2018, pp. 95-104.

[4] Yong Chen, Ting-Zhu Huang, Liang-Jian Deng, Xi-Le Zhao, Min Wang, “Group sparsity-based regularization model for remote sensing image stripe noise removal”, Elsevier- Neurocomputing, Volume 267, 6 December 2017, pp. 95-106.

[5] Jing Dong, Zifa Han, Yuxin Zhao, Wenwu Wang, Ales Prochazka, Jonathon Chambers, “Sparse analysis model based multiplicative noise removal with enhanced regularization”, Elsevier- Signal Processing, Volume 137, August 2017, pp. 160-176.

[6] Lianghai Jin, Zhiliang Zhu, Xiangyang Xu, Xiang Li, “Two-stage quaternion switching vector filter for color impulse noise removal”, Elsevier- Signal Processing, Volume 128, November 2016, pp. 171-185.

[7] De-Yong Lu, “A hybrid optimization method for multiplicative noise and blur removal”, Elsevier- Journal of Computational and Applied Mathematics, Volume 302, 15 August 2016, pp. 224-233.

[8] Xianquan Zhang, Feng Ding, Zhenjun Tang, Chunqiang Yu, “Salt and pepper noise removal with image inpainting”, Elsevier- AEU - International Journal of Electronics and Communications, Volume 69, Issue 1, January 2015, pp. 307-313.

[9] Jielin Jiang, Jian Yang, Yan Cui, Lei Luo, “Mixed noise removal by weighted low rank model”, Elsevier- Neurocomputing, Volume 151, Part 2, 5 March 2015, pp. 817-826.

[10] Tudor Barbu, “Robust Anisotropic Diffusion Scheme for Image Noise Removal”, Elsevier- Procedia Computer Science, Volume 35, September 2014, pp. 522-530.

[11] Xia Chen, Chen Tang, Xiusheng Yan, “Switching degenerate diffusion PDE filter based on impulse like probability for universal impulse noise removal”, Elsevier- AEU - International Journal of Electronics and Communications, Volume 68, Issue 9, September 2014, pp. 851-857.

[12] Xiubao Sui, Qian Chen, Guohua Gu, "Adaptive grayscale adjustment-based stripe noise removal method of single image", Elsevier- Infrared Physics & Technology, Volume 60, September 2013, pp. 121-128.
[13] Floris Gisolf, Anwar Malgoezar, Teun Baar, Zeno Geradts, "Improving source camera identification using a simplified total variation-based noise removal algorithm", Elsevier- Digital Investigation, Volume 10, Issue 3, October 2013, pp. 207-214.
[14] Jourabloo, A. H. Feghahati, M. Jamzad, "New algorithms for recovering highly corrupted images with impulse noise", Elsevier- Scientia Iranica, Volume 19, Issue 6, December 2012, pp. 1738-1745.
[15] Lianghai Jin, Caiquan Xiong, Hong Liu, "Improved bilateral filter for suppressing mixed noise in color images", Elsevier- Digital Signal Processing, Volume 22, Issue 6, December 2012, pp. 903-912.
[16] Wei Ding, "A New Method for Image Noise Removal using Chaos-PSO and Nonlinear ICA", Elsevier- Procedia Engineering, Volume 24, December 2011, pp. 111-115.
[17] Tom Mélange, Mike Nachtegael, Stefan Schulte, Etienne E. Kerre, "A fuzzy filter for the removal of random impulse noise in image sequences", Elsevier- Image and Vision Computing, Volume 29, Issue 6, May 2011, pp. 407-419.
[18] Ovidiu Ghita, Paul F. Whelan, "A new GVF-based image enhancement formulation for use in the presence of mixed noise", Elsevier- Pattern Recognition, Volume 43, Issue 8, August 2010, pp. 2646-2658.
[19] K. Kannan, B. Rajesh Kanna, C. Aravindan, "Root Mean Square filter for noisy images based on hyper graph model", Elsevier- Image and Vision Computing, Volume 28, Issue 9, September 2010, pp. 1329-1338.
[20] Huaping Liu, Yulong Liu, Fuchun Sun, "Traffic sign recognition using group sparse coding", Elsevier- Information Sciences, Volume 266, January 2014, pp. 75-89.
[21] Zhan-Li Sun, Han Wang, Wai-Shing Lau, Gerald Seet, Danwei Wang, "Application of BW-ELM model on traffic sign recognition", Elsevier- Neurocomputing, Volume 128, October 2014, pp. 153-159.
[22] Fatin Zaklouta, Bogdan Stanciulescu, "Real-time traffic sign recognition in three stages", Elsevier- Robotics and Autonomous Systems, Volume 62, August 2012, pp. 16-24.
[23] Shuihua Wang, Hangrong Pan, Chenyang Zhang, Yingli Tian, "RGB-D image-based detection of stairs, pedestrian crosswalks and traffic Signs", Elsevier- Journal of Visual Communication and Image Retrieval, Volume 25, November 2014, pp. 263-272.
[24] Jonathan J. Kay, Peter T. Savolainen, Timothy J. Gates, Tapan K. Datta, "Driver behavior during bicycle passing maneuvers in response to aShare the Road sign treatment", Elsevier-Accident Analysis and Prevention, Volume 70, April 2014, pp. 92-99.
[25] Jesmin Khan, Sharif Bhuiyan, Reza Adhami, "Hierarchical clustering of EMD based interest points for road sign detection", Elsevier- Optics & Laser Technology, Volume 57, October 2014, pp. 271-283.
[26] Zong-Yao Chen, Wei-Chao Lin, Shih-Wen Ke, Chih-Fong Tsai, "Evolutionary feature and instance selection for traffic sign recognition", Elsevier- Computers in Industry, Volume 74, September 2015, pp. 201-211.
[27] Samuele Salti, Alioscia Petrelli, Federico Tombari, Nicola Fioraio, Luigi Di Stefano, "Traffic sign detection via interest region extraction", Elsevier- Pattern Recognition, Volume 48, June 2014, pp. 1039-1049.
[28] J.M. Lillo-Castellano, I. Mora-Jiménez, C. Figuera-Pozuelo, J.L.Rojo-Álvarez, "Traffic sign segmentation and classification using statistical learning methods", Elsevier- Neurocomputing, Volume 153, November 2015, pp. 286-299.
[29] Haojie Li, Fuming Sun, Lijuan Liu, Ling Wang, "A novel traffic sign detection method via color segmentation and robust shape matching", Elsevier- Neurocomputing, Volume 169, May 2015, pp. 77-88.
[30] Zhenyu An, Zhenwei Shi, Ying Wu, Changshui Zhang, "A novel unsupervised approach to discovering regions of interest in traffic images", Elsevier- Pattern Recognition, Volume 48, February 2015, pp. 2581-2591.
[31] Yingying Zhu, Chengquan Zhang, Duoyou Zhou, Xinggang Wang, Xiang Bai, Wenyu Liu, "Traffic sign detection and recognition using fully convolutional Network guided proposals", Elsevier- Neurocomputing, Volume 214, July 2016, pp. 758-766.
[32] Ayou Ellahyani, Mohamed El Ansari, Ilyas El Jaafari, "Traffic sign detection and recognition based on random forests", Elsevier- Applied Soft Computing, Volume 46, February 2016, pp. 805-815.
[33] Selcan Kaplan Berkaya, Huseyin Gunduz, Ozgur Ozsen, Cuneyt Akinlar, Serkan Gunal, "On circular traffic sign detection and recognition", Elsevier- Expert Systems with Applications, Volume 48, 2016, pp. 67-75.
[34] Yongtao Yu, Jonathan Li, Chenglu Wen, Haiyan Guan, Huan Luo, Cheng Wang, "Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data", Elsevier- ISPRS Journal of Photogrammetry and Remote Sensing, Volume 113, January 2016, pp. 106-123.
[35] Hamed Habibi Aghdam, Elnaz Jahani Heravi, Domenec Puig, "A practical approach for detection and classification of traffic signs using Convolutional Neural Networks", Elsevier- Robotics and Autonomous Systems, Volume 84, July 2016, pp. 97-112.
[36] Mario Soilan, Belen Riveiro, Joaquin Martinez-Sanchez, Pedro Arias, "Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory", Elsevier- ISPRS Journal of Photogrammetry and Remote Sensing, Volume 114, February 2016, pp. 92-101.
[37] Y.Ouerhani, A. Alfalou, M. Desthieux, C. Brosseau, "Advanced driver assistance system: Road sign identification using VIAPIX system and a correlation technique", Elsevier- Optics and Lasers in Engineering, Volume 89, May 2016, pp. 184-194.
[38] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs", IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 3, June 2015, pp. 1360-1369.
[39] Nadra Ben Romdhane, Hazar Mliki, Mohamed Hammami, "An Improved Traffic Signs Recognition and Tracking Method for Driver Assistance System", in the Proceedings of IEEE International Conference on Information Security (ICIS'2016), June 26-29, 2016, Okayama, Japan.
[40] Ruben Laguna, Ruben Barrientos, L. Felipe Blazquez, Luis J. Miguel, "Traffic sign recognition application based on image processing techniques", in the Proceedings of the Elsevier- Proceedings of the 19th World Congress, The International Federation of Automatic Control, Cape Town, South Africa. August 24-29, 2014, pp. 104-109.
[41] Tingting Yang, Xiang Long, Arun Kumar Sangaiah, Zhigao Zheng, Chao Tong, "Deep detection network for real-life traffic sign in vehicular networks", Elsevier- Computer Networks, Volume 136, May 2018, pp. 95-104.
[42] S.M. Bascon, J.A. Rodriguez, S.L. Arroyo, A.F. Caballero, F. Lopez-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs", Computer Vision Image Understanding, Volume 114, Issue 3, 2010, pp. 373-383.
[43] C.C. Chang, Y.P. Hsieh, "A fast VQ codebook search with initialization and search order", Information Sciences, Volume 183, Issue 1, 2012, pp. 132-139.
[44] Ciresan, U. Meier, J. Mascim, J. Schmidhuber, "A committee of neural networks for traffic signs classification", in Proceedings of International Joint Conference on Neural Networks (IJCNN), July 2011, pp. 1918-1921.
[45] S.Escalera, O. Pujol, P. Radeva, "Traffic sign recognition system with b-correction", Elsevier- Machine Vision Applications, Volume 21, Issue 2, 2010, pp. 99-111.
[46] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition", IEEE Transactions on Intelligent Transportation Systems, Volume 11, Issue 4, 2010, pp. 917-930.
[47] Y. Gu, T. Yendo, M.P. Tehrani, T. Fujii, M. Tanimoto, "Traffic sign detection in dual-focal active camera system", in the Proceedings of IEEE Intelligent Vehicles Symposium (IV), 2011, pp. 1054-1059.
[48] Haojie Li, Fuming Sun, Lijuan Liu, Ling Wang, "A novel traffic sign detection method via color segmentation and robust shape matching", Elsevier- Neurocomputing, Volume 169, December 2015, pp. 77-88.
[49] Y. Huang, K. Huang, Y. Yu, T. Tan, "Salient coding for image classification", in the Proceedings of Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1753-1760.
[50] Alexander Shustanov, Pavel Yakimov, "CNN Design for Real-Time Traffic Sign Recognition", Elsevier- September 2017, pp. 718-725.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 561-573 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Lightweight, Hardware-Based Support for Isolation in Mixed-Criticality Network-on-Chip Architectures

Giacomo Valente, Paolo Giammatteo, Vittoriano Muttillo, Luigi Pomante*, Tania Di Mascio

Center of Excellence DEWS, Università degli Studi dell'Aquila, 67100 L'Aquila, Italy

ARTICLEINFO

Article history:
Received: 15 June, 2019
Accepted: 31 July, 2019
Online: 25 August, 2019

Keywords:
Network on Chip
Isolation
Mixed-Criticality
Hardware Support

ABSTRACT

Spatial and temporal isolation is a crucial issue in embedded systems executing multiple tasks with several levels of criticality. This is considerably significant in the context of multi-processor (or multi-core) embedded systems running multiple mixed-criticality applications in parallel. This work deals with the issue of isolation of different application classes on Network on Chip (NoC) architectures and proposes a lightweight hardware mechanism able to support mixed-criticality requirements and specifically designed to be introduced into existing network interfaces. This mechanism supports the execution of different and contemporary applications with several criticality levels by supervising the messages exchange among network nodes, with the introduction of limited hardware and software overhead on the monitored network. The proposed solution is described and evaluated by means of logical simulations and an implementation on reconfigurable logic, using a reference NoC architecture with mesh topology. Scalability of the proposed approach is also discussed and evaluated by means of network simulations. Results show an area occupation less than 1% in a 3x3 mesh NoC, and a good scalability of the proposed mechanism in an 8x8 mesh network, indicating it as a valid lightweight solution able to enforce isolation in NoCs.

1. Introduction

It is possible to define the criticality of the generic component of a system as the level of assurance needed for that element [1]. Embedded system is a clear example of mixed-criticality system if different software elements executing on the same hardware platform possess two or more levels of criticality. The fundamental problems in mixed-criticality system management are (i) the simultaneous satisfaction of discordant requirements regarding advantageous access to shared resources, in order to make the most of performance of the system, (ii) and a strict partitioning to avoid disorder between different elements.

Regarding single-processor embedded systems, it is essential to guarantee the isolation between tasks in terms of time, and several techniques able to ensure temporal determinism of running have been presented in scientific literature [2]. On the other hand, about multi-processor systems, different applications run in parallel on different processors. The applications must compete in accessing shared resources, taking advantage of the specific communication architectures of different hardware implementations. Recently, it was possible to observe a growing interest towards the usage of Network on Chip (NoC) architectures [3] as a platform for systems with high mixed criticality. Compared to traditional shared (hierarchical) bus solutions, NoCs can more efficiently support the implementation of multi and many core systems. Table 1 (inspired by the works presented in [4] and [5]) provides a qualitative comparison of the two approaches.

In this prospective, the present work shows a mechanism with lightweight hardware for the management of mixed criticality in a multi-processor embedded system conceived on a Network on Chip. In particular, the aim is to give a hardware-based support in order to control the flow of the messages through the NoC, with the goal of guaranteeing a reserved access to the NoC itself on the base of the criticality level of the different tasks running on the NoC nodes. In such a way, the proposed mechanism supports spatial and temporal isolation, since, in a given period, only a subset of the tasks can access the NoC and the connected resources. Indeed, temporal isolation ensures no interference in the time domain among tasks to access to a shared resource (i.e., mutual exclusive access), while spatial isolation protects a shared resource

*Luigi Pomante, Email: luigi.pomante@univaq.it

Table 1: Shared bus vs. NOC: a qualitative comparison

	Concurrent access to different resources	Throughput	Physical implementation Overhead	Scalability	Design complexity	Power dissipation	Area occupation	Max frequency
Shared Bus	-	-	=	-	+	-	-	-
NoC	+	+	=	+	-	+	+	+

so that tasks are able to freely perform access only to a part of it, or only a subset of existing tasks are able to access to that resource [6].

The main characteristics of the proposed mechanism are the following ones:

- limited effect on system performance, thanks to a hardware-based approach;
- no limit on the number of criticality levels in the system;
- supporting both spatial and temporal isolation.

In particularly, the innovation points of the proposed approach are the followings:

- independency from NoC topology, i.e., it can be adapted to each NoC scenario;
- portability among different networks.

It is worth noting that this paper extends the one presented in [7], by providing:

- detailed HW design, described in Hardware Description Language (HDL), together with logical simulation (sec. 4);
- extended scalability analysis by means of OMNET++ simulations (sec. 5 and 6).

The paper is outlined in this way. Section 2 describes the most important concepts related to the management of isolation in the context of mixed-criticality in multi-processor and NoC based systems; the section also briefly summarizes the current main existing strategies for the effective management of the different criticality levels. Section 3 shows the proposed mechanism and details its advantages. Section 4 illustrates the detailed HDL-based HW design and simulation issues. Section 5 and 6 show the analysis of a simulated platform that helps to test the effectiveness of the proposed mechanism for NoC with more nodes. Finally, Section 7 summarizes conclusions and discusses future works.

2. Isolation in Mixed-Criticality Systems

Spatial and temporal isolation is a fundamental issue for embedded systems. In particular, for those systems that perform multiple activities with different levels of criticality, or better to say, mixed critical embedded systems. The topic is particularly significant for areas such as aeronautics and automotive. Several standards have the specific purpose of finding a solution to this problem and two significant examples are the AUTOMOTIVE OPEN SYSTEM ARCHITECTURE (AUTOSAR, [8]), a software architecture in the automotive area, and the Aeronautical Radio INCorporated (ARINC, [9]), which gives the specifications for spatial and temporal sectioning in avionics applications critical for safety (ARINC 653).

