ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 3-ISSUE 2|MAR-APR 2018

www.astesj.com
ISSN: 2415-6698

EDITORIAL BOARD

Editor-in-Chief

Editorial Board Members

Regional Editors

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 36 accepted papers in Computer Science and Artificial Intelligence domains.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 3 Issue 2 **March-April 2018**

CONTENTS

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 3, No. 2, 1-7 (2018)
www.astesj.com
Special issue on Advancement in Engineering Technology

ASTES Journal
ISSN: 2415-6698

The Internet of Things ecosystem: the blockchain and data protection issues

Nicola Fabiano[*]

Studio Legale Fabiano, 00179, Roma (Italy)

ARTICLE INFO

Article history:
Received: 12 November, 2017
Accepted: 19 February, 2018
Online: 08 March, 2018

Keywords:
Internet of Things
Legal issues
Data Protection and privacy Law
Security
Blockchain
Risks
Legal framework
Privacy standard

ABSTRACT

The IoT is innovative and important phenomenon prone to several services and applications such as the blockchain which an emerging phenomenon. We can describe the blockchain as blockchain as a ser- vice because of the opportunity to use several applications based on this technology. We, indeed, should take into account the legal issues related to the data protection and privacy law to avoid breaches of the law. In this context, it is important to consider the new European General Data Protection Regulation (GDPR) that will be in force on 25 May 2018. The contribution describes the main legal issues related to data protection and privacy focusing on the Data Protection by Design approach, according to the GDPR. Furthermore, I resolutely believe that is possible to develop a global privacy standard framework that organisations can use for their data protection activities.

1 Introduction

To define the Internet of Things (IoT) could be a challenge [1] due to its technical and conceptual complexity [2]. The IoT is a phenomenon founded on a network of objects linked by a tag or microchip that send data to a system that receives it. The IoT includes every connection among objects, so we have machine-to-machine (M2M) systems, where each machine talks with other machine(s), communicating real-time data and information. Nowadays we are faced with several devices but mainly such as smartphones and apps, sensors, chip, and any other electronic system. We read [3]

> *The concept was simple but powerful. If all objects in daily life were equipped with identifiers and wireless connectivity, these objects could communicate with each other and be managed by computers.*

In 2012 the Global Standards Initiative on Internet of Things (IoT-GSI) the Internet of Things (IoT) defined the IoT as "the infrastructure of the information society[1]." Not that the IoT phenomenon is realised only when two or more objects are linked to each other in a network such as the Internet. Apart from this kind of connection, an object could also be indirectly linked to a person, thereby setting up a ring network among objects and people. Its very simple, for example, to imagine a ring network that could link a person with one or more objects (a clock, a chair, a lamp, etc.) equipped with a technological system (RFID, near field communication NFC, etc.). However, the IoT is a virtual reality that reproduces exactly what happens in the real world. Lets imagine that our clock, chair, and lamp all contain chips and are used by a person with special needs. From a medical point of view, it may be crucial, for instance, to know how many times he uses the chair. At the same time, it is necessary to help him by automatically turning on the lamp when he sits in the chair. Using chips, it is possible for the objects to communicate among themselves (e.g., the lamp turning on when the chair sends data that the man is sitting down) and at the same time send data over the Internet for, say, medical analysis. The information provided by each object can be aggregated, thereby creating a profile for him. The profile may contain sensitive information about the man, which raises the possibility of

[*] Corresponding Author Nicola Fabiano, ROME (ITALY), info@studiolegalefabiano.eu

[1] The Internet of Things (IoT) has been defined in Recommendation ITU-T Y.2060 (06/2012) as a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies [4].

https://dx.doi.org/10.25046/aj030201

his being monitored. This is a very important point for privacy. This scenario could present a lot of legal issues related to privacy and data protection law. The main goal is to evaluate the impact of the IoT phenomenon on the fundamental rights such as the right to respect for private and family life according to the European Convention on human rights [5]. There are others legal aspects to take into account developing a project on IoT. It is quite important to highlight that there are differences between data protection and privacy especially in Europe where the Charter of Fundamental Rights of the European Union identifies as fundamental rights privacy (article 7) and data protection (article 8).

2 The correct approach: privacy is not security

The main focal point to address a correct approach to any evaluation of data protection and/or privacy risks, in general, is to understand the differences between security and privacy. The correct equation is the following one:

$$security \neq privacy \tag{1}$$

where security is different from privacy. In fact, according to this principle, it is possible to adopt very high-security measures, but this can not mean to respect privacy law either protect users privacy. Often this concept is every indication that it is necessary to intervene on the security systems to be compliance with privacy law. Obviously, this is a big misunderstanding and could create confusion on the privacy approach and its consequences. Adopting security measures is certainly a value, but it is not the correct way to deal with privacy issues. To address privacy and data protection correctly, it is necessary to start from the privacy by design (or da- ta protection by design and by default) approach as further and better clarified below. Privacy is embedded into design2. More clearly Privacy, having been embedded into the system before the first element of information being collected, extends throughout the entire life-cycle of the data involved, from start to finish.

2.1 Protecting privacy through the privacy by design approach

The Internet of Things represents a global revolution: the objects that people use in the real world can talk to other objects and at the same time to the data subjects themselves. This scenario goes from the Internet of People (IoP) - because of the connection among people - to the Internet of Everything. This consciousness is the real engine that has pressed politicians and regulators to intervene in the IoT realm. In fact, there is a growing desire to create a general, comprehensive, and structured legal framework for the Internet of Things to protect users and consumers. In October 2010, the 32nd International Conference of Data Protection and Privacy Commissioners adopted a resolution on Privacy by Design (PbD) [6] that is a landmark and represents a turning point for the future of privacy. Instead of relying on compliance with laws and regulations as the solution to privacy threats, PbD takes the approach of embedding privacy into the design of systems from the very beginning.

The primary goal is to draw up two concepts: a) data protection and b) user. Regarding privacy, we have always thought in term of compliance with laws, failing to evaluate the real role of the user (and his or her personal data). To develop an adequate data protection and privacy approach, we must start any process with the user the person who has to be protected putting him or her at the centre. That means that during the design process, the organisation always has to be thinking of how it will protect the users privacy. By making the user the starting point in developing any project (or process), we realise a PbD approach.

This methodological approach is based on the following seven foundational principles [7]:

1. **Proactive not reactive; preventative not remedial;**
2. **Privacy as the default setting;**
3. **Privacy embedded into design;**
4. **Full functionality positive-sum, not zero-sum;**
5. **End-to-end security full lifecycle protection;**
6. **Visibility and transparency keep it open;**
7. **Respect for user privacy keep it user-centric.**

We can see why the Privacy by Design approach is so important in the IoT environment. In fact, the Internet of Things should adopt the PbD principles and statements, always placing the user at the centre. The European Data Protection Supervisor (EDPS) has promoted PbD, touting the concept in its March 2010 Opinion of the European Data Protection Supervisor on Promoting Trust in the Information Society by Fostering Data Protection and Privacy [8] as *a key tool for generating individual trust in ICT*. It was not long after this endorsement that the 32nd International Conference of Data Protection and Privacy Commissioners adopted the PbD concept as well. In Europe, this approach became *”Data Protection by Design and by Default”* (DPbDabD) in the EU Regulation 679/2016 [9] and indeed establishing this concept in the law is a welcome development. Nevertheless, it is kind of interesting to notice that the EU legislator used a different expression (i.e., data protection by design and by default) from the one adopted in the international context (i.e., Privacy by Design). These two expressions represent two different methodological approaches. Privacy by Design is structured in a trilogy of applications (information technology, accountable business practices, physical design) and the seven principles quoted above. The EU formulation is more descriptive and not based on a method; also, the by default concept is autonomous, whereas the PbD approach embeds the same concept into by design. According to the text of Article 25 of the Regulation 679/2016, it is clear that the EU legislator considers by design and by default as different concepts, even though the words by design comprehend the concept by default, making the latter phrase redundant. The EU formulation is more descriptive and not based on a method; also, the by default concept is autonomous, whereas the PbD approach embeds the same concept into by design. Furthermore, the EU Regulation 679/2016 seems to pay a lot of attention to the technical and security aspects instead of the legal concerns, as seen in highlighting of the term ”security”. Hence, the Internet of Things should adopt the Privacy by Design principles and statements, always placing the user at the centre.

3 IoT evolution and its applications: a challenge

The IoT phenomenon makes to spring several applications in different sectors (Personal, Home, Vehicles, Enterprise, Industrial Internet) [10]. This is a continuously evolving system, and we see to the development of many application in each sector. In the last few years, it arose the interest (the needing) to guarantee highest security levels both for the Industries and the users.

The fields of Big Data and Blockchain are the leading emerging phenomena in the IoT ecosystem, but people paid attention more to the technical and security issues than the privacy ones.

Certainly, the security aspects are relevant to avoid or reduce the risks for data privacy. However, from a privacy point of view, we cannot dismiss the right approach, according to the PbD principles. In the first phase of analysis, any project has to be evaluated also thinking how to protect privacy data and personal information applying the PbD principles. In concrete, after the evaluation process, the project has to comply with the law and not after starting it. Once the project starts, it does not need any process of compliance with the law because, according to the PbD principles, the same project has to be already in compliance with the privacy law before starting it. In this case, (during the life cycle of the project) it is not required any evaluation of compliance with the law. In fact, any assessment it is necessary during the design phase of the project, just for the nature of the approach ”by design”, applying the PbD principles correctly.

Several IoT applications have been developed in the field named ”smart”, such as smart grid, smart city, smart home, smart car, etc. This indeed represents what is the IoT evolution that it will continue to grow and develop creating a lot of fields of action. In the ”smart context,” we cannot dismiss from the privacy and security risks related to the communication among objects especially in the case of processing personal data.

The main questions are:

- ”Where are the users’ personal data stored?”
- ”Who manage the users’ personal data?”
- ”What kind of security measures has been adopted?”
- ”Can it be considered a smart system compliance with the privacy law?”

The answers depend on the design model used during the developing preliminary phase. In fact, the ”design” is a fundamental topic as we illustrate in the following considerations.

However, in the IoT echo-system are emerging two relevant and complex aspects in part closely related between them: **Big Data** and **Blockchain**. In the last few years, these have been items of interest in the IoT phenomenon, intensifying the interest by whom deals with it, especially because of the implications both from the side of the developers and from the users. Hence, Big Data and Blockchain represent the new challenge and the new applications of the IoT phenomenon.

This scenario entails the need to deepen these aspects, especially regarding the privacy and data protection issues.

3.1 Big Data: privacy issues and risks in the Internet of Things

Despite its many potential benefits, the Internet of Things poses important privacy and security risks because of the technologies involved.

According to the Gartner Newsroom [11], 6.4 Billion Connected Things will be in use in 2016, up 30 percent from 2015 and the device online are estimated to reach 20.8 billion by 2020. This represents a scenario to be monitored not only for the big data phenomenon but also for threats and risks to privacy and security.

A recent study on the threats to our privacy, security and safety, under the ”Cyberhygiene” project [12], carried out the report (not yet published) named ”*Understanding end-user cyber hygiene in the context of the Internet of Things: A Delphi-study with experts*”. This report, in the beginning, says that ”*This study aimed to establish expert consensus concerning the 1) key malicious IoT threats, 2) key protective behaviours for users to safeguard themselves in IoT environments, and 3) key risky user behaviours that may undermine cyberhygiene in IoT environments*”[2]. In conclusion, this report says ”*There was consensus on the need to consider behaviours across IoT lifecycles. By considering behaviour across each lifecycle, we have been able to identify key behaviours that users need to adopt when using IoT devices. Furthermore, we have been able to identify key threats that can, for example, put users sensitive information at risk and risky behaviours that may lead users to be at risk of a successful attack*”.

[2] See also ”Review of Cyber Hygiene practices” - ENISA - `https://www.enisa.europa.eu/publications/cyber-hygiene`

No doubt, therefore, that even in the IoT ecosystem there are important risks and threats to privacy and it should take appropriate precautions.

On the one hand, we can control devices such as vending machines and stereo speakers with our smartphones, manage devices in our homes (domotics for energy saving, security, comfort, communication) by remote control, and use smartphone apps to book reservations or purchase services. Larger-scale IoT applications might include public security systems or warehouse inventory control systems. It is evident the acceleration of the technological evolution in the last few years and the IoT phenomenon it is not exempt[3]. IoT considers the pervasive presence in the environment of a variety of things, which through wireless and wired connections and unique addressing schemes can interact with each other and cooperate with other things to create new applications/services and reach common goals. In the last few years IoT has evolved from being simply a concept built around communication protocols and devices to a multidisciplinary domain. Devices, Internet technology, and people (via data and semantics) converge to create a complete ecosystem for business innovation, reusability, interoperability, that includes solving the security, privacy and trust implications.

On the other hand, we have seen the fast and exponential data growth, data traffic and, hence, another paradigm well-known as Big Data[4]. Big data implies data analysis and data mining procedures but working on big data values[5]. Nowadays it is very simple to develop apps that, by accessing to data, can execute data mining activities with every possible consequence. In this context, the primary goal is to protect data because of their highest value. Among the main risks we can indeed present the following:

- ***Identification of Personal Information***

The IoT system allows you to transfer data on the Internet, including personal data. Personal information may be transmitted only when the object in which the microchip installed is linked to a person. This connection may be direct or indirect.

We could have a direct link when the user is aware of the possible transmission of his or her personal data and gives consent. Alternatively, let us suppose that a person buys something. Alternatively, the connection may be indirect when the object is not linked directly to a person but only indirectly through the use of information that belongs to that person. For example, if we have x objects linked together by the Internet, I might know information about object nr. 1, but I cannot know to whom this information belongs. I can know, however, that objects nr. 2, 3, and so on are connected among themselves and to a person named Jane. In this way, it is possible to link every piece of information provided by the objects (2, 3, etc.) to Jane. Furthermore, if I know that it is possible to link object nr. 1 to the others (2, 3, etc.), I can also indirectly know that the information provided by object nr. 1 likewise belongs to Jane.

- ***Profiling***

There are several risks and threats in the Internet of Things, but the main one is probably profiling [13], [14]. If objects are linked to a person, it will be possible to obtain personal information about that person through the information transmitted over the Internet by each of those objects. Furthermore, these transmitted data may be stored in one or more servers. When a person can be identified through the use of credit or loyalty cards, its very simple to know the types of products purchased and so on to profile the person, learning about his or her habits and lifestyle. The person may have previously provided consent for the dissemination of data related to his or her purchases for advertising purposes. Regarding privacy, is it possible to protect a person? Who manages the personal data? Where will this data be stored? Profiling can also be an issue with the movement toward smart grids and cities, a phenomenon that is close in nature to the Internet of Things. For some years now, there has been an interest in modernising the existing electrical grid by introducing smart meters, which can communicate a consumers energy consumption data to the relevant utilities for monitoring and billing purposes. From a legal perspective, there is the need to consider the privacy issues arising from these initiatives, such as consumer profiling, data loss, data breach, and lack of consent (consent is mandatory by law).

- ***Geolocation***

Geolocation is another risk because nowadays, by our device (first of all the smartphones) it is very simple to find precise details on the location, for instance, digital photos. Inside each picture file there are some fields among them EXIF and GPS that contain the technical information about the photo and also the location where the picture was taken. If the user has not previously deactivated the geolocation service in the camera or smartphone, and the pictures have been published on a website or social network, anyone who views the photo can know exactly where the picture was taken and see who was there.

In this way, privacy could be compromised. When smartphones and other mobile devices are connected

[3] IoT is a concept and a paradigm with different visions, and multidisciplinary activities

[4] Big data is a term for datasets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualisation, querying, updating and information privacy. The term ”big data” often refers only to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. *”There is little doubt that the quantities of data now available are indeed large, but thats not the most relevant characteristic of this new data ecosystem.”* (Wikipedia)

[5] Is well-known the Four Vs of Big Data: Volume, Velocity, Variety and Veracity (IBM). Considering data as value it is possible extend the approach to 5 V (last V as value).

to the Internet, as they typically are, they contribute every day to the Internet of things, sending data ready to be used by other people.

- ***Liability for Data Breaches***

In Europe, there are numerous national and European Community (EC) laws relating to personal data breaches. Hence, the Internet of Things also has effects on liability in cases where the data being collected and transmitted lacks the appropriate security measures. For example, Directive 2002/58/EC [15] states that: *In case of a particular risk of a breach of the security of the network, the provider of a publicly available electronic communications service must inform the subscribers concerning such risk and, where the risk lies outside the scope of the measures to be taken by the service provider, of any possible remedies, including an indication of the likely costs involved.*

Another risk is the loss of data during processing. The consequences entail, of course, liability for the data controller and data processor related to each particular situation. In fact, because the processing of personal data involves risks to the data in question (such as the loss of it), the EU Regulation n. 679/2016 on data protection contains an article requiring data controllers to conduct a data protection impact assessment (DPIA)[6] an evaluation of data processing operations that pose particular risks to data subjects.

According to the Article 35, paragraph 1, of the EU Regulation n. 976/2016 (GDPR) ”*Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks*”.

This is the law prescription on the need to conduct a data protection impact assessment (DPIA) in some cases. This preventive action could avoid or reduce risks for the fundamental rights such as data protection and privacy. This demonstrates as is crucial to pay attention to security and privacy in any project development.

3.2 Blockchain: what about privacy?

The blockchain ”*is a shared, immutable ledger for recording the history of transactions*” [16]; it is a ledger of records. The blockchain was imagined by Satoshi Nakamoto [17]. Blockchain works as a distributed database, and its structure guarantees any modification or alteration due to the strong link and timestamp among each block.

The blockchain was theorised thinking to the web of trust principles and, hence, based on the consensus according to the proof of work concept related to the computational power made available by each one. Subsequently, the participation in the blockchain processes has been revised in terms of proof of stake according to the amount of cryptocurrency hold of each one. Given that, nowadays we can have different kind of participation in the blockchain consensus (proof of work - proof od stake), and according to the solution adopted, it will determined the blockchain governance. Furthermore, it is possible to distinguish three different kinds of blockchain:

1. Public blockchain;
2. Consortium blockchain;
3. Private blockchain [18]

The main differences are essentially related to the permission to write the blocks and participate in the consensus processes. The blockchain can be used to provide services, such as the guarantee about the personal identity. It is possible to implement an application for the electronic identification (eID) through which to store, securely in each block of the block- chain, the personal information that could be used to know with certainty the identity of a person. The blockchain can be also used for public services by the Public Bodies or to storage securely and without risks of modification electronic documents or any other resource.

The blockchain poses problems related to data protection and privacy, considering them as two different concepts (but they are rights). It is a field well-known by the engineers and developers, but often ad- dressed only increasing the security measures instead of considering the data protection and privacy law. Regarding privacy, Satoshi Nakamoto [17] argues that ”*privacy can still be maintained by breaking the flow of information in another place: by keeping public keys anonymous*”. However, the author says also that ”*The risk is that if the owner of a key is revealed, linking could reveal other transactions that belonged to the same owner*”. That represents a significant chink in the privacy perspective. Ensuring privacy and data protection is one of the main aims of any project which has to address ”by design”, not leaving any possibility to compromise personal data and/or personal information. Axon [19] argues that privacy issues can be dealt with ”privacy-awareness” enabling ”*two levels of anonymity: total anonymity, and anonymity to the neighbour group level*”. However, ”privacy-awareness” do not seem a valid solution because this way it is not enough to be compliance with the EU GDPR, according to the Article 25 (Data protection by design and by default).

In another technical contribution [20] you read ”*Maintaining privacy on the blockchain is a complicated issue*”. The authors propose ”*A couple of ways to mitigate but not completely eliminate this issue, if privacy is important for the considered application*”. Privacy is certainly important on the blockchain, and for this reason, it would be better to address the issue finding a ”legal” solution to be compliance with the law.

[6] In the rest of the world this is well-known as Privacy Impact Assessment (PIA).

Other authors [21] say *"Despite the benets provided by these services, critical privacy issues may arise. That is because the connected devices (the things) spread sensitive personal data and reveal behaviours and preferences of their owners. Peoples privacy is particularly at risk when such sensitive data are managed by centralised companies, which can make an illegitimate use of them . . . "*. It is very appreciable these authors' approach [21] because they propose a technical solution presenting it in terms of "private-by-design IoT" [7]. Despite the fact that this proposed solution highlights the concept "by design", from a legal point of view it does not seem to take on the issue related to the obligation required by the EU GDPR.

This short scenario shows how on the blockchain there are certainly privacy issues addressed only providing technical solutions, without any legal reference. Apart from the high technical solution, hence, we cannot dismiss the law obligations, where they are applicable, like in Europe, according to the GDPR mentioned above. This panorama confirms the equation according to security is different from privacy; a system could be very secure but not compliance with the privacy law. On the contrary, a system could be compliance with the privacy law and, hence, very secure (obviously if it has been adopted the security measures).

This is an obligation for the controller. Giving the structure of the blockchain, it seems that any subject or person or owner (as defined by Nakamoto) should be a controller and consequently bound to respect the EU GDPR. From this scenario arise many consequences for the "owner" regarding law obligations.

In fact, according to the GDPR, it is mandatory to "implement appropriate technical and organisational measures" [8]

In certain cases, if there are high risks to the rights and freedom of natural person, it is mandatory to provide a Data Protection Impact Assessment (DPIA) to carry out, prior to the processing, *an assessment of the impact of the envisaged processing operations on the protection of personal data* (Article 35). Another aspect is related to the consent according to the GDPR. In fact, according to the Article 6, paragraph 1.a, *the data subject has given consent to the processing of his or her personal data for one or more specific purposes*. Moreover, according to the Article 7, paragraph 3, *the data subject shall have the right to withdraw his or her consent at any time*. Against this background, it is clear that the technical infrastructure of the blockchain prevents the concrete withdrawal of the consent by the data subject because this solution implies the node deletion. The consent issue represents undoubtedly another legal point related to the blockchain and the data protection law.

According to the technical structure of the blockchain, all the system has not any controller because of the lack of a central controller (a general "supervisor") who is responsible for all the nodes. Each owner, hence, is a controller for the data processing of his node. In this perspective each owner, apart from the general securities profiles of the blockchain, has to respect the law and he is himself is a data processing controller. Due to the blockchain technical configuration, in the event of a node was compromised, it is possible to amount to a controller's liability and, in this case, there are certainly other consequences for the owner's node.

4 Conclusion

The Internet of Things involves all stakeholders from companies to consumers. Focusing on the user (consumer) is particularly important to guarantee a level of confidentiality that will earn the users trust. This solution is made possible by adopting the maximum level of security through the Privacy by Design (PbD or DPbDabD) approach and performing PIAs to evaluate the privacy risks of data collection and processing.

The industries may be wary of efforts to regulate the Internet of Things, as it regards the IoT phenomenon as a source of enormous business opportunities. For example, changes in lifestyle such as the use of more technological services like domotics applications can certainly increase the consumers quality of life (and industrys profits). It will be up to consumers, regulators, and privacy professionals to convince the business sector that understanding the risks related to the IoT will produce the same business opportunities to protect privacy and increase the quality of life.

As I hope I have shown, it is crucial to set up a privacy standard to facilitate a methodological approach to privacy and data protection. With the Internet of Things reaching ever more deeply into peoples lives, it would be beneficial to have an international privacy standard for processing personal data in the same way throughout the world using the forward-looking PbD (or DPbDabD) approach.

From a legal point of view, the main difficulty in setting up and using a privacy standard relates to existing laws, which are different in each nation (and even in different states and provinces within those nations). It is possible to develop a standard privacy framework that organisations can use for their data protection activities, adapting it to national legislation while keeping the central framework for all nation-

[7] You read "With the purpose of preventing this situation, the goal of our research is to encourage a decentralized and private-by-design IoT, where privacy is guaranteed by the technical design of the systems. We believe that this can be achieved by adopting Peer-to-Peer (P2P) systems."

[8] Article 25, par. 1, says *"Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects"*.

states.

Since the Privacy by Design (or DPbDabD) approach is the foundational methodological approach to privacy protection, the privacy standard should be adopted according to PbD principles and statements. At the moment we have no record of international privacy standard model.

A Privacy Management System (PMS) could be a reference model or a software system working on the PbD principles. To develop a PMS confers a benefit to all the stakeholders because in this way it is possible to automate every process guaranteeing a good data protection level, by reducing the privacy and security risks. Furthermore, it is feasible to use the Artificial Intelligence and Machine Learning principles to develop a software based on a PMS to facilitate professionals, public body, Industries and Organizations in their activities.

References

[1] Hahn Jim: The Internet of Things (IoT) and Libraries - Library Technology Reports; Chicago 53.1 (Jan 2017): 5-8,2.

[2] AA.VV.: River Publishers, Digitising the Industry Internet of Things Connecting the Physical, Digital and Virtual Worlds, 2016

[3] Cisco.com. San Francisco, California: Lopez Research, An Introduction to the Internet of Things (IoT), November 2013. Retrieved 23 October 2016 http://www.cisco.com/c/dam/en_us/solutions/trends/iot/introduction_to_IoT_november.pdf

[4] ITU - Global Standards Initiative on Internet of Things (IoT-GSI): The Internet of Things (IoT) - http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

[5] European Convention on human rights - http://www.echr.coe.int/Documents/Convention_ENG.pdf

[6] Resolution on Privacy by Design. 32nd International Conference of Data Protection and Privacy Commissioners, Jerusalem - https://edps.europa.eu/sites/edp/files/publication/10-10-27_jerusalem_resolutionon_privacybydesign_en.pdf

[7] 7 Foundational Principles. "Privacy by Design" - https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

[8] EDPS: Opinion of the European Data Protection Supervisor on Promoting Trust in the Information Society by Fostering Data Protection and Privacy. European Data Protection Supervisor (EDPS) - https://edps.europa.eu/sites/edp/files/publication/10-03-19_trust_information_society_en.pdf

[9] 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) - http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[10] What's The Big Data? - Internet of Things Market Landscape - https://whatsthebigdata.com/2016/08/03/internet-of-things-market-landscape/

[11] Gartner: Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015 - http://www.gartner.com/newsroom/id/3165317

[12] Cyberhigiene project - https://www.petrashub.org/portfolio-item/cyberhygiene/

[13] Ann Cavoukian: Springer, Identity in the Information Society. Identity in the Information Society, 2010

[14] Mireille Hildebrandt: FIDIS. Behavioural Biometric Profiling and Transparency Enhancing Tools, 2009

[15] Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector - http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058&from=en

[16] IBM, Understand the fundamentals of IBM Blockchain - https://www.ibm.com/blockchain/what-is-blockchain.html

[17] Satoshi Nakamoto, Bitcoin: A peer-to-peer electronic cash system - https://bitcoin.org/bitcoin.pdf

[18] Vitalik Buterin, On Public and Private Blockchains. Retrieved from https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/

[19] Louise Axon, University of Oxford - Privacy-awareness in Blockchain-based PKI (2015) - https://ora.ox.ac.uk/objects/uuid:f8377b69-599b-4cae-8df0-f0cded53e63b

[20] Konstantinos Christidis and Michael Devetsikiotis, Blockchains and Smart Contracts for the Internet of Things - http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7467408

[21] Conoscenti Marco, Vetr Antonio; De Martin Juan Carlos - Peer to Peer for Privacy and Decentralization in the Internet of Things - In: 39th International Conference on Software Engineering, Buenos Aires (AR), May 20-28, 2017. pp. 1-3 - http://porto.polito.it/2665723/1/peer_to_peer_for_privacy_and_decentralization_in_the_internet_of_things.pdf

[22] Ahmed Kosba, Andrew Miller, Elaine Shi, Zikai Wen, Charalampos Papamanthou - Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts (2016) - https://eprint.iacr.org/2015/675.pdf

[23] Castelluccia, Claude in: Privacy, Accountability and Trust - Challenges and Opportunities - https://www.enisa.europa.eu/publications/pat-study

[24] Ann Cavoukian, Jules Polonetsky, Christopher Wolf: Identity in the Information Society, Smart Privacy for the Smart Grid: Embedding Privacy into the Design of Electricity Conservation.