In order to examine the issue of mixed criticality, different analytical models can be followed [1]. Usually, applications are modelled as a group of elements. Each one consists of a finite number of tasks. Each task is expected on a processing resource, that can be shared among different tasks. For the present analysis, each task executes a specific activity by performing a certain number of jobs and it is periodically scheduled on a shared processing resource.

Focusing on criticality, several schemes have been presented in literature defining of levels of criticality, from the easiest one, where only two levels are allowed (i.e., non-critical and critical), up to configurations where the number of allowed levels is potentially unlimited [10]. Focusing on isolation, the tasks belonging to components with a lower criticality level shall not be able to interfere with higher criticality ones.

In literature, different studies have analyzed the problem of the management of mixed criticality in single processor systems, from the isolation point of view [2]. A static allocation of memory during compilation phase is an appropriate strategy in order to reach spatial isolation. Contrarily, a Memory Management Unit and a Memory Protection Unit could support isolation for dynamic memory allocation. Another appropriate strategy in order to reach isolation between tasks with several criticality levels is virtualization. Anyway, full virtualization is generally not satisfactory for embedded systems (especially if exists real-time constraints), indeed the required overhead may be impactful on the temporal constraints of the application. The usage of Hypervisors [11] in these situations permits to run ate the same time several operating systems upon a platform in sharing with low overhead, but still maintaining the isolation of time and space [12].

It is fundamental in a multi-processor system with shared bus architecture, administer the access to shared communication elements. This could be done by partitioning the system in order to eliminate disturbances among applications executing on different cores, or on peripherals device with DMA. The issue of partitioning in multi-processor systems is already addressed by Pellizzoni et al. [13], with the definition of the Architectural Analysis and Design Language, which is a form of Architectural Description Language used for mixed-critical systems that supports by construction the monitoring and optimization of the communication and processing phase. The time-triggered model [14] is a different method of partitioning, where a high level of criticality is related to the time-triggered traffic, while traffic

Table 2: Routing arbitration policies for NoCs.

Arbitration Policy	Round Robin	Time Division Multiplexing	Fixed Priority
Description	Usually aimed at obtaining fairness on arbitration, but not used in real-time systems, given the problems in calculating the worst-case latencies for the various transactions.	The various data flows are statically allocated to separated time slots. This arbitration policy is largely adopted in real-time and mixed criticality implementations, but determining a reservation scheme for the various time slots is not trivial.	The various transactions are managed based on predefined priority levels.

triggered by events is generally the best effort. Time-Triggered Ethernet or TTP/C are kind of protocols that support this kind of communication mechanism called Time-Division Multiple Access (TDMA). If no isolation-oriented strategies are implemented, the management of shared resources is usually demanded to the specific bus arbitration.

Finally, in a multi-processor system based on NoC, router arbitration schemes are fundamental. Generally, NoCs are conceived on the arbitration policies shown in Table 2.

2.1. State of Art

Schoeberl [15] suggested a regular structured time-triggered NoC (TDMA-based) capable to support foreseeable communications both on-chip and off-chip. This architecture uses a pseudo-static communication schedule implemented in a Cyclone II EP2C35 *Field-Programmable Gate Array* (FPGA) on the Altera DE2 board. The underlying network topology is simple (e.g., ring structure) and optimized for easy routing (wire routing, not message routing). They implement a simple demo application (a voting triple modular redundancy sensor), but they evaluate the scheduling policy offline for each different scenario. An advanced tool to calculate and verify the schedule is important to render the proposed NoC useful.

Tobuschat [16] developed a NoC capable to support natively a mixed criticality. This system is conceived on a methodology (namely, back suction) capable to maximize the bandwidth allocated for low importance messages, guarantying that the most critical messages are delivered by satisfying the related deadlines. The authors said that sufficient independence is reached, and worst-case behavior can be predicted through the usage of virtualization, monitoring infrastructure, and control mechanisms. Programming of the monitors is only possible by the system controller, so its correct implementation becomes an essential point of this work, introducing additional load to the system. Furthermore, the knowledge of all possible interference enables timing predictable behavior of the whole system, but this assumption is not simple to verify with respect to real scenarios.

Burns [17] described the Wormhole Protocol for Mixed Criticality (WPMC). This protocol points out the dual-criticality, fixed priority NoCs. Furthermore, if an infraction in transaction deadlines is catched, it permits to limit the use of communication elements in favor of high criticality transactions. Successively, WPMC has been updated in order to advance the low criticality packets average latency and the latency of the worst-case of the high criticality ones [18]. The main limitations of this work are related to the maximum number of criticality levels considered (not more than two), the lack of a mode change protocol among several criticality levels, and the study of mixed-criticality end-to-end latency analysis (i.e., considering task execution as well as traffic-flows).

More in general, state of art solutions to provide isolation converge on two main areas: Time Division Multiplexing, that implies a conservative design with increased resource requirements [15], and Monitoring of the System, to react in case of unexpected situations.

The mechanism proposed in this paper falls in the second category, with the novelty that the monitoring mechanism is moved into the NoC itself, involving only NoC interconnection elements and not NoC nodes. This specific aspect will be analyzed, in more detail, in the following section. In this way, the introduced mechanism is independent from the NoC topology, as the control action can be applied independently from the specific NoC architecture. The mechanism is also easily portable, as it is simple to introduce custom communication control strategies into existing network interfaces. These are the two main advances respect to the state of art.

3. Proposed HW Support for Isolation

This paper proposes a mechanism able to consider the different criticality levels of tasks running on a NoC, and to regulate the network traffic basing on specific network parameters. The system model is firstly introduced, then the proposed mechanism is detailed and analyzed.

3.1. System Model

In the proposed approach, a NoC consists of, at least, two *Nodes* (N) and one or more *Routers* (R). Every node includes one or more processors/cores, memories, and other peripherals. Intra-node communications can entrust on several approaches (generally a hierarchical bus and shared-memory structure). These internal details are not a constraint for the proposed structure, so they will be considered abstracted from now on. Inter-node communications are conceived on message passing: this implies that every node possess a *Network Interface* (*NI*), exploited to send/receive messages, and linked to a single router port. Every single router can be linked, to other *NI* and/or to other routers, considering on the NoC topology. The routers transmit the messages to final nodes according to the used routing protocol.

In Figure 1, it is shown a schematic reference NoC, consisting of four nodes and routers connected in a mesh topology. Every node of the NoC run one or more tasks T_i. In our case a task could represent an OS process, thread or simply a generic abstraction of a piece of software that executes a specific function). Each task is characterized at least by a *task criticality level* (c_i), i.e., the level of insurance associated to the task itself. Each task has a default criticality statically assigned at design-time and can deliver a

message (M) over the NoC through the NI of the node on which it is running. Every message is then outlined by a *priority* equal to the sender task criticality.

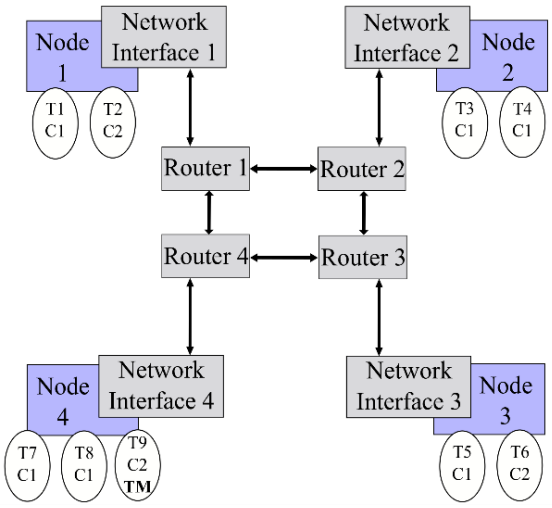


Figure 1: NoC with 4 nodes in a mesh topology.

3.2. Proposed HW mechanism

In order to define a mechanism supporting isolation, the network status is characterized by a global parameter named *Severity*. At certain time, the NoC will perform with a level of severity S in the range [1, Smax]. The isolation mechanism relies on the following two hypotheses:

- for a given temporal instant, only the messages sent by tasks with a criticality level $c \geq S$ can be transmitted by the NI and forwarded to the routers;
- the severity level can be modified, at run-time, by tasks that have specific privileges. These tasks (*Task Masters*, indicated as TM in Figure 1) are statically defined at design time.

At start-up, the system works in the least conservative mode, with a minimum severity level. If an anomaly condition is detected, and tasks with high criticality need to be executed, the system may switch to higher severity level, causing those tasks with higher criticality to access the network without competition with lower criticality ones. The following additional hypotheses are assumed:

- each NI can store the severity level of the network, to allow the transmission of new messages only if their criticality is greater or equal than S;
- the severity level of the network could be modified (raised or lowered) by a TM, by introducing inside the network a message which change the *severity*. This message is propagated to all network nodes and processed inside the NI. The reception of a message with a change in severity will activate an update of the severity level memorized inside the interface.

Being only the TM able to change the NoC Severity, the configuration of the system results to be protected: indeed, an untrustworthy entity is not able to change this configuration and possibly corrupt the system itself. The condition to prompt for a Severity change depends on the presence of anomalies: metrics able to indicate when an anomaly is verified can be defined, depending on the application executed on the NoC; then, by monitoring these metrics, it is possible to relate the TM transmission of a Severity change with the anomaly identification.

Observing Figure 1 and supposing that it is related to a scenario with two levels of criticality and two levels of severity, when the NoC possess a severity level equal to one, all the tasks can send and receive messages through the network. The routers could use a simple *First-in First-out* (FIFO) policy to manage the message forwarding toward the proper ports: when the FIFO is full, the message is rejected by the router, waiting for space in the queue. When a TM (T9 in the example of Figure 1) starts the procedure to change the level of severity of the NoC, all the NI shall be notified in order to update the corresponding value. When the severity level reaches the value of two, only T2, T6 and T9 are enabled to send messages while all the tasks will be able only to receive them. It is important noting that the designer owns the responsibility in order to eliminate, or to keep tolerable, conditions where a task cannot react to the message of a more critical one due to the NoC severity level possessed in that moment.

The reference HW design of the severity management mechanism, able to implement the isolation mechanism above described, is shown in Figure 2. It has an *Input Buffer* (called Message Buffer in the figure) to manage message traffic from the node: when a message from the node reaches the input of the buffer, the *Comparator* module checks if the criticality of the input message is greater or equal than the current network severity (stored in the *Severity Register*). If that is the case, the output of the comparator will be low, and the message will be stored into the *Message Buffer*. Otherwise, the output of the comparator will be high, and the message will be rejected: specifically, the loading of the message buffer will be inhibited, and a reject notification will be sent to the node. This functionality is described in the flowchart shown in Figure 3, where the methods *fillInputBuffer(), Analyze(), StoreMessageBuffer()* and *Reject()* implement, respectively, the reading from the node, the analysis of the received *Message*, the storage into the message buffer to transmit along the NoC and the rejection of the *Message*.

As above indicated, the aim of the proposed hardware mechanism is to assist the NoC design where, at a certain time, only the packets sent by a task with criticality greater or equal than the severity of the NoC are transmitted. The NI of nodes that executes the sending tasks blocks all other messages. It should be highlighted that this degradation is often tolerable in systems with mixed-criticality [17], as it eliminates any influence between lower and higher criticality flow. It can be noticed that the suggested solution does not reduce the number of criticality network levels. Also, it can be noticed that the proposed mechanism supports both spatial and temporal isolations: the former is ensured by the fact that only tasks with criticality greater than S can use the network resources to send messages. The latter is ensured by the fact that Severity can be changed over time, so giving the opportunity to all the tasks to access the network resource in specific temporal slots while still being able to assure that the most critical ones, when needed, can run without interferences due to less critical tasks.

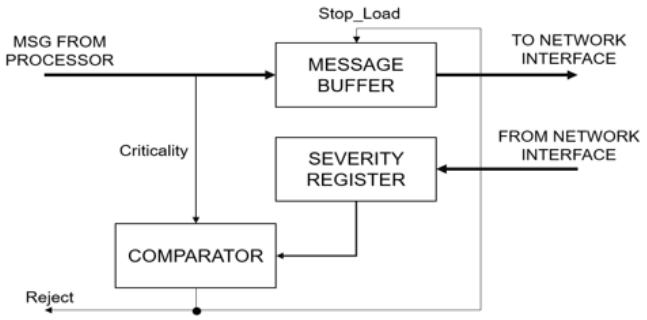


Figure 2: Implementation of the mechanism in a Network Interface (NI).

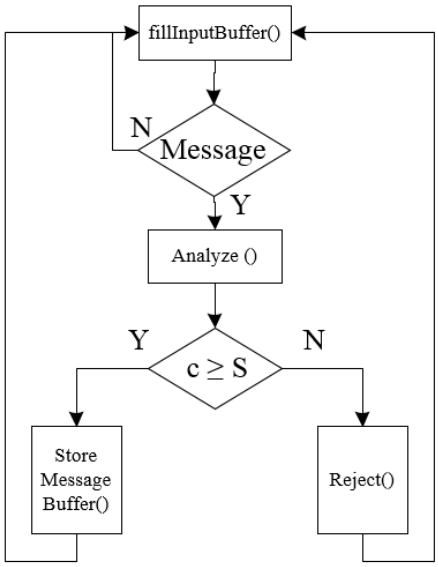


Figure 3: Functionality of the proposed HW mechanism.

4. HW Mechanism Design and Validation

This section presents the approach adopted to design and validate the proposed HW mechanism. The interests have been focused on three features: behavior validation, practicability check, and evaluation of the scalability. Regarding the first one, it allows to check if the developed mechanism behaves as expected. It is mainly based on HDL-based design and simulations. The second feature allows to check if this mechanism is possible to realize on real systems. It is based on a *Hardware-In-the-Loop (HIL)* approach applied to a simple NoC implementation on an FPGA platform. The final and third one (described in section 5) allows to check if it is possible to exploit the mechanism in real-size NoC. So, it is based on a *Network Simulation* approach to validate behavior and evaluate scalability, without the need to develop very complex NoCs. So, the following paragraphs, presents all the details needed to understand the performed HDL-based design and simulations activities.

4.1. Selected Reference NoC

A reference NoC has been considered for validation activities. It is an open-source NoC [19] described in HDL, that can be simulated and synthetized on FPGA. Such a NoC is of interest, since its design is very simple, allowing to reduce the number of required logic units for prototypal implementations. Moreover, it is provided with a MATLAB program to monitor the network traffic at run-time.

The reference NoC architecture consists of a network of nodes able to send and receive data, in order to complete their actions. Devices in nodes can be of different types, such as processors, memories and input/output devices. A network adapter is used to connect a device to a router, and a router connects the node to the rest of the network. Packet switching is used as a communication method across the network, with packets used as the communication medium.

Specifically, there are three types of packets in the network: *write*, *read_request* and *read_return*, as shown in Figure 4. All of them have a header, to indicate their type, and other fields depending on their goal.

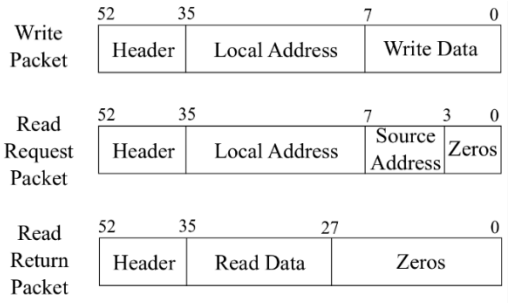


Figure 4: Packets of the reference NoC.

The *write* packet is sent by a master node to write data to a slave, while a *read_request* is sent by a master node to a slave one, and the latter replies with a *read_return* packet. *write* and *read_request* packets require a destination address: the reference NoC has a 32-bit address space, where the first 4 bits are used as the unique node IDs, while the remaining 28 bits are used for local addresses. The *read_request* packet also contains the source ID used by the receiver to send the *read_return*. *read_return* packets only contain the destination node ID of the source where they are replying to. A *write_packet* contains 8-bits of write data, while a *read_return* contains 8-bits of read data. The packets are sent over one clock cycle, rather than broken up over multiple clock cycles, to keep the hardware and logic simple.

Focusing on *network adapter* (NA), it represents interface between the node and the router, and its purpose is to convert signals from the local bus into a suitable packet format for the network, and back again. There are two types of network adapters and devices: *master* and *slave*. The *master network adapter* receives the following signals from the master device: *write_address*, *write_enable*, *write_data* and *read_request*. A master device can connect to the network adapter and the network should be totally transparent. The NA sends back *not_ready*, *read_return* and *read_*data. Any device wishing to connect to the network needs to handle a *not_ready* signal from the network adapter. Its output interface with a router is a *packet_data_out*. It can receive a busy signal from the router and a *read_return* packet.