[25] Guy Zyskind, Oz Nathan, Alex Sandy Pentland - Enigma: Decentralized Computation Platform with Guaranteed Privacy (2015) - https://arxiv.org/pdf/1506.03471.pdf

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 3, No. 2, 8-20 (2018)
www.astesj.com
Special issue on Advancement in Engineering Technology

ASTES Journal
ISSN: 2415-6698

Cancer Mediating Genes Recognition using Multilayer Perceptron Model- An Application on Human Leukemia

Sougata Sheet[*,1], Anupam Ghosh[2], Sudhindu Bikash Mandal[1]

[1]*A. K. Choudhury School of Information Technology, University of Calcutta, Kolkata, 700098, India*

[2]*Department of Computer Science & Engineering, Netaji Subhash Engineering College, Kolkata, 700152, India*

ARTICLEINFO

Article history:
Received: 14 November, 2017
Accepted: 22 January, 2018
Online: 08 March, 2018

Keywords:
Artificial Neural Network
Multilayer perceptrons
p-value
t-test

ABSTRACT

In the present article, we develop multilayer perceptron model for identification of some possible genes mediating different leukemia. The procedure involves grouping of gene based correlation coefficient and finally select of some possible genes. The procedure has been successfully applied three human leukemia gene expression data sets. The superiority of the procedure has been demonstrated seven existing gene selection methods like Support Vector Machine (SVM), Signal-to-Noise Ratio (SNR), Significance Analysis of Microarray (SAM), Bayesian Regularization (BR), Neighborhood Analysis (NA), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) is demonstrated, in terms of the affluence of each Go attribute of the important genes based on p-value statistics. The result are properly validated by before analysis, t-test, gene expression profile plots. The proposed procedure has been capable to select genes that are more biologically significant for mediating of leukemia then those obtained by existing methods.

1 Introduction

In the body, cancer is the uncontrolled enhancement of unusual cell. When the bodys normal control procedure stop the working, then cancer are build-up [1]. Old cells do not expire and in its place develop out of control, establishing fresh, abnormal cells. These additional cells may create a mass of tissue, called a tumor [2]. Leukemia is a cancer which starts in blood forming tissue, usually the bone marrow [3]. It leads to the over production of abnormal white blood cells [4]. However, the abnormal cells in leukemia do not function in the same way as normal white blood cells [5]. There are several types of leukemia, based upon how speedily the disease is growing up and the type of abnormal cells are generate. Peoples are may be affected at all stage by cancer [6]. In 2015, 54,270 people are prospective to be diagnosed with leukemia. The overall five-year correlative anointing rate for leukemia has more than four fold since 1960. From 1960 to 1965, the five-year correlative anointing rate among whites (only data available) with leukemia was 14%. From 1975 to 1980, the five-year correlative anointing rate for the total population with leukemia was 34.20% [7], and from 2004 to 2010, the overall correlative anointing rate was 60.30% [8]. From 2005-2010, the five-year relative correlative anointing overall were Chronic Myeloid Leukemia (CML) is 59.90% [9], Chronic Lymphocytic Leukemia (CLL) is 83.50%, Acute Myeloid Leukemia (AML) is 25.40% overall and 66.30% for children and adolescents younger than 15 years and Acute Lymphocytic Leukemia (ALL) is 70% overall [10], 91.80% for children and adolescents younger than 15 years, and 93% for children younger than 5 years. In 2015, 24,450 people are expected to die from leukemia with 14,040 males and 10,050 females [11]. In Us 2007-2011, leukemia was the fifth most common cause of cancer deaths in men and the sixth most common in women.

In the field of microarray data analysis, one of the most challenging remittance is gene selection [12]. In gene expression data normally contains a large number of variables genes compared to the numbers of samples [13]. The conventional data mining methodology cannot be directly used to the data due to this identity problem [14]. In this reason analysis of gene expression data used dimension reduction procedure [15]. The gene selection which deducts the genes

[*]Corresponding Author: Sougata Sheet & sougata.sheet@gmail.com

https://dx.doi.org/10.25046/aj030202

extremely related to the pattern of every types disease in order to escape such problems [16]. Parametric and non-parametric tests are the statistical approach [17]. For example t-test and Wilcoxon rank sum test have been thoroughly used for searching differentially revealed genes since they are instinctive to understand and implement [18]. But they have a restriction to propagate, if more than two classes and require time swallowing coordination to solve the problem of multiple testing [19]. For three or more groups, the Kruskal-Wallis test can be used. But it may be created biased result because of reliance on the number of samples, when it is used to microarray data whose sample size are normally unbalanced [20]. Many diseases, they are reason by the problems such as chromosomal disequilibrium and gene mutations, which give away abnormal gene expression patterns. These pattern get the information about underlying genetic process and states of several types disease. If these patterns can be analyzed appropriately, they can be effective for recognize the disease sample and identifying the extent to which a patient is affliction from the disease and which can be help in the supervision of disease.

The microarray gene expression data have been collected to underlying biological process of a number of diseases [21]. It is very essential to narrow down from thousands of genes to a few disease genes and gene ranking [18]. In microarry data analysis genes selection is most important phase. For classification of data analysis, several forms of technique have been proposed for gene ranking [22]. These are classified into three several types: filter, wrapper and embedded process [23]. Each of these categories has its personal advantages and disadvantages. For example, filter process are computationally useful and simple but minor performance than the other process. On the other hand, wrapper and embedded procedure are comparatively much complicated and computationally costly but it usually gives excellent classification performance as they mainly apply classifier characteristics in gene ranking. Filter procedure include T-score, which is t-statistic standardized interrelation between input and output class labels [15]. On the other hand, wrapper and embedded procedure include Support Vector Machine (SVM) and its variants, random forest-RFE, elastic net, guided regularized random forest [24], balanced iterative random forest [25] etc . Main distinction of filter process and wrapper or embedded procedure is how they behave samples when ranking genes. For example, in filter procedure, all the samples are usually used for gene ranking but the quality and relevance data samples are ignore [26]. On the other hand wrapper or embedded procedure, classifier such as boosting algorithm, logistic regression, Support Vector Machine (SVM) etc., are used to gene ranking [27].

For complicated data analysis, we can used an Artificial Neural Network (ANN) model [28]. An ANN was externally used for solve the problem such as diagnosis of different types of cancer such as speech recognition, breast cancer [28] [29]and cervical cancer. Several types of effective researches on blood cells using neural network model has been committed [30]. Ongun *et al* enhancement a completely automated classification of bone marrow and blood using several types of way such as neural network and support vector machine (SVM). The best performance of SVM with accuracy of 91.05% as parallelism to Multilayer Perceptron (MLP) network using Conjugate Gradient Descent, Linear Vector Quantization (LVQ), and k-Nearest Neighbors (KNN) classifier which generate accuracy of 89.74%, 83.33% and 80.76% respectively.

In this paper, we proposed a method based on neural network models for identify a set of possible genes mediating the development of cancerous growth in cell. We said this model is Multilayer Perceptron Model-1 (MLP-1) and Multilayer Perceptron Model-2 (MLP-2) [31]. At first we form group of genes using correlation coefficient, then we select the most important group. Finally, we present a set of possible genes get by this method, which may be responsible for cancerous growth in human cell. In this article, we consider three human leukemia gene expression data sets. The usefulness of the procedure, along with its excellent result over several others procedure, has been demonstrated three cancer related human leukemia gene expression data sets. The results have been compared seven existing gene selection methods like Support Vector Machine (SVM), Signal-to-Noise Ratio (SNR), Significance Analysis of Microarray (SAM), Bayesian Regularization (BR), Neighborhood Analysis (NA), Gaussian Mixture Model (GMM), and Hidden Markov Model (HMM). The results are appropriately validated by some previous investigations and gene expression profiles, and compared using t-test, p-value, and number of enriched attributes. Moreover, the proposed procedure has get more number of true positive genes than the existing ones in identifying responsible genes.

2 Related Work

In this article, we have proposed procedure based neural network models for identification of cancer mediating genes. On existing gene selection methods, we have made a survey for comparative analysis [32]. Among them, we have select some existing gene selection methods like SVM, SNR, SAM, BR, NA, GMM, HMM.

SVM is a one type of machine learning procedure which is differ two classes by maximizing the margin between them [33]. For cancer classification, support vector machines (SVMs) is used to identify important genes [34]. The Lasso (L1) SVM and standard SVM are often considered using quadratic and linear programming procedure. A recurrent algorithm is used to solve the Smoothly Clipped Absolute Deviation (SCAD) SVM efficiently. Almost all the cases, it is noticed that with smaller standard errors the SCAD-SVM selects a smaller and a more stable number of genes than the L1-SVM. Another algorithm of gene selection using the weight magnitude as ranking criterion is Recursive Feature Elimination (RFE) SVM [35]. The SVM-RFE procedure ranks all the genes according to some scoring operation, and remove one or more genes with the lowest score values [36]. When the maximal classification accuracy is achieved, then the procedure will be stop [37].

Signal-to-noise ratio (SNR) is applied to rank the correlative genes according to their discriminative power. The procedure starts with the evaluation of a single gene, and frequently finding for other genes based on some statistical criteria [38]. The high SNR genes scores are select as the significant ones. Measurement of SNR score are affected by the size of variables [39]. When there are more than variables, the average and disunity of the other variables of another classes are dependent on the number of variables and data dispersion, which effect SNR ranking of the important variables due to the enhancement in noise in the data. The procedure is more efficient of finding and ranking a smaller number of important variables, when the number of variables can be decrease significantly.

Significance analysis of microarray (SAM) is a one type of gene selection procedure which use a set of gene specific *t*-test and identify genes with statistical significant changing in expression values [40]. The basis of change in the gene expression values, every gene is assigned a score value. If the gene score value the greater than a threshold value which indicate potentially significant [41]. False Discovery Rate (FDR) is the percentage of such genes identified by chance [42]. In order to calculate FDR, insignificance genes are identified by analyzing layout of the measurements. Identify the smaller or larger sets of genes can be adjusted by using threshold value, and FDRs are computed for every set. The main problem is that, in permutation step where entire gene group are put in one group for evaluation. This needs an expensive computation and it likely confuses the analysis because of the noise in gene expression data.

A simple Bayesian procedure has been used to remove the regularization parameter [43]. The value of a regularization parameter is determined by degree of sparsely, which is get an optimal result. Normally this include a model selection step, and calculate the minimization of cross validation error based on intensive search.

Neighborhood analysis (NA) is a procedure for clustering multivariate data analysis based on a given distance metric over the data. Functionally, purpose of NA and K-nearest neighbor algorithm are same [44]. Any significant correlation cannot detect and this is the main disadvantage. This problem may be due to the few number of genes and also likely that the phenotype is too complicated to be connected with a cluster of genes, and a more extend relationship may present in gene expression.

A Gaussian mixture model (GMM) is represented as a weighted sum of gaussian component densities which is based on a parametric probability density function [45]. For parameter selection GMM has been used. We have designed GMM on microarray gene expression data for gene selection. On the other hand, for genes identification, we have designed Hidden Markov Model (HMM) on microarray gene expression data sets. Normally HMMs provide an intuitional framework for representing genes with their several functional properties, and proficient algorithms can be creating to use these models to identify genes.

3 Methodology

Let us assume a set $G = (g_1, g_2, ..., g_i)$ of i genes are known which is hold the normal samples for first m expression values and diseased samples for subsequent n expression values. Now correlation coefficient of gene based normal samples are calculated. Therefore, correlation coefficient R_{qr} within q^{th} and r^{th} is given by [46] [47]

$$R_{qr} = \frac{\sum_{l=1}^{m}(g_{ql} - y_q) * (g_{rl} - y_r)}{\sqrt{(\sum_{l=1}^{m}(g_{ql} - y_q)^2)} * \sqrt{(\sum_{l=1}^{m}(g_{rl} - y_r)^2)}} \tag{1}$$

Here y_q and y_r are the mean of expression values of q^{th} and r^{th}, gene respectively in normal samples.

Similarly, for diseased samples the iteration coefficient R_{qr} between q^{th} and r^{th} genes is given by

$$R'_{qr} = \frac{\sum_{l'=1}^{n}(g'_{ql'} - y'_q) * (g'_{rl'} - y'_r)}{\sqrt{(\sum_{l'=1}^{n}(g'_{ql'} - y'_q)^2)} * \sqrt{(\sum_{l'=1}^{n}(g'_{rl'} - y'_r)^2)}} \tag{2}$$

Using equation 1 and 2 each pair of genes is computed. The genes are located in the similar group if $R_{qr} \geq 0.5$. Now we have used interrelation coefficient to narrow down hearted the invention space by searching genes of a comparable behavior in terms of related expression patterns. The set of responsible genes mediating certain cancers are recognized in this procedure. The choice of 0.50 as a threshold value has been done through extensive experimentation for which the distances among the cluster center have become maximize [48]. The main set of genes is obtained in this pathway [49].

An extremely simplified model of biological structures is an Artificial Neural Networks that imitative the conduct of the human brain. A huge number of interrelated processing components of the layered structure is composed and intended to imitative biological neurons. MLP model is the one of the most popular ANN types with back-propagation algorithm. Figure 1, show the architecture of MLP model containing of interconnected neurons of three layers [50]. This three layers are input layer, hidden layer, and output layer [51]. This model is represented as $n \times p \times q$. Where input layer consist of n number of neurons, hidden layer consist of p number of neurons and output layer consist of q number of neurons. In the higher layer every neuron in every layer is completely attached to all neurons and every link has a weight connected with it. In interconnected neurons, these weights are define the nature and strength of the influence. The output signals from one layer are conducted to the consequent layer by using links that amplify the signals based on the interconnected weights [52]. Exception of the input layer neurons, the total input of every neuron is the sum of weighted outputs of the previous layer neurons. In hidden and output layers neurons, we can calculate the output by using sigmoid logistic function such as an activation function.

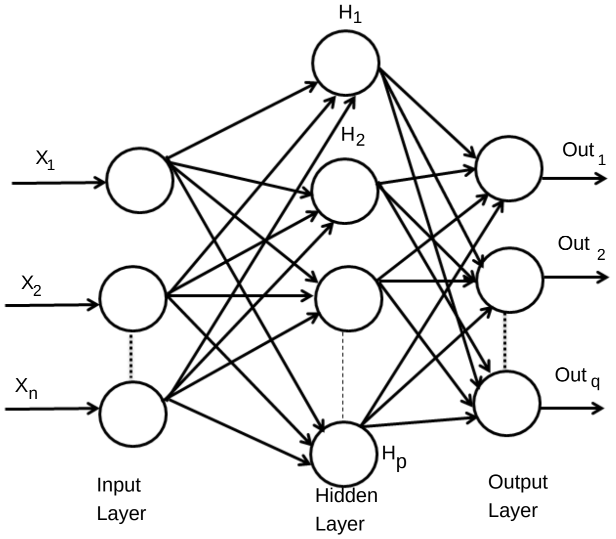


Figure 1: Architecture of an MLP network

Hidden layer neurons execute a non-linear alteration that qualifies the MLP model to simulate a more difficult and non-linear structure within the constraints of a three layer MLP model and more than one hidden layer are used. By employing an incremental adaptation approach, the MLP model has generalized curve fitting capability. Input patterns and output patterns was carried out by MLP model and specified proportion is randomly selected. A learning procedure to correct the linking weights recurrently and to reduce the system error mechanism by every forward processing of the input signal by using back-propagation algorithm. In the initial stage of the learning procedure, to create a input pattern from the input layer to the output layer. The error of every output neuron was calculated from the difference between the calculated and desired outputs. $\varepsilon(p)$ is the system error of p^{th} training pattern, which is defined as

$$\varepsilon(p) = \frac{1}{2}\sum_{k=1}^{q}(D_k(p) - x_k(p))^2 \tag{3}$$

Where $D_k(p)$ and $x_k(p)$ are the k^{th} element of the desired output and calculated output respectively. On the other hand number of neurons in output layer is q. Readjustment of the weights in the hidden layers and output layers is the next step by using a generalized delta rule which is minimizing the distinction between the desired outputs and calculated outputs. Every interconnection weight of the incremental correction can be calculated by

$$\Delta\omega_{kj} = -\gamma\frac{\partial\epsilon(p)}{\partial\omega_{kj}} + \rho\Delta\omega_{kj}(p-1) \tag{4}$$

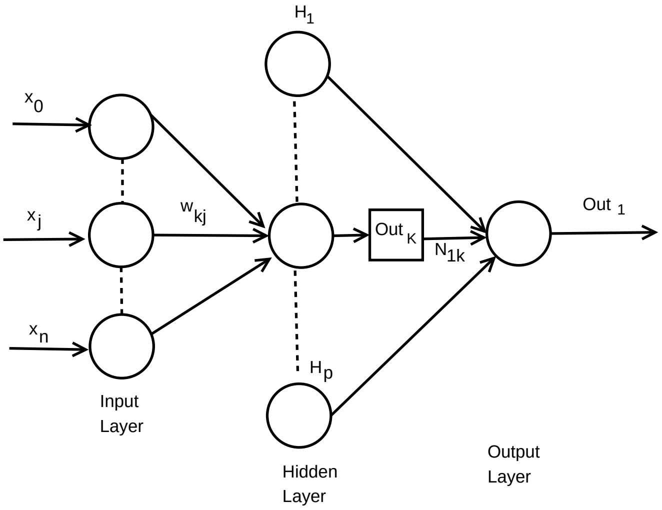


Figure 2: $n \times p \times 1$ MLP model architecture.

The incremental correction of the interconnection weight between j^{th} neurons and k^{th} neuron is $\Delta\omega_{kj}(p)$, in the last iteration, the incremental correction is $\Delta\omega_{kj}(p-1)$ learning rate is γ and the range of $0 < \gamma < 1$, momentum factor is the ρ and range is $0 \le \rho < 1$. By using learning rate the updated weight is control and improve of the effectiveness of learning procedure by using momentum. Training set carried when the squared errors of average sum are over and all training patterns was globally reduced. On training period completion, the testing period was conduct those input pattern which was not present in the training set. Now we consider a MLP model with $n \times p \times 1$ network. Input layer, hidden layer and output layer are n, p, 1 respectively. Show in Figure 2. For both hidden and output neurons was select sigmoid function as the activation function. The mapping function of input-output are realize by the network Out_1 can be computed as

$$Out_1 = \phi(U_1) = \frac{1}{1 + exp^(-U_1)} \tag{5}$$

$$Out_k = \phi(u_k) = \frac{1}{1 + exp^(-u_k)} \tag{6}$$

Where

$$U_1 = \sum_{k=0}^{p} N_{1k} Out_k \tag{7}$$

$$u_k = \sum_{j=0}^{p} \omega_{kl} x_j \tag{8}$$

Algorithm 1 Training procedure - MLP-1

```
Step 1: Initialize all weight and bias.
Step 2: While terminating condition is not satisfied;
    Step 2.1 For each training tuple; propagate the input forward,
Step 3: For each output layer of unit $q$,
    Step 3.1: Output of an input unit, which is actual input values,
Step 4: For every hidden layers or output layers,
    Step 4.1: Compute the net input of unit $q$ with respect to the previous layer $p$;
    Step 4.2: Compute the output of $q$, layer.
Step 5: Compute the error of output layer of unit $q$;
Step 6: Compute the error with respect to the next higher layers.
Step 7: Increment the weight of every layers.
Step 8: Update the weight of every layers.
Step 9:Increment the bias of every layers.
Step 10: Update the bias of each layers.
Step 11: After some iteration, when all $\Delta\omega_{pq}$ in the previous epoch were so small as to bellow specified threshold 0.05, then the procedure will be stop.
```

Algorithm 2 Training procedure - MLP-2

```
Step 1: Initialize the weight and bias.
Step 2: Perform a vector of row which is the weight interconnection between the hidden layers ($p$ nodes) and output layer.
Step 2: Perform $n \times p \times$ matrix and interconnection weights between the input nodes ($n$) and hidden nodes ($p$) is $\omega$.
Step 3: Row vector are computed.
Step 4: Relative weight of input node are computed.
Step 4: After some iteration, when all weight values in the previous epoch were so small as to bellow specified threshold 0.05, then the procedure will be stop.
```

Interconnection weight between the k^{th} hidden neuron and output neuron is N_{1k} interconnection weight between the k^{th} hidden neuron and j^{th} input neuron is ω_{kj}. In the j^{th} hidden neuron the output value is Out_k. x_j is the input of the j^{th} input neuron, where $Out_0 = -1$. The case belongs to class 1, where $Out \geq 0.05$ and case belongs to class 0 where $Out_1 < 0.05$. A node can recognized an output of that node given an input or set of inputs by using activation function. A linear access can be produce 1 or 0 output, but non-linear process the activation function can be generate the output in the specific limit. The activation function can be accepted large forms construct on the data sets. A set of training samples, comparing the networks prediction for each sample with the actual known class level are frequently generating by using back-propagation learning method. The weights are changes as to minimize the mean squared error between the actual class and the prediction network for each sample. The weight and every bias are initialized to small random number of the network. Now we can calculate the total input and output of each unit in the output layer and hidden layer. At first the sample are fed to the input layer of the network. Note that for unit q in the input layer, its output is equal to its input, that is $Out_q = \eta_q$ for input q. Given an unit q in a hidden layer or output layer, the net input η_q, to unit q is

$$\eta_q = \sum \omega_{pq} Out_p + \beta_q \tag{9}$$

Where ω_{pq} is the weight of the connection from unit p in the previous layer to unit q; Out_p is the output of unit p from the previous layer and β_q is the bias of the unit. Given the net input η_q to unit q and output of unit q is computed as

$$Out_q = \frac{1}{1 + \lambda^{-\eta_q}} \tag{10}$$

The error is created backwards by updating the weight and bias in the network. For a unit q in the output layer, the error κ_q is computed by

$$\kappa_q = Out_q(1 - Out_q)(\chi_q - Out_q) \tag{11}$$

Where Out_q is the actual output of unit q and χ_q is the true output. The error of hidden layer of unit of unit q is

$$\kappa_q = Out_q(1 - Out_q) \sum \kappa_r \alpha_{qr} \tag{12}$$

Where α_{qr} is the weight of the connection from unit q to unit r and κ_r is the error of unit r. Now the weight of every layer are incremented and update the weight value of each layer.

$$\Delta\omega_{pq} = (\mu)\kappa_q Out_q \tag{13}$$

$$\omega_{pq} = \omega_{pq} + \Delta\omega_{pq} \tag{14}$$

Now the bias of every layer are incremented and update the all bias value of each layer.

$$\Delta\beta_q = (\mu)\kappa_q \tag{15}$$

$$\beta_q = \beta_q + \Delta\beta_q \tag{16}$$

After some iteration, when all $\Delta\omega_{pq}$ in the previous epoch were so small as to bellow specified threshold 0.05, then the procedure will be stop. The procedure will be describe in algorithm 1 and algorithm 2.

4 Description of the Data sets

In this work, we can select three types of leukemia gene expression data sets. The data sets ID is GDS-2643, GDS-2501, GDS-3057 and title of the data set is Waldenstrom's macroglobulinemia (B lymphocytes and plasma cells), B-cell chronic lymphocytic leukemia, and Acute Myeloid Leukemia respectively. The data base web link is http://ncbi.nlm.nih.gov/projects/geo/.

The data set (GDS-2643) consists of 22,283 numbers of genes with 56 samples. Among them, there are 13 normal samples which consist of 8 normal for B lymphocytes and 5 normal plasma cells and 43 diseased samples which consist of 20 Waldenstrom's macroglobulinemia, 11chronic lymphocytic leukemia, 12 multiple myeloma samples.

The dataset (GDS-2501) data set consists of 22,283 numbers of genes with 16 samples. In this sample analysis of B-cell chronic lymphocytic leukemia (B-CLL) cells that express or do not express zeta-associated protein (ZAP-70) and CD38. The prognosis of patients with ZAP-70-CD38- B-CLL cells is good, those with ZAP-70+CD38+B-CLL cells is poor.

The dataset (GDS-3057) content 26 Acute Myeloid Leukemia (AML) patients with normal hematopoietic cells at a variety of different stages of maturation from 38 healthy donors The total data set consist of 22,283 number of genes with 64 samples. Among them, there are 38 normal samples which contents 10 normal for bone marrow, 10 normal samples for peripheral blood, 8 normal samples for bone marrow CD34 plus and 10 normal samples for Primed peripheral blood hematopoietic stem cells (PBSC) CD34 plus. On the other hand there are 26 leukemia samples which contents 7 bone marrow and 19 peripheral bloods.

5 Comparative Performance Evaluation of the Models

In this section, the usefulness of the procedure has been demonstrated three types of human leukemia gene expression data sets. Now we can apply comparative analysis with seven existing methods like SVM, SNR, SAM, BR, NA, GMM and HMM. We have applied this procedure on the gene expression data sets for selecting important genes. We have found two classifier groups. One is normal class and another is disease class. After some iteration, we have found normalized value of every gene. Here we have considered a threshold value which is 0.05. After normalization if the gene value is grater then 0.05, then which types of gene is normal gene. After normalization if the gene value is less than 0.05, then which types of gene is disease gene. We consider several genes that are most significant of our experiment. The gene expression values are significantly changes from normal samples to diseased samples. Applying this process on the first data set (GDS-2643), we have found that genes like CYBB, TPT1 and PRDM2 among the most important genes which are over the expressed the diseased samples. On the other hand CRYAB minimize the expression value and fully significant in diseased samples. The gene are recognize as an under expressed gene. In order to limited size of manuscript, we have showed only the profile plots of genes of GDS-2643 data set (depicted in Figure 3). In the case of GDS-2501 data set, the genes like ACTB, CENPN, ALCAM, PXN have changed their expression values for normal samples to diseased samples. Similarly, the dataset GDS-3057, like TDG, CTIF, LLS, NAB2 have changed their expression values for normal samples to diseased samples. The usefulness of the methodology has been shown three types of leukemia gene expression data. We have applied the methodology on the aforesaid gene expression data sets for selecting some important gene intercede diseases. Now we have applied the procedure on the previous gene expression data sets for selecting some important diseases mediating genes. In this procedure, at first based on correlation values the genes are placed into group. For GDS-2643, we have found six groups which holding 1869, 1131, 1033, 601, 537 and 1208 number of genes (Table 5). The most important group by both MLP-1 and MLP-2 has been selected 1869 number of group genes. Now we have considered two classes for both MLP-1 and MLP-2. One is normal samples class and other is diseased samples class. ω is a weight coefficient value which is initialized by random numbers. Both MLP-1 and MLP-2, we have found 28 and 30 number of genes respectively, when grouping of genes and most important groups selection are completed. Now we have found 20 numbers genes that are present in both procedure. Among them, based on their ω values we have selected 18 number of genes.

Similarly in GDS-2501, we have found eight groups which containing 652, 1031, 1217, 1301, 816, 539, 741 and 912 number of genes. Similarly in GDS-3057, eight group of genes which contain 2521, 2624, 2241, 2341, 2471, 803, 2191, and 2238. Now we have found 1869 genes for GDS-2643, 1217 genes for GDS-2501 and 2624 genes for GDS-3057 respectively by applying both MLP-1 and MLP-2. Finally we have found 24, 21 and 20 number of most important genes corresponding to the three data set by using MLP-1. On the other hand we have found 25, 21 and 19 number of most important genes corresponding to the three data set by using MLP-2. The number of genes which is found by both MLP-1 and MLP-2 are 18, 21 and 17 to these data sets. Table (5) show that for different sets of genes the number of functionally enriched attributes corresponding to these methods. It has been found that both procedure MLP-1 and MLP-2 performed the best results for all data sets. These results show that

Data set ID	Selected group	No of selected groups from selected group	Group	No genes in each group
GDS-2643	1	18	1	1869
			2	1131
			3	1033
			4	601
			5	537
			6	1208
GDS-2501	2	21	1	652
			2	1031
			3	1217
			4	1301
			5	816
			6	539
			7	741
			8	912
GDS-3057	4	17	1	2521
			2	2624
			3	2241
			4	2341
			5	2471
			6	803
			7	2191
			8	2238

Table 1: Selection of groups and genes for different data set

the both MLP-1 and MLP-2 procedure has been able to select the more number of important genes responsible for mediating a disease than the other seven existing procedure.

5.1 Validation of the Result

In this part, we have analysis the results which is founded by different types of procedure including MLP-1 and MLP-2. In this comparison we have using p-value, t-test, biochemical pathways, F-test and sensitivity.

5.1.1 Statistical Validation

Prosperity of every GO attributes of every genes has been computed by its p-value. When the p-value is low, it means that the genes are biologically significant. Now we have create a comparative analysis, by using some other procedure like SVM, SNR, SAM, BR, NA, GMM and HMM. For different sets of genes Table 2 show that number of functionally enriched attributes corresponding to these procedure. For all data set we have been show that MLP-1 and MLP-2 performed the best result. The more important genes are select by these procedure are show in these results. Both NFM-1 and NFM-2 are accomplished of getting more number of true positive genes in terms of identifying the GO attributes and cancer attributes with respect to other existing procedure are depicted in Figure(4). Now we have 478 GO attributes of 375 gene set are identified among them 102 GO attributes of cancer related are identified.

Now we validate the results statistically, we have performed t-test on the genes identified by DLM on each data sets. t-test is the statistical significance which indicate whether or not the difference between two groups average most likely reflects an original difference in the population from which the group wear sampled. The t-value show the most significant genes (99.9%) which p-value <0.001. For this three types of data set we can apply t-test and we get corresponding t-value. We have identify some important genes like IARS (4.78), MMP25 (5.68), TYMS (4.96), HPS6 (5.24), MLX (4.12), CALCA (4.32), HIC2 (4.12), ANP32B (5.16), TFPI (4.51), CRYAB (3.08), NCF1C (3.71), HNRNPH1 (3.12), etc. The number in the bracket shows t-value of the corresponding gene. The t-value of this genes exceeds the value for $p = 0.001$. This means that this gene is highly significant (99.9% level of significance). Similarly genes like ERCC5 (3.52), PRDM2 (3.45), PRIM2 (2.54), TPT1 (3.35), RPS26 (2.41), EFCAB11 (3.71), PRPSAP2 (3.43), PRKACA (2.44), etc exceed the t-value for $P = 0.01$. It indicated that this gene is significant at the level of 99%. Similarly genes like MED17 (2.55), MAPK1 (2.35), PIK3CB (2.45), NMD3 (2.15), ARG2 (2.16), EXOC3 (2.56), WHSC1 (2.71), RFC4 (2.35), GLB1L (2.71), HNF1A (2.41) etc exceeds the value for $P = 0.05$. It indicate that this genes significant at the level of 95%. Similarly genes like FLG (1.77), TXNL1 (1.24), RIN3 (1.34), CYBB (2.31), ZNF814 (1.45), KLF4 (1.18) etc exceeds the value for $P = 0.1$. It indicate that this type of genes significant at the level of 90%.

In GDS-2643, we have found cancer pathway for non-small cell and small cell leukemia. In these two

Data set	Gene Set	MLP-1	MLP-2	SVM	SNR	SAM	BR	NA	GMM	HMM
GDS-2643	First 5	60	58	59	10	12	15	10	27	20
	First 10	72	73	65	17	18	20	13	32	25
	First 15	77	79	71	21	22	26	12	41	29
	First 20	79	92	76	30	26	30	16	45	34
GDS-2501	First 5	83	78	85	15	55	48	33	39	48
	First 10	88	85	88	24	62	55	41	44	60
	First 15	101	98	95	31	68	72	52	51	55
	First 20	107	106	103	39	84	65	66	68	77
GDS-3057	First 5	45	52	37	21	24	26	24	19	20
	First 10	63	62	46	35	30	29	33	25	28
	First 15	67	65	57	37	29	31	39	27	35
	First 20	69	70	55	41	31	35	42	36	39

Table 2: Result of several sets of genes on number of attributes

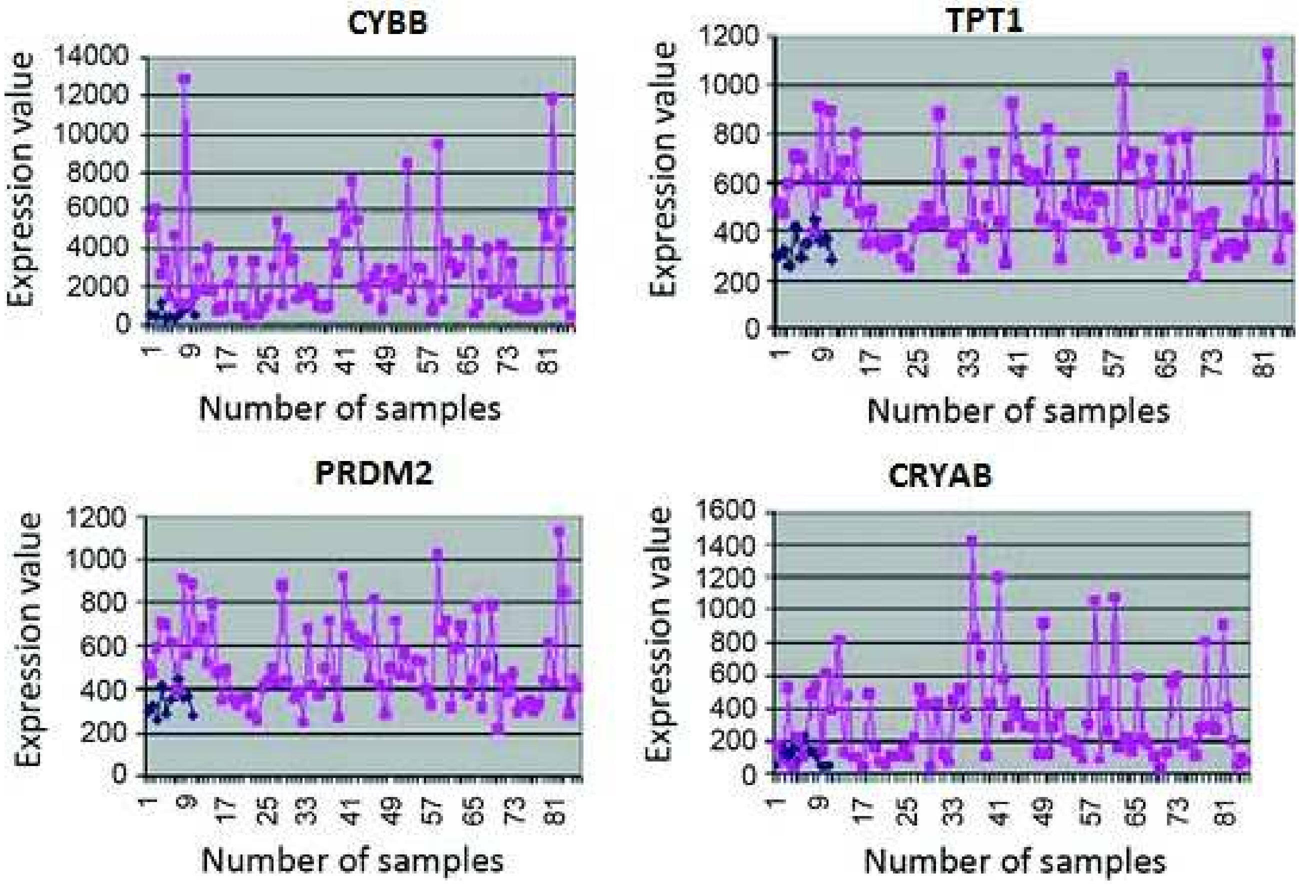


Figure 3: Expression profiles of some over-expressed genes (CYBB, TPT1, PRDM2) and under-express (CRYAB) in normal (shown by blue points) and disease (shown by red points) sample of human leukemia expression data.

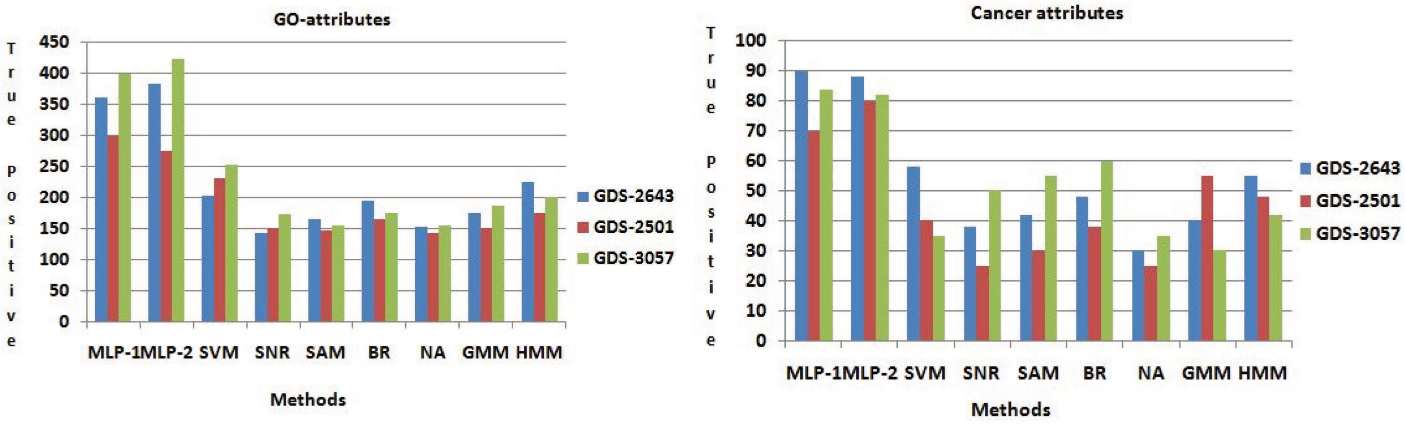


Figure 4: Identification of GO attributes and Cancer attributes the Performance comparisons of MLP-1 and MLP-2 with other seven existing methods.

Comparison among the methods in biochemical pathway

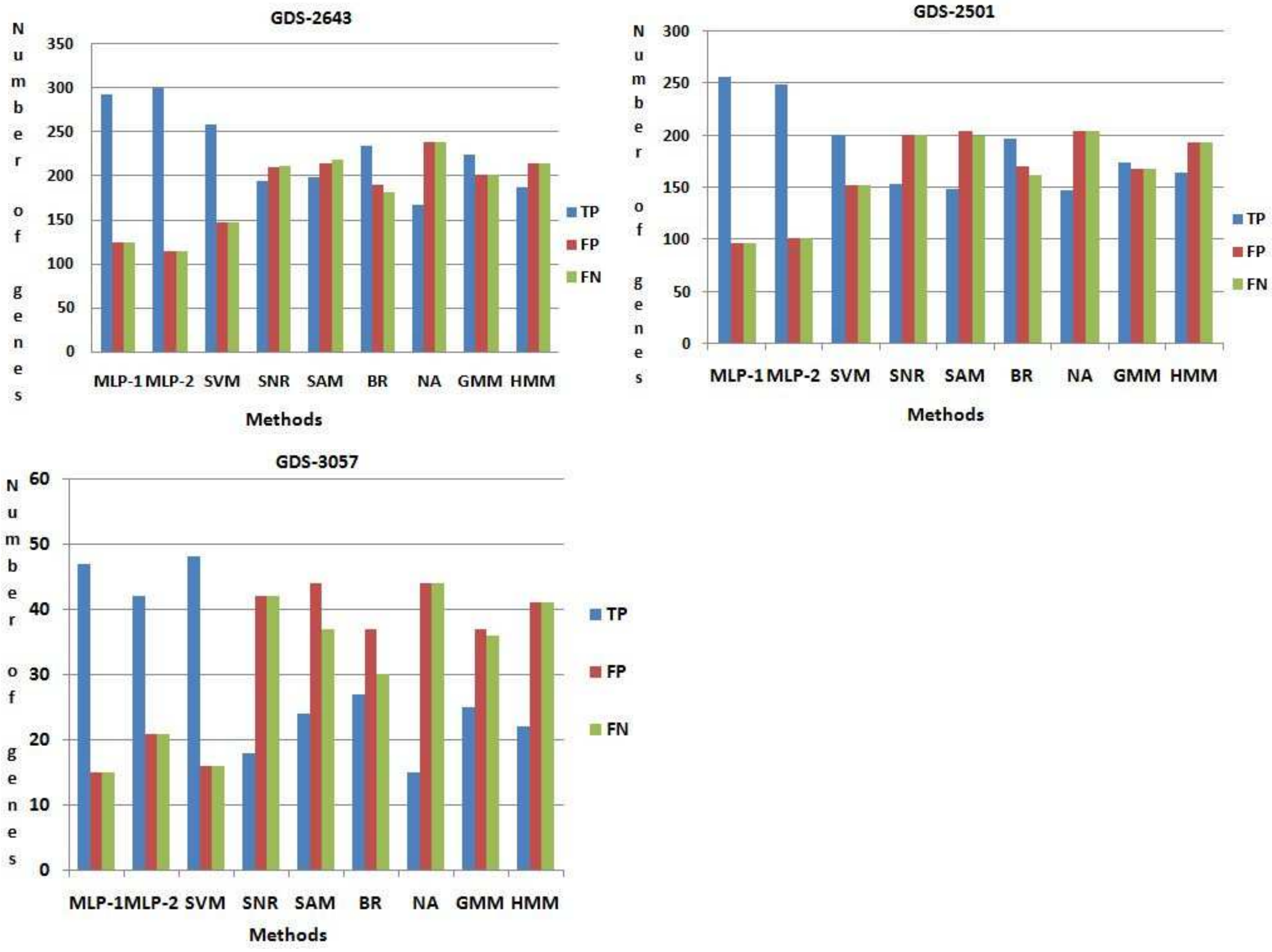


Figure 5: Comparison among the methods. Here TP, FP, FN indicate the number of $truepositive$, $falsepositive$, $falsenegative$ respectively.

pathway we have found 407 number of genes. Now this set of genes, are compared with those obtained by seven existing procedures. Here 293 and 300 number of genes are identified which is common in database information and the results of both MLP-1 and MLP-2, respectively. These genes are called $truepositive(TP)$ genes. In top rank 407 genes, we have found 116 and 110 number of genes are present respectively in both procedure MLP-1 and MLP-2 but not present these pathway. These genes are called $falsepositive(FP)$ genes. Similarly we have found 116 and 110 number of genes are $falsenegative(FN)$ for both MLP-1 and MLP-2 procedure respectively. Now we have calculate the number of $truepositive(TP)$, $falsepositive(FP)$ and $falsenegative(FN)$ genes for other seven existing methods. From Figure (5) show that both NFM-1 and NFM-2 procedure have been able to identify more number of $truepositive$ genes, but less number of $falsepositive$ and $falsenegative$ genes compared to all the other procedures. Similarly in GDS-2501 and GDS-3057, we have been able to identify more number of $truepositive$ genes, but less number of $falsepositive$ and $falsenegative$ genes compared to all the other procedures and result are show in Figure (5).

5.1.2 Biological Validation

The disease mediating gene list corresponding to a specific disease can be founded by a gene database namely NCBI (http://www.ncbi.nlm.nih.gov/Database). We have found several set of genes for GDS-2643, GDS-2501 and GDS-3057 respectively. For GDS-2643, by using both MLP-1 and MLP-2, we have identified 351 numbers of genes. Now we have compared this set of genes with 351 genes from NCBI and both MLP-1 and MLP-2 can be identified 247 and 241 number of genes respectively, which is common both sets. These genes said $truepositive(TP)$. On the other hand, (351-247) = 104 and (351-241) = 110 number of genes are not present the list of genes which is found from NCBI for both procedure MLP-1 and MLP-2 respectively. These numbers of gene are called $falsepositive(FP)$ genes. Similarly (351-247) = 104 and (351-241) = 110 number of genes are present NCBI database but not present in the set of genes which is found by both procedure MLP-1 and MLP-2 respectively. These numbers of gene are called $falsenegative(FN)$ genes. Likewise other procedure, viz., SVM, SNR, SAM, BR, NA, GMM, HMM are compared our result on GDS-2643, GDS-2501 and GDS-3057 respectively. Figure (6) show that both MLP-1 and MLP-2 found the more number of true positive (TP) genes then the other existing procedure for GDS-2643, GDS-2501 and GDS-3057 respectively.

Now we can order to validate our result and calculate the Sensitivity of gene expression data sets. The Sensitivity is computed by using following formula:

$$Sensitivity = \frac{TP}{TP + FN} \tag{17}$$

At first we have computed the total number of true positive genes for every data set for every procedure. As our result, Sensitivity of both MLP-1 and MLP-2 procedure is more than the other existing procedure. Figure (7) show that Sensitivity of both MLP-1 and MLP-2 has performed the best for GDS-2501 data set.

6 Conclusion

In this article, we have proposed a model based on multilayer perceptron, which will select the genes that have been changed quite significantly from normal stage to disease stage. Base on vale of correlation, the different types of genes can be obtained and we have found most important groups. The gene of these groups are evaluated by using both procedure MLP-1 and MLP-2. The most important genes are gained by the procedure have also been corroborated by p-values of genes. The best result of the procedure compression too few standing once has been demonstrated. The output have been corroborated using biochemical pathway, p-value, t-test, sensitivity and some existing result expression profile plots. It has been obtained that the procedure has been capable to the genes are most significant.

Appendix

F-score: F-score are a statistical method for determining accuracy accounting for both precision and recall. The formula for traditional F-score is, F-score= $2*(precision*recall/precision+recall)$. Where precision = $TP/TP+FP$, Recall = $TP/TP+FN$.
True Positive: A true positive test result is one that detect the condition when the condition is present. True positive rate = $TP/(TP+FN)$.
False Positive: A false positive is an error in some rating method in which a condition tested for is badly found to have been detected. A false positive test result is one that detect the condition when the condition is absent. False positive value = $FP/(FP+TN)$.
False Negative: A result that appears negative when it should not. A false negative test result is one that does not detect the condition when the condition is present. False negative value = $FN/(TP+FN)$.
True Negative: A true negative test result is one that does not detect the condition when the condition is absent. True negative value = $TN/(TN+FP)$.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This work has been supported by the University Grant Commission (UGC) RGNF-2014-15-SC-WES-69906, Govt of India and TEQIP-II Project, University of Calcutta.

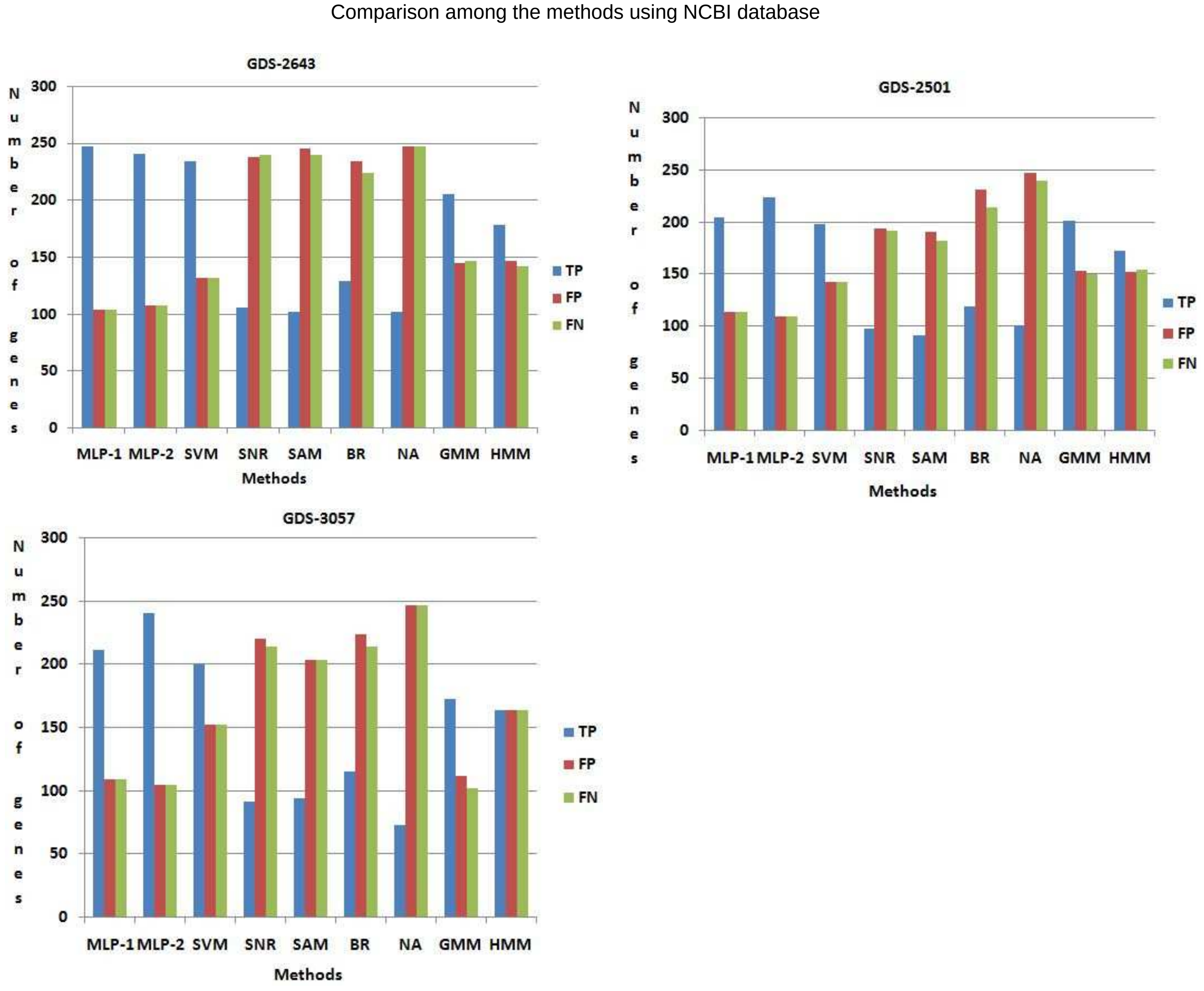


Figure 6: Comparison among the methods using NCBI database. Here TP, FP, FN indicate the number of $truepositive$, $falsepositive$, $falsenegative$ respectively.

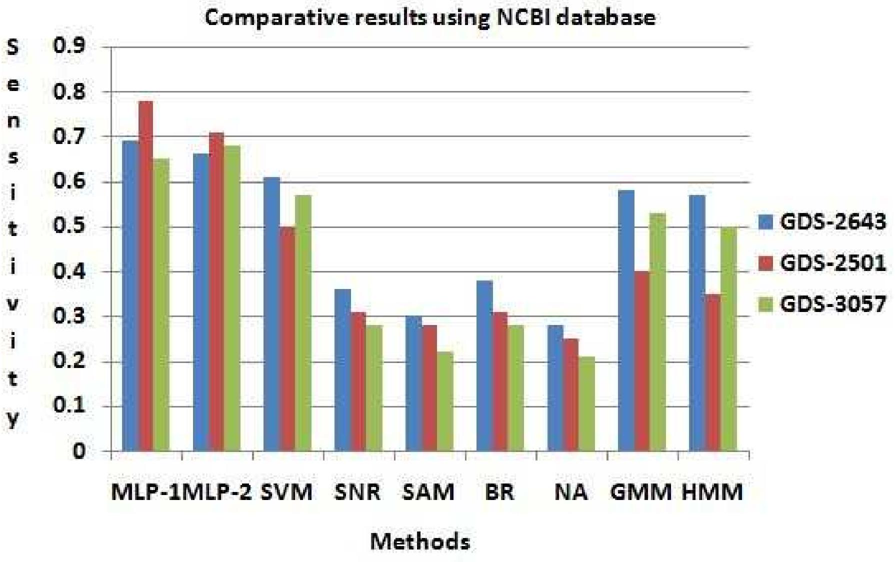


Figure 7: Comparison among the methods using NCBI database.