Routers have five sets of channels connected to them, as shown in Figure 5. As the mesh organization is used in the network, four directions can be identified: *north*, *east*, *south*, *west*, and one additional channel going toward the *local* network adapter. The router is clocked with the global clock of the system. In order to

move the packets around the network from sources to destinations, routers look at a special field of the incoming packets, called *xy-counters* [19]: they give priority to *y-direction*, firstly check in what vertical direction they should forward the packet, and then checking for the horizontal one. If a router receives a packet in which both x and y counters are set to 0, then it indicates that the packet was destined for that router and that it has now reached its destination (and should be forwarded to the local network adapter).

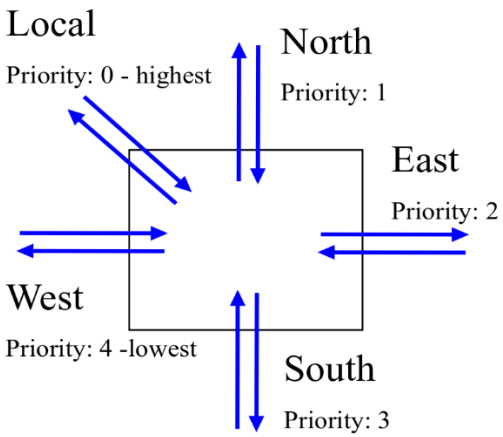


Figure 5: Router.

In order to support on debugging actions, a hardware monitoring system is offered together with the reference NoC [20] [21]. The monitoring system interfaces with an UART controller, allowing the communication with a host computer. A MATLAB program is also provided to be executed on host computer, providing a visual representation of the current traffic on the network, which is updated several times per second.

The proposed HW isolation mechanism has been integrated in an instance of the reference NoC discussed in the previous sub-section, composed of 16 nodes in a mesh topology, shown in Figure 6. Each device within a node is represented by a simple *finite state machine* that acts as a master or slave processor (in the following TP — *Test Processor*), executing some specific instructions.

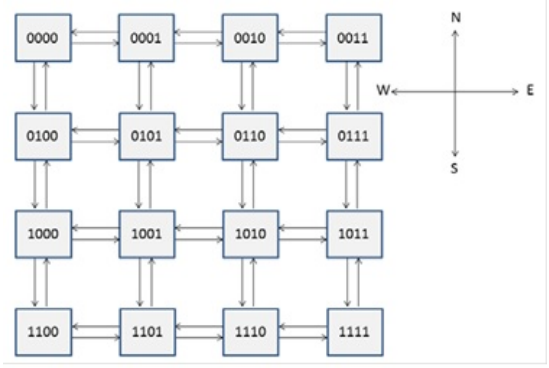


Figure 6: Reference 16 meshed nodes NoC.

In such a reference NoC, *Severity* and *Criticality* concepts have been introduced. If a TP tries to send a message with criticality less than current network severity, the message is blocked by NA. It is worth noting that messages are not blocked by routers, since, in this case, they are unaware of the isolation mechanisms. This improves portability but also means that, when a severity change occurs, all the packets already inside the networks are still able to freely circulate. In this case, they will be managed accordingly to their priority, as expected by the type of NoC. For validation purposes, *Criticality*, *Severity* and *Priority* have been set in the interval [0, 7]. In order to support the proposed approach, the NoC has been modified to introduce a fourth type of packet, called *severity_change,* shown in Figure 7. The goal of the packet is to allow a *master* task to trigger a severity change.

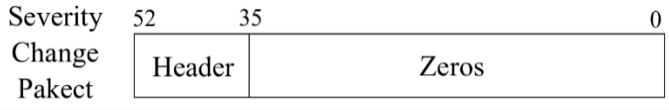


Figure 7: Packets managed by the reference NoC.

Together with the introduction of a new packet type, further assumptions have been done to reduce the management complexity of the network, focusing on the purpose of the tests:

- TM has been associated to the node in the upper-right corner of NoC shown in Figure 6. The corresponding TP is the only one able to change the severity of the network.
- Severity change requests are forwarded by following a snake-coil path, as shown in Figure 8. Routers forward such a packet only in one fixed direction

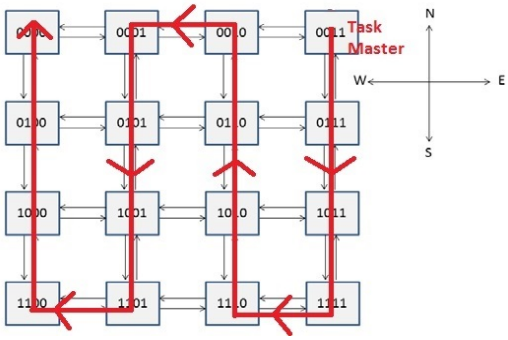


Figure 8: Snake-coil path.

Master and slave NA adopt a two-levels input buffer at both TP and Router side. By means of them and a proper busy signal, it is possible avoid losing input packets. If two packets are concurrently coming from TP and Router, they are managed at the same time, if possible, otherwise they are serialized giving priority to the one with greater criticality. As said before, the criticality/severity check is performed only in Master NA, specifically by checking the input buffer at TP side. This represents the major change with respect to the reference NoC. Another important one is that, if a message coming from TP is not accepted by NA, TP is notified by means of a proper *msg_rejected* signal. A detailed view on NA modifications is shown in Figure 9 and Figure 10.

The router is identical to the reference NoC one, apart from the need of improving the existing priority-based routing approach. In fact, in the case of concurrent packets forwarding, greater priority shall be given to criticality instead than to the port. It is worth noting that such a policy is simply priority-based, i.e., the router does not need to know about the criticality/severity mechanisms (in fact it is designed to work without knowing anything about the current severity level).

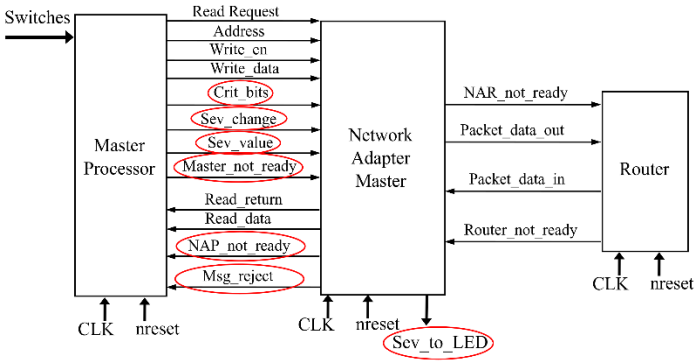


Figure 9: Modified master NA.

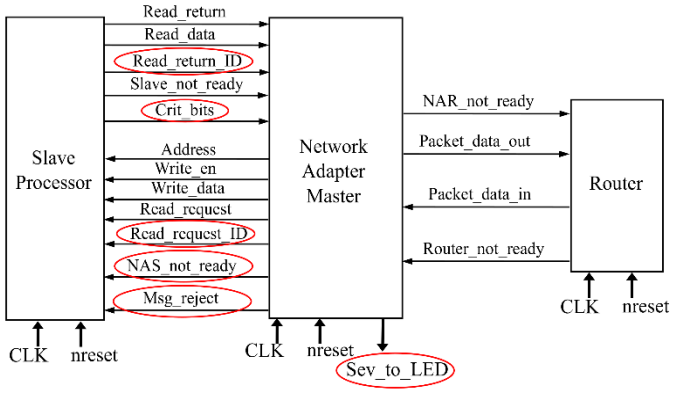


Figure 10: Modified slave NA.

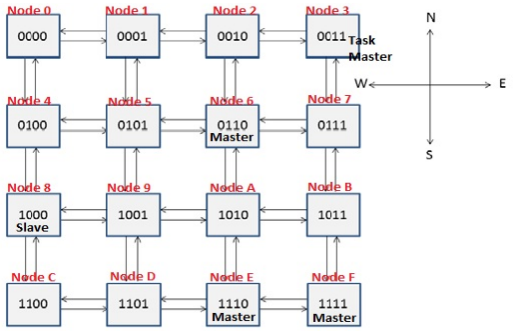


Figure 11: Simulated NoC.

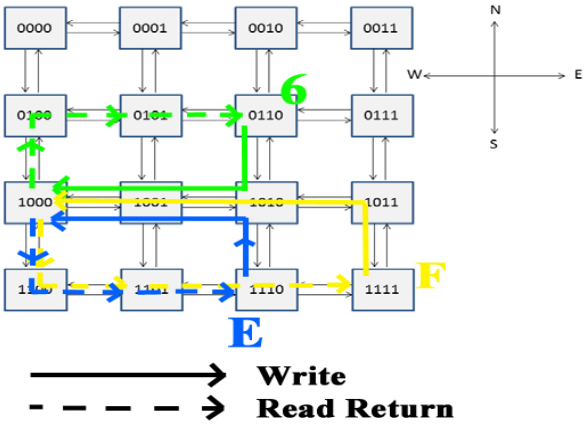


Figure 12: Write and read_return paths.

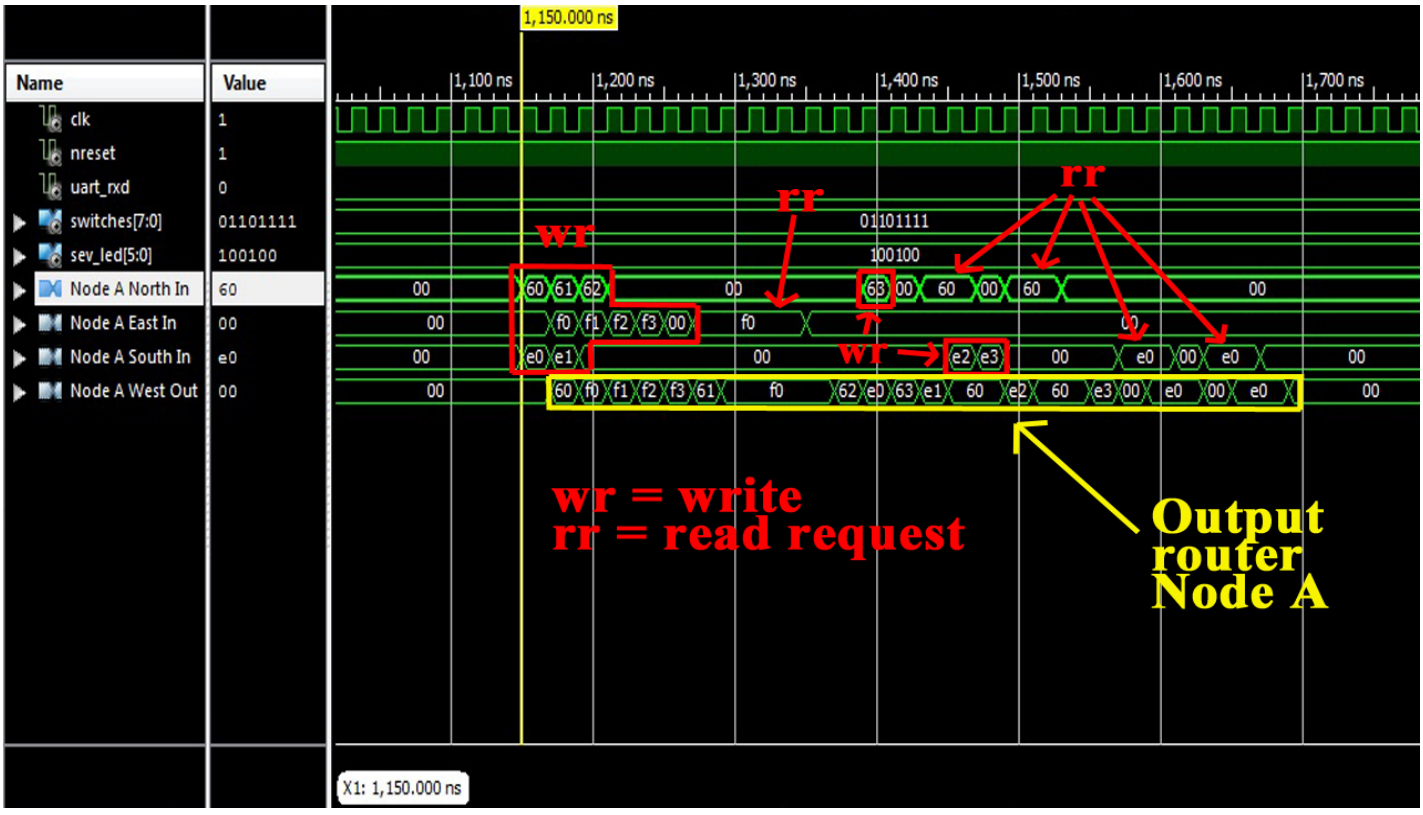


Figure 13: Packet forwarding in router close to node A

4.1. Validation by means of logical simulation

The above discussed NoC has been described in VHDL and simulated by means of Xilinx *ISim* [22]. In this way, it has been possible to verify the right forwarding of packets and the behavior of the proposed mechanism. In Figure 11, the simulated NoC is shown: it is composed of 16 nodes with different IDs (represented both in binary and hexadecimal base). Nodes 6, E and F are the master ones (i.e., they have a master NA and a master TP). In the considered test case, they perform some writings and readings by involving node 8 (*slave*). Written data (*write_data*) are composed of node ID and a progressive number (from 0 to 3, since, in the considered test case, each node performs 4 consecutive writing). In Figure 12, the main communication paths between the masters and slave node number 8 are shown.

Test 1 – Router packet forwarding

The first proposed test has the goal of verifying the correct packets forwarding done by routers. In Figure 13, it is shown the router close to master Node A managing input from ports N, E, and S, and forwarding them towards W (since all the write operations are related to node 8). The reported values represent the *write_data* field in hexadecimal base. In the presented test case, master NAs make use of their 3-bits ID LSB as criticality values for their sent messages; with this assumption, the following criticality values have been considered: criticality 6 for node 6 (110), criticality 6 for node E (110), criticality 7 for node F (111).

In Figure 13, it can be noticed that packets from nodes 6 and E reach the router with 1 clock cycle anticipation (since node F is far than the others). It is also possible to notice that the first two *write* packets (60 and e0), that requires the same output port, are correctly forwarded giving priority to the packet 60 (same criticality, but N port). Then, the newly arrived f0 is quickly forwarded, since it has greater criticality with respect to e0. The output is the sequence f0, f1, f2, f3. Since also the *read_request* packets form F have criticality 7, they are always forwarded in a single clock cycle (f0 is high for 4 clock cycles).

Test 2 – Isolation

In Figure 14, the isolation capability of the proposed mechanism is verified. Considering a severity level set to maximum value (i.e., 7) and stored by each NA, a TP tries to send a message with criticality less than the current severity. In particular, the TP tries to send 4 *write* and 4 *read_request* packets with criticality 6 (110); it receives the *msg_rejected* signal for 8 clock cycles. It is worth noting that, in this test, *msg_rejected* notifications are not managed in any way by the TP (e.g., no tentative resending).

Test 3 - Severity change

In Figure 15, a severity change in the NoC is shown. It is possible to notice the *severity_change* packet (highlighted in red) firstly sent by node 3 (the only allowed to do it, since it is supposed to be the only one to have a TM) to its NA. Then, such a command is forwarded to the whole NoC (following the path already shown in Figure 8), i.e., each NA updates its internal current severity value and setups a new message to forward the severity_change command to the next node.