References

[1] S. K. Veettil, K. G. Lim, N. Chaiyakunapruk, S. M. Ching, M. R. Abu Hassan, "Colorectal cancer in Malaysia: Its burden and implications for a multiethnic country, Asian Journal of Surgery, 2016. 10.1016/j.asjsur.2016.07.005

[2] L. Seewald, J. W. Taub, K. W. Maloney, E. R.B. McCabe, "Acute leukemias in children with Down syndrome, 107(1), 25-30, 2012. 10.1016/j.ymgme.2012.07.011

[3] J. F. Zeidner, J. E. Karp, "Clinical activity of alvocidib (flavopiridol) in acute myeloid leukemia, Leukemia Research, 39(12), 1312-1318, 2015. 10.1016/j.leukres.2015.10.010

[4] R. L. Sielken Jr, C. V.-Flores, "A comprehensive review of occupational and general population cancer risk: 1,3-Butadiene exposureresponse modeling for all leukemia, acute myelogenous leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, myeloid neoplasm and lymphoid neoplasm, Chemico-Biological Interactions, 241, 50-58, 2015. 10.1016/j.cbi.2015.06.009

[5] T. Ripperger, B. Schlegelberger, "Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome, European Journal of Medical Genetics, 59(3), 133-142, 2016. 10.1016/j.ejmg.2015.12.014

[6] G. Mezei, M. Sudan, S. Izraeli, L. Kheifets, "Epidemiology of childhood leukemia in the presence and absence of Down syndrome, Cancer Epidemiology, 38(5), 479-489, 2014. 10.1016/j.canep.2014.07.006

[7] S. Izraeli, "The acute lymphoblastic leukemia of Down Syndrome Genetics and pathogenesis, European Journal of Medical Genetics, 59(3), 158-161, 2016. 10.1016/j.ejmg.2015.11.010

[8] H. Suzuki, A. Shigeta, T. Fukunaga, "Death resulting from a mesenteric hemorrhage due to acute myeloid leukemia: An autopsy case, Legal Medicine, 16(6), 373-375, 2014. 10.1016/j.legalmed.2014.07.003

[9] J. F. Zeidner, J. E. Karp, "Clinical activity of alvocidib (flavopiridol) in acute myeloid leukemia, Leukemia Research, 39(12), 1312-1318, 2015. 10.1016/j.leukres.2015.10.010

[10] 10. P. H. Lin, C. C. Lin, H. I. Yang, L. Y. Li, L. Y. Bai, C. F. Chiu, Y. M. Liao, C. Y. Lin, C.Y. Hsieh, C. Y. Lin, C. M. Ho, S. F. Yang, C. T. Peng, F. J. Tsai, S. P. Yeh, "Prognostic impact of allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia patients with internal tandem duplication of FLT3, Leukemia Research, 37(3), 287-292, 2013. 10.1016/j.leukres.2012.10.005

[11] L. Seewald, J. W. Taub, K. W. Maloney, E. R.B. McCabe, "Acute leukemias in children with Down syndrome, Molecular Genetics and Metabolism, 107(1), 25-30, 2012. 10.1016/j.ymgme.2012.07.011

[12] A. Ghosh, R. K. De, "Fuzzy Correlated Association Mining: Selecting altered associations among the genes, and some possible marker genes mediating certain cancers, Applied Soft Computing, 38, 587-605, 2016. 10.1016/j.asoc.2015.09.057

[13] A. Ghosh, R. K. De, "Development of a fuzzy entropy based method for detecting altered gene-gene interactions in carcinogenic state, Journal of Intelligent & Fuzzy Systems, 26, 2731-2746, 2014. 10.3233/IFS-130942

[14] A. Ghosh, R. K. De, "Linguistic Recognition System for Identification of Some Possible Genes Mediating the Development of Lung Adenocarcinoma, Inf. Fusion, 10, 260-269, 2009. 10.1016/j.inffus.2008.11.007

[15] P. A. Mundra, J. C. Rajapakse, "Gene and sample selection using T-score with sample selection, Journal of Biomedical Informatics, 59, 31-41, 2016. 10.1016/j.jbi.2015.11.003

[16] A. Ghosh, R. K. De, "Identification of certain cancer-mediating genes using Gaussian fuzzy cluster validity index, Journal of Biosciences, 40(4), 741-754, 2015. 10.1007/s12038-015-9557-x

[17] S. Tabakhi, A. Najafi, R. Ranjbar, P. Moradi, "Gene selection for microarray data classification using a novel ant colony optimization, Neurocomputing, 168, 1024-1036, 2015. 10.1016/j.neucom.2015.05.022

[18] S. Saha, D. B. Seal, A. Ghosh, K. N. Dey, "A Novel Gene Ranking Method Using Wilcoxon Rank Sum Test and Genetic Algorithm, Int. J. Bioinformatics Res. App, 12, 263-279, 2016. 10.1504/IJBRA.2016.078236

[19] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system, "Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on, 4, 2461-2466, 2001. 10.1109/IJCNN.2001.938753

[20] H. H. Zhang, J. Ahn, X. Lin, C. Park, "Gene Selection Using Support Vector Machines with Non-convex Penalty, Bioinformatics, 22, 88-95, 2006. 10.1093/bioinformatics/bti736

[21] A. Ghosh, R. K. De, "Interval based fuzzy systems for identification of important genes from microarray gene expression data: Application to carcinogenic development, Journal of Biomedical Informatics, 42, 1022-1028, 2009. 10.1016/j.jbi.2009.06.003

[22] A. Ghosh, B. C. Dhara, R. K. De, "Comparative Analysis of Cluster Validity Indices in Identifying Some Possible Genes Mediating Certain Cancers, Molecular Informatics, 32(4), 347-354, 2013. 10.1002/minf.201200142

[23] V. Elyasigomari, M.S. Mirjafari, H.R.C. Screen, M.H. Shaheed, "Cancer classification using a novel gene selection approach by means of shuffling based on data clustering with optimization, Applied Soft Computing, 35, 43-51, 2015. 10.1016/j.asoc.2015.06.015

[24] R. D. Uriarte, S. A. de Andres, "Gene selection and classification of microarray data using random forest, BMC Bioinformatics, 7(1), 1-13, 2006. 10.1186/1471-2105-7-3

[25] H. Deng, G. Runger, "Gene selection with guided regularized random forest, Pattern Recognition, 46(12), 3483-3489, 2013. 10.1016/j.patcog.2013.05.018

[26] A. Anaissi, P. J. Kennedy, M. Goyal, D. R. Catchpoole, "A balanced iterative random forest for gene selection from microarray data ,BMC Bioinformatics , 14(1), 1-10, 2013.10.1186/1471-2105-14-261

[27] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines, Machine Learning, 46(1), 389-422, 2002. 10.1023/A:1012487302797

[28] L. A. Menndez, F. J. de Cos Juez, F. S. Lasheras, J. A. A Riesgo, "Artificial neural networks applied to cancer detection in a breast screening programme, Mathematical and Computer Modelling, 52, 983-991, 2010. 10.1016/j.mcm.2010.03.019

[29] M. C.Sharma, G. P.Tuszynski, M. R.Blackman, M. Sharma, "Long-term efficacy and downstream mechanism of anti-annexinA2 monoclonal antibody (anti-ANX A2 mAb) in a pre-clinical model of aggressive human breast cancer, Cancer Letters, 373(1), 27-35, 2016. 10.1016/j.canlet.2016.01.013

[30] T. V. da Silva, R. V. A. Monteiro, F. A. M. Moura, M. R. M. C. Albertini, M. A. Tamashiro, G. C. Guimaraes, "Performance Analysis of Neural Network Training Algorithms and Support Vector Machine for Power Generation Forecast of Photovoltaic Panel, IEEE Latin America Transactions, 15(6), 1091-1100, 2017. 10.1109/TLA.2017.7932697

[31] P. Daz-Rodrguez, J. C. Cancilla, G. Matute, D. Chicharro, J. S. Torrecilla, "Inputting molecular weights into a multilayer perceptron to estimate refractive indices of dialkylimidazolium-based ionic liquidsA purity evaluation, Applied Soft Computing, 28, 394-399, 2015. 10.1016/j.asoc.2014.12.004

[32] M. Dashtban, M, Balafar, "Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts, Genomics, 109(2), 91-107, 2017. 10.1016/j.ygeno.2017.01.004

[33] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, N. Cristianini, "Gene selection for cancer classification using support vector machines, Machine Learning, 389-422, 2002.

[34] C. D. A. Vanitha, D. Devaraj, M. Venkatesulu, "Gene Expression Data Classification Using Support Vector Machine and Mutual Information-based Gene Selection, Procedia Computer Science, 47, 13-21, 2015. 10.1016/j.procs.2015.03.178

[35] S. Mishra, D. Mishra, "SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm, Karbala International Journal of Modern Science, 1(2), 86-96, 2015. 10.1016/j.kijoms.2015.10.002

[36] Y. Tang, Y. Q. Zhang, Z. Huang, "Development of Two-Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4(3), 365-381, 2007. 10.1109/TCBB.2007.70224

[37] W. H. Chan, M. S. Mohamad, S. Deris, N. Zaki, S. Kasim, S. Omatu, J. M. Corchado, H. Al Ashwal, "Identification of informative genes and pathways using an improved penalized support vector machine with a weighting scheme, Computers in Biology and Medicine, 77, 102-115, 2016. 10.1016/j.compbiomed.2016.08.004

[38] Y. Tang and Y. Q. Zhang and Z. Huang and X. Hu and Y. Zhao, "Recursive Fuzzy Granulation for Gene Subsets Extraction and Cancer Classification, Trans. Info. Tech. Biomed., 12(6), 723-730, 2008. 10.1109/TITB.2008.920787

[39] D. Du, K. Li, X. Li, M. Fei, "A novel forward gene selection algorithm for microarray data, Neurocomputing, 133(6), 446-458, 2014. 10.1016/j.neucom.2013.12.012

[40] V. de Schaetzen, C. Molter, A. Coletta, D. Steenhoff, S. Meganck, J. Taminau, C. Lazar, R. Duque, H. Bersini, A. Nowe, "A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9(4), 1106-1119, 2012. 10.1109/TCBB.2012.33

[41] S. Chakraborty, "Simultaneous cancer classification and gene selection with Bayesian nearest neighbor method: An integrated approach, Computational Statistics & Data Analysis, 53(4), 1462-1474, 2009. 10.1016/j.csda.2008.10.012

[42] Y. Tang, Y. Q. Zhang, Z. Huang, "Development of Two-Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis, IEEE/ACM Trans. Comput. Biol. Bioinformatics, 4(3), 365-381, 2007. 10.1109/TCBB.2007.70224

[43] Y. Tang, Y. Q. Zhang, Z. Huang, X. Hu, Y. Zhao, "Recursive Fuzzy Granulation for Gene Subsets Extraction and Cancer Classification, IEEE Transactions on Information Technology in Biomedicine, 12(6), 723-730, 2008. 10.1109/TITB.2008.920787

[44] P. A. Mundra, J. C. Rajapakse, "SVM-RFE With MRMR Filter for Gene Selection, IEEE Transactions on NanoBioscience, 9(1), 31-37, 2010. 10.1109/TNB.2009.2035284

[45] F. Ojeda, J. A.K. Suykens, B. De. Moor, "Low rank updated LS-SVM classifiers for fast variable selection, Neural Networks, 21(2), 437-449, 2008. 10.1016/j.neunet.2007.12.053

[46] A. Ghosh, B. C. Dhara, R. K. De, "Selection of Genes Mediating Certain Cancers, Using a Neuro-fuzzy Approach, Neurocomput., 133, 122-140, 2014.

[47] S. Sheet, A. Ghosh, S. B. Mandal, "Selection of Genes Mediating Human Leukemia, Using Boltzmann Machine, Advanced Computing and Communication Technologies: Proceedings of the 10th ICACCT, 83-90, 2018. 10.1007/978-981-10-4603-2-9

[48] S. Sheet, A. Ghosh, S. B. Mandal, "Identification of influential biomarkers for human leukemia - An Artificial Neural Network approach, International Journal of Soft Computing & Artificial Intelligence, 4, 27-32, 2016.

[49] A. Ghosh, R. K. De, "Neuro-fuzzy Methodology for Selecting Genes Mediating Lung Cancer, Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence, 388-393, 2011.

[50] S. Sheet, A. Ghosh, S. B. Mandal, "Selection of genes mediating human leukemia, using an Artificial Neural Network approach, Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2010-2014, 2017. 10.1109/AEEICB.2017.7972415

[51] A. Narayanan, E.C. Keedwell, J. Gamalielsson, S. Tatineni, "Single-layer artificial neural networks for gene expression analysis, Neurocomputing, 61, 217-240, 2004. 10.1016/j.neucom.2003.10.017

[52] H. Q. Wang, H. S. Wong, H. Zhu, T. T.C. Yip, "A neural network-based biomarker association information extraction approach for cancer classification, Journal of Biomedical Informatics, 42(4), 654-666, 2009. 10.1016/j.jbi.2008.12.010

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 21-27 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Community Detection in Social Network with Outlier Recognition

Htwe Nu Win*[,1], Khin Thidar Lynn[2]

[1]*Web Mining, University of Computer Studies, Mandalay, ZIP Code 05071, Myanmar*

[2]*Faculty of Information Science, University of Computer Studies, Mandalay, ZIP Code 05071, Myanmar*

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Social Network
Community
Outlier
Edge Structure
Overlapped Neighbor Node

ABSTRACT

Exploring communities and outliers in Social Network is based on considering of some nodes have overlapped neighbor node within the same group as well as some nodes have no any link to the other node or have no any overlapped value. The existing approaches are based on the overlapping community detection method were only defined the overlap nodes or group of overlap nodes without thinking of which nodes might have individual communities or which nodes are outliers. Detecting communities can be used the similarity measure based on neighborhood overlapping of nodes and identified nodes so called outliers which cannot be grouped into any of the communities. This paper proposed method to detect communities and outliers from Edge Structure with neighborhood overlap by using nodes similarity. The result implies the best quality with modularity measurement which leads to more accurate communities as well as improved their density after removing outliers in the network structure.

1. Introduction

This paper is an extension of work originally presented in 18thIEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and parallel/Distributed Computing (SNPD 2017) [1] is used undirected and unweighted graph data based on overlapping neighborhood. It is explicitly focused on edge structure to detect outliers and significant communities with nodes similarity. Unfortunately, the number of outliers that implemented is higher than that we expected. Therefore, we continue improve our proposed approach to get the significant result by detecting outliers and community.

Social networks can be considered in graph theory point of view. Thinking of detection community in Social networks plays the important role in recent year. It is defined the presence of groups of nodes that are high tightly links connected with each other than with less links connected to nodes of different groups. The challenges in considering of community detection method are become popular in Social Network. The previous authors had finished thinking about them in various points of corners. In overlapping community detection method were only defined the overlap nodes or group of overlap nodes without thinking of which nodes might be included in its own individual communities . Then, the consideration of community detection does not force each node into a certain group, some independent nodes, which cannot be grouped into any communities, are allowed far outside the detected groups can be measured by the predefined threshold of minimal valid size (mvs) of communities as outliers. However, there are still challenges in considering of some nodes have no any common node within the same group as well as some nodes have no any link to the other node.

It can be used similarity measure based on neighborhood overlapping of nodes to organize communities and to identify outliers which cannot be grouped into any of the communities.

In this paper, we detect communities and outliers from Edge Structure with neighborhood overlap by using nodes similarity. This paper explores the use of neighborhood overlapping by using vertex similarity method for detecting outlier and significant

*Corresponding Author:Htwe Nu Win, Web Mining, University of Computer Studies, Mandalay, ZIP Code 05071, Myanmar, htwenuwin99@gmail.com

https://dx.doi.org/10.25046/aj030203

community. The heart of this approach is to represent the underlying dataset as an undirected graph, where a user refers to each node and friendship between two users represents each edge. Before we measure the similarity among neighborhood overlap, finding seed node by using the degree centrality is necessary which is designed to find nodes that are most “central” to the graph. We operate similarity from the most centrality node and its neighborhood nodes. The values of zero similarity are then used to identify as outliers. To illustrate, consider tiny graph which contain8 nodes and 11 edges as shown in Figure 1. Upon applying communities, rounded with two circles are group, and node h so called outlier is saturated with outside them. It can be seen clearly the significant communities and outliers in this toy example.

This paper proposed the method to detect communities, nodes which are high tightly linked each other as community and outliers which do not have links overlap values with another.

The rest of paper is organized as follows. We briefly surveyed related work in section 2. In section 3, we describe the background methodology of our work. And then, we briefly describe about our propose system in section 4. In section 5, we discuss about the experiment and evaluation of our work. Then, we conclude our work in section 6 and talk about our future idea.

2. Related Work

As described in section 1, this paper is an extension of work originally presented in 18thIEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and parallel/Distributed Computing (SNPD 2017) [1] is used undirected and unweighted graph data based on overlapping neighborhood. It was explicitly focused on edge structure to detect outliers and significant communities with nodes similarity. Unfortunately, the number of outliers that implemented is higher than that we expected. Therefore, we continue improve our proposed approach to get the significant result by detecting outliers and community.

Many approaches of community and outlier detection algorithm have been proved over years. Each trend has efficient and effective in their ways. The proposal of a community discovery algorithm for large networks that iteratively finds communities based on only on local information only without considering the global information which was described in [2], this work is cited it in section. The paper based on considering of defining and evaluation networks community based on ground-truth [3] was discussed how to evaluate and measure the proposed system by comparing the previous and extracted community. In [4], the authors used proximity and random walks, to assess the normality of nodes in bipartite graphs. On the other hand, in [5], the authors proposed the one could use to compute various measures associated with the nodes in the given graph structure, dyads, triads, egonets, communities, as well as the global graph structure. The paper which was proposed an algorithm to detect anomalous nodes in weighted graphs is called OddBall was discussed in [6] that easily reveal nodes with strange behavior. However, it was only considered the side of outlier without community. This work is used some methods in the synthesis Lectures on Data Mining and Knowledge Discovery which is describe in [7] was the series publications on topics pertaining to data mining, web mining, text mining, and knowledge discovery, including tutorials and case studies.

In [8] , the authors proposed the efficient algorithm combined the hierarchical and spectral clustering in non-overlapping community. Then, [9] was represented the new community detection method on network model. In [10], the authors described modularity and community structure in networks based on the problem of detecting and characterizing this community structure was one of the outstanding issues in the study of networked systems. The proposed method for detecting such communities, built around the idea of using centrality indices to find community boundaries was described in [11] . [12] was presented several key features of Gephi in the context of interactive exploration and interpretation of networks. In [13], the authors proposed the method of detecting outliers with community by using minimum valid size (mvs). If the minimum size was 2, it would be chosen the single node and marked it as outlier. In [14], the authors surveyed on Social Community Detection which was also specially focus on to community evaluation since this step becomes important in social data mining. The results of proposed method of applying the metric, modularity, and several popular community quality metrics to two real dynamic networks was described in [15]. Continually, in [16], the authors discussed the attempting of a thorough exposition of the topic, from the definition of the main elements of the problem, to the presentation of most methods developed, with a special focus on techniques designed by statistical physicists, from the discussion of crucial issues like the significance of clustering and how methods should be tested and compared against each other, to the description of applications to real networks. Then, the paper in [17] was presented the graph-based approaches to uncovering anomalies in domains where the anomalies consist of unexpected entity/relationship deviations that resemble non-anomalous behavior. Using synthetic and real-world data, it was evaluated the effectiveness of these algorithms at discovering anomalies in a graph-based representation of data. The original citation of dataset, Zachary Karate Club was described in [18]. The paper was presented by the algorithms OCNS for detecting community overlapping base on node similarity was discussed in [19]. The following sections were described the method of the above references in detail which will be the background methodology for this paper.

3. Basic Definitions and Notations

3.1. Outliers

Nodes which have no overlapped values to its adjacent nodes as well as each node in a graph cannot be grouped into any of the communities is defined outliers. This system is based on edge structure approach in the graph to remove outliers before detecting community.

3.2. Node Degree and Its Neighborhood

In network G, the degree of any node i is the number of nodes adjacent to i. Generally, the more degree that the node has, the more important it will be. Two vertices v and u are called neighbors, if they are connected by an edge. Let N_i be the neighborhood of vertex i in a graph, i.e., the set of vertices that are directly connected to i via an edge.

3.3. Community

Communities in Social network can be defined as group of nodes which have more links connecting nodes of the same group and comparatively less links connecting nodes of different groups. Communities may be groups of related individuals in social networks. Identifying communities in a network can be provided valuable information about the structural properties of network, the interactions among nodes in the communities, and the role of the nodes in each community. In an undirected graph (V, E), where the total number of node, $|V|=n$ and total number of edges,$|E|=m$ are defined. We can identify set of communities such that $Coms=\{V_1',V_2',\ldots\ldots,V_{cn}'\}$ where $\bigcup_{\mathrm{i=1}}^{\mathrm{cn}} \mathrm{V_i}' \subseteq \mathrm{V_i}$and cn is the total number of communities Coms should satisfy, $V_i' \cap V_j'=\phi$.

Thus, the goal of community detection is to identify sets of nodes with a common (often external/latent) function based only the connectivity structure of the network. Then it can be considered an axiomatic approach and define four intuitive properties that communities would ideally have. Intuitively, a "good" community is cohesive, compact, and internally well connected while being also well separated from the rest of the network. This allows us to characterize which connectivity patterns a given structural definition detects and which ones it misses.

3.4. Degree Centrality

The importance of a node is determined by the number of nodes adjacent to it. The larger the degree of node, the more important the node is. Those high-degree nodes naturally have more impact are considered to be more important. The degree centrality is defined as

$$(v_i) = d_i = \Sigma A_{ij} \tag{1}$$

When one needs to compare two nodes in different networks, a normalized degree centrality should be used,

$$CD'(v_i) = d_i/(n-1) \tag{2}$$

Here, n is the number of nodes in a network. It is the proportion of nodes that are adjacent to node v_i.

Let N_i denote the neighbors of node v_i. Given a link $e(v_i, v_j)$the neighborhood overlap is defined as;

$$Overlap(v_i, v_j) = \frac{number\ of\ shared\ friends\ of\ both\ v_i\ and\ v_j}{number\ of\ friends\ who\ are\ adjacent\ to\ at\ least\ v_i\ and\ v_j} = \frac{|N_i \cap N_j|}{|N_i \cup N_j|} - 2; \tag{3}$$

We have -2 in the denominator just to exclude $v_i\ and\ v_j$ from the set $N_i \cup N_j$. If there are no overlap vertices in any two $N_i\ and\ N_j$ means $|N_i \cap N_j| = \emptyset$, we can identified N_j as outliers of N_i. Assuming like that, our work identify outliers are appeared with among separated communities.

3.5. Vertex Similarity

It can be assumed that communities are groups of vertices similar to each other. We can compute the similarity between each pair of vertices after searching seed nodes. Most existing similarity method are based on the measurement of distance called Euclidean, Manhattan and etc., Although, to considered the similarity between selected node and is neighborhood, Jaccard Similarity is more convenient in this work which we will measure the similarity based on the neighborhood overlap of seed nodes.

3.6. Graph

The definition of Social network can be imaged as graphs, users or things might be nodes and their relationship might be edges. This system will be represented their social network as an undirected and unweighted graph which mean no distinction between the two vertices associated with each edge. The notation of a graph $G = (V,E)$ consists of two sets V and E. The elements of $V = \{v_1, v_2, \ldots, v_N\}$ are the nodes or vertices of the graph G where each vertex vi is associated with the instance x_i from the input data X and the cardinality of $|V|$ is N. The elements of $E = \{e_1, e_2, \ldots, e_M\}$ are links or edges between nodes and the cardinality of $|E|$ is M. An edge connecting the vertices $v_i$and v_j is denoted by e_{ij}.

4. Community Detection Approach

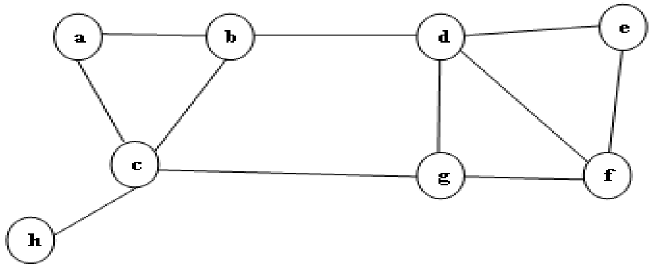
There are two main components will be used to detect communities. They are (i) Finding seed nodes before detecting community and (ii) Detecting community using similarity measure.

4.1. Finding Seed nodes

This part presents in detail finding for detecting communities and making clear the processes of selecting the initial seed node, associating nodes which incident upon a seed node, and electing new seed nodes. The basic idea is inspired by the well-known degree centrality method. It can be used the centrality of nodes to measure which node is seed within that network. It is the most suitable centrality measure for the default measure which yields the most accurate results and also is easy to compute from which we experimented with our results. It is a simple centrality measure that counts how many neighbors a node has. The importance of a node is determined by the number of nodes adjacent to it. The larger the degree of node, the more important the node is. Those high-degree nodes naturally have more impact are considered to be more important. Degree centrality is defined in section 3.5. For example: by viewing Figure 1, the degree centralities become node c and node d by using equation (2).

4.2. Detecting community using similarity measure

In this component, we used the existing methods which is convenient in our work namely Jaccard Similarity which we will measure the similarity based on the neighborhood overlap of seed nodes. This system will be assumed that communities are groups of vertices which of their neighbor nodes are overlapped to each other. We can compute the similarity between each pair of vertices after searching seed nodes to detect community. This method is shown in section 3.6. The example of vertex similarity is shown in the following.



Node ={a,b,c,d,e,f,g,h};

Edge ={(a,b),(a,c),(b,c),(b,d),(c,h),(c,g),(d,e),(d,f),(d,g),(e,f),(f,g)};

Figure 1: Tiny Graph

Table1: Finding Node Centrality

Node	Degree	Normalize Degree
a	2	2/7=0.28
b	3	3/7=0.42
c	4	4/7=0.57
d	4	4/7=0.57
e	2	2/7=0.28
f	3	3/7=0.28
g	3	3/7=0.28
h	1	1/7=0.14

The most centralities are node d and node , firstly we choose node c.