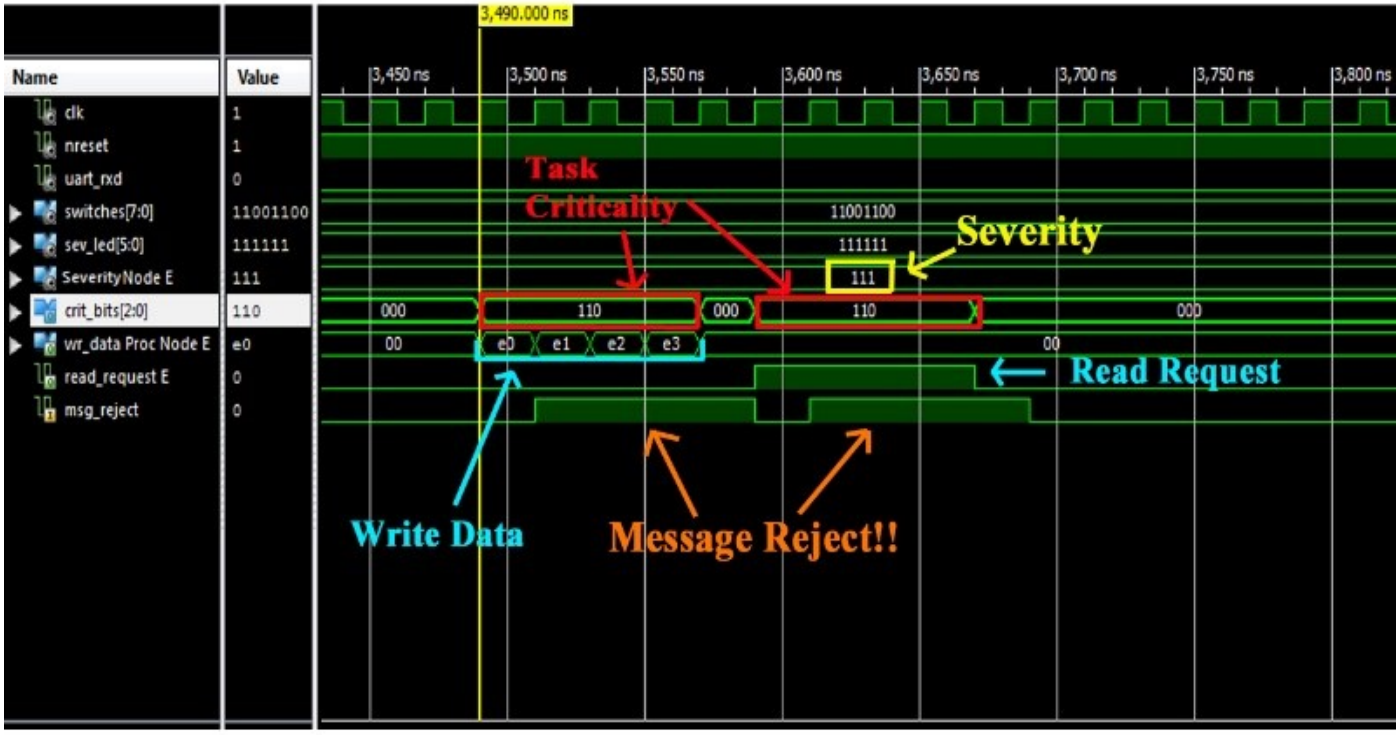


Figure 14: Criticality/Severity check.

4.2. HIL-based validation

In order to validate a prototypal implementation of the proposed mechanism, an instance of the reference NoC, enriched with that mechanism, has been implemented on a Xilinx Spartan-3 FPGA [23] using Xilinx *ISE Design Suite 14.7* [24]. The testbed is shown in Figure 16. The prototype consists in a NoC with a 3x3 mesh topology, where each node device is represented by a master or slave *Dummy Processor* (DP), namely a simple finite state machine that acts as a processor executing some specific instructions. Such a network implements an 8-level Severity and Criticality system. To keep low the complexity, while focusing on proposed mechanism features, the TM has been associated to the node in the upper-right corner so that the corresponding DP is the only one able to change the severity of the NoC (actually, the corresponding dummy processor has been designed so that the user can manually decide when to do severity changes and at what value, by using board switches). Aside from the node at the top right that holds the Task Master, all other nodes have been randomly distributed between the two master and slave types, and the associated Dummy Processors have been configured to create a decent amount of traffic packets at all criticality levels in the network. Severity change requests are forwarded to the other nodes by following a fixed snake-coil path, as shown in Figure 17.

All packets are forwarded by routers with a *Packet-Criticality* based priority policy [19]: the routers first check in which vertical direction they should forward the packet, and then check for the horizontal one. If a router receives a packet in which both x and y counters are set to 0, then it indicates that the packet was destined for that node and that it has now reached its destination. If, at a given time, two or more packets from the input ports are destined to the same output port, a conflict occurs, and priority is given to the packet with the highest criticality, whereas other packets are blocked before they can be sorted out later. In the case of equal criticality, priority is given according to the input port of the conflict packets, by following the descending order: Local, North, East, South and West. It is worth noting that such a policy is simply priority-based, i.e. the router does not need to know about the criticality/severity mechanism (indeed it is designed to work without knowing nothing about the current severity level).

Master and slave NA adopt a two-level input buffer at both DP and Router side. By means of them and a proper busy signal, it is possible avoiding losing input packets.

After the system implementation on Spartan 3 FPGA, the monitoring system and the MATLAB program have been used to check the NoC status. Red wires indicate connections that have a lot of traffic, yellow ones indicate a small amount of traffic while white ones indicate no traffic at all. Other than the traffic, it is also possible to check the amount of *busy* and *msg_rejected* events. In Figure 18, it is possible to notice that with a low severity level (000) traffic is quite high, since all the nodes are able to send messages

After the setting of a severity change to the highest severity level (111) (the changing of severity using the TM is shown in Figure 17), the traffic appears to be heavily reduced, as shown in Figure 19. The area occupations are reported in Table 3. The impact is very small, so the adjective “lightweight”.



Figure 15: Severity Change.

Table 3: Area occupation.

	FFs	LUTs	f_{MAX}
Complete NoC with the proposed mechanism	37 %	70 %	125 MHz
Proposed mechanism (area per each NI)	1 (<1%)	2 (<1%)	-
Proposed mechanism (area in 3x3 mesh)	9	18	-

5. Scalability evaluation

In this section, a model of the reference NoC has been developed using the Omnet++ simulator [25] with the aim to evaluate the scalability of the proposed mechanism. The model of the NoC supports *n*-levels Severity and Criticality. Multiple standard (unprivileged) tasks and Master tasks can be statically associated to each node of the network, and different traffic patterns can be associated with each task. The simulator includes a statistic unit able to investigate the traffic sequences, and the possibility to use a flexible time base module able to model synchronous and asynchronous NoC. In the case of synchronous NoC, the model of the network makes use of a module (named the Time-base synchronizer) able to generate clock messages, ensuring synchronous communication between the various nodes. Other model features are the following:

- in a fixed, predetermined device, a single TM executes on the platform
- each router uses the *xy* routing strategy shown previously and the broadcasting of the Severity Change Packet;
- the TM can determine the traffic status and eventually its congestion, by inserting a special broadcast packet called the Inquire Packet (IP);

The TM will modify the severity by sending a *Severity Change Packet* if the measured congestion status exceeds a predefined threshold, or if there is a timeout elapse. This severity administration policy needs an additional overhead for the development of the routers, as every router should be capable to follow the number of queued packets and deal with the IP. Furthermore, the suggested enquiry strategy needs the insertion of additional packets inside the network, possibly disturbing with device traffic.

In Figure 20, there is the result of a simulation of the modelled system in case of a synchronous 8x8 NoC. This figure shows the total number of packets inserted into the network over time, reported in generic time units (i.e., number of clock cycles) and two indices of the network status: the number of queued packets at a certain time and the current network severity. It should be observed that the value of the two network indices is multiplied by one hundred for a better clarity of the figure. It can be seen that the traffic regulation mechanism is adequately capable to limit the traffic of the network in case of congestion, allowing only the successful transactions of high criticality messages in a timely manner: this shows the feedback that the proposed mechanism can provide when unexpected behavior from the tasks are verified.

In Figure 21, it is reported the number of total sent messages (red line) and the number of dropped messages (blue line). The number of dropped messages is lower than the total messages that go through the network. In particular, the number of total messages in the tested time interval is $4.14 \cdot 10^6$, while the number of dropped messages is $3.00 \cdot 10^6$.

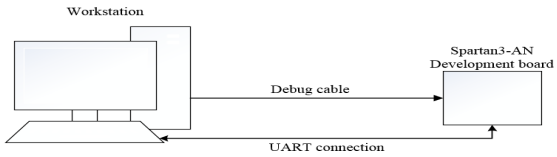


Figure 16: Testbed.

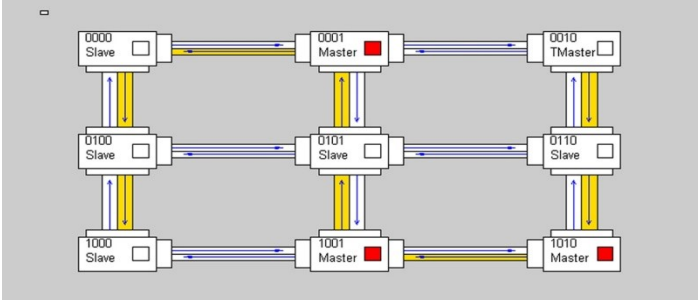


Figure 17: Snake-coil path.

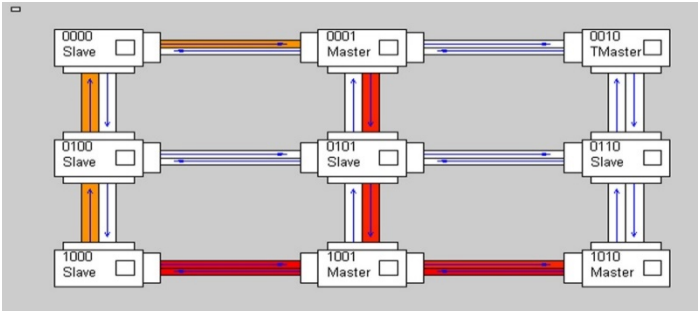


Figure 18: Network traffic with low severity (000).

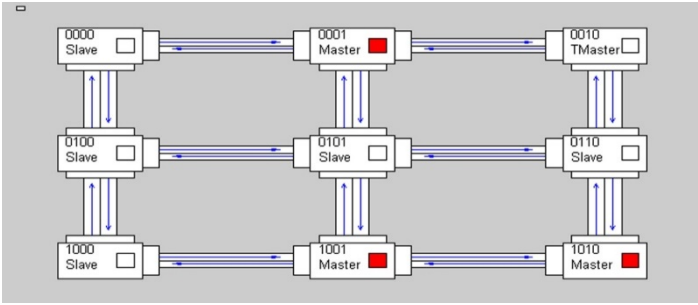


Figure 19: Network traffic with high severity (111).

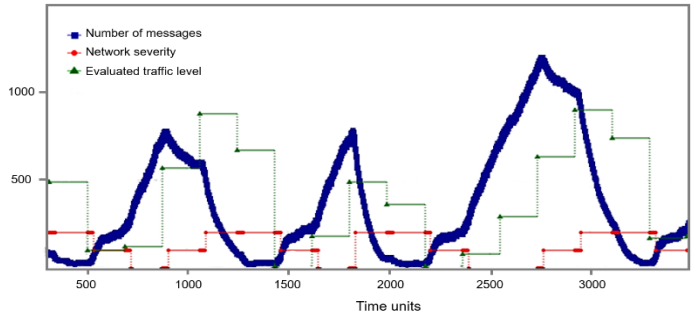


Figure 20: Simulation of a synchronous 8x8 mesh network.

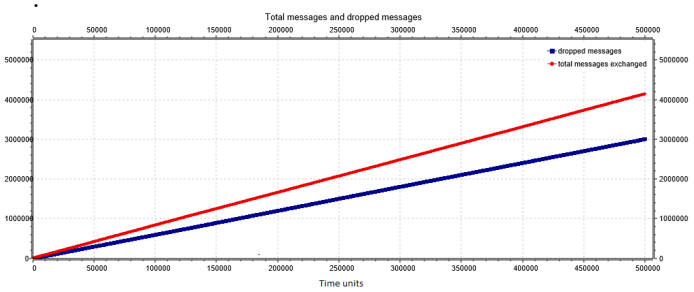


Figure 21: Number of total messages (red line) and number of dropped messages (blue line) for a synchronous 8x8 mesh network.

6. Further Analysis

Thanks to effort spent on creating a model, further analysis has been done by means of system simulations, introducing other aspects and evaluating a different broadcast strategy and a different management of the control traffic messages.

6.1. Different broadcast strategy

In this test, routers have the possibility of changing the broadcast of the Severity Change Packet. There are two possibilities:

- *snake-coil strategy*: the message, with TM in the node at the top right of the network, is broadcast one router at a time in a north-south direction, then moves one step to the left when the message reaches a router located at the northern or southern limit in the network itself and resume the journey in the north-south direction. The method is already described and used in the previous two sections.
- *star-broadcast strategy*: it takes its name from the design formed by the overall path travelled by the messages. The origin node of the messages can be in any position. The first router forwards the packet to all the connected ports in order to reach the maximum number of routers. Routers located along the north-south axis respect to the first router forward the message to the east, west and north, if they received the message from the south, or forward it to east, west and south, if they received it from the north. Finally, the routers that receive the packet from the east [or west] forward the broadcast packet along the west [or east] direction in order to make the broadcast packet continue along the east-west axis (see Figure 22).

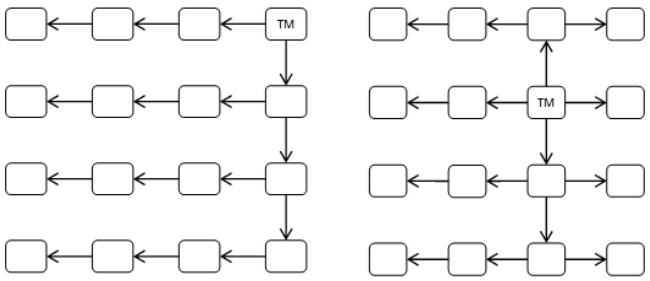


Figure 22: Star-broadcast strategy.

In Figure 23, it is shown the behavior of the system in case of a synchronous 8x8 NoC with the star-broadcast strategy instead of snake-coil one. As in previous section, the value of the two network indicators is scaled by a factor of one hundred for better clarity of the graph. Here, the TM periodically sends a control signal to which all routers respond by entering their own maximum queue value. In this period, the TM waits a maximum time equal to 10 clock cycles between one message and another before deciding to change the severity of the network. After this time, if the TM does not receive all the answers, then the network severity increases, trying to moderate the circulation of messages. Vice versa, when all the messages are received by the TM, the latter can decide to decrease the severity of the network. When all the messages are received, the TM in this test evaluates the average of the received answers.

The test shows that the proposed mechanism manages to control the flow of messages within the network, although the network itself is flooded with many control messages. The peak of messages reached within the network exceeds one thousand units, a situation in which the severity of the network is zero. In this case, there are several messages in the network and the TM fails to receive all the control messages, so the network severity is raised to one (one hundred in Figure 23).

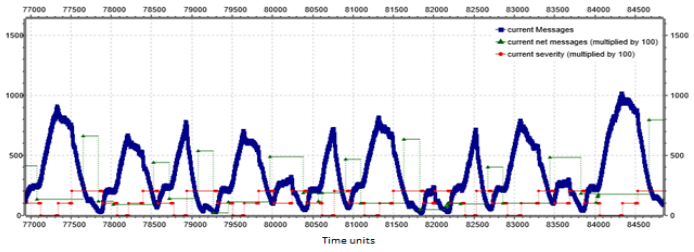


Figure 23: Simulation of a synchronous 8x8 mesh network without traffic control and with star-broadcast strategy

This leads to a decrease of messages in the network, but not enough so the TM causes another increase in severity. With severity two (two hundred in Figure 23), the severity does not decrease until the value answered by all nodes is less than a threshold value. The response received from all the nodes is averaged in order to obtain an indicative value of the network state. The value is considered only when the responses are received from all the routers. Subsequently, two tests are conducted on this type of broadcast strategy. The first one, shown in Figure 24, uses an exact average value obtained from all the responses received. In the second case, shown in Figure 25, the average is modified in correspondence of a minimum value in order to be higher.

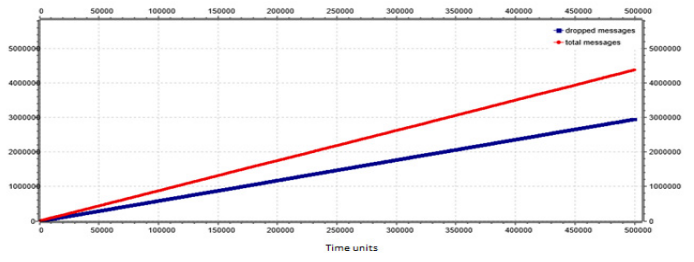


Figure 24: Number of total messages (red line) and number of dropped messages (blue line) for a synchronous 8x8 mesh network with star-broadcast strategy, exact average case.