$c = \{a, b, h\};$

$$c\ and\ a = \frac{(c \cap a)}{(c \cup a) - 2} = \frac{1}{4 - 2} = 0.5;$$

$$c\ and\ b\ = \frac{(c \cap b)}{(c \cup b) - 2} = \frac{1}{4 - 2} = 0.5;$$

$$c\ and\ h\ = \frac{(c \cap h)}{(c \cup h) - 2} = \frac{0}{0 - 2} = 0;$$

We got, there is no overlapped value between c and h. Therefore, we identified node h as outlier, then $node\ a$ and $node\ b$ are the members of community corresponding by their node centrality ($node\ c$). The community is shown in Figure 2. Then, choose the most centrality value from the rest of the graph. Now, $node\ d$ is the most centrality.

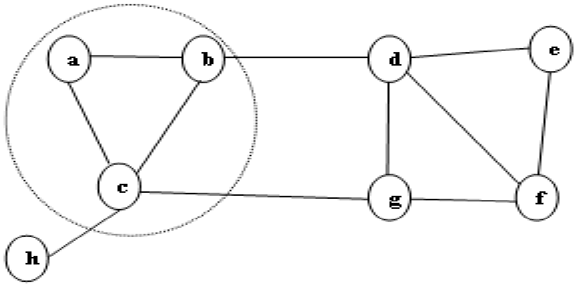


Figure 2: Community of node c Centrality

$d = b, g, e, f;\ b = a, c, d;\ e = d, f;\ f = d, e, g;\ g =\ c, d, f;$

$$d\ and\ e = \frac{(d \cap e)}{(d \cup e) - 2} = \frac{1}{5 - 2} = 0.33;$$

$$d\ and\ f = \frac{(d \cap f)}{(d \cup f) - 2} = \frac{2}{5 - 2} = 0.667;$$

$$d\ and\ g = \frac{(d \cap g)}{(d \cup g) - 2} = \frac{1}{6 - 2} = 0.25;$$

We got, $node\ e$, $node\ f$ and $node\ g$ are the members of community corresponding by their node centrality ($node\ d$). Figure 3 and Figure 4 are the extracted communities after identifying outlier.

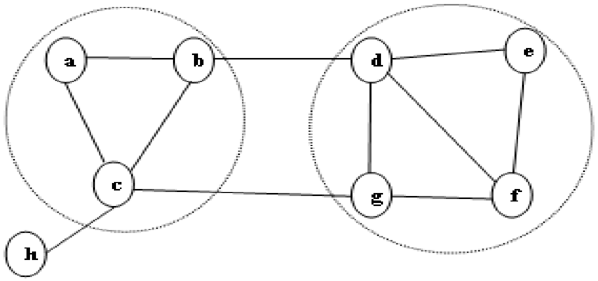


Figure 3: Community of node d Centrality

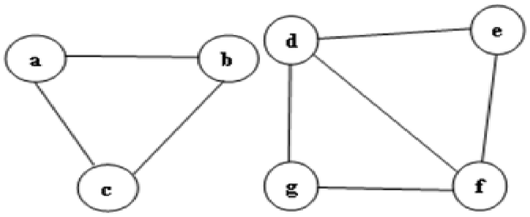


Figure 4: Two Communities after detecting and removing outliers

4.3. Descriptions of Algorithm

For undirected and un-weighted networks dataset, outliers is determined by a node which have no common values then identifying the communities based on the similarity measure in this approach. The following procedures are the steps of detecting outliers and communities:

Step1: Determine seed node by using vertex centrality method.
Step2: Compute the neighborhood overlap of seed nodes by using Jaccard Similarity measure.
Step3: If a node is adjacent to seed node and have overlapped value, determine it as a member of that seed node.
Step4: One node, the member of that seed node within the same community, has another node which is adjacent to its member node and there is overlapped value between them, then identify the linked node as the member of that related community.
Step5: One node which is adjacent to seed node but there is no overlapped values to that seed node is defined as outlier.
Step6: If there is no more adjacent node for that seed node; find another seed node which has the highest centrality value in the left stack of the degree centrality measurement and then repeat the process until there is no more node to be considered.

5. Experiments

5.1. *Description of Dataset*

In this paper, real undirected networks, Zachary Karate Club Dataset is used. In this Dataset statistics, "nodes" represents the number of friends; "edges" represents the number of friendship in the network. Zachary's karate club network is one of the popular studies in social network analysis and has been used as one of the typical test examples by many researchers to detect community structures in complex network. There are 34 member nodes, 78 edges and splits into two clubs, one is indicated as circles and the other is indicated as squares which are shown in Figure 5.

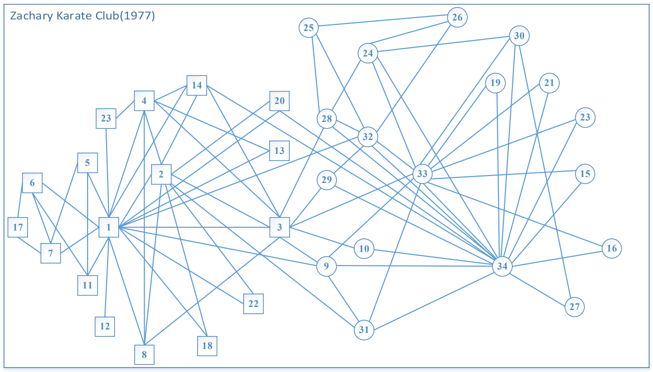


Figure 5: Zarchary Karate Club, Squares and circles indicate two communities observed by Zarchary (1977)

5.2. *Evaluation Method*

In this section, we show the results of communities and outliers. Before showing our experimental result, we first introduce the evaluation methods that are used in our proposed approach.

When there's no ground-truth for the communities to assess the quality of detected communities, it could be used quality measurement so called quality scoring functions which characterize how community like is the connectivity structure of a given set of nodes. The idea is that given a community scoring function, one can then find sets of nodes with high score and consider these sets as communities. All scoring functions build on the intuition that communities are sets of nodes with many connections between the members and few connections from the members to the rest of the network. And it can be formalized the intuition that "good" communities are both compact and well-connected internally while being relatively well-separated from the rest of the network. In quality measurement, it can be showed the quality of detecting communities in different point of view. There are many possible ways to mathematically formalize this intuition.

Among them it will be used the most useful five kinds of quality functions are as follows:

- Scoring functions based on internal connectivity
- Scoring functions based on external connectivity
- Scoring functions based on community goodness metrics
- Scoring functions based on a network model

5.2.1. Scoring functions based on internal connectivity

Internal density is defined as ratio between the intra-community edges and all the edges in the graph and is one of the simplest measures for community quality which is biased towards coarse-grained communities.
Internal density: $f(S) = \frac{m_S}{n_S(n_S-1)/2}$ is the internal edge density of the node set S where S be the set of nodes, m_S is the number of edges between the members of S and n_s is the number of nodes in S. The greater the value is the better for internal density.

5.2.2. Scoring functions based on external connectivity

The fraction of existing edges (out of all possible edges) leaving the cluster $f(S) = \frac{c_S}{n_S(n-n_S)}$ where n be the number of nodes in the network and c_S be the number of edges on the boundary of S. The smaller value is the better for this external connectivity compare with the internal density.

5.2.3. Scoring functions based on community goodness metrics

A clustering coefficient is a measure of the relationship to which people in a network tend to group together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes. It is a real number between zero and one that is zero when there is no community, and one for maximal community, which happens when the network consists of disjoint cliques.

Clustering coefficient (CC) $CC = \frac{2N_V}{K_V(K_V-1)}$ is based on the premise that network communities are manifestations of locally inhomogeneous distributions of edges, because pairs of nodes with common neighbors are more likely to be connected with each other.

5.2.4. Scoring function based on a network model

Modularity is defined as having more internal edges and less external edges as is defined as modularity measures the strength of each partition by considering the degree distribution. One main problem with modularity approach is that it cannot detect well defined small communities when the graphs are extremely large. let us talk about the brief definition of modularity. The Q value is between 0 and 1 and the real network modularity function value is generally between 0.3 and 0.7. Community structure is more obvious with the greater value of modularity.

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j),$$

5.3. Results

To prove the proposed approach we first test it on the well-known karate friendship studied by Zachary, which has been become a classical studies workbench by many researchers for community detection algorithm testing. There are 34 member nodes, 78 edges and splits into two communities.

As shown in figure, node 34 and node 1 are seed nodes, the most important node among their neighboring is the main part of our proposed approach. We had to start those nodes to detect communities and identify outliers by using similarity measurement based on edge structure. We found that nodes which have no any overlapped value among their neighbors are determined as outliers which are node 10 and node 12. Table 2 is shown about the detected node community memberships better correspond to ground-truth node community memberships.

Table 2: Communities of Ground-Truth and Proposed Approach, their members and number of members in corresponding communities

Karate Dataset	Communities	Member (Node) in Community	No. of Members
Ground-Truth	Community1	34, 9,10, 15, 16, 19, 21, 23, 24, 25, 26,27, 28, 29, 30, 31, 32, 33	18
	Community2	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22	16
Proposed Approach	Community1	34, 9, 15, 16, 19, 21, 23, 24, 25, 26 27, 28, 29, 30, 31, 32, 33	17
	Community2	1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 14, 17, 18, 20, 22	15

Table 3: Community C, Number of Intra-Nodes (Intra-N), Number of Intra-Edges(Intra-E), Internal Connectivity(IC), External Connectivity(EC) and Average Clustering Coefficient (ACC)

C	Intra-N	Intra-E	IC	EC	ACC
1	17	34	0.25	0.12	0.7033
2	15	32	0.3	0.11	

Some methods based on the idea that nodes can be the member of two or more communities. But some condition, in thinking of which user should be situated on individual community. Our proposed approach intends to split the multiple communities clearly and remove the nodes which are not necessary to group into any communities. In OCNS method had been proved that they can detect the overlapping node definitely. However, they could not be considered which overlapped node is the member of which community exactly. In considering of detection community without studying the overlapping node combines spectral methods with clustering techniques, and uses the concept of modularity in order to develop a working algorithm and the quantitative of individual communities are different from the ground truth community. Moreover, even removing the outliers, the modularity value (Q) of the proposed approach is better than the other system as shown in Table4.

Table 4: Comparison with other methods in Modularity Value(Q).

Algorithms	Q	No. of Communities	Outlier (node)
GN	0.4013	5	-
Detecting network communities: a new systematic and efficient algorithm	0.4	5	-
OCNS	0.4304	2	-
Gephi	0.416	4	-
Proposed Approach	0.534	2	10,12

6. Conclusion

This proposed approach was used neighborhood overlapping with vertex similarity to detect community and outlier based on edge structure. It showed the steps of detecting outliers and communities in detail. Then, it was discussed about the evaluation measurement in different point of view because of different community criteria. The experiments on real Zachary Karate club network show that our algorithm outperforms other community based algorithms in terms of modularity value, number of communities and members in communities. On the other hand, our approach gets the high quality measurement in the assumption of no ground-truth community. In similarity measurement, nodes which have the overlapped values are detected as communities correspond with its vertex centrality. Nodes which have no any overlapped values in the communities or which need not be necessary to group into the community is defined by outliers. However, in case of node which has no any overlapped value and is connected with multiple communities are still leaving as our future work.

References

[1] Htwe Nu Win and Khin Thidar Lynn, "Community Detection in Facebook with Outlier Recognition", 18thIEEE/ACIS International Conference on Softerwar Engineering, Artificial Intelligence, Networking and Parallel /Distrubuted Computing (SNPD 2017) , June 26-28, 2017, Kanazawa, Ishikawa, Japan, Pages 155-159.

[2] J.Y. Chen, O. R. Zaiane, R. Goebel, "Detecting Communities in Large Networks by Iterative Local Expansion", International Conference on Computational Aspects of Social Networks, 2009.

[3] Jaewon Yang and Jure Leskovec,"Defining and Evaluating Network Communities based on Ground-truth", the Proceedings of 2012 IEEE International Conference on Data Mining (ICDM), 2012.

[4] Jimeng Sun, HuimingQu, DeepayanChakrabarti, and Christos Faloutsos, "Neighborhood formation and anomaly detection in bipartite graph", ICDM, November 27-30 ,2005.

[5] Keith Henderson, Tina Eliassi-Rad, Christos Faloutsos, Leman Akoglu, Lei Li, Koji Maruhashi, B. AdityaPrakash, and Hanghang Tong, "Metricforensics: A multi-level approach for mining volatile graphs", In Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Washington, DC, 2010, pages 163–172.

[6] Leman Akoglu, Mary McGlohon and Christos Faloutsos , "Anomaly Detection in Large Graphs", November 2009, CMU-CS-09-173.

[7] Lei Tang and Huan Liu, "Community Detection and Mining in Social Media", ISBN:9781608453559, A Publication in the Morgan & Claypool Publishers series, 2010.

[8] Luca Donetti and Miguel A Mu˜noz,"Detecting network communities: a new systematic and efficient algorithm", 2004 IOP Publishing Ltd PII: S1742-5468(04)87880-4 , doi:10.1088/1742-5468/2004/10/P10012, Journal of Statistical Mechanic: Theory and Experiment, an IOP and SISSA Journal, http://stacks.iop.org/JSTAT/2004/P10012.

[9] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", Phys. Rev. E, vol. 69, p. 026113, Feb 2004.

[10] M. E. J. Newman, "Modularity and community structure in networks", Proceedings of the National Academy of Sciences, vol. 103, no. 23, pp.8577–8582, 2006.

[11] M.Girvan, M.E.J.Newman, "Community structure in social and biological networks", Proc.Natl. Acad. Sci. USA 99, 7821-7826, 2002.

[12] Mathieu Bastian, SebastienHeymann and Mathieu Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks", Copyright 2009, Association for the Advancement of Artificial Intelligence (www. aaai.org)

[13] Meng Wang, Chaokun Wang, Jeffrey Xu Yu and Jun Zhang, "Community Detection in Social Networks: An In-depth Benchmarking Study with a ProcedureOriented Framework", 41st International Conference on Very Large Data Bases, August 31st September 4th 2015.

[14] Michel Plantie' and Michel Crampes, "Survey on Social Community Detection", Social Media Retrieval, Springer Publishers, Computer Communications and Networks, 978-1-4471-4554-7, pp.65-85, 2013.

[15] Mingming Chen, Tommy Nguyen and Boleslaw K. Szymanski, "On Measuring the Quality of a Network Community Structure", Proceedings of the IEEE Social Computing Conference, Washington DC, September 8-14, pp. 122-127,2013.

[16] S. Fortunato, "Community Detecction in Graph", Phys. Rep. 486, 75 (2010).

[17] William Eberle and Lawrence B. Holder."Discovering structural anomalies in graph-based data", In ICDM Workshops, 2007, pages 393– 398.

[18] Zachary, W. W.: "An information how model for conflict and fission in small groups", J. Anthropol. Res, 33 452-473. 1977.

[19] Zuo Chen, MengyuanJia , Bing Yang, and Xiaodong Li," Detecting Overlapping Community in Complex Network Based on Node Similarity", Computer Science and Information Systems 12(2):843–855 DOI: 10.2298/CSIS141021029C , 2014.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 28-39 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTES

ASTESJ

ISSN: 2415-6698

Software and Hardware Enhancement of Convolutional Neural Networks on GPGPUs

An-Ting Cheng*[,1], Chun-Yen Chen[1], Bo-Cheng Lai[1], Che-Huai Lin[2]

[1]*Institute of Electronics Engineering, National Chiao Tung University, Hsinchu, 300, Taiwan*

[2]*Synopsys Taiwan Co., Ltd., Hsinchu, 300, Taiwan*

ARTICLE INFO

Article history:
Received: 30 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Convolutional Neural Network
GPGPU
Design and Optimization

ABSTRACT

Convolutional Neural Networks (CNNs) have gained attention in recent years for their ability to perform complex machine learning tasks with high accuracy and resilient to noise of inputs. The time-consuming convolution operations of CNNs pose great challenges to both software as well as hardware designers. To achieve superior performance, a design involves careful concerns between exposing the massive computation parallelism and exploiting data reuse in complex data accesses. Existing designs lack comprehensive analysis on design techniques and decisions. The analytical discussion and quantitative proof behind the design criterion, such as choosing proper dimensions to parallelize, are not well studied. This paper performs a series of qualitative and quantitative studies on both the programming techniques and their implications on the GPU architecture. The observations reveal comprehensive understanding on the correlation between the design techniques and the resulting performance. Based on the analyses, we pinpoint the two major performance bottlenecks of CNN on GPGPU: performing computation and loading data from global memory. Software and hardware enhancements are proposed in this paper to alleviate these issues. Experimental results on a cycle-accurate GPGPU simulator have demonstrated up to 4.4x performance enhancement when compared with the reference design.

1. Introduction

Convolutional Neural Networks (CNNs) have gained attention in recent years for their ability to perform complex machine learning tasks. Followed by winning the 2012 ImageNet competition, CNNs have demonstrated superior results in a wide range of fields including image classification, natural language processing and automotive. In addition to high accuracy in object recognition, systems using CNNs are more robust and resilient to noise in the inputs when compared to conventional algorithmic solutions. However, the enormous amount of computing power required by CNNs poses a great challenge to software as well as architecture engineers. The most time-consuming operation in a CNN is the convolution operation, which takes up over 90% of the total runtime. Therefore, the convolution operation becomes one of the most important concerns when implementing CNNs.

GPGPUs have demonstrated superior performance on CNN by exploiting the inherent computation parallelism. Due to the scaling architectures as well as ease-of-programming environment, GPGPUs are among the most widely adopted platforms for CNN. However, it is not a trivial task to have an efficient CNN design on a GPGPU. To achieve superior performance, a design involves careful concerns between exposing the massive computation parallelism and exploiting data reuse in complex data accesses.

This paper is an extension of work originally presented in ICASI 2017 [1]. In this paper, we perform a series of qualitative and quantitative studies on both the programming techniques and their implications on the GPU architecture. The observations reveal comprehensive understanding on the correlation between the design techniques and the resulting performance. There exist several frameworks and libraries that provide solutions to performing convolution on GPGPUs, such as cuDNN [2], Caffe [3], fbfft [4], and cuda-convnet2 [5]. Among the existing solutions, cuda-convnet2 is one of the widely used open-source implementations that enable superior performance on a variety of CNN schemes [6]. It employs design techniques and optimization strategies mainly for NVidia GPGPU architectures. However, while providing a solid implementation, cuda-convnet2 lacks

*Corresponding Author: An-Ting Cheng, National Chiao Tung University, Email: ericcorter78.ee01@g2.nctu.edu.tw

https://dx.doi.org/10.25046/aj030204

comprehensive analysis on its design techniques and decisions. The analytical discussion and quantitative proof behind the design criterion, such as choosing proper dimensions to parallelize, are not well studied. In addition, current GPGPUs are not designed specifically for convolution. There exist potential architecture enhancements that could significantly enhance the computation efficiency with minor hardware and software cost.

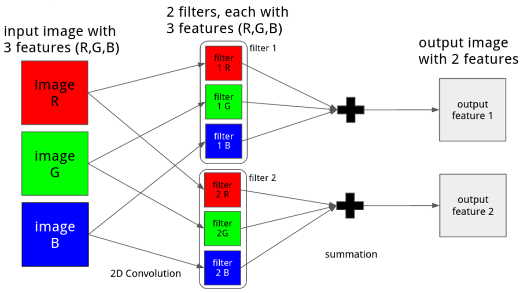


Figure 1: Convolution in CNN

This paper performs a series of qualitative and quantitative studies on both the programming techniques and their implications on the GPU architecture. The studies focus on the widely adopted NVidia GPGPU architecture and CUDA programming environment. The observations reveal comprehensive understanding on the correlation between the design techniques and the resulting performance. Based on the analyses, we pinpoint the two major performance bottlenecks of CNN on GPGPU: performing computation and loading data from global memory. Software and hardware enhancements are proposed in this paper to alleviate these issues. In the computation part, we demonstrate how to avoid excessive local memory accesses, the operations that severely degrade performance, by applying loop unrolling. We then propose two simple yet effective hardware accelerators to speed up the computation of the partial sums. In the data-loading part, we identify that a significant fraction of the time is spent on calculating addresses in the inner loops of CNN. We propose two software techniques to considerably reduce the computation. A low-cost address generator is then introduced to speed up the address calculation. Experimental results on a cycle-accurate GPGPU simulator, GPGPU-sim [7], have demonstrated up to 4.4x performance enhancement when compared with the original cuda-convnet2 design.

The rest of the paper is organized as follows. Section 2 discusses the implementation of the convolution kernel. The techniques of exposing parallelism and data reuse will also be discussed in Section 2. Section 3 and 4 describe the proposed software and hardware improvements to the convolution kernel. Section 5 discusses previous related works and Section 6 presents the conclusions.

2. Implementing Convolution on GPGPU

2.1 Convolution in CNN

Convolution is the basic building block as well as the most time-consuming operation in CNNs. The convolution operation in CNN does more than just convolving two 2D matrices. It takes a set of trainable filters and apply them to the input images, creating one output image for each input image. Each input image consists of one or more multiple feature maps, which means that every pixel in an image contains several features. For example, each pixel in an RGB picture contains three features: red, green and blue. Each filter also has the same number of features that correspond to the input images. When applying the filters to an input image, the filters are convolved across the width and height of the image, and the product of each feature is summed up to produce an output feature map. The number of features in each output image is therefore equal to the number of filters. Figure 1 illustrates an example of the convolution between one image and 2 filters with 3 features, producing one output with 2 features.

The high-level algorithm of the convolution operation in CNN is listed in Figure 2, where conv2 represents computing the 2D convolution between two 2D matrices. The inputs and outputs of the algorithm are all arranged in 4-dimensional arrays. Inputs to the algorithm are the image array images (image_count, height, width, image_features) and the filter array filters (image_features, filter_size, filter_size, filter_count). The output is the array outputs (image_count, height, width, filter_count). Each of the parameters used in the algorithm is described as below.

```
for (i = 0; i < image_count; i++) {
  for (j = 0; j < filter_count; j++) {
    result = zeros(height, width);
    for (k = 0; k < image_features; k++)
      result += conv2(images(k, :, :, i),
                 filters(k, :, :, j));
    outputs(j, :, :, i) = result;
  }
}
```

Figure 2: Pseudocode of convolution operations

image count. This parameter is the size of input mini-batch. A mini-batch contains multiple independent input images to be processed. Each image in the mini-batch will be processed by the same set of filters to produce one output image.

width, height. These two parameters are the width and height of the input images. All input images in the mini-batch have the same size. In this paper, the size of output images is the same as the size of the input images.

image_features. This parameter indicates the number of features maps in an input image. This is also the number of feature maps in a filter.

filter_size. In the context of CNNs, filters are always square-shaped. Therefore, we use only one parameter, which is filter_size, to represent both the width and height of a filter. As a result, the number of pixels in a filter is (filter_size * filter_size).

filter_count. This parameter is the total number of filters. Because each filter produces an output feature map, filter_count is also the number of output feature maps.

2.2 Exploiting Parallelism and Data Reuse in Convolution

This paper uses cuda-convnet2 [5] as the reference implementation of CNN on GPGPUs. Cuda-convnet2 is one of the widely used open-source implementations that enable superior performance on a variety of CNN schemes [6]. It employs design techniques and optimization strategies mainly for NVidia GPGPU architectures. This section will discuss how to implement the convolution algorithm efficiently with CUDA [8].

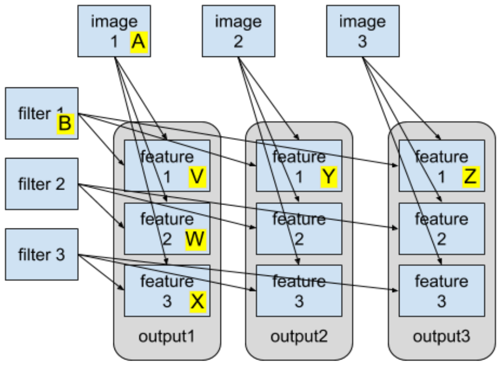


Figure 3: The relation between the input data (images and filters) and the output images.

CUDA is a design environment for massively parallel applications. CUDA applications exploit parallelism provided by the GPGPU by breaking down the task into blocks containing the same number of threads. Functionally speaking, blocks are independent to each other. Although some of them may be assigned to the same computing tile (a.k.a. SM (Streaming Multiprocessor in NVIDIA GPU), the computing model dictates that one block cannot communicate with another. Threads in a block are assigned to the same computing tile so that they can communicate and share data. This computing model leads to two important design decisions: 1) which dimensions to parallelize; and 2) what data to share and reuse between threads within a block.

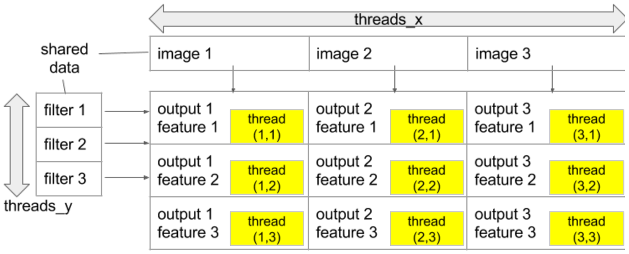


	image 1		image 2		image 3	
filter 1	output 1 feature 1	thread (1,1)	output 2 feature 1	thread (2,1)	output 3 feature 1	thread (3,1)
filter 2	output 1 feature 2	thread (1,2)	output 2 feature 2	thread (2,2)	output 3 feature 2	thread (3,2)
filter 3	output 1 feature 3	thread (1,3)	output 2 feature 3	thread (2,3)	output 3 feature 3	thread (3,3)

Figure 4: Configuration of threads in a block

Parallelizing one dimension means that the task is divided into smaller pieces by splitting at that dimension. For example, if we choose to parallelize the image_count dimension, computation for different images are divided into smaller tasks. In this case, each task is responsible for a small number of images. There are six dimensions in the inputs: image_count, image_features, width, height, filter_size and filter_count. A designer needs to decide which of these dimensions should be parallelized.

One limitation of the CUDA programming model is that different blocks cannot communicate with each other. Therefore, we can only parallelize dimensions that divide the problem into smaller independent tasks. In other words, outputs of each task should be stored separately without being combined into larger results. Based on this criterion, we will separately examine each of the dimensions to determine if it can be parallelized.

The dimension image_count represents the number of images. Because each image produces its own output independent of other images, we can parallelize the image_count dimension. The width and height dimensions can also be parallelized because computation of each output pixel is independent. Next, each filter produces its own output feature map, so the filter_count dimension can be parallelized. Now we are left with two dimensions to examine. The features dimension represents the number of feature maps in input images as well as filters. Because the results from different feature maps are summed to form a single output feature map (see Figure 1), we do not parallelize the feature dimension. Finally, we also do not parallelize filter_size because the products of filter pixels and image pixels are summed to form a single output pixel. To summarize, the dimensions that we are going to parallelize are image_count, width, height and filter_count.

The next step is to decide what data is reused and shared by different threads within a block. Figure 3 depicts the relation between the input data (images and filters) and the output images. Some input/output data set are labeled with capital letters for ease of explanation. The arrows in the figure indicates the input-output relation of the data. For example, the arrow pointing from A (image 1) to V (feature 1 of output 1) means that computation of V depends on A. From the figure we can see that different features in the same output image share the same input image (e.g. V, W, X all depend on A). Also, the same feature in different output images share the same filter (e.g. V, Y, Z all depend on B). These observations reveal some data-reuse opportunities.

To benefit from reusing both images and filters, a block should load pixels of multiple images and filters from the global memory. The threads also need to efficiently reuse the loaded data. This can be done by arranging the threads into a 2D configuration as show in Figure 4. The total number of threads in a block is threads_x * threads_y, and each thread is responsible of computing a single output pixel. All threads have an x and y index, where the x index determines which image the thread uses and the y index determines the filter. By doing this, the kernel only needs to load threads_x image pixels and threads_y filter pixels to compute threads_x * threads_y outputs. Each loaded image pixel is reused by threads_y threads, and each loaded filter pixel is reused by threads_x threads. As a result, this design reduces the memory access required to load the images to 1 / threads_y and that of the filters to 1 / threads_x.

```
__shared__ float
    images_pixel[features][filter_size*filter_size][threads_x];
__shared__ float
    filter_pixel [features][filter_size*filter_size][threads_y];
<collaboratively load images and filters>
float result = 0;
for (int i = 0; i < filter_size*filter_size; i++)
    for (int f = 0; f < features; f++)
        result +=
        image_pixel[f][i][threadIdx.x]*
filter_pixel[f][i][threadIdx.y];
<write result back to global memory>
```

Figure 5: High-level structure of the CUDA kernel

The high-level structure of the CUDA kernel is listed in Figure 5. Shared memory arrays are allocated to store the image and filter pixels to be reused. At the beginning, all threads work together to load the image and filter pixels required by the block. Then, each thread computes the output pixel it is responsible for. Finally, the computed results are written back to global memory.

```
__shared__ float
    images_pixel [features][cached_pixels][threads_x];
__shared__ float
    filter_pixel [features][cached_pixels][threads_y];

<collaboratively load images and filters>

float result = 0;
for (int p = 0; p < filter_size*filter_size; p += cached_pixels){
    for (int i = 0; i < cached_pixels; i++)
        for (int f = 0; f < features; f++)
            result += image_pixel[f][i][threadIdx.x] *
            filter_pixel[f][i][threadIdx.y];
}

<write result back to global memory>
```

Figure 6: Modified kernel that loads data in chunks

This kernel is functional, but there are some details that can be improved. The first one is that this kernel loads the entire filter into shared memory, resulting in higher shared memory usage. This can be improved by loading the filter pixels in fixed-size chunks instead of loading as a whole. Due to the commutativity and associativity of summation, the final result is equal to the sum of the partial result of each chunk. Now we can modify the program by adding a parameter cached_pixels to set the size of the chunk. The modifications are listed in Figure 6. Note that both images and filters are loaded in chunks, or tiles. This technique is normally referred as tiling. As listed in (1), the shared memory usage after tiling (*SharedMem$_{tile}$*) can be greatly reduced from the original cost (*SharedMem$_{original}$*).

$$SharedMem_{tile} = x = \frac{\text{cached\_pixels}}{\text{filter\_size}^2} \times SharedMem_{original} \qquad (1)$$

Currently, each thread in the kernel computes only one output pixel. We can generalize the kernel to compute multiple output pixels per thread. An output pixel is computed from one image and one filter, so we will add two extra kernel parameters images_per_thread and filters_per_thread to make the kernel compute multiple output pixels. These two parameters decide how many images and filters should each thread use. Because each pair of image and filter generate one output pixel, the number of outputs of each thread is images_per_thread * filters_per_thread. The modification to the program is listed in Figure 7. By adjusting images_per_thread and filters_per_thread, we can make each thread compute multiple outputs and therefore reduce the total number of blocks for the same problem size. For example, if the original kernel (equivalent to the modified kernel with images_per_thread = filters_per_thread = 1) has N blocks in total, the new kernel with images_per_thread = filters_per_thread = 2 only has N/4 blocks.

This kernel is used as the baseline program on which we propose enhancements. In the next section, we will describe how we choose the input sizes to use in our experiments.

2.3 Choosing the Input Size for Experiments

GPGPU-sim can obtain detailed execution behavior of CUDA programs on a GPU architecture similar to GTX-480. However, performance simulation in GPGPU-sim is very slow compared to a physical GPU. For example, a program that takes 100ms on an NVidia GTX480 can take up to 3 hours when running on GPGPU-sim. With such long simulation periods, changes to the program or the simulator itself cannot be quickly tested. Therefore, we think it is beneficial to reduce the size of the input dataset for shorter simulation time. But reducing the input data size, if done improperly, might produce inaccurate simulation results that deviates from the behavior of the original data. In this section, we will discuss a way to choose the size of the reduced data.

```
__shared__ float
    images_pixel [images_per_thread][features]
[cached_pixels][threads_x];
__shared__ float
    filter_pixel [filters_per_thread][features] [cached_pixels][threads_y];
<collaboratively load images and filters>
float result[images_per_thread][filters_per_thread] = {};
for (int p = 0; p < filter_size*filter_size; p += cached_pixels) {
    for (int i = 0; i < cached_pixels; i++)
        for (int f = 0; f < features; f++)
            for (int ii=0; ii<images_per_thread; ii++)
                for (int if=0; if<filters_per_thread; if++)
                    result += image_pixel[ii][f][i][threadIdx.x] *
                    filter_pixel[if][f][i][threadIdx.y];
}
<write result back to global memory>
```

Figure 7: Modified kernel with multiple output in each thread

Computation in CUDA is broken down into independent blocks. Each block is assigned to a streaming multiprocessor (SM) so that every thread in the block can be run concurrently via fine-grain context switching of warps (groups of 32 threads). Also, each SM is capable of executing multiple blocks at the same time. On GTX480, the maximum number of blocks that a SM can run concurrently is limited by the following limiting factors:

1. An SM has 32768 registers.
2. An SM has 48kB of shared memory.
3. There should be no more than 8 blocks assigned to the same SM at the same time.

Each thread in a block has its own set of registers, so the number of registers used by each block is block *size*×*register per thread*. The limitation of blocks per SM due to the limiting factor 1 mentioned above is

$$\left\lfloor \frac{32768}{registers\ per\ block} \right\rfloor \qquad (2)$$

Another limitation is the size of shared memory. Because an SM only has 48kB of shared memory, the limiting number of blocks per SM due to limiting factor 2 is

$$\left\lfloor \frac{48\text{kB}}{shared\ memory\ per\ block} \right\rfloor \qquad (3)$$

At last, due to scheduler hardware limitations (limiting factor 3), a SM can only run at most 8 blocks concurrently. Concluding

(2), (3) and the hardware limitation, the actual maximum number of blocks per SM can be obtained by (4).

$$\min\left(\left\lfloor\frac{32768}{registers\ per\ block}\right\rfloor,\left\lfloor\frac{48\text{kB}}{shared\ memory\ per\ block}\right\rfloor\right) \tag{4}$$

In GTX480, there are 15 SMs. If each SM can run eight blocks in parallel, then the GPU can run 120 blocks concurrently. While using very large datasets, the total number of blocks will be much larger than 120 so that most of the time, all SMs on the GPU is doing some work. However, if the reduced dataset has less than 120 blocks, then some of the SMs on the GPU will be idle all the time, making occupancy lower than it should have been in larger datasets. Also, if the number of blocks is slightly more than 120 blocks, the GPU will execute the first 120 blocks in its full capability, and then execute the remaining blocks using only some of the SMs while leaving other SMs idle. This also makes measurements inaccurate in the same way.

Therefore, it is preferable to adjust the reduced data size so that blocks fill in all SMs during the entire simulation period. In other words, the total number of blocks divided by the maximum number of concurrent blocks should be a whole number or slightly less than a whole number (ex. 1.99). By reducing the data size like this, measurements will be closer to that of larger data sets.

Input parameters		Kernel parameters	
image_count	128	images_per_thread	4
image_width	5	filters_per_thread	4
image_height	6	threads_x	16
image_features	3	threads_y	4
filter_count	32	cached_pixels	4
filter_width, filter_height	32		

Table 1: Parameters used in the experiments

The actual input data size and kernel parameters we use in the experiment are listed in Table 1. According to these parameters, we can compute the total number of blocks using (5).

$$number\ of\ blocks = \left\lceil\frac{image\_count}{thread\_x \times\ images\_per\_thread}\right\rceil \times \left\lceil\frac{image\_width\ \times\ image\_height\ \times\ filter\_count}{thread\_y\ \times filters\_per\_thread}\right\rceil \tag{5}$$

After compiling this program, the compiler outputs the following information:

1. Each thread uses 28 registers.
2. Each block uses 3900 bytes of shared memory.

Now we compute the number of blocks per SM. The limitation caused by registers is

$$\left\lfloor\frac{registers\ per\ SM}{registers\ per\ thread \times block\ size}\right\rfloor = \left\lfloor\frac{32768}{28 \times (16 \times 4)}\right\rfloor$$

$$= 18\ blocks\ per\ SM \tag{6}$$

The limitation caused by shared memory is

$$\left\lfloor\frac{\text{shared memory per SM}}{shared\ memory\ per\ block}\right\rfloor = \left\lfloor\frac{48\text{kB}}{3900}\right\rfloor$$

$$= 12\ blocks\ per\ SM \tag{7}$$

Because both (6) and (7) exceed the hardware limitation of eight blocks per SM, maximum numbers of blocks per SM in this case is 8. Multiplying with the total numbers of SMs on GTX480, we get 8×15=120 blocks that can be executed on the GPGPU in parallel. Since there are exactly 120 blocks to be executed, all the blocks can be run in parallel without any SM stalling.

2.4 Summary of Design Concerns

In this section, we discussed an implementation of convolution in CUDA step-by-step. This convolution kernel employs a 2D block configuration to enable sharing of onboard data between threads through the shared memory, reducing accesses to the global memory. The kernels are also parameterized to enable adjusting the amount of work done by each thread. We also explained how we choose a relatively small input size that utilize all SMs. This method of choosing input sizes reduces the error of the experiment results caused by idle SMs. In all the following experiments, we will use the input sizes listed in Table 1.

According to simulation using GPGPU-sim, we identified that the two bottlenecks of the convolution kernel are computation of partial sums and loading of image and filter data. The following two sections, Section 3 and Section 4, will elaborate the proposed software and hardware improvements to speed up these two bottlenecks.

3. Accelerating Computation Part

The result of profiling the baseline implementation is shown in Figure 8. According to the profiling result, the most time-consuming part in the entire kernel is the computation part, which takes up over 97% of the overall runtime. In this section, we will focus on reducing the computation cost by various techniques including software and hardware modifications.

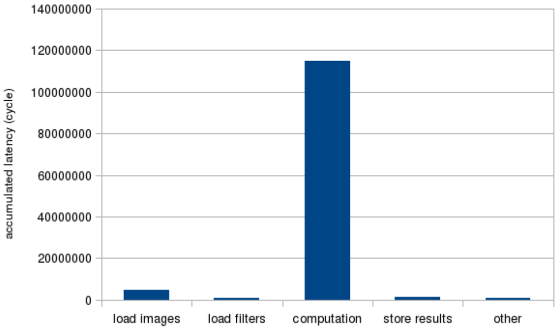


Figure 8: Initial Performance Breakdown

3.1 Avoid Local Memory Access

In order to improve the performance of the computation part, we need to understand what is responsible for its relatively long latency. Instruction-level breakdown of the profiling result listed in Figure 9 reveals that the latencies of instructions accessing local memory are orders of magnitude greater than other instructions. In this section, we will discuss why there is local memory accessing in the code and how can we avoid it.

latency	instruction
…	…
900003	add.u64 %rd55, %rd54, %rd51;
825928	mul.lo.u64 %rd56, %rd55, 4;
909892	add.u64 %rd57, %rd13, %rd56;
92553	ld.shared.f32 %f47, [%rd57+0];
25725679	ld.local.f32 %f48, [%rd53+0];
1059732	mad.f32 %f49, %f43, %f47,%f48;
84348	st.local.f32 [%rd53+0], %f49;
26358148	ld.local.f32 %f50, [%rd53+4];
1059915	mad.f32 %f51, %f44, %f47, %f50;
84135	st.local.f32 [%rd53+4], %f51;
26520275	ld.local.f32 %f52, [%rd53+8];
1061173	mad.f32 %f53, %f45, %f47, %f52;
84381	st.local.f32 [%rd53+8], %f53;
26074621	ld.local.f32 %f54, [%rd53+12];
1062764	mad.f32 %f55, %f46, %f47, %f54;
80641	st.local.f32 [%rd53+12], %f55;
168702	add.u32 %r98, %r98, 1;
650839	add.u64 %rd53, %rd53, 16;
…	…

Figure 9: Local memory access latency

3.1.1 Local Variables in CUDA

Before going into discussion, we will first briefly introduce how local variables (also known as automatic variables) are handled in NVidia GPGPUs. In the CUDA programming language, local variables are normally placed in the stack frame of the current function call. But compilers are also allowed to put them in registers if the architecture permits. Execution stack of CUDA threads are placed in a special memory space called local memory. Local memory resides in the global memory but is partitioned and allocated to each thread. Each thread can only see its own copy of local memory. Because global memory is much slower than registers, it is often preferable to put local variables in registers instead of local memory.

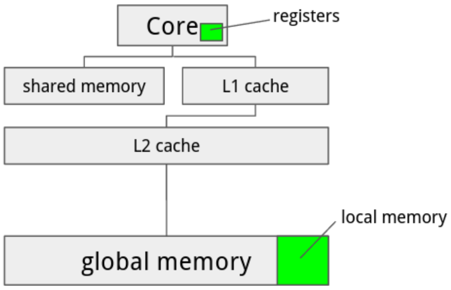


Figure 10: Registers and Local Memory

However, there are two limitations in CUDA regarding the use of registers. One limitation is that Fermi GPGPUs only have 32768 register in each core, so the total number of registers used by all threads in a block cannot exceed 32768. This limitation is less of a problem in convolution because the number of threads per block is relatively small (less than 100), and the number of register required for each thread is around 60. The other limitation is that registers have no addresses. This implies that an array can be put in registers only if no indexing is performed on them.

3.1.2 Loop Unrolling

In the convolution kernel, the partial sums computed by each thread are stored in a local array and accumulated over all pixels as shown in Figure 11. The loop counter f and g are used with the subscript operator to access the array element for each image and filter. In this case, the compiler needs to put the array in local memory because registers cannot be indexed. This will cause the GPU to access global memory in every iteration of the inner loop.

```
float result[filters_per_thread][images_per_thread];
for (int pixel = 0; pixel < filter_pixels; pixel += cached_pixels)
{
 <load filter pixels>
  <load image pixels>
    for (int i =0; i < cached_pixels*image_features; i++)
    {
      for (int f=0; f<filters_per_thread; f++)
     {
        for (int g=0; g<images_per_thread; g++)
       {
          result[f][g] += image_pixel[i][g] * filter_pixel[i][f];
        }
      }
    }
}
```

Figure 11: Local Array in the Convolution Kernel

It can be seen in the figure that the overall latency is dominated by local memory access (highlighted in boldface). The bottleneck can be totally avoided if the array elements are put in registers instead of local memory. However, the compiler cannot put the array in register because the array needs to be dynamically indexed.

One way to get around this limitation is to apply loop unrolling, which effectively eliminates all dynamic indexing on the array by expanding the loop and replacing the indices with constants. As long as the array is not dynamically indexed, the compiler can allocate registers for the array elements.

```
for (int i = 0; i < cached_pixels * image_features; i++) {
 #pragma unroll
   for (int f = 0; f < filters_per_thread; f++)
   {
#pragma unroll
      for (int g = 0; g < images_per_thread; g++)
      {
         result[f][g] += image_pixel[i][g] * filter_pixel[i][f];
      }
   }
}
```

Figure 12: Applying Loop Unrolling

Loop unrolling in CUDA can be enabled by setting up the preprocessing hint during compilation. A directive #pragma unroll is provided to let the programmer issue unrolling hints so that the compiler knows which loops should be unrolled. Using the unroll directive, we can apply loop unrolling to the original program (shown in Figure 12). By using loop unrolling on the two inner-most loops, the `result` array is expanded into multiple independent registers. As a result, the inner loop no longer requires local memory access. As shown in the profiling result in

Figure 13, latency of local memory accesses is completely eliminated after unrolling the loop.

latency	instruction
189822	ld.shared.f32 %f59, [%rd13+0];
1681	ld.shared.f32 %f60, [%rd16+0];
17345	mov.f32 %f61, %f2;
25172	mad.f32 %f62, %f59, %f60, %f61;
8650	mov.f32 %f63, %f62;
1681	ld.shared.f32 %f64, [%rd13+4];
18890	mov.f32 %f65, %f4;
24716	mad.f32 %f66, %f64, %f60, %f65;
6933	mov.f32 %f67, %f66;
1699	ld.shared.f32 %f68, [%rd13+8];
17569	mov.f32 %f69, %f6;
24335	mad.f32 %f70, %f68, %f60, %f69;
5724	mov.f32 %f71, %f70;
1684	ld.shared.f32 %f72, [%rd13+12];
17055	mov.f32 %f73, %f8;
24019	mad.f32 %f74, %f72, %f60, %f73;
192675	mov.f32 %f75, %f74;
1707	ld.shared.f32 %f76, [%rd16+4];
16959	mov.f32 %f77, %f10;
...	…

Figure 13: Profiling result after applying loop unrolling

Before applying loop unrolling, it takes 524k cycles for the program to finish. The letter k is a postfix indicating 1,000. After applying loop unrolling, the number of cycles is reduced to 148k, resulting in a 71% improvement on the overall performance. The latency breakdown in Figure 14 shows that the computation part is dramatically improved by loop unrolling.

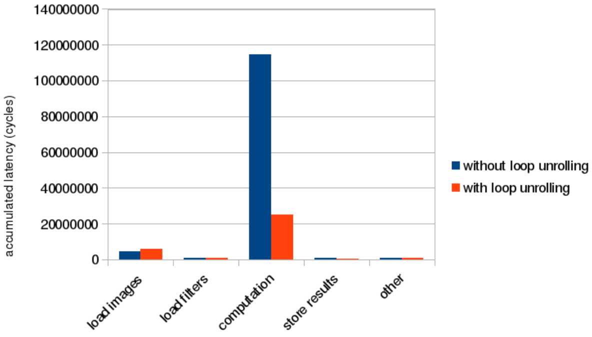


Figure 14: Loop unrolling performance improvement

3.2 Adding Inner Product Engine

The computation part (shown in Figure 15) is still the bottleneck even after applying loop unrolling, taking up about 69% of the overall execution time. Looking at the loop as a whole, what it does is computing the product between each image and filter pixel and sum them together. The partial sums are then accumulated in the array `result`.

```
01: for (int i = 0; i < cached_pixels * image_features; i++) {
02:   #pragma unroll
03:    for (int f = 0; f < filters_per_thread; f++)
04:   {
05:    #pragma unroll
06:       for (int g = 0; g < images_per_thread; g++)
07:       {
08:          result[f][g] += image_pixel[i][g]* filter_pixel[i][f];
09:       }
10:   }
11: }
```

Figure 15: Computation loop

If we exchange the order of the outmost loop with the two inner loops in Figure 16, the inner loops (Line 05 to Line 11) becomes an inner product operation between the two arrays. This modification does not change the behavior of the original function because there is no dependency between each loop iteration. Since the two outer loops are unrolled, the inner product becomes the only bottleneck.

```
01:  #pragma unroll
02:  for (int f=0; f < filters_per_thread; f++) {
03:     #pragma unroll
04:     for (int g=0; g < images_per_thread; g++) {
05:        for (int i=0;  i<cached_pixels* image_features; i++)
06:        {
07:          result[f][g] += image_pixel[i][g] * filter_pixel[i][f];
08:        }
09:    }
10: }
```

Figure 16: Inner product in the computation loop

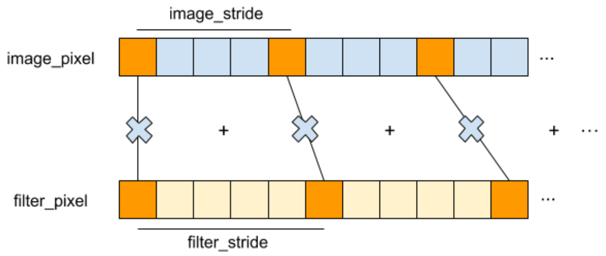


Figure 17: Inner Product

To reduce the bottleneck, we propose adding a hardware inner product accelerator to the cores, one unit for each thread. The accelerator loads pairs of image and filter pixels from shared memory, compute the product of each pair, and then sum all the products together. The program passes the starting addresses and strides of both arrays and the number of elements to the accelerator. These arguments are then stored in the internal registers of the accelerator and are reused until their values are changed again. The final result is also stored in the unit and can be retrieved in the program.

The hardware architecture of the inner product accelerator unit is shown in Figure 18. For each iteration, it loads two elements from shared memory and accumulate the product of them to a register. It requires one multiplier, one adder and a register to store the partial results. In the actual implementation, the multiplier and adder can be fused into a fused multiply-add (FMA) circuit. We assume that it requires 2 cycles to load the two elements from shared memory, and the latency of the FMA is also assumed to be 2 cycles. A 2-stage pipeline can then be used to repeatedly load elements and compute FMA. In our work, we model the latency of the inner product unit as 2 * N, where N is the number of elements to compute.

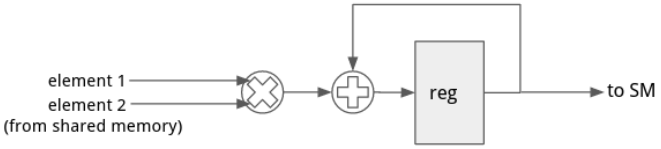


Figure 18: Inner Product Unit

Modeling the accelerator requires modifying both GPGPU-sim and the application. We modeled the inner product accelerator in GPGPU-sim by intercepting memory access to specific addresses. We modified the load (ld) and store (st) handlers to check if the source or target address matches any of the designated addresses. List of the addresses we use and their functions are listed in Table 2. If the address matches, the original memory access is skipped and the corresponding function of the accelerator is performed. For example, as soon as the thread writes 16 to address 0xffffffc8, GPGPU-sim sets the stride register of the accelerator to 16.

Table 2: Addresses used by the inner product engine

address	**direction**	**function**
0xffffffc0	write	address of image array
0xffffffc8	write	stride of image array
0xffffffd0	write	address of filter array
0xffffffd8	write	stride of filter array
0xfffffff0	write	number of elements
0xfffffff8	read	start computing and retrieve the result; stall the thread for N cycles

To use this inner product accelerator, the application needs to use the addresses to manipulate the registers inside the accelerator. First, the application should pass the starting address of the image and filter array to 0xffffffc0 and 0xffffffd0 respectively. The image stride is threads_x * images_per_thread, and the filter stride is threads_y * filters_per_thread. These two values are constant as the dimensions of the arrays are known in compile time, so it is only necessary to write them once at the beginning of the kernel. The fifth parameter, number of elements, is also known in compile time and only needs to be write once. Finally, after all parameters are set, the program needs to read from address 0xfffffff8 to retrieve the result of the inner product. The modified program is listed in Figure 19.

```
// beginning of the kernel
*(unsigned long long*)(0xffffffc8)  =  sizeof(shm_images[0]);
*(unsigned long long*)(0xffffffd8)  =  sizeof(shm_filters[0]);
*(unsigned long long*)(0xfffffff0)  =  cached_pixels * image_features;
// inside the computation part
#pragma unroll
for (int f = 0; f < filters_per_thread; f++) {
   *(float**)(0xffffffd0) = &shm_filters[0][threadIdx.y * filters_per_thread
     + f];
#pragma unroll
   for (int g = 0; g < images_per_thread; g++) {
      float partial_sum;
         *(float**)(0xffffffc0) = &shm_images[0][threadIdx.x *
          images_per_thread + g];
       partial_sum = *(float*)0xfffffff8;
       result[f][g] += partial_sum;
   }
}
```

Figure 19: Program modified to use the inner product engine

The profiling result of the modified program is listed as follows in Figure 20. As shown in the figure, the computation part is improved by 88% and results in 66% improvement on the overall performance. The bottleneck of the program is no longer the computation part.

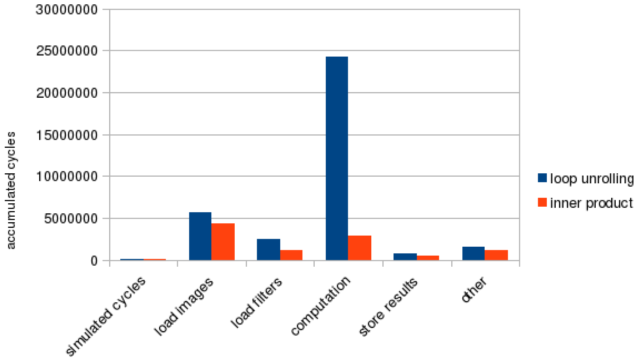


Figure 20: Loop unrolling performance improvement

3.3 Outer Product Engine

In this section, we will discuss an alternative way to accelerate at the computing part. The two inner loops (shown in line 03~09 in Figure 21) can also be viewed as computing the products of each pair of elements in image_pixel[i] and filter_pixel[i]. This operation is called the outer product, which can be represented as multiplying a column matrix with a row matrix as shown in the figure. Inputs to the outer product are two arrays image_pixels[i] and filter_pixel[i], each with length images_per_thread and filters_per_thread. Each element in the first array is multiplied with each element in the second array, producing a total of images_per_thread * filters_per_thread numbers. The outer product is performed cached_pixels * image_features times, and the output of each time is accumulated to produce the partial sums.

```
01: for (int i = 0;  i < cached_pixels * image_features;  i++) {
02:    for (int f = 0; f < filters_per_thread; f++) {
03:       for (int g = 0; g < images_per_thread; g++)
04:       {
05:          result[f][g] += image_pixel[i][g] * filter_pixel[i][f];
06:       }
07:    }
08: }
```

Figure 21: Outer product in the computation loop

We propose an accelerator to speed up the computation of the outer product. The accelerator has at least images_per_thread * filters_per_thread internal registers to store the computed partial sums. It loads two arrays, compute the product of each pair of elements, and accumulate the results to the internal registers. Values of the registers can be read or reset to zero by the program.