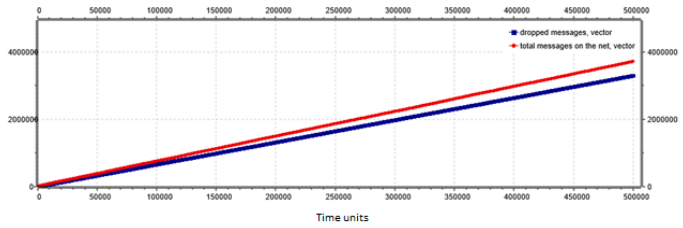


Figure 25: Number of total messages (red line) and number of dropped messages (blue line) for a synchronous 8x8 mesh network with star-broadcast strategy, modified average case.

Considering these two results, the network that performs best for message delivery time is the one that uses the modified average of traffic. However, this method results in a lower number of total messages sent over the network. Furthermore, with the use of the modified average, the maximum capacity of the routers results decreased in messages of a 20%.

6.2. Control traffic messages management

In this situation, devices and routers use two types of states to differentiate the type of messages processed, by extending the functionalities already present in the implemented NoC. Here, a single slot is dedicated to control traffic messages management. Two states have been identified:

- State zero indicates classic operation, i.e. read, write and reply messages are exchanged.
- State one only messages intended for traffic control are exchanged.

The traffic control messages in this type of network have a priority equal to that of the TM, in order to do not mix service messages with those of communication useful between the cores. Currently this functionality is supported only for a synchronous NoC. For a synchronous 8x8-node networks with control traffic messages management, the best behavior is obtained with the star-broadcast strategy (see Figure 26 and Figure 27). Here, there is a better delivery time for all messages.

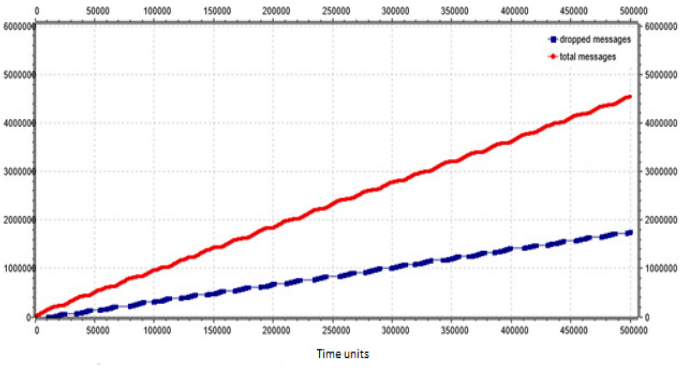


Figure 26: Number of total messages (red line) and number of dropped messages (blue line) for an 8x8 mesh network with control traffic messages management and snake coil strategy.

Analyzing all the experimental results for 64-node networks without control traffic messages management, the broadcast strategy that behaves best is the snake coil, unless the star-broadcast strategy modifies the average of the messages received. Therefore, if we consider the modified average, the star-broadcast improves and shortens the delivery time of the messages.

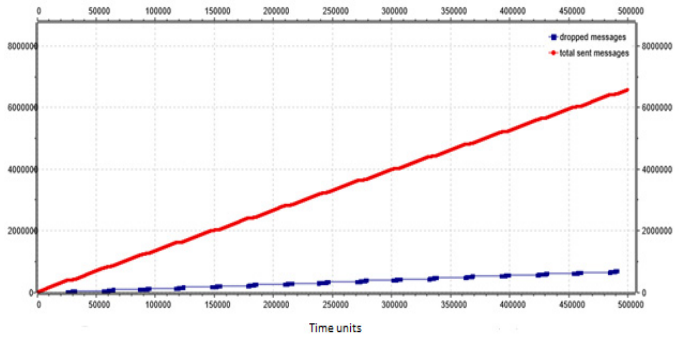


Figure 27: Number of total messages (red line) and number of dropped messages (blue line) for an 8x8 mesh network with control traffic messages management and star-broadcast strategy.

For the 64-node network with control traffic messages management, the best behavior is given by the star-broadcast. It must be remembered that, in the network with state, a fixed delay is introduced in the maximum time of delivery of the messages equal to the time of control necessary to probe the network. Despite this, the star-broadcast strategy keeps the average delivery time of messages low, even if it has a high delivery time.

7. Conclusions

Isolation is an important issue for embedded systems on which multiple tasks with different level of criticality are running. This paper has suggested a lightweight isolation mechanism to be introduced into existing Networks on Chip. This mechanism supports the execution of multiple applications with different criticality levels by supervising the packet exchange between network nodes. It does not reduce the criticality levels and it supports both spatial and temporal isolation. The system main innovation is its autonomy from the topology of the NoC and its easily flexibility among different NoCs. Small NoC implementations have been provided, showing the small impact in the area occupation, and motivating the adjective “lightweight”. Simulation on a network simulator has been proposed to evaluate the behavior of the mechanism on a NoC with more elements, and the feedback applied when there are unexpected situations is shown.

8. Future Works

Future works involve a further analysis to precisely characterize the overhead of proposed mechanism in a real existing NoC solution, and to investigate the best severity change policy. Moreover, given the outlined results, the possibility of adding a lightweight support for network status analysis will be explored as well.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

Authors would like to thank you Fabio Federici, Mattia Micozzi, Riccardo Cardinali and Biagio Iorio for the support on solving technical issues and precious feedbacks.

References

[1] A. Burns, R.I. Davis, "A survey of research into mixed criticality systems", ACM Computer Surveys, 50(6):1–37, 2017.
[2] S.K. Baruah, S. Vestal, "Schedulability analysis of sporadic tasks with multiple criticality specifications," in 2008 Proceedings of the 20th Euromicro Conference on Real-Time Systems (ECRTS), Prague, Czech.
[3] A. Agarwal, C. Iskander, R. Shankar, "Survey of Network on Chip (NoC) Architectures & Contributions," Journal of Engineering, Computing and Architecture, 3(1), 2009.
[4] R. Kamal, N. Yadav, "NoC and Bus Architecture: a comparison", International Journal of Engineering Science and Technology (IJEST), 4(4), April 2012.
[5] A. Achballah, S. Saoud "A Survey of Network-On-Chip Tools", International Journal of Advanced Computer Science and Applications (IJACSA), 4(9), 2013.
[6] A. Esper, G. Nelissen, V. Nelis, E. Tovar, "How realistic is the mixed-criticality real-time system model?", In Proceedings of the 23rd International Conference on Real Time and Networks Systems, RTNS '15, pages 139–148.
[7] F. Federici, M. Micozzi, V. Muttillo, L. Pomante and G. Valente, "Simulation-Based Analysis of a Hardware Mechanism to Support Isolation in Mixed-Criticality Network on Chip," 2017 European Modelling Symposium (EMS), Manchester, 2017, pp. 185-190.
[8] "AUTOSAR". [Online]. Available: https://www.autosar.org/ [Accessed: 13-June-2019].
[9] "ARINC". [Online]. Available: https://www.aviation-ia.com/ [Accessed: 13-June-2019].
[10] P. Ekberg, M. Stigge, N. Guan, W. Yi, "State-based mode switching with applications to mixed criticality systems," in 2013 Proceedings of the 1st International Workshop on Mixed Criticality Systems (WMC), Vancouver, Canada.
[11] Andrew S. Tanenbaum and Herbert Bos. 2014. Modern Operating Systems (4th ed.). Prentice Hall Press, Upper Saddle River, NJ, USA.
[12] V. Muttillo, G. Valente, L. Pomante. 2018, "Criticality-driven Design Space Exploration for Mixed-Criticality Heterogeneous Parallel Embedded Systems," In Proceedings of the 9th Workshop and 7th Workshop on Parallel Programming and RunTime Management Techniques for Manycore Architectures and Design Tools and Architectures for Multicore Embedded Computing Platforms (PARMA-DITAM '18).
[13] R. Pellizzoni, P. Meredith, M-Y. Nam, M. Sun, M. Caccamo, and L. Sha, "Handling mixed-criticality in SoC-based real-time embedded systems," in 2009 Proceedings Of the 7th ACM international conference on Embedded software (EMSOFT), Grenoble, France.
[14] H. Kopetz, A. Ademaj, P. Grillinger, and K. Steinhammer, "The time-triggered ethernet (TTE) design," in 2005 Proceedings of the Eighth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 05), Seattle, WA, USA.
[15] M. Schoeberl, "A time-triggered network-on-chip," in 2007 Proceedings of the International Conference on Field-Programmable Logic and its Applications (FPL 2007), Amsterdam, The Netherlands.
[16] S. Tobuschat, P. Axer, R. Ernst, J. Diemer, "IDAMC: A NoC for mixed criticality systems," in 2013 Proceedings of the IEEE 19th International Conference on Embedded and RealTime Computing Systems and Applications (RTCSA), Hakodate, Japan.
[17] A. Burns, J. Harbin and L. S. Indrusiak, "A Wormhole NoC Protocol for Mixed Criticality Systems," in 2014 IEEE Real-Time Systems Symposium (RTSS), Rome, Italy.
[18] L. S. Indrusiak, J. Harbin and A. Burns, "Average and Worst-Case Latency Improvements in Mixed-Criticality Wormhole Networks-on-Chip," in 2015 27th Euromicro Conference on Real-Time Systems (ECRTS), Lund, Sweden.
[19] G. Best, M. Birman, O. Rahnama, W. Pawlak. "Design and implementation of a simple mesh Network-on-Chip". [Online]. Available: https://github.com/mattbirman/Network-on-Chip-in-VHDL [Accessed: 13-June-2019].
[20] G. Valente et al., "A Flexible Profiling Sub-System for Reconfigurable Logic Architectures," 2016 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP), Heraklion, 2016, pp. 373-376.
[21] A. Moro, F. Federici, G. Valente, L. Pomante, M. Faccio and V. Muttillo, "Hardware performance sniffers for embedded systems profiling," 2015 12th International Workshop on Intelligent Solutions in Embedded Systems (WISES), Ancona, 2015, pp. 29-34.
[22] "ISIM". [Online]. Available: www.xilinx.com [Accessed: 13-June-2019].
[23] "Xilinx Spartan3AN" Data Sheet [Online]. Available: https://www.xilinx.com/support/documentation/data_sheets/ds557.pdf, [Accessed: 24-May-2019].
[24] "Xilinx ISE Design Suite 14.7" [Online]. Available: https://www.xilinx.com/products/design-tools/ise-design-suite.html, [Accessed: 28-July-2019].
[25] "Omnet++". [Online]. Available: https://omnetpp.org/, [Accessed: 13-June-2019].

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 4, 574-584 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Multiple-Optimization based-design of RF Integrated Inductors

Houcine Marouani[*,1], Amin Sallem[1], MondherChaoui[1], Pedro Pereira[2], Nouri Masmoudi[1]

[1]*LETI-ENIS, University of Sfax, Tunisia*

[2]*CTS-UNINOVA, FCT NOVA, Portugal*

A R T I C L E I N F O

Article history:
Received: 22 May, 2019
Accepted: 11 August, 2019
Online: 25 August, 2019

Keywords:
Metaheuristics; Particle Swarm Optimization; Gravitational Search Algorithm; Differential Evolution;
Multi-objective MOPSO;
PESAII;
MOEA/D;
RF Integrated Inductors;

A B S T R A C T

In this paper, a *multiple-objective Metaheuristics study is discussed. Initially, three mono-objective metaheuristics will be explored in order to design and optimize Radio-Frequency integrated inductors. These metaheuristics are: An evolutionary algorithm called The Differential Evolution (DE), An algorithm supported on Newton's laws of gravity and motion called the Gravitational Search Algorithm (GSA) and, finally, A swarm intelligence algorithm called the Particle Swarm Optimization (PSO). The performances of these three mono-objective metaheuristics are evaluated and compared over three benchmark functions and one application to optimize the layout of a RF silicon-based planar spiral inductor, the double π-model is adopted. Secondly, three references multi-objective metaheuristics using Pareto front are used respectively the multi-objective PSO (MOPSO), the Pareto envelope-based selection algorithm-II (PESAII) and the multi-objective evolutionary algorithm based on decomposition (MOEA/D). The performances of these multi-objective optimization algorithms are evaluated and compared over two bi-objective benchmark functions and the same application used in the first section. Two conflicting performances were optimized, namely the quality factor 'Q' (to be maximized) and the device area 'dout' (to be minimized) for the RF inductor. It was concluded that the multiple-objective PSO are significantly more efficient and robust for difficult problems than the other metaheuristics.*

1. Introduction

Nowadays, the advances in the integration technology agree the conception and the realization of complex integrated electronic systems. Therefore, the decision-making must be broached in an optimal way. Decision making can be seen as a set of the following steps: "Formulation", "Modelling", "Optimization", and "Implementation" of the problem [1]. In order to increase the accuracy of the optimization model, or algorithm, up to an adequate solution is establish, the decision making process could be repeated. The block "Optimization of the problem" is a vital part of this process and can be found in many areas, like engineering among others, and particularly in the topic of Radio Frequency (RF) circuit design [2-4].

The mono-objective optimization algorithms are classified into two categories: the heuristics and the metaheuristics algorithms. In fact, the first category is developed to solve a specific problem [5]. The second category represent a group of optimization techniques where theses popularity are increased during the last two decades and being considered the most successful and promising optimization algorithms [6–8]. The most used ones in the literature are the Evolutionary Algorithms (EA) such as local search (LS), simulated annealing (SA), tabu search (TS), genetic algorithms (GA) [9-14] etc.

Recently, a new group of nature inspired metaheuristic algorithms were published in literature. These metaheuristics are part of Swarm Intelligence. These techniques are based on animal behaviour and insect conduct, aiming to mimic their skills solving daily problems. In the area of Radio Frequency (RF) circuit design, Swarm Intelligence (SI) techniques are broadly applied, and among such techniques the PSO "Particle Swarm Optimization" [15–17], and the ACO "Ant Colony Optimization" [18–21], are the most used.

In this paper a multiple-objective metaheuristics study is discussed. Firstly, we propose to evaluate and compare such references mono-objective metaheuristics optimizing such

[*]Corresponding Author: Houcine Marouani, LETI-ENIS, Université of Sfax, Road Soukra km 4 3038 Sfax, Tunisia, +21652930505
houcine.marouani@gmail.com

benchmark functions and one Radio-Frequency application in terms of convergence, robustness and time computing. Three mono-objective algorithms are considered: An evolutionary algorithm entitled the Differential Evolution (DE), an algorithm based on Newton's laws of gravity and motion entitled the Gravitational Search Algorithm (GSA) [22, 23], and a swarm intelligence algorithm entitled the Particle Swarm Optimization (PSO). The drive issue for this comparative study is to give some sort of substantiation for the choice of the "best" metaheuristic among others [24]. Secondly, multi-objective problems are discussed. A study was conducted to the maximization of the quality factor Q and the minimization of the device area d_{out} for the Radio-Frequency (RF) integrated inductor circuit [25-30]. Three algorithms for multi-objective Pareto optimization [31-39] are used. An evaluation concerning the performances of the three abovementioned multi-objective metaheuristics: the convergence and the robustness analyses are offered [40-41].

The structure of this work is the following. Section 2 presents a summary of the mono-objective metaheuristics used. Results and analysis regarding the convergence and the robustness study of the three algorithms over three benchmarks functions are offered in Section 3. Section 4 reports the application of these mono-objective optimization applied to the layout of a RF silicon-based planar spiral inductor. Section 5 presents the uses of multi-objective algorithm to optimize both the maximum quality factor and to minimize the device area of the RF inductor with three multi-objective algorithms: the multi-objective particle swarm optimization (MOPSO), the Pareto Envelope-based Selection Algorithm II (PESAII) and the Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D). A comparison of its performance's meters is concluded. Finally, Section 6 wrap up the final remarks, conclusion and future works.

2. An overview on the mono-objective metaheuristics

2.1. Differential Evolution algorithm

In 1997, Storn and Price developed the differential evolution (DE), which belongs to evolutionary algorithm category. It is an effective, robust and simple global optimization algorithm [13].

Differential evolution algorithm is similar to the genetic algorithm [25], uses similar genetic operators like crossover, mutation and selection operators. Comparing both algorithms, the best solutions achieve are dependent of the process, the genetic algorithm relies on crossover operators while the differential evolution algorithm on the mutation operation (differential strategy) [26]. DE generates new candidates (trial vectors) of its population according to target vector manipulation and difference vector. Afterwards, it compares the performance of the resulting candidate against the original members, and if it shows to have a better solution, it changes it; else, keep the original candidate. The differential evolution algorithm technique uses three operators: mutation, crossover and selection.

2.1.1. Mutation Operator

The *mutation* operator is defined as follows:

$$V_i^{k+1} = X_{r1}^k + F\ (X_{r2}^k - X_{r3}^k) \tag{1}$$

Where V^{k+1} is the i^{th} trial vector produced by the mutation strategy; k represents the generation to which the population belongs; $r1$, $r2$, and $r3$ are mutually exclusive integers arbitrarily generated from 1 to N (N is the population size) and F is a scaling factor with optimal value in the range of *0.5 to 1.0*.