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \begin{bmatrix} w & x & y & z \end{bmatrix} = \begin{bmatrix} aw & ax & ay & az \\ bw & bx & by & bz \\ cw & cx & cy & cz \\ dw & dx & dy & dz \end{bmatrix}$$

Figure 22: Outer product

The hardware architecture of the outer product accelerator unit is shown in Figure 23. For each iteration, it loads four elements from image array and one element from filter array multiply them to a register, and the adder adds previous partial sum to the original register. It consumes four multiplier, one adder and number of image array * filter array registers to save the partial results. Suppose the multiplier requires 1 cycle to produce the result, storing result to register and loading pixel from shared memory also consumes 1 cycle. Then, a 3-stage pipeline can be applied to repeatedly load elements, compute outer product and store results. In our implementation, we model the latency of the outer product unit which assume that the accelerator can load one pixel from the shared memory each cycle. It will need to preload some pixels from the image array before using a 3-stage pipeline, and the number of preloaded pixels depends on image array length. Therefore, we can assume that the total latency of the accelerator is image array length + (2*N)-2 ,where N is the number of elements to compute.

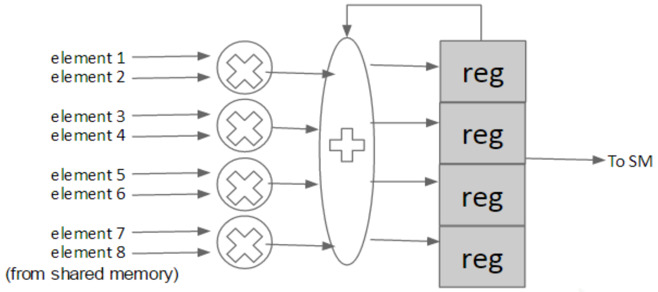


Figure 23: Outer Product Unit

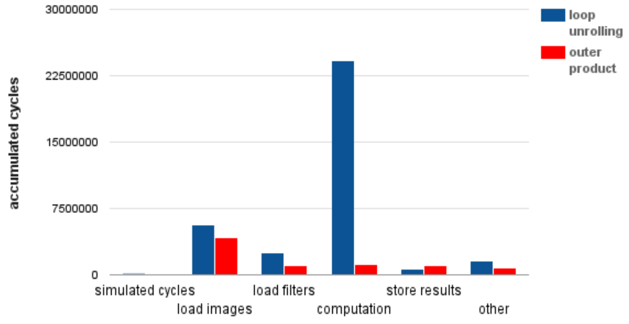


Figure 24: Outer product performance improvement

4. Accelerating Data-Loading

After applying the improvements described in Section IV, the computation part is improved a lot. As illustrated in *Figure 25*, the data-loading part, including loading of images and filters, becomes the bottleneck of the program.

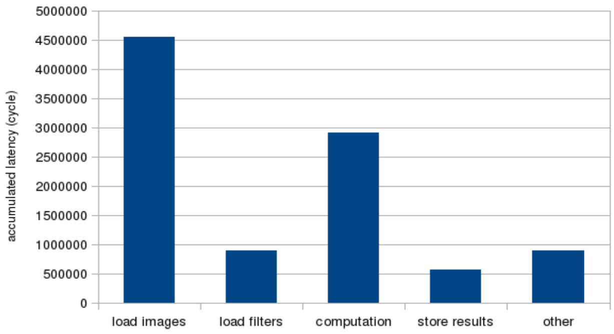


Figure 25: Performance breakdown after improvements

The data-loading part is responsible for loading image and filter pixels from global memory to shared memory. It is composed of deeply nested control structures of loops and conditionals. The control structures themselves also take time to execute, especially for the inner loops. Any subtle overhead inside inner loops can build up and become major bottlenecks. In this section, we propose two software approaches to reduce the overhead inside inner loops. We also propose a hardware accelerator to speed up address calculation in the image-loading part.

4.1 Strength Reduction

One of the major bottlenecks in inner loops is the computation of array indices for each iteration. In the data-loading loops, array indices in inner loops can contain complex arithmetic expressions that translates into larger number of instructions. Because the arithmetic instructions are executed in the inner loops, latencies of them can quickly build up and become a major bottleneck.

Take the program in Figure 26 as an example. To compute the array index for images, it needs to compute two multiplications and two additions for each iteration (line 4). If we work out the total number of operations, we will find that the program needs to carry out image_features * images_per_thread multiplications and additions in total.

```
01: for (f = 0; f < image_features; f++) {
02:    for (i = 0; i < images_per_thread; i++) {
03:       image_pixel[f][i] = images[ base + f * stride + i * threadIdx.x];
04:    }
05: }
```

Figure 26: Arithmetic operation in the inner loop

We propose a method based on strength reduction to improve the performance of this program. This method takes advantage of the fact that some arithmetic operations can be reduced to successive simpler operations. For example, multiplication can be done with repeated addition of the multiplier. Instead of computing the multiplication in each iteration, we can use a separate counter variable idx to accumulate the index throughout the entire loop and update it according to the following rule.

The arithmetic expression for computing the array index can be broken down into 3 parts:

a. loop invariants (terms that does not change throughout the nested loops)
b. multiples of the outer loop counter f
c. multiples of the outer inner loop counter i

Terms belonging to Part a is constant with respect to both the inner and outer loops. Therefore, the term `base` is used to initialize idx before entering the loops. Part b contains the term f * stride, whose value increases by stride whenever f is incremented. Therefore, stride is added to idx at the end of the outer loop. Part c contains the term i * threadIdx.x. The value of this term goes from 0 to (images_per_thread-1) * threadIdx.x in the inner loop and returns to zero again. Therefore, we will first save the value of idx before entering the inner loop, increment idx in each iteration, and restore idx after leaving the inner loop. The resulting program is listed in Figure 27.

```
for (int f = 0; f < image_features; f++) {
  for (int i = 0; i < images_per_thread; i++) {
    if (image_index + i < image_count) {
      /* load image `image_index + i` */
    }
  }
}
```

Figure 27: Applying strength reduction

In this transformed program, it only needs to compute one addition for each iteration in the inner loop and one addition for each iteration in the outer loop. In total, we get image_features * (images_per_thread+1) additions. The total number of operations is cut in half compared to the original program, so the modified program should run faster in the data-loading part. Profiling results in Figure 28 supports this prediction by showing that the performance of the image-loading and filter-loading parts are improved significantly. The impact of strength reduction is 25.9% on the overall performance.

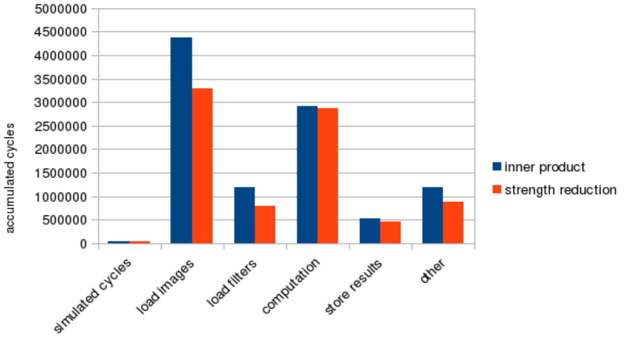


Figure 28: Strength reduction performance improvement

4.2 Extract Conditionals

In addition to strength reduction, we also identified another useful optimization strategy in the data-loading loop. The idea is to prevent unnecessary condition checks in inner loops by checking whether the conditional is necessary in the loop before entering the loop. To illustrate the case, consider the following code snippet taken from the convolution kernel:

```
int idx = base;
for (f = 0; f < image_features; f++) {
  int idx_save = idx;
  for (i = 0; i < images_per_thread; i++) {
    image_pixel[f][i] = images[idx];
    idx += threadIdx.x;
  }
  idx = idx_save;
  idx += stride;
}
```

Figure 29: Boundary checking in the data-loading loop

In the inner-most loop, it checks whether the index of image to load (image_index + i) is within bounds and only load the image if so. Because the total number of image is not always a multiple of images_per_thread, the boundary condition check is necessary here to ensure that the index to load is within bounds. However, checking for the boundary condition every time in an inner loop degrades performance. To eliminate redundant checks in inner loops, the loop is duplicated and modified into two variants: one with boundary checking, and the other without them. Without the checking overhead, the one without boundary checking will run faster. The problem we are left with is how to choose between these two variants.

Because the loop variable i is always less than or equal to images_per_thread - 1, image_index + i will always be less than or equal to image_index + images_per_thread - 1. If given image_index + images_per_thread - 1 < image_count, we will automatically get image_index + i < image_count. In other words, if image_index + images_per_thread <= image_count, there is no need to check for the boundary condition. As a result, we will choose the faster loop without boundary checking if image_index + images_per_thread <= image_count, or the slower loop otherwise. The resulting code is listed in Figure 30.

```
if (image_index + images_per_thread <= image_count) {
  for (int f = 0; f < image_features; f++) {
    for (int i = 0; i < images_per_thread; i++)
    { // no boundary check
      // load image `image_index + i`
    }
  }
} else {
  for (int f = 0; f < image_features; f++) {
    for (int i = 0; i < images_per_thread; i++)
    {
      if (image_index + i < image_count)
      { // boundary check
        // load image `image_index + i`
      }
    }
  }
}
```

Figure 30: Reduced boundary check code

Profiling result before and after applying this technique is shown in Figure 31. This technique improves the overall performance by 2.6% compared to the previous version using only strength reduction.

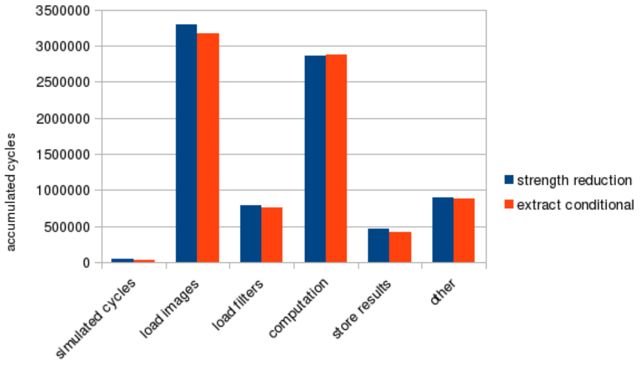


Figure 31: Extract conditionals performance improvement

4.3 Index Conversion Accelerator

In the data-loading part of the convolution kernel, filter and image pixels are loaded to shared memory one after another. Before loading each pixel from global memory, the program must compute the index of the pixel in the input arrays. In each iteration of the loop, the program first loads several filter pixels. While loading the filter, there is no need to compute the index because the loop counter itself represents the index of the filter pixel to load. However, for each filter pixel, it is still necessary to load the corresponding image pixel. Computing the index of image pixels is more involved. Sometimes the filter pixel is placed outside the image and doesn't overlap with any pixel in the input. When this happens, the loaded image pixel should be set to zero because zero-padding is used outside the edges of the image. If the filter pixel is placed within the image, the filter pixel index should be converted into the image pixel index, which is then used to load the image pixel from global memory.

```
for (int p = 0; p < cached_pixels; p += threads_y) {
  int pixel_index = pixel + p + threadIdx.y;
  int x = image_pixel_x – filter_size / 2 + pixel_index % filter_size;
  int y = image_pixel_y - filter_size / 2 + pixel_index / filter_size;
  if (y >= 0 && y < image_height && x >= 0 && x < image_width)
  {
    int image_pixel_index = (y * image_width + x) * image_count;
    <load image pixel from image_pixel_index>
  } else {
    <set image pixel to 0>
  }
}
```

Figure 32: Index conversion in the data-loading loop

Computing the image pixel index and checking for the boundary condition takes up about 1/3 of the image-loading time. Therefore, we propose adding an accelerator to speed up this two tasks at the same time. Before describing what this accelerator should do, we will first look at the code to convert filter pixel index to image pixel index.

From the code listed above, we can see that there are three steps involved in converting filter pixel index to image filter index. The first step is break down filter pixel index to its x and y component and offset the coordinates by the location of the kernel. Then, boundary check is performed on the (x, y) point to ensure that there is a corresponding pixel in the input. Finally, the (x, y) is converted to the image pixel index. These steps are translated to tens of instructions and slows down the program.

```
for (int p = 0; p < cached_pixels; p += threads_y) {
<invoke index converter using inline assembly>
// the result is stored in image_pixel_index
  if (image_pixel_index >= 0) {
  <load image pixel from image_pixel_index>
  } else {
  <set image pixel to 0>
  }
}
```

Figure 33: Modified program using the index conversion accelerator

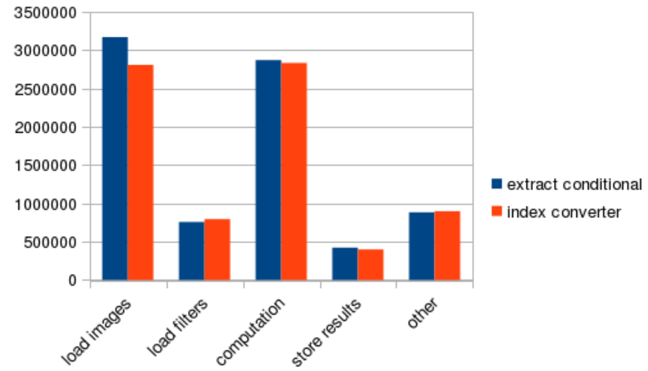


Figure 34: Index converter performance improvement

We propose adding a hardware accelerator to do the index conversion. The accelerator we propose will do the three steps altogether in one instruction. If the (x, y) point is within bounds, it returns the index of the image pixel. Otherwise, it returns -1. We implement the accelerator in GPGPU-sim as an instruction and use inline assembly in the convolution kernel to invoke the instruction. The modified program is listed in Figure 33. The performance improvement of using the index converter is listed in Figure 34. The overall performance is improved by 5%.

5. Related Work

Accelerating convolutional neural networks is a very popular research topic. Accelerators have been developed in different hardware technologies. Eyeriss [9][15] developed by Yu-Hsin Chen et al. is an ASIC CNN accelerator that can run AlexNet at 35fps with only 278mW of power consumption. There are other ASIC accelerators proposed to exploit the redundancy of CNN networks [16][17][18]. Cheng Zhang [10] implemented a CNN accelerator on FPGA and achieved 61.62 GFLOPS under 100MHz clock frequency. He also proposed an analytical design scheme using the roofline model.

The GPGPU is also a widely-used platform for CNN. NVidia developed a software library named cuDNN [2] that uses GPGPU to speed up convolution. When integrated with the Caffe framework, it can improve the performance by up to 36%. However, the cuDNN library is proprietary and cannot be studied by the community. The fbfft library [4] also uses GPGPU to speed up CNN, but it employs a different algorithm (FFT) to compute convolution. Cuda-convnet2 is an efficient implementation of CNN for NVidia GPGPU. It is by far the fastest open-source CNN implementation. However, it lacks analysis on the techniques it uses to improve the performance.

6. Conclusion

This paper describes an implementation of the convolution operation on NVidia GPGPU and analyze the techniques in the implementation. We propose software and hardware enhancements to the program to speed up the computation of partial sums and loading the input data, which are the two major bottlenecks. The experiments have shown that the proposed modifications have achieved 4.4x speedup compared with the baseline implementation.

References

[1] Lin, C.-H.; Cheng, A.-T.; Lai, B.-C.; "A Software Technique to Enhance Register Utilization of Convolutional Neural Networks on GPGPUs," IEEE International Conference on Applied System Innovation, May 2017. http://ieeexplore.ieee.org/document/7988499/

[2] Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., & Shelhamer, E. (2014). cudnn: Efficient primitives for deep learning. arXiv preprint arXiv:1410.0759. https://arxiv.org/abs/1410.0759

[3] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 675-678). ACM.

[4] Vasilache, N., Johnson, J., Mathieu, M., Chintala, S., Piantino, S., & LeCun, Y. (2014). Fast convolutional nets with fbfft: A GPU performance evaluation. arXiv preprint arXiv:1412.7580. https://arxiv.org/abs/1412.7580

[5] Alex Krizhevsky. cuda-convnet2. https://code.google.com/p/cuda-convnet2/, 2014. [Online; accessed 23-January-2015].

[6] Lavin, A. (2015). maxDNN: an efficient convolution kernel for deep learning with maxwell gpus. arXiv preprint arXiv:1501.06633. https://arxiv.org/abs/1501.06633

[7] Bakhoda, A., Yuan, G. L., Fung, W. W., Wong, H., & Aamodt, T. M. (2009, April). Analyzing CUDA workloads using a detailed GPU simulator. In Performance Analysis of Systems and Software, 2009. ISPASS 2009. IEEE International Symposium on (pp. 163-174). IEEE. http://ieeexplore.ieee.org/abstract/document/4919648/

[8] Nickolls, J., Buck, I., Garland, M., & Skadron, K. (2008). Scalable parallel programming with CUDA. Queue, 6(2), 40-53.

[9] Chen, Y. H., Krishna, T., Emer, J., & Sze, V. (2016, January). 14.5 Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. In 2016 IEEE International Solid-State Circuits Conference (ISSCC) (pp. 262-263). IEEE. http://ieeexplore.ieee.org/document/7738524/

[10] Zhang, C., Li, P., Sun, G., Guan, Y., Xiao, B., & Cong, J. (2015, February). Optimizing fpga-based accelerator design for deep convolutional neural networks. In Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (pp. 161-170). ACM. https://dl.acm.org/citation.cfm?id=2689060

[11] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105). https://dl.acm.org/citation.cfm?id=2999257

[12] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://arxiv.org/abs/1409.1556

[13] Chen, T., Du, Z., Sun, N., Wang, J., Wu, C., Chen, Y., & Temam, O. (2014, February). Diannao: A small-footprint high-throughput accelerator for ubiquitous machine-learning. In ACM Sigplan Notices (Vol. 49, No. 4, pp. 269-284). ACM. https://dl.acm.org/citation.cfm?id=2541967

[14] Cavigelli, L., Magno, M., & Benini, L. (2015, June). Accelerating real-time embedded scene labeling with convolutional networks. In Proceedings of the 52nd Annual Design Automation Conference (p. 108). ACM. http://ieeexplore.ieee.org/document/7167293/

[15] Y.-H Chen, T. Krishna, J. S. Emer, and V. Sze, “Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks,” IEEE Journal of Solid-State Circuits, 2017.

[16] D. Kim, J. Ahn, and S. Yoo, “A novel zero weight/activation-aware hardware architecture of convolutional neural network,” in Design, Automation, and Test in Europe (DATE), 2017.

[17] A. Parashar, M. Rhu, A. Mukkara, A. Puglielli, R. Venkatesan, B. Khailany, J. Emer, S. W. Keckler, and W. J. Dally, “SCNN: An accelerator for compressed-sparse convolutional neural networks,” in 44th Annual International Symposium on Computer Architecture (ISCA), 2017, pp. 27–40.

[18] S. Zhang, Z. Du, L. Zhang, H. Lan, S. Liu, L. Li, Q. Guo, T. Chen, and Y. Chen, “Cambricon-X: An accelerator for sparse neural networks,” in 49th IEEE/ACM International Symposium on Microarchitecture (MICRO), 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 40-48 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

An Advanced Algorithm Combining SVM and ANN Classifiers to Categorize Tumor with Position from Brain MRI Images

Rasel Ahmmed*,[1], Md. Asadur Rahman[2], Md. Foisal Hossain[3]

[1]*Department of Electronics and Communication Engineering, East West University, Dhaka, Bangladesh*

[2]*Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh*

[3]*Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh*

ARTICLE INFO

Article history:
Received: 30 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Magnetic resonance imaging (MRI)
Temper based K-means & modified fuzzy C-means clustering (TKFCM)
Artificial Neural Network (ANN)
Support Vector Machine (SVM)
Tumor Size; and Tumor Position

ABSTRACT

Brain tumor is such an abnormality of brain tissue that causes brain hemorrhage. Therefore, apposite detections of brain tumor, its size, and position are the foremost condition for the remedy. To obtain better performance in brain tumor and its stages detection as well as its position in MRI images, this research work proposes an advanced hybrid algorithm combining statistical procedures and machine learning based system Support Vector Machine (SVM) and Artificial Neural Network (ANN). This proposal is initiated with the enhancement of the brain MRI images which are obtained from oncology department of University of Maryland Medical Center. An improved version of conventional K-means with Fuzzy C-means algorithm and temper based K-means & modified Fuzzy C-means (TKFCM) clustering are used to segment the MRI images. The value of K in the proposed method is more than the conventional K-means. Automatically updated membership of FCM eradicates the contouring problem in detection of tumor region. The set of statistical features obtained from the segmented images are used to detect and isolate tumor from normal brain MRI images by SVM. There is a second set of region based features extracted from segmented images those are used to classify the tumors into benign and four stages of the malignant tumor by ANN. Besides, the classified tumor images provide a feature like orientation that ensures exact tumor position in brain lobe. The classifying accuracy of the proposed method is up to 97.37% with Bit Error Rate (BER) of 0.0294 within 2 minutes which proves the proposal better than the others.

1. Introduction

The brain controls all psychological and physiological activities of human body. These functional activities can be disrupted or damaged due to the abnormal cell division or growing tumor in our brain that causes miscellaneous problems to the malfunction of our body. A human brain is divided into several major areas and these major areas (see Fig. 1) are related to different functional part of our body. Especially, these major areas are known as frontal lobe (marked with 1), central lobe (marked by 2, 3, & 4), parietal lobe (marked by 5 & 6), occipital lobe (marked with 7), and temporal lobe (marked with 8). Frontal lobe functions to control our thinking, emotion, innovation, and other cognitive works. Central lobe is a part of the frontal lobe and this part controls our movement related functions. The temporal lobe is responsible for listening and it helps to avoid the uncertainty principle. Occipital lobe helps us to see or observe anything. Another major part that controls the speech processing through reading compression area, sensory speech area and motor speech area of Broca is parietal lobe. These major areas can be affected by tumors which are definitely threat to our normal living. Therefore, the proper detection is the first priority for the remedy. This paper presents an efficient method for the detection of the tumor size as well as the position with classified tumor stages from MRI images and this work is an extension of our conference paper [1] that was presented in ECCE-2017.

*Rasel Ahmmed, East West University, Email: rsa@ewubd.edu

https://dx.doi.org/10.25046/aj030205

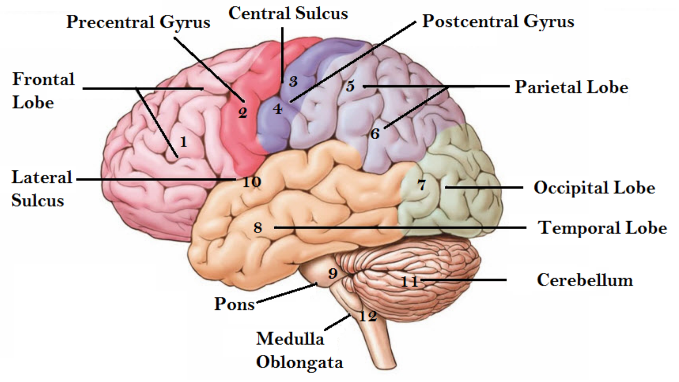


Figure 1: Major functional areas of human brain

In the recent era, there is a radical change in the field of medical science because it has experienced a great invention to the brain imaging. Previously it was predicted by the experts about brain hemorrhage on the basis of symptoms of a patient. But in recent decades the imaging techniques ease the way to detect any abnormality in the brain. There are several techniques of brain imaging like Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Ultrasound Imaging, Positron Emission Tomography (PET), and so on. Among the aforementioned imaging techniques, MRI is a gold accepted modality by which the brain can properly be imaged avoiding any harm.

In case of imaging by MRI, water density in soft tissue is considered and we know that water density in brain tissue is comparatively high than other tissue like bone. Due to the inhomogeneity of brain structure, their contrast values differ randomly in MRI and that is why the proper detection of tumor or tumor size becomes often difficult for the general observer [2]. Recently, proficient brain detection has been a remarkable challenge for medical engineering science. In this case, MRI gets more attention because of its grayscale images. Though MRI provides good contrast value than the other techniques, a suitable segmentation of brain MRI image is ostensible for abnormality detecting from the brain. Since the brain structure is quite complicated, MRI image segmentation needs the sophisticated method and precise results [3]. The segmentation of an MRI image describes notable image regions to attain region(s) of interest (ROI's) like as tumors, edema, legions, necrotic tissues, etc. from brain MRI images [4].

In medical diagnosis, most of the doctors rely on MRI images due to its robustness and that is why the accuracy of the prediction algorithms is very important because the result is crucial for patient treatment. Region growing algorithm [3] is used to categorize brain tumor through statistical feature based brain MRI image classification. The results are attained through some predefined criteria such as intensity information and/or edges. Consequently, the connected regions of an image are being extracted. The primary limitation of this method is finding seed points through manual interaction. The principal component analysis (PCA) algorithm with K-means discussed in [4] can be used to define the tumor class on the basis of some correlated pixel of the MRI images. The increased number of features and samples cause more time to consume and increase inaccuracy in results of PCA based K-means algorithm [5].

The classification with K-means clustering is claimed for CT-Liver image. Along with experimental application in brain MRI explained that proper segmentation can be possible with exact thresholding [6]. Again the *K*-value does not exceed greater than 3 if there is any gray level intensity more than that. The research work presented in [7], has introduced a masking algorithm for the classification with the aid of any classifier algorithm. It is effective in finding automatic seed point and neighbors but the dimension of the mask is to change manually for different brain MRI images [7]. In [8]-[10], the Fuzzy C-means algorithm (FCM) was proposed for segmentation. After that, an expert system was introduced with predefined membership and clustered centroid to trace a landmark tissue comparing with a prior model. On the other hand, FCM is described in [8] has limitation due to its noise sensitivity and inadequacy in the detection of abnormality in brain MRI images like a tumor, edema, and cyst. One of the most acceptable and used techniques for brain MRI image classification is Artificial Neural Network (ANN). An ANN technique is discussed in [11] with convincing results. Nonetheless, the procedure of ANN actually requires a perfect pixel classifier, high dimensional training data, and long time to attain the results those are susceptible conditions for the acute patients.

Most of the methods mentioned above are good in some specific point of view like better accuracy but time-consuming or low accuracy with less time consumption. To enhance the overall performance there be requisite of hybridization of those methods in a way that can be able to overcome these limitations.