2.1.2. Crossover Operator

One the *mutation* operation is done, the crossover operator is applied to maintain the miscellany of the population. The crossover operator generates the offspring individual U_i at the k^{th} generation by selecting the solution component values from X_i or V_i, using the following equation (2):

$$U_{ij}^{k+1} = \begin{cases} V_{ij}^{k+1}\ if\ \ rand \le CR \\ X_{ij}^k\ otherwise, \end{cases} \tag{2}$$

where CR: the crossover probability $\in [0, 1]$.

2.1.3. Selection Operator

The selection operator is adopted to compare the trial vector U_i^k with the target vector X_i^k if the fitness of the trial vector is more suitable than the target vector. Such operation, which occurs every k^{th} generation, is presented in equation (3):

$$X_i^{k+1} = \begin{cases} U_i^k\ if\ f(U_i^k) \le f(X_i^k) \\ X_i^k\ otherwise, \end{cases} \tag{3}$$

where f: the fitness function.

The flowchart of DE algorithm is given in Figure 1.

2.2. Gravitational Search Algorithm

In [22, 23], the authors proposed the gravitational search algorithm as a new swarm intelligence metaheuristic. This metaheuristic is based on Newton's laws of gravity and motion. In GSA, a cluster of agents (solutions) named masses are used to solve and find the optimal solution (best agent). Based on the gravitational force, where all objects are attracted among each other, and this force causes a global movement of such objects towards objects with heavier masses [23]. The weightier masses, representing the best solutions, are known for moving slowly compared to those less heavy. The GSA is represented by N agents, where the position of the i^{th} object is defined with (4):

$$X_i = (x_i^1, ..., x_i^d, ..., x_i^N) \text{ for } i = 1, 2, ..., N \tag{4}$$

where x_i^d is the position of ith object in the d^{th} dimension and N is dimension of the problem to be optimized.

The mass values of every objects are defined by (5):

$$M_i(t) = \frac{fit_i(t) - worst(t)}{\sum_{j=1}^{N} (fit_j(t) - worst(t))} \tag{5}$$

where $fit_i(t)$, $M_i(t)$ represents the fitness value and the mass value of the object *i* at *t*, and $worst(t)$ is evaluated by (6):

$$worst(t) = \max_{j\in\{1,\ldots,N\}} fit_j(t) \tag{6}$$

To estimate the total forces and the acceleration of an agent, it should be evaluated over gravity law, equation (7), then equation (8). Resulting, the new velocity is calculated by equation (9) and the new position is computed by equation (10).

$$F_i^d(t) = \sum_{j\in kbest, j\neq i} rand_j G(t)\frac{M_j(t)M_i(t)}{R_{ij}(t)+\varepsilon}(x_j^d(t)-x_i^d(t)) \tag{7}$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \\ = \sum_{j\in kbest, j\neq i} rand_j G(t)\frac{M_j(t)}{R_{ij}(t)+\varepsilon}(x_j^d(t)-x_i^d(t)) \tag{8}$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{9}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{10}$$

where $R_{ij}(t)$ is the Euclidian distance between two agents, *i* and *j* defined as $R_{ij}(t) = \| X_i(t), X_j(t) \|_2$. k_{best} is the set of first *K* agents with the best fitness value and biggest mass. k_{best} is a function of time, which is initialized to k_0 at the beginning and is decreased with time. k_0 is adjusted to *N* and reduced linearly to 1. Moreover, *G* is reduced in a exponentially way from G_0 toward zero by time.

The flowchart of GSA algorithm is given in Figure 2.

2.3. Particle swarm optimization

The particle swarm optimization algorithm is a swarm-intelligent metaheuristic proposed by Kennedy and Eberhart [15]. PSO is an optimization algorithm that describes the social behavior of animals like birds and fish in search of food.

In PSO, each particle (solution) flies through the design/search space and moves randomly in the following three simple rules:

- Cohesion: the particles are attracted to the average position of the group or best position 'p_{best}';
- Alignment: the optimal value attained until here via every particle in the social neighbourhood 'l_{best}',
- Separation: to avoid collisions, particles keep a certain distance between them, the optimal particle position found in all the swarm 'g_{best}'.

Therefore, the new velocity and the particle position on the next iteration are performed through equations (11) and (12):

$$\vec{v}_{k+1}^{i} = \omega_k \vec{v}_k^{i} + c_1 r_1(\vec{p}_k^{i} - \vec{x}_k^{i}) + c_2 r_2(\vec{p}_k^{g} - \vec{x}_k^{i}) \tag{11}$$

$$\vec{x}_{k+1}^{i} = \vec{x}_k^{i} + \vec{v}_{k+1}^{i} \tag{12}$$

ω_k is an inertia weight, defined via a decreased inertia function by equation (13) in the training progress for objective the reduction of the influence of the past velocities.

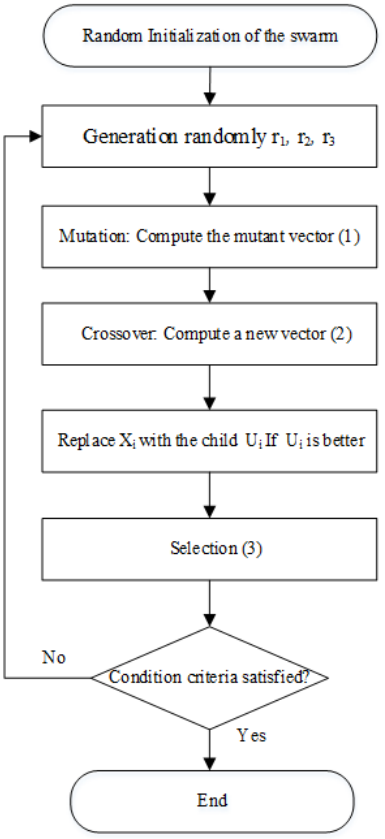


Figure 1: DE Flowchart

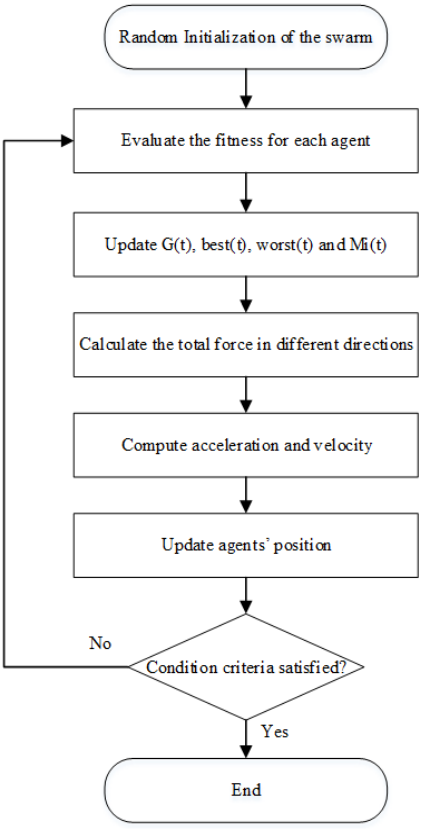


Figure 2: GSA Flowchart

Where

c_1 : the cognition learning rate,

c_2 : the social learning rate,

p_k^i : the best individual position,

p_k^g : the best global position,

$r_{1,2}$: two random parameters $\in [0,1]$.

$$\omega_k = \omega_k^{max} - \frac{\omega_k^{max} - \omega_k^{min}}{total_iter} iter \quad (13)$$

where

ω_k^{max} : the initial weight,

ω_k^{min} : the final weight,

total_iter: the number of the total iteration,
iter: the number of the current iteration.

The flowchart of PSO algorithm is given in Figure 3.

3. Application examples: the mono-objective problems

A. Benchmark function

Three mono-objective metaheuristics are used respectively the DE, GSA and PSO over three benchmark functions. A robustness study regarding the aforementioned metaheuristics is presented. The parameters used to characterize the three algorithms are shown in Tables 1, 2 and 3. For simulation and study proposes, the three metaheuristics were tested 100 times for each one of the benchmark functions.

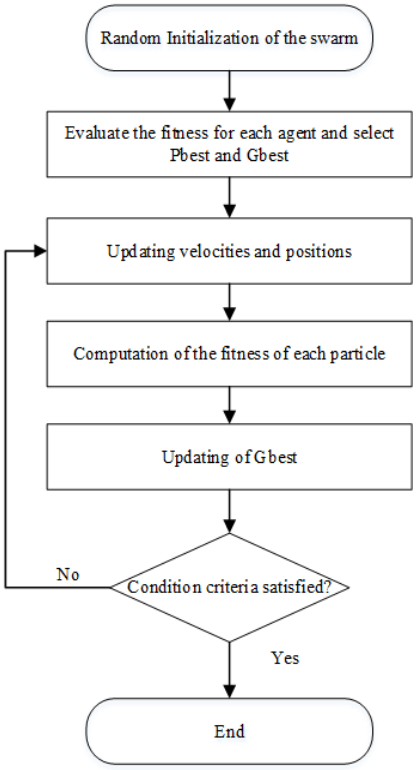


Figure 3: PSO Flowchart

Table 1: The DE parameters

Population size	*Number of Iterations.*	*beta_min*	*beta_max*	*CR*
100	1000	0.2	0.8	0.2

Table 2: The GSA parameters

Number of agents	*Number of Iterations.*	*Elitist Check*	*Rpower*	*min_flag*
100	1000	1.0	1.0	1.0

Table 3: The PSO parameters

Swarm size	*Number of Iterations.*	*w*	*c1*	*c2*
100	1000	1.0	1.5	2.0

3.1. Benchmark function application

In object to check performances of such mono-objective metaheuristics, a bunch of commonly used benchmark functions, considering this sort of problems/algorithms, were used [24]. These problems permit testing these mono-objective algorithms to confirm their convergence, concerning the approximation to the global solution [8, 18, 24]. The set of three benchmark functions are offered in Table 4.

Table 4: The benchmark functions

Item Function	*Objective functions to minimize*
Function #1	$F1 = \sum_{i=1}^{n} x_i^2$ Range: [-100; 100]; Dim n=2
Function #5	$F5 = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$ Range: [-30; 30]; Dim n=2
Function #7	$F7 = \sum_{i=1}^{n} [1:dim].*(x_i^4) + rand$ Range: [-1.28; 1.28]; Dim n=2

3.2. Optimization results

Function F1

The results of the benchmark function F1 are compared via the correspondent three optimization algorithms, as illustrated in Figure 4.

A Boxplots representation is used (due to probabilistic characteristic of these metaheuristics) to display the distribution of obtained measures [27]. In the following, 100 runs are considered for this function. The robustness analysis using the box-plot for the benchmark function F1 is presented in Figure 5.

Function F5

The results of the benchmark function F5, via the correspondent three optimization algorithms, are represented in Figure 6.

In the following, 100 runs are considered for function F5. The robustness analysis using the box-plot for the benchmark function is presented in Figure 7.

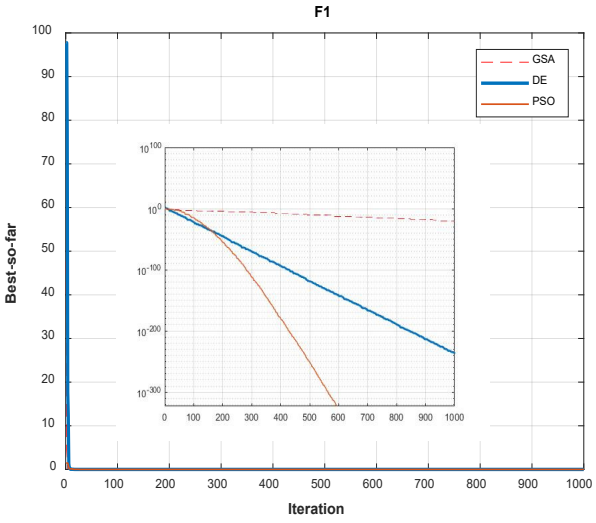


Figure 4: Convergence results of F1 with GSA, DE and PSO.

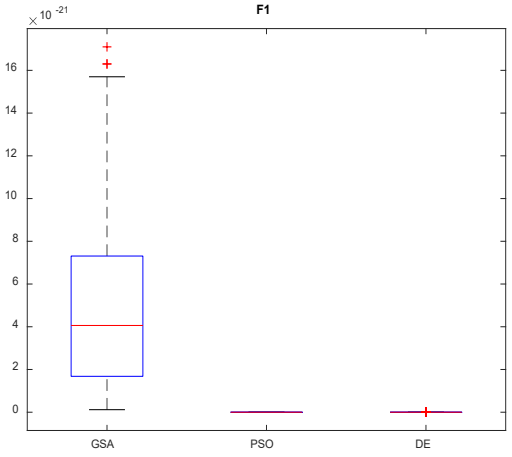


Figure 5: Robustness analysis of F1

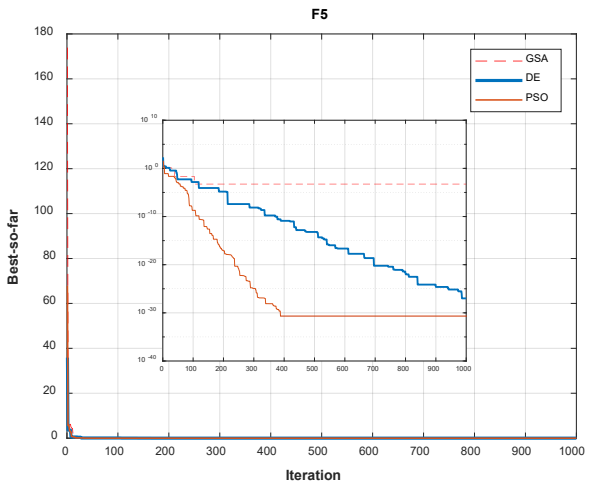


Figure 6: Convergence results of F5 with GSA, DE and PSO.

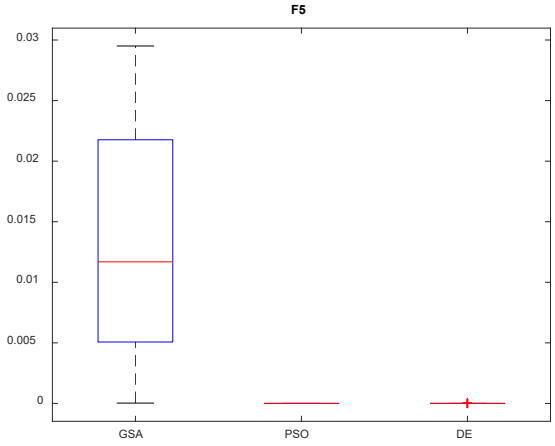


Figure 7: Robustness results of F5.

Function F7

The optimization results of the benchmark function F7, via the corresponding three algorithms, is illustrated in Figure 8. 100 runs are considered for function 7. The robustness analysis using the box-plot for the benchmark function F7 is presented in Figure 9.

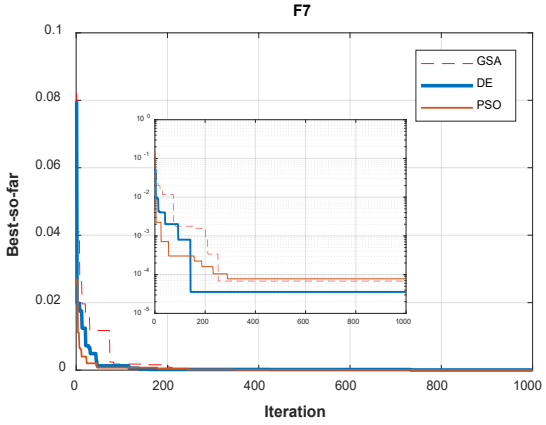


Figure 8: Convergence results of F7 with GSA, DE and PSO.

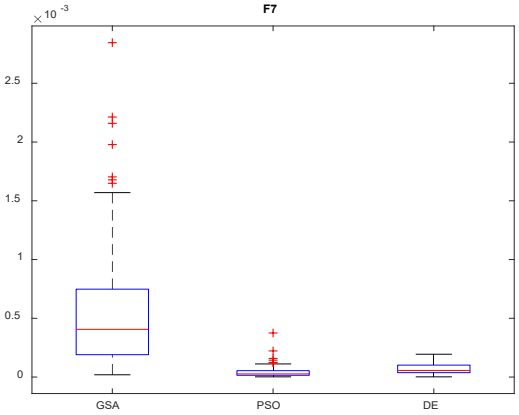


Figure 9: Robustness results of F7.

Figure 10 presents a Radio Chart representation of the average (AVG) execution times over 100 runs of three algorithm GSA, DE and PSO for the three benchmark functions.

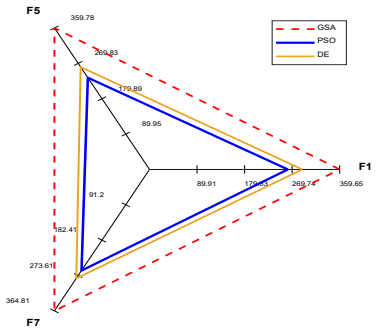


Figure 10: The average computing time of GSA, DE and PSO.

In accordance with the convergence of results regarding the three mono-objective metaheuristics according to the three benchmarks functions (Figures 4, 6 and 8) and the robustness results displayed in Figures 5, 7 and 9, we notice that DE and PSO have the better results compared to GSA.

The average executing time is illustrated in Figure 10. This figure shows that PSO is faster than DE and GSA.

Therefore, and according to the results achieved, the PSO algorithm is the most efficient algorithm, as it presents a best compromise among convergence, robustness and computational time.

B. Application example: RF spiral inductors

a. An overview on the integrated spiral inductors

CMOS technology is nowadays the only option able to fit the low-cost demand and to integrate digital, as well as analog radio-frequency (RF) modules on a single chip to realize the RF system on-chip (SoC). The efficiency and reliability of CMOS-RF blocks can be improved by means of on-chip passive devices (same substrate), such as voltage-controlled oscillators (VCO), low-noise amplifiers (LNA) and especially CMOS spiral inductor. Spiral inductor is broadly used and their applications in millimeter-wave circuit are investigated. High performance on-chip inductor has become increasingly important with their increasingly frequencies operation of the circuits.

Integrated spiral inductors models are based on the layout parameters, as displayed in Figure 11. These parameters are:

- w: the inductor width,
- s: the spacing between track,
- n: the number of turns,
- N_{side}: the inductor shape (square, hexagonal, octagonal),
- d_{in}: the internal diameter,
- d_{out}: the external diameter.

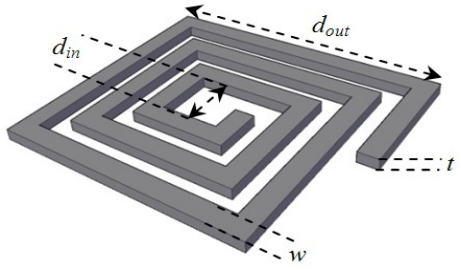


Figure 11: Layout characteristics for a square inductor.

For assessing the quality of the inductor design, different characteristics may be considered: the quality factor (Q), the operation frequency for the maximum Q, the self-resonant frequency f_{SR}, the inductance value L or the area. In this work, two parameters are selected: the quality factor Q and the device area d_{out}. Therefore, all the abovementioned characteristics are strongly dependent on the spiral inductor geometric parameters, technological constraints parameters and the application of this RF inductor.

At high frequencies, the simple but widely used single-π equivalent-circuit model can't accurately characterize the electrical performance of silicon-based planar spiral inductor. Many other equivalent-circuit inductor models have been reported in past decade [28]. In this paper, the double-π model is adopted and illustrated in Figure 12. This model lumped-element is evaluated with a wide range of equations and these values are specified in [29-30]. This model is represented by:

- DC parameters (Rs, Rp, Lo, Lp);
- Crossover capacitance, Cs: referred as the capacitance between the spiral and the underpass required to connect the outside to the inner turn of the planar inductor;
- Metal-to-metal capacitance, Cc: effect due to the proximity of inductor tracks;
- Metal-to-substrate capacitance, Cox;
- Rsub and Csub: the ohmic losses in the conductive silicon substrate;
- Rsc: the electric coupling between lines over the conductive substrate.

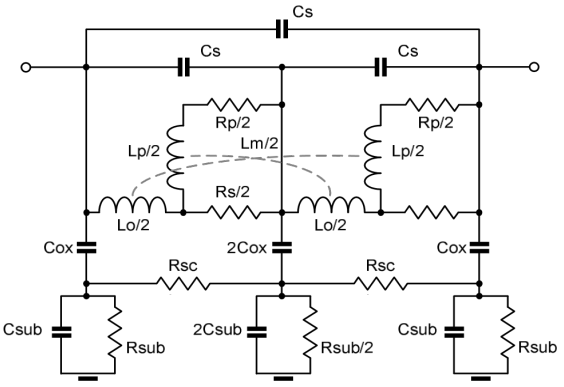


Figure 12: Inductor double π-model.

Table 5: Physical parameters

Parameters	Value
Metal Thickness (t)	2.8 µm
Space between turns (s)	2.5 µm
Sheet Resistance	10 mΩ/□
Oxide Thickness	5.42 µm
Oxide Thickness between spiral and underpass	0.26 µm
Oxide Permittivity (ε_r)	4
Substrate Thickness	700 µm
Substrate Permittivity (ε_r)	11.9
Substrate Resistivity	28 Ω cm

b. Optimization results

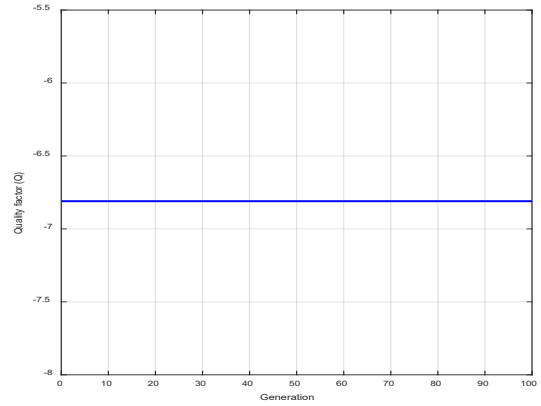
In this section, results highlight the ability of the PSO algorithm to perform the spiral inductor on Silicon technologies, the UMC130 technology is addressed. The conception for two inductors with 5.0 nH value, 6 nH value operating at 0.7 and 1 GHz, respectively, is considered. The technological and physical parameters are presented in Table 5, whereas the determination of the layout parameters is enabled with the uses of the constraint's parameters or technology constraints represented in Table 6. In addition, is considered a minimum space between tracks (s) of 2.5 µm and a maximum output diameter dout of 250 µm.

Table 6: Constraints parameters

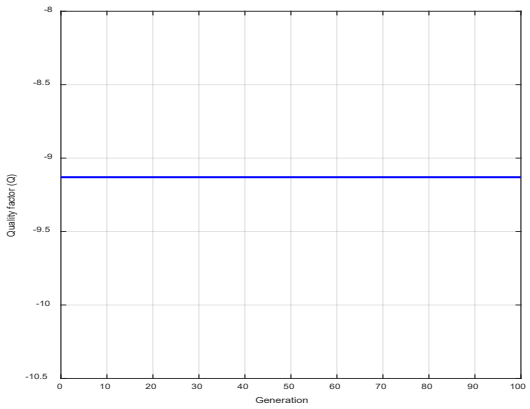
Parameters	Min	Max
w (µm)	5.0	20.0
d_{in}(µm)	20.0	80.0
n	1.5	15.5

The inductor quality factor Q, desired to be higher as possible, is the objective function to be optimized.

Figures 13 displays the optimization results with PSO algorithm (13.a, 13.b) for two inductors: the first inductor of 5.0 nH value operating at 0.7 GHz and the second inductor of 6 nH value operating at 1 GHz. The maximization of the quality factor Q is addressed.



(a)



(b)

Figure 13: Optimization results with PSO (Quality factor Q vs. Generation): (a) 5 nH, frequency= 0.7 GHz, (b) 6 nH, frequency= 1 GHz.

Table 7 gives the optimization results values (inductor sizes) with PSO algorithm.

Table 7: Optimization and simulation results

L = 5nH, Frequency = 0.7 GHz					
Algorithm	w (µm)	d_{in} (µm)	n	$d_{out\text{-}Op}$ (µm)	Q_{Op}
PSO	10.88	60.40	5.67	206.91	6.81
L = 6nH, Frequency = 1.0 GHz					
Algorithm	w (µm)	d_{in} (µm)	n	$d_{out\text{-}Op}$ (µm)	Q_{Op}
PSO	9.00	61.90	6.08	196.90	9.13

4. Application examples: the multi-objective Optimization Algorithms

4.1. The multi-objective particle swarm optimization

Multi-Objective Particle Swarm Optimization (MOPSO) is proposed by Coello *et al*., [31]. The goal of the MOPSO and generally all the multi-objective algorithms is to provide a set of Pareto optimal solutions (with non-dominated solutions) of the aforementioned problem. All the non-dominated particles, this means all solutions, in the swarm are then collected into a sub-swarm named *repository*, and each particle selects its global best target. A group of MOPSO can be found in related literature, like:

- The non-dominated sorting PSO [32],
- The MOPSO using the Crowding Distance mechanism associated with a mutation operator which increasing the exploration ability of the particles and maintains the diversity of non-dominated solutions in the external archive by MOPSO-CD [33-35],
- The MOPSO using the Crowding Distance mechanism associated with a mutation operator and the Roulette Wheel selection technique to select social leader or global best to avoid an excessive number of non-dominated solutions in the external archive MOPSO-CDR [36].

The MOPSO version used in this paper use the Grid Making technique, instead of the crowing-distance computation which requires sorting the population consistent with each objective function value in ascending order of magnitude. This version uses the Grid Making technique with the mutation operator and the Roulette Wheel selection called MOPSO-GMR [37]. The flowchart of MOPSO algorithm is given in Figure 14.

4.2. Multi-Objective Evolutionary Algorithm based on Decomposition

The multi-Objective Evolutionary Algorithm supported on Decomposition (MOEA/D) is an evolutionary algorithm. The MOEA/D has adopted a technique that consists in decomposing the multi-objective optimization problems into several single-objective, called sub-problems. Each sub-problem has its own best solution ever found which is determined associating all of solutions found by algorithm. There are several methods to define the decomposed objective functions, for example: the weighted sum,

the distance or the norm of difference vector, the predefined ideal point in the objective space. The MOEA/D algorithm is developed by Zhang and Li [38]. The flowchart of MOEA/D algorithm is given in Figure 15.

4.3. Pareto Envelope-based Selection Algorithm II

The third multi-objective algorithm is an evolutionary algorithm: Pareto Envelope-based Selection Algorithm II (PESA-II). The PESA-II is based in the same genetic algorithm mechanism associated with the selection based on Pareto envelope. The PESA-II is configured with an external archive to save the approximate Pareto solutions. This means that the Parents and the mutants are selected based on the grids created as well as in the geographical distribution of archive members (external archive) [39].

The general process of the PESAII algorithm is as follows, with IP is the size of the internal population and EP is the size of the external populations(archive). The flowchart of PESAII algorithm is given in Figure 16.

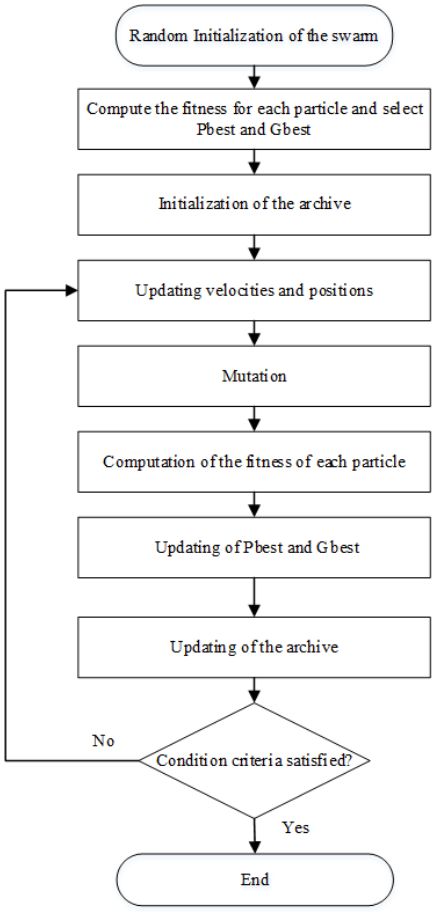


Figure 14: MOPSO Flowchart

4.4. Performances metrics: Hypervolume (HV)

The quality of the Pareto Fronts obtained by the multi-objectives algorithms can be evaluated through a few metrics. The most used performance metric was the *hypervolume* (HV) metric, also known as *S-metric* or *hyper-area*. The hypervolume indicator considers all three aspects: accuracy, diversity and cardinality.

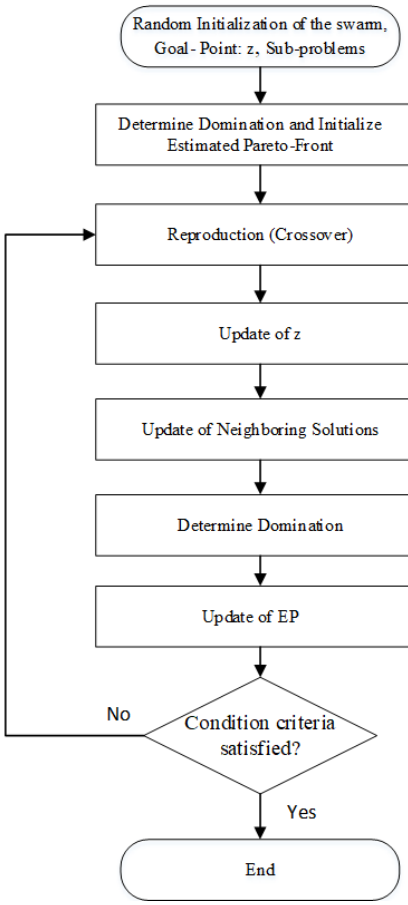


Figure 15: MOEA/D Flowchart

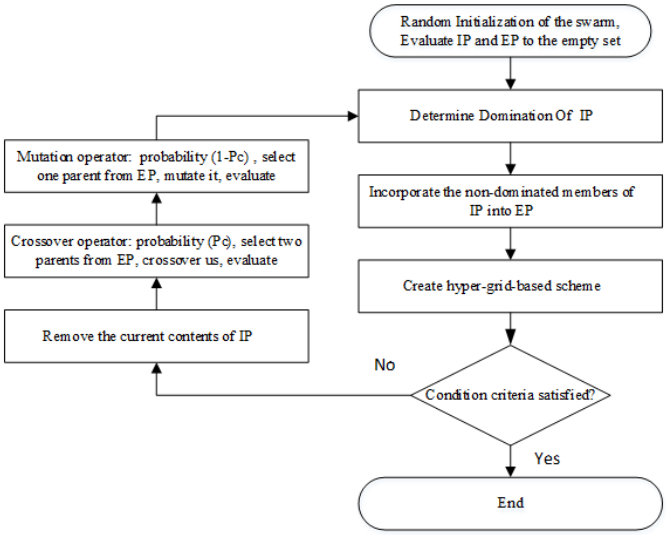


Figure 16: PESAII Flowchart

Thus, it is possible to measure the size of the objective space enclosed by an approximation set. To compute such enclosed space a reference point must be adopted. Figure 17 display a graphical representation of this performance indicator. Whenever one approximation set completely dominates another approximation set, the first HV will be superior than the second. Therefore, the HV is supposed to be *Pareto compliant* [40-41].

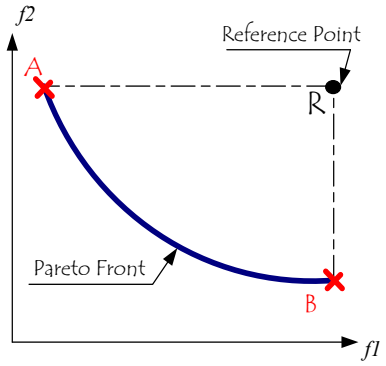


Figure 17: Example of the dominated Hypervolume in 2D.

4.5. Results of the test function and RF inductor applications

The aim consists of optimizing two conflicting performances: maximizing the Quality factor Q, and minimizing the device area, d_{out} by generating the trade-off surface (Pareto front). We give optimization results using the well-known PESAII and MOEA/D techniques for a RF inductor with value of 6 nH at 1 GHz. Also, two test function of ZDT family are used for benchmarking the performance of multi-objective Pareto optimization methods are illustrated in table 8.

Table 8: Multi-objective benchmark functions of Zitzler (ZDT)

Item Function	Objective functions to minimize
ZDT1 Objectives=2 Range: [0; 1]; Dim n=30	$f_1(x) = x_1$ $f_2(x) = g(x)\left[1-\sqrt{\frac{x_1}{g(x)}}\right]$ $with\ g(x) = 1+9\left[\frac{\sum_{i=2}^{n} x_i}{(n-1)}\right]$
ZDT6 Objectives=2 Range: [0; 1]; Dim n=10	$f_1(x) = 1-e^{(-4x_1)}\sin^6(6\pi x_1)$ $f_2(x) = g(x)\left[1-(\frac{f_1(x)}{g(x)})^2\right]$ $with\ g(x) = 1+9\left[\frac{\sum_{i=2}^{n} x_i}{(n-1)}\right]^{\frac{1}{4}}$

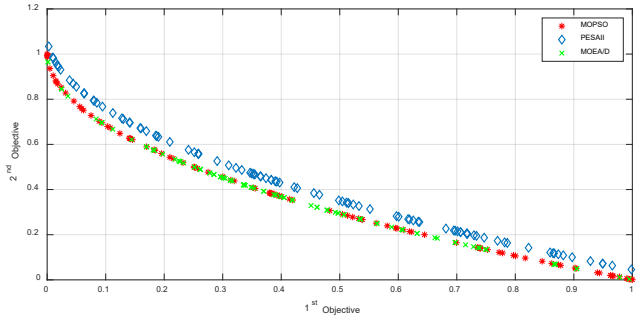


Figure 18: Pareto Front (MOPSO vs. PESAII vs MOEA/D) for ZDT1.

Figure 18, 19 and 20 represents the Pareto fronts obtained using MOPSO, PESAII and MOEA/D for ZDT1, ZDT6 and RF inductor application, respectively. Where it can be noticed that the better distribution given by MOPSO, when compared to PESAII and MOEA/D.

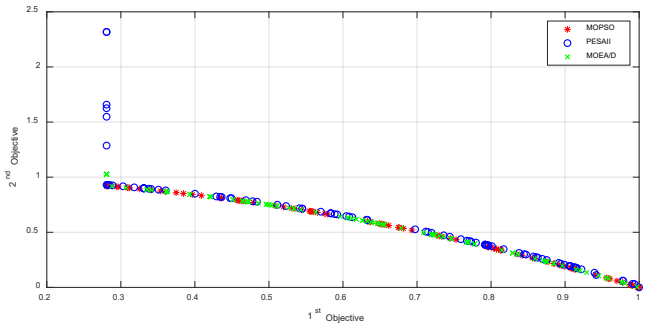


Figure 19: Pareto Front (MOPSO vs. PESAII vs MOEA/D) for ZDT6.

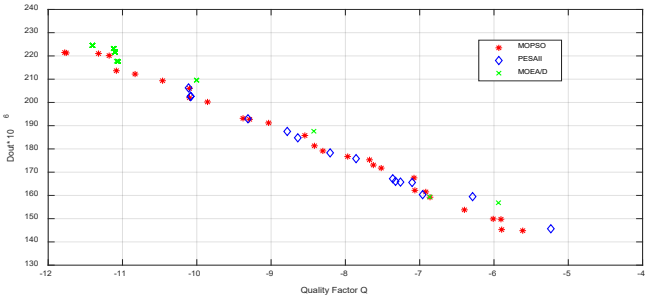


Figure 20: Pareto Front (MOPSO vs. PESAII vs MOEA/D) for RF inductor.

4.5.1. The hyper-volume indicator HV

Table 9 presents the average, maximum, minimum, and deviation values of the hypervolume indicator, over 20 runs of each algorithm, for two test function ZDT1, ZDT6 and the RF inductor application.

Table 9: Average, max, min, and standard deviations of the hypervolume indicator

		PESAII	MOEA/D	MOPSO
ZDT1	Avg	0.6918	0.7291	0.7587
	Max	0.7087	0.7563	0.7609
	Min	0.6643	0.6488	0.7558
	Std	0.0139	0.0346	0.0019
ZDT6	Avg	0.4782	0.4506	0.4869
	Max	0.4873	0.4843	0.4893
	Min	0.4675	0.3840	0.4849
	Std	0.0063	0.0577	0.0016
RF Inductor	Avg	0.5235	0.3272	0.5496
	Max	0.5260	0.4113	0.5508
	Min	0.5196	0.2663	0.5476
	Std	0.0025	0.0577	0.0017

Figure 21 represents the average value of the HV indicators for the 3 multi-objective problems. As a consequence, the MOPSO algorithm admits the highest average value of the HV indicator.

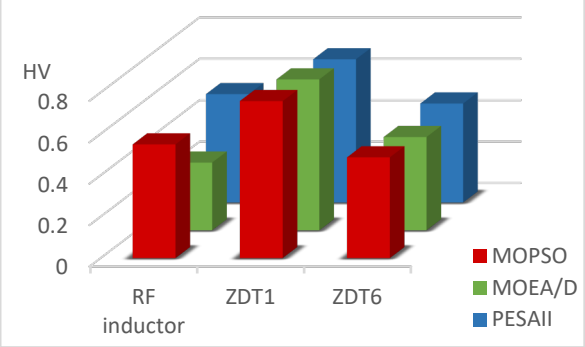


Figure 21: Average values of the HV indicators for the 3 multi-objective problems

4.5.2. *Computing time*

The maximum and the minimum execution times over 20 runs of each algorithm, is given in Table 10. Figure 16 displays the average value of the execution times.

Table 10: Average, Max and Min of the execution time

Time(S)		PESAII	MOEA/D	MOPSO
ZDT1	Avg	152.69	278.48	62.94
	Max	157.56	281.77	73.90
	Min	149.64	275.02	55.67
ZDT6	Avg	130.46	281.47	58.55
	Max	137.07	294.05	66.35
	Min	125.49	268.88	54.36
RF Inductor	Avg	105.91	166.25	54.71
	Max	106.53	221.28	58.07
	Min	105.28	141.38	51.58

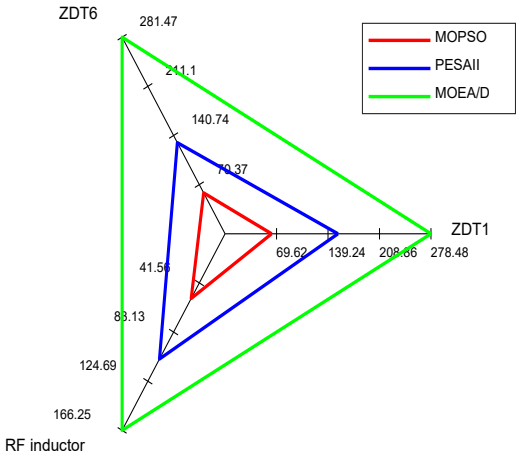


Figure 22: AVG execution time considering the three multi-objective problems

A comparison of the execution times is illustrated using the Radio Chart representation (figure 12) aiming to highlight the speediness of each algorithm. It can easily conclude that MOPSO is the faster when compared with other metaheuristics.

5. Conclusion

An objective comparison among the optimization results obtained via multiple-objective metaheuristics, dealing with performance optimization of RF circuits, is presented in this work. In first part, mono-objective metaheuristics are used. Among these metaheuristics the Swarm Intelligence metaheuristics like the Particle Swarm Optimization (PSO) and the Evolutionary Algorithms such us the Gravitational Search Algorithm (GSA) and the differential evolution algorithm (DE). These three algorithms were used to optimize three benchmark functions and the RF integrated inductors configured with the inductor double π-model. Results obtained shows that the use of swarm-based techniques in sizing and designing RF circuit offers a good argument, principally the PSO algorithm. In second part of this paper, we applied the multi-objective technique to solve two conflicting performances problem: the quality factor (Q) maximization as well as the minimization of the device area (dout) of the same RF application example using three multi-objective algorithms: firstly the multi-objective particle swarm optimization (MOPSO); secondly the multi-objective Evolutionary Algorithm based on decomposition (MOEA/D) and, a third one, the Pareto Envelope-based Selection Algorithm-II (PESA-II) by generating the trade-off surface (Pareto-front). The results obtained, which are supported by two performance metrics, a) the hypervolume indicator (HV) and b) the execution time, shows the following: MOPSO is better than other metaheuristics according to HV indicators, also MOPSO is the faster than other metaheuristics according to the CPU computing times.

References

[1] E.-G. Talbi, Metaheuristics: from design to implementation, Wiley, 2009.
[2] J. Dréo, A. Pétrowski, P. Siarry, E. Taillard, Metaheuristics for hard optimization: methods and case studies, Springer, 2006.
[3] M. Köppen, G. Schaefer, A. Abraham, Intelligent Computational Optimization in Engineering, Springer, 2011.
[4] F. Padilla, A. Torres, J. Ponce, M. D. Torres, S. Ratté, E. Ponce-de-Leon, "Evolvable metaheuristics on circuit design", in: Advances in analog circuits, E. Tlelo-Cuautle (Ed.), Intech, 2011. http://dx.doi.org/10.5772/14688
[5] G. Polya, How to Solve It, Princeton University Press, Princeton, NJ, 1945.
[6] A. Daniel, Evolutionary Computation for Modeling and Optimization, Springer, 2006.
[7] A. E. Eiben, J. E. Smith, Introduction to evolutionary computing, Natural Computing Series, Springer, 2007.
[8] P. Siarry, Z. Michalewicz, Advances in metaheuristics for hard optimization, Natural Computing Series, Springer, 2008.
[9] J. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, University of Michigan Press, 1975.
[10] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Educational Publishers Inc, 1989.
[11] K. Y. Lee, M. A. El-Sharkawi, Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems, IEEE Press Series on Power Engineering, 2008.
[12] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, ''Optimization by Simulated Annealing'', Science, New Series, 220(4598), 671-680, 1983. https://doi.org/10.1126/science.220.4598.671
[13] K.V. Price, R. M. Storn, J. A. Lampinen, Differential Evolution: a practical approach to global optimization, Springer, 2006.
[14] M. Fakhfakh, Y. Cooren, A. Sallem, M. Loulou, P. Siarry, ''Analog circuit design optimization through the particle swarm optimization technique'',

Analog integrated circuits and signal processing, 63(1), 71--82, 2009. https://doi.org/10.1007/s10470-009-9361-3
[15] J. Kennedy, R. C. Eberhart, ‘‘Particle swarm optimization’’, Proceedings of the IEEE international conference on neural networks,1942-1948, 1995.
[16] R. A. Vural, T. Yildirim, ‘‘Analog circuit sizing via swarm intelligence’’, AEU - International journal of electronics and communications, 66(9), 732--740, 2012. https://doi.org/10.1016/j.aeue.2012.01.003
[17] B. Benhala, P. Pereira, A. Sallem, Focus on Swarm Intelligence Research and Applications, Nova Science Publishers, 2017.
[18] A. Sallem, B. Benhala, M. Kotti, M. Fakhfakh, A. Ahaitouf, M. Loulou, ‘‘Application of swarm intelligence techniques to the design of analog circuits: evaluation and comparison’’, Analog Integrated Circuits and Signal, Processing, 75(3), 499-516, Springer, 2013. https://doi.org/10.1007/s10470-013-0054-6
[19] M. Kotti, B, Benhala, M. Fakhfakh, A. Ahaitouf, B. Benlahbib, M. Loulou, A. Mecheqrane, ‘‘Comparison between PSO and ACO techniques for analog circuit performance optimization’’, The international conference on microelectronics, 2011. https://doi.org/10.1109/ICM.2011.6177367
[20] B. Benhala, A. Ahaitouf, M. Fakhfakh, A. Mechaqrane, B. Benlahbib, ‘‘Optimal analog circuit sizing via ant colony optimization technique’’, International journal of computer science and network security, 11(6), 223-231, 2011.
[21] B. Benhala, A. Ahaitouf, M. Kotti, M. Fakhfakh, B. Benlahbib, A. Mecheqrane, M. Loulou, F. Abdi, E. Abarkane, “Application of the ACO technique to the optimization of analog circuit performances”, In: Analog Circuits: Applications, Design and Performance, E. Tlelo Cuautle (Ed.), NOVA Science Publishers, Inc, 2011.
[22] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, ‘‘GSA: A Gravitational Search Algorithm’’, Information Sciences, 179, 2232--2248, 2009. https://doi.org/10.1016/j.ins.2009.03.004
[23] S. Jianga, Y. Wanga, Z. Jiaa, ‘‘Convergence analysis and performance of an improved gravitational search algorithm’’, Applied Soft Computing, 24, 363-384, 2014. https://doi.org/10.1016/j.asoc.2014.07.016
[24] https://www.sfu.ca/~ssurjano/optimization.html(Accessed 13May 2019).
[25] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Educational Publishers Inc, 1989.
[26] M. Ali, A. Torn, ‘‘Population set based global optimization algorithms: some modifications and numerical studies’’, Computer Operation Research, 31(10), 1703-1725, 2004. https://doi.org/10.1016/S0305-0548(03)00116-3
[27] J. M. Chambers, W. S. Cleveland, B. Kleiner, P. A. Tukey, Graphical methods for data analysis, Duxbury Press, 1983.
[28] Yu Cao, R. A. Groves, N. D. Zamdmer, J.-O. Plouchart, R. A. Wachnik, Xuejue Huang, T. -J. King, Chenming Hu, ‘‘Frequency-independent equivalent-circuit model for on chip spiral inductors’’, IEEE Journal of Solid-State Circuits, 38(3), 419-426, 2003. https://doi.org/10.1109/JSSC.2002.808285
[29] P. Pereira, A. Sallem, M. Fakhfakh, M. H. Fino, F. Coito, A Technology Aware Optimization of RF Integrated Inductors, In: Tlelo-Cuautle E. (ed.) Analog Circuits: Applications, Design and Performance, Nova Science Publishers, Inc., 213-234, 2012.
[30] P. Pereira, M. H. Fino, F. Coito, M. Ventim-Neves, ‘‘RF Integrated Inductor Modeling and its Application to Optimization-Based Design’’, In: Analog Integrated Circuits and Signal Processing, 73(1), 4-55, Springer Netherlands, 2011. https://doi.org/10.1007/s10470-011-9682-x
[31] A. C. Coello and M. S. Lechuga, “MOPSO: A Proposal for Multiple Objective Particle Swarm (2002),” Proc. 2002 Congr. Evol. Comput. CEC’02 (Cat. No.02TH8600), 2, 1051–1056, 2002.
[32] X. Li, “A Non-dominated Sorting Particle Swarm Optimizer for Multiobjective Optimization,” 37–48, 2007,2003. https://doi.org/10.1007/3-540-45105-6_4
[33] M. R. Sierra and C. A. Coello, “Improving PSO-Based Multi-objective Optimization Using Crowding, Mutation and ∈-Dominance,” in EMO’2005, LNCS 3410. 505–519, 2010. https://doi.org/10.1.1.79.8066
[34] Deb, K., Pratap, A., Agarwal, S.; Meyarivan, T., “A fast and elitist multi-objective genetic algorithm: NSGA-II”, IEEE Transactions on Evolutionary Computation, 6(2): 182-197. https://doi.org/10.1.1.17.7771
[35] A. Sallem, B. Benhala, M. Kotti, M. Fakhfakh, A. Ahaitouf, and M. Loulou, “Application of swarm intelligence techniques to the design of analog circuits: Evaluation and comparison,” Analog Integr. Circuits Signal Process., 75(3), 499–516, 2013. https://doi.org/10.1007/s10470-013-0054-6
[36] R. A. Santana, M. R. Pontes, and C. J. A. Bastos-Filho, “A multiple objective particle swarm optimization approach using crowding distance and roulette wheel,” in ISDA 2009 - 9th International Conference on Intelligent Systems: Design and Applications, 1, 237–242, 2009. https://doi.org/10.1109/ISDA.2009.73
[37] http://yarpiz.com/category/multiobjective-optimization(Accessed 13 May 2019).
[38] Z. Qingfu and L. Hui, “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition,” IEEE Trans. Evol. Comput., 11(6), 712–731, 2007. https://doi.org/10.1109/TEVC.2007.892759
[39] D. W. Corne, J. D. Knowles, and M. J. Oates, “The Pareto Envelope-Based Selection Algorithm for Multiobjective Optimization,” no. Mcdm, 839–848, 2000. https://doi.org/10.1007/3-540-45356-3_82
[40] L. Thiele and E. Zitzler, “Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach,” IEEE Trans. Evol. Comput., 3(4), 257–271, 1999.
[41] E. Zitzler, D. Brockhoff, and L. Thiele, “The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration,” Evolutionary Multi-Criterion Optimization. 862–876, Springer, 2007. https://doi.org/10.1007/978-3-540-70928-2_64