An advanced algorithm combining SVM and ANN for tumor classification is introduced in this research work. Brain MRI images with normal and abnormal behavior are firstly enhanced through some filter and preprocessing steps. Thereafter, for the detection and classification of the tumor in the brain, proposed segmentation processes namely temper based K-means and modified Fuzzy C-means (TKFCM) clustering algorithm is used. In this technique, the K-values vary from 1 to 8 those are limited to only 1 to 3 in conventional *K*-means and the automatically updated membership function eradicates limitation of FCM. Then, two kinds of features are extracted from these segmented images. One is used to classify the tumor with SVM as it is easy to classify two kinds of dataset in this method and another is used to classify tumor with ANN into five categories along with four malignant stages. Again, the extracted features provide the classification through ANN and the orientation of tumor define the exact position of the tumor in the lobe (i.e. right, left &center) of the brain.

This paper is structured to present that in section 2 conventional K-means and Fuzzy C-means algorithm is presented, the proposed algorithm is described step by step in section 3, the results and discussions are in section 4 and finally, total work is concluded with a few words in section 5.

2. State of the Arts

2.1. K-means clustering

The conventional K-means is discussed with proper explanation in [6] and based on that idea, in this section this method is represented with a slight modification concerning the proposed work. Suppose, a data set $\{x_1, \ldots, x_N\}$ contains N number of observations where x is D-dimensional Euclidian variable. The basic intention of *K*-means is to divide the data set into N numbers of clusters, where the value of K is given. Naturally, this can be considered that a cluster including a group of data points and their inner side distances are small compared to the outer sided distances of the cluster. This concept can be formulated by familiarizing a set of *D*-dimensional vectors c_k, where *k*=1, …, *K*, in which c_k is an example related to the k^{th} cluster. Therefore, it is considered that c_k is representing the center of the cluster. Now it is consequence of the previous technique to find an assignment of those data points of clusters and a set of vectors $\{c_k\}$, in such an approach that the summation of the squares of each data point distance to its closest vector c_k could be minimum.

Now, it is suitable to define some symbolization to designate the assignment of data points to the clusters. For each data point x_n, a set of binary indicator variables $b_{nk}\in\{0,1\}$ can be introduced, where, *k*= 1, …, *K* represents the *K* clusters. The data point x_n is assigned to cluster *k* then b_{nk}=1 for $j\neq k$. This is recognized as the 1-of-*k* coding scheme. Consequently, definition of an objective function for distortion measurement [6] can be written as,

$$J=\sum_{n=1}^{N}\ \sum_{k=1}^{K} b_{nk}\left\|x_n-c_k\right\|^2 \tag{1}$$

The relation (1) represents the summation of the squares of each data point distance to its allotted vector c_k. The objective is to determine the values of $\{b_{nk}\}$ and $\{c_k\}$ so that the system can minimize *J*. It is usually determined through an iterative procedure in which each iteration comprises two successive steps related to successive optimizations regarding of the values of $\{b_{nk}\}$ and c_k. At first, some preliminary values of c_k are chosen and then the first phase *J* is being minimized with respect to the $\{b_{nk}\}$, maintaining the values of c_k, fixed. In second phase, *J* is to minimize in regard to the c_k, maintaining the values of $\{b_{nk}\}$, fixed. This two-stage optimization process is repeated until convergence. These two stages of updating $\{b_{nk}\}$ and c_k correspond respectively to the *E* (expectation) and *M* (maximization) steps of the EM algorithm in [12], and to emphasize this and EM is used the terms *E* step and *M* step in the context of the *K*-means algorithm.

Here, $\{b_{nk}\}$ is considered as first determination because *J* in (1) is a linear function of $\{b_{nk}\}$. This optimization is generally evaluated to offer a closed form solution. The values of *n* are independent and that is why this can be optimized for each *n*, separately. By selecting $\{b_{nk}\}$ as 1 gives the minimum value of $\left\|x_n-c_k\right\|^2$ for whatever the value of *k*. In other words, it can be merely allotted the n^{th} data point to the neighboring cluster center. More strictly, this can be stated as [12],

$$b_{nk}=\begin{cases}1 & if\ k=\arg\ \min_a\left\|x_n-c_a\right\|^2,\ a=1,\dots,k\\ 0 & otherwise\end{cases} \tag{2}$$

Consequently in this situation, it is to consider that the optimization procedure of the c_k with the values of $\{b_{nk}\}$ is occurred immovable. Here, *J* is the objective function which is actually quadratic function of c_k, and it is commonly minimized through setting its derivative regarding the values of c_k to be zero and this consideration gives the following mathematical relation given in (3).

$$2\sum_{n=1}^{N} b_{nk}(x_n-c_k)=0 \tag{3}$$

From (3), the values of c_k can be easily evaluated as,

$$c_k=\frac{\sum_n b_{nk}x_n}{\sum_n b_{nk}} \tag{4}$$

The denominator of (4) is equal to the number of points allotted to the cluster *k* and subsequently this outcome has a simple explanation, explicitly set c_k equal to the mean of all the data points, x_n those are being assigned to the cluster *k*. Hence, this technique is recognized as the K-means algorithm.

2.2. Fuzzy c-means

The Fuzzy c-means (FCM) algorithm for image clustering with FORTAN code was first introduced in [14]. In this paper, we have presented FCM algorithm with an improvement of earlier clustering methods which actually followed by the explanation given in [13]. According the approach of this paper, suppose *R* is the set of real number where R^P and R^+ are the set of *p* tuples of real number and set of nonnegative real number, respectively. Here, W_{cn} is a matrix of order *c* × *n* which is called feature space where feature element, $x\in R^P$ & feature vector $x=(x_1, x_2, \dots, x_p)$ is consists of *p* real numbers.

Delineation 1: If *X* is a subset of R^P and every function $u:X\rightarrow[0,1]$ is considered to be assigned to each $x\in X$, its grade of membership should be in the Fuzzy set *u*. The function *u* is termed a Fuzzy subset of *X*. It can be noted that there could be infinite Fuzzy sets related to the set *X*. It is anticipated to make "partition" *X* by the means of Fuzzy sets. Normally, it is executed by defining a number of Fuzzy sets on *X* such that for each $x\in X$. The summation of the Fuzzy memberships of *x* in the previously considered Fuzzy subsets is one.

Delineation 2: It is given that, a finite set $X\subseteq R^P$, $X=(x_1,x_2,\dots,x_p)$, and an integer *c* ($2\leq c\leq n$) originate a Fuzzy *c* partition of *X* that can be represented by a matrix $U\in W_{cn}$ whose entries satisfy the following conditions:

i) The number of row *i* of *U* or $U_i=(u_{i1},u_{i2},\dots,u_{in})$ exhibits the i^{th} membership function of *X*.

ii) The number of column *j* of *U* or $U_i=(u_{1j},u_{2j},\dots,u_{cj})$ revelations the values of the *c* membership functions of the j^{th} data in *X*.

iii) The term, u_{ik} will be construed as u_i (x_k) which actually represents the value of the membership function of the i^{th} Fuzzy subset for the k^{th} data.

iv) The summation of the membership values for each x_k will be always one.

v) No Fuzzy subset will be empty.

vi) No Fuzzy subset will contain all elements of *X*.

M_{fc} denotes the set of the partitions of *X* in case of Fuzzy *c*. Here, the distinctive subset $M_c\subseteq M_{fc}$ of *X* in every u_{ik} is 0 or 1. In addition, the subset is the discrete set of non-Fuzzy *c* partitions of *X*. The solution space, *M* is for the conventional clustering algorithms. The Fuzzy c-means algorithm followed by this proposed work uses the iterative optimization method in order to approximating an objective function minimization which measures similarity on $R^P\times R^P$.

Delineation 3: Suppose that, $U\in W_{fc}$ is a Fuzzy *c* partition of *X*, and consider that *v* is the *c* tuple (v_1, v_2, …,v_c), $v_i\in R^P$. Therefore, $J_m : M_{\text{fcd}} \in R^P \rightarrow R^+$ is described by the following relation:

$$J_m = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{jk})^m (d_{ik})^2\ ;\ \ v=(v_1,v_2,...,v_c)\in R^{cP} \quad (5)$$

Additionally, $v_i\in R^P$ is considered to be the cluster center or prototype of class *i*, $1\leq\ i\leq c$, and consequently,

$$d_{ik}^2 = \|x_k - v_i\|^2 \quad (6)$$

Here, $\|\,.\,\|^2$ represents any inner product norm metric that defines the Euclidian distance [6], and $m\in[1,\infty]$. This distance calculates the distance from cluster centroid to each object. If we take the Euclidean distance with the distance matrix at null iteration, we get following relationship:

$$D=\begin{bmatrix}\alpha_1 & \beta_1 & \chi_1 & \delta_1\\ \alpha_2 & \beta_2 & \chi_2 & \delta_2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \quad (7)$$

Each column in *D* symbolizes the object. The first row of *D* relates to the distance of the each object to the 1st centroid and similarly, the second row of *D* is the distance of the each object to the 2nd centroid. As for example, distance from cluster center c= (*x*, *y*) to the first centroid c_1= (α_1, α_2) is $(x-\alpha_1)^2+(y-\alpha_2)^2$, and its distance to the second centroid c_2 = (β_1, β_2) is $(x-\beta_1)^2+(y-\beta_2)^2$, and so on.

The FCM algorithm by the iterative optimization method produces a J_m Fuzzy C-partition of the data set $X=\{x_1, . . . , x_n\}$. The basic steps to implement the algorithm can be represented by the following consecutive mathematical phases (followed the explanation given in [13]).

1) Fix the cluster numbers *c*, $2\leq\ c\leq n$ where, *n* is the number of data items. Besides, it is to fix, *m* ($1< m <\infty$). Choose any inner product induced norm metric $\|\,.\,\|$ e.g., $A\in W_{pp}$.

2) Initialize the Fuzzy *c* partition $U^{(0)}$,

3) At step *b*, *b*=0, 1, 2, ...

4) Compute the *c* cluster for the centers $\{v_i^{(b)}\}$ with $U^{(b)}$ and the formula for the *i*th cluster center is usually consider as the following relation:

$$v_{il}=\frac{\sum_{k=1}^{n}(u_{ik})^m x_{kl}}{\sum_{k=1}^{n}(u_{ik})^m},\qquad (l=1,2,...,p) \quad (8)$$

5) Bring up-to-date, $U^{(b)}$: determine the memberships in $U^{(b+1)}$ as the following steps a) & b). For k = 1, 2, 3, …, *n*.

a) Evaluate, I_k and $\tilde{I}_\text{k}$:

$$I_k=\{i\,|\,1\leq i\leq c,\ d_{ik}=\|x_k-v_i\|=0\},$$
$$\tilde{I}_k=\{1,2,.....,c\}-I_k, \quad (9)$$

b) For data item *k*, determine the new membership values so that,

i) if I_k = 1

$$u_{ik}=\left[\sum_{j=1}^{C}\left\{\frac{d\left(x_k,v_i\right)}{d\left(x_j,v_k\right)}\right\}^{\frac{-2}{m-1}}\right] \quad (10)$$

ii) else, u_{ik} = 0 for all $i\in I_k$ and $\sum_{i\in I_k} u_{ik}=1$, next *k*.

6) Compare $U^{(b)}$ and $U^{(b+1)}$ in a convenient matrix norm; if $\left\|U^{(b)}-U^{(b+1)}\right\|\leq\varepsilon$, (where $\varepsilon=\{0\ to\ 1\}$) stop; otherwise, set *b*=*b*+1, and go to step 4.

The use of FCM algorithm necessitates the determination of several parameters, *i.e.*, $c\in m$, the inner product norm $\|.\|$, and a matrix norm. In addition to that, the set $U^{(0)}$ of initial cluster centers should have to be defined for sure. Although no necessary rules for choosing a good value of *m* are available in the literature. In most of the cases, the value of *m* is typically reported as the useful range of values as $1\leq m\leq 5$. The objective of the algorithm mentioned above is to reduce the computational burden imposed by iterative looping between (9) and (10) when *c*, *p*, and *n* are large.

3. Methodology

3.1. MRI image collection

The data of brain MRI images are collected from internet public repository. The images of normal and tumorous brain with Lower Grade Glioma or Glioblastoma Multiforme are collected from the sources [15]-[17]. The number of the used data for each MRI image for classification of normal and tumor brain through SVM is of 39 images. There are 37 images for the classification of benign and malignant tumor stages.

3.2. Image processing

The enhanced images are achieved from the raw MRI images through some steps described as follow.

- **Image conversion and orientation setting**: The images MRI images are converted from *.mha* format and *.dicom* format into *.jpg* by using MATLAB conversion tools application. On the consequence of the conversion, the sizes and directions are reset. This step is conducted in order to have the same size and direction for all the MRI images. This process is performed automatically by using MATLAB with 256×256 pixels for the betterment of image usage.
- **Image enhancement:** The transformed images from the previous step, at first they are converted to L*a*b* images for the comparatively better view and quality. That is why the values of the luminosity of the images can be spanned with a range from 0 to 100 which should be scaled to [0 1] range (appropriate for MATLAB intensity images of class double) before applying the three contrast enhancement techniques like adjusted, adaptive thresholded, and histogram imaging. In this method, for smoothing the images hybridization of both weiner2 (image, [40, 40]) and median2 filter is assured and acquired good results.

3.3. Proposed TKFCM algorithm for tumor detection

In this segmentation process, K-means algorithm is used to segment MRI images on the basis of gray level. This gray level is selected depending on the temper of the image. Then the modified Fuzzy c-means algorithm which depends on the updated membership is applied to segment the temper based K-means segmented image. The membership of modified Fuzzy c-means is updated with the cluster distances from centroid defined by the features of the tumor MRI image. The TKFCM algorithm is the combination of the K-means algorithm and Fuzzy c-means algorithm with some important modifications. The temper is added in the proposed approach along with the conventional K-means algorithm which is identified by the temper or gray level intensity in the brain MRI images. Besides, the Fuzzy c-means membership and Euclidian distance are also modified by the image features.

Here, the coarse image $B(x_i, y_i)$ which is marked and describing the desired tempers for the K-means could be found through convolution of gray level based temper and image given as,

$$B(x_i, y_j) = \sum_{i=n+1}^{M+n} \sum_{j=n+1}^{N+n} P(x_i, y_j) \oplus T_{MN(resize)} \tag{11}$$

Temper based window is selected by T_{MN} that is calculated as,

$$T_{MN} = \sum_{i=n+1}^{M-n} \sum_{j=n+1}^{N-n} P(x_i, y_j) \tag{12}$$

In (12), there is presented a temper based matrix of image with a number of gray level intensities, G and number of bins, S those are used to detect the temper of the images $P(x_i, y_j)$. Where n is defined as $n = (window\ Size-1)/2$. With exact value of the temper, row and column, the desired temper is obtained.

Separately temper based K-means and modified Fuzzy c-means clustering algorithm for segmentation can be written in equation as below:

$$J_k = \sum_{i=1}^{C} \sum_{j=1}^{K} B(x_i, y_j) \left\| x_i - c_j \right\|^2 \tag{13}$$

$$J_m = \sum_{j=1}^{N} \sum_{i=1}^{K} (U_{ij})^m (d_{ij})^2 \tag{14}$$

Here, M and N are the row and column of the binary coarse image matrix $B(x_i, y_j)$. The number of data points in clusters, centroid of the cluster, and number of clusters are defined by C, N, and K, respectively.

Then, from (13) and (14) the contour through which the exact location of tumor portion in any image can be find is

$$J_{km} = \oint_c \left(J_k, J_m \right) \tag{15}$$

The relation given in (15) shows a contour integral of temper based K-means image and updated membership based FCM image, where c is the contour value. The whole method that has been proposed for tumor detection from brain MRI image using Temper based K-means and modified fuzzy C-means is described by the flow chart given in Figure 2.

3.4. Feature extraction

The system extracts the first and second order statistical features as in [18]. The first order statistical features like contrast, correlation, entropy, energy, and Homogeneity are used to detect exact tumor and its position in the brain MRI image through SVM. On the other hand, second order region based statistical features provide area, eccentricity, and the perimeter is required for distinguishing the malignant and benign tumor. These second order feature values are used as the input of the ANN and provide desired tumor categories. The feature extraction procedure is mentioned step by step in Figure 2.

3.5. Methodology of combining SVM and ANN

In this proposed work, the combination of SVM and ANN is used to classify the tumor and its stages. The hybridization of the method is used firstly to classify normal and tumor dataset using Linear SVM Kernel and then classify the tumor data into different stages through ANN. In SVM there will be a hyperplane between the set of data points as the decision boundary. In this case, there are two classified data of normal and abnormal (Tumor) brain images and the hyper planes of SVMs are used to separate these classified data as normal data and tumor data. The input towards the ANN is the information of feature extraction; the first and second order statistic features. Then, the ANN will generate the output results as of benign and malignant I-IV tumors. The process of using ANN to attain the proposed goal of this research work is explained by the flow diagram presented in Figure 3.

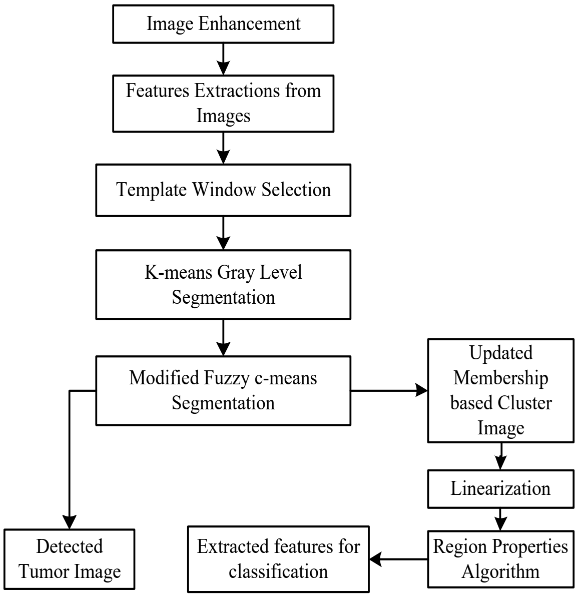


Image Enhancement
Features Extractions from Images
Template Window Selection
K-means Gray Level Segmentation
Modified Fuzzy c-means Segmentation
Updated Membership based Cluster Image
Linearization
Region Properties Algorithm
Extracted features for classification
Detected Tumor Image

Figure 2: Methodology steps for proposed TKFCM algorithm in tumor detection

4. Results and Discussions

To implement the proposed methodology, first of all, a database of 46 multifaceted brain tumor images is created. It is aforementioned that, the images are collected from the sources of [15]-[17]. To make the quality of the images acceptable for the proposed methodology, some necessary steps were to do and the database was set for the network. To do so for the previous consequence, a thresholding method is applied with threshold level 0.8 and the morphological operation was performed. In addition to that median and hybrid filters were used to remove the primary noises from the MRI images. The filtered MRI images are presented in Figure 4 (a) & 4(d).

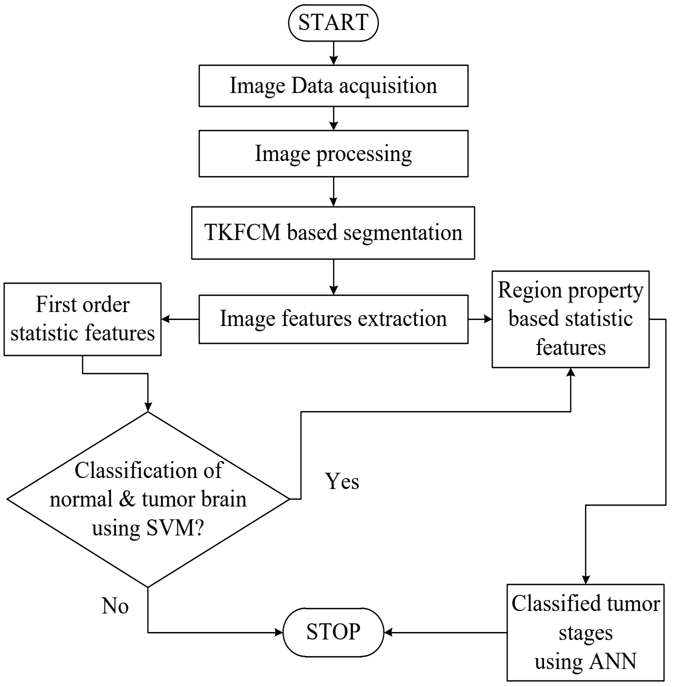

START
Image Data acquisition
Image processing
TKFCM based segmentation
First order statistic features
Image features extraction
Region property based statistic features
Classification of normal & tumor brain using SVM?
Yes
No
STOP
Classified tumor stages using ANN

Figure 3: The methodology with their step by step description to implement the proposed hybridization of SVM & ANN algorithms in the filed tumor classification.

The preprocessing steps include filtration, increase of luminosity, and adjust of contrast. After the previous processing, all the images are segmented by the proposed TKFCM based segmentation process. The effects of this segmentation process on images are shown in Figure 4(b) & 4(e) and 4(c) & (f). Temper based K-means (TK) segmentation method shows 8-gray level intensity-based images in Figure 4(b) & 4(e). These figures describe that the input image is segmented by the combined effects of TK-segmentation and the modified approach of FCM algorithm.

The updated membership function with proper Euclidian distance for modified FCM represents the detected tumors in Figure 4(c) & 4(f). These are some examples of the optimum result that can differ this modified and hybrid TKFCM method from the conventional methods. The extracted tumors according to the proposed technique are marked with red color in Figure 4(c) & 4(f). Therefore, these segmented areas can be used to determine the region property based statistical features from the images.

In Figure 5, we presented ten images (given in Figure 5 (a) & (d)) as the input of classification by the proposed TKFCM scheme. These ten images are taken from the created database and used for the classification of the tumor and its area. The enhanced input images for TKFCM (shown in Figure 5(a) & 5(d)) are preprocessed by the steps described by the visual explanation in Figure 4. In Figure 5 (b) & 5(e), the detected brain tumor images marked as the red color those are the outcomes of the previous images by the application of the TKFCM method. Classified brain tumors through linearization of TKFCM are presented in Figure 5 (c) & 5(f). From which the region based features are obtained by the level thresholding, updated membership function, and region properties algorithm.

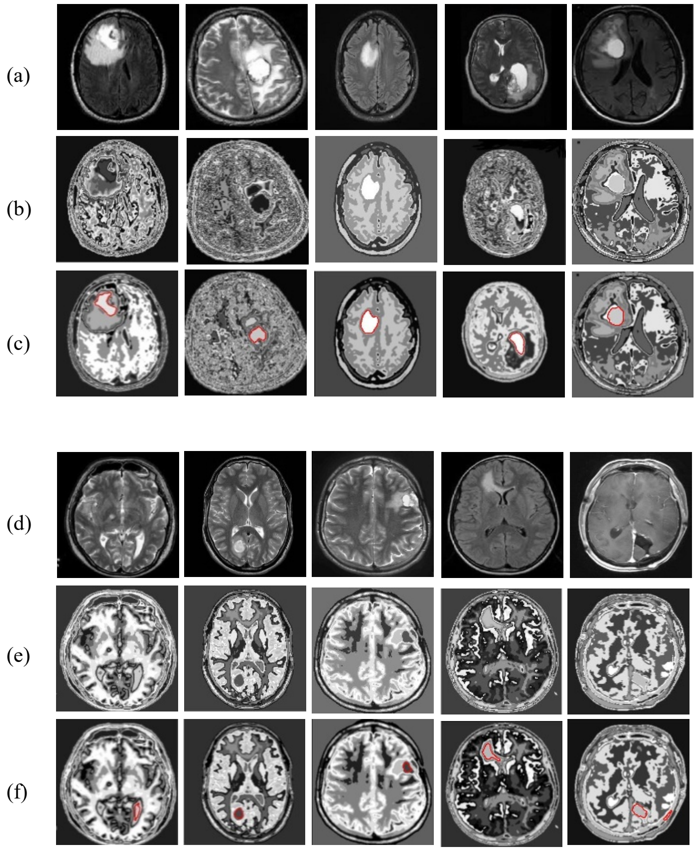


Figure 4: (a) & (d): Filtered images for TKFCM, 4(b & e): Temper based segmented images, 4(c & f): TKFCM based detected tumor images.

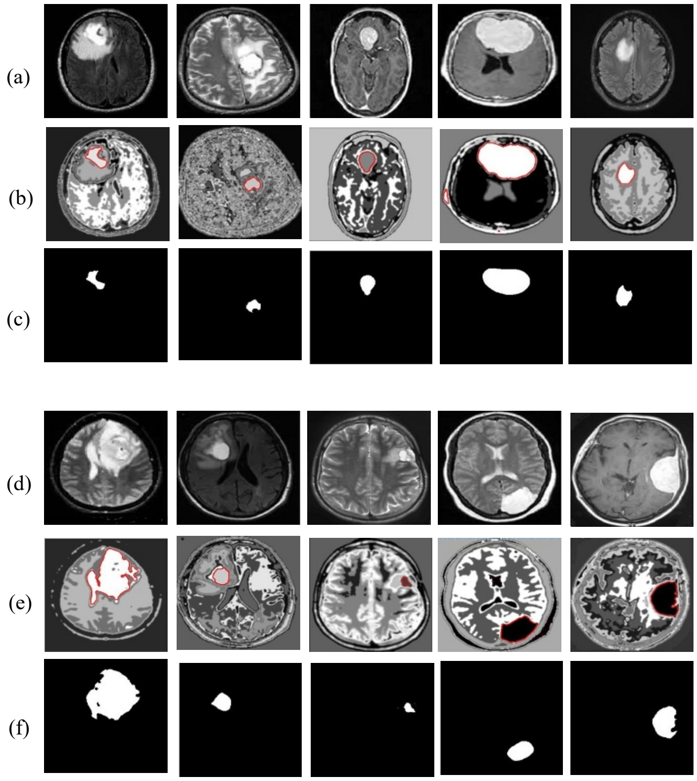


Figure 5: Enhanced Input images for TKFCM in 5(a & d), TKFCM based red marked tumor images in 5(b & e), and classified brain tumor through linearization of TKFCM in 5(c & f).

From the segmented image, two kinds of features are extracted: one is statistical features and other is region property based features. The first kind of statistical features *i.e.*, contrast, correlation, entropy, homogeneity, energy are used to classify whether there is any normal brain or tumorous brain using SVM. Therefore, SVM classifies the images into two categories- tumorous and normal. For this significance, two major kinds of MRI image data of normal and abnormal brain are fed to SVM network and the corresponding results of classification property of the SVM are presented in Figure 6(a). There are 46 images used to classify whether the system classified 37 tumor images, 8 normal images and rest 1 is misclassified. As a result, the accuracy of SVM method is 97.44% which is very convincing to move to the next step c.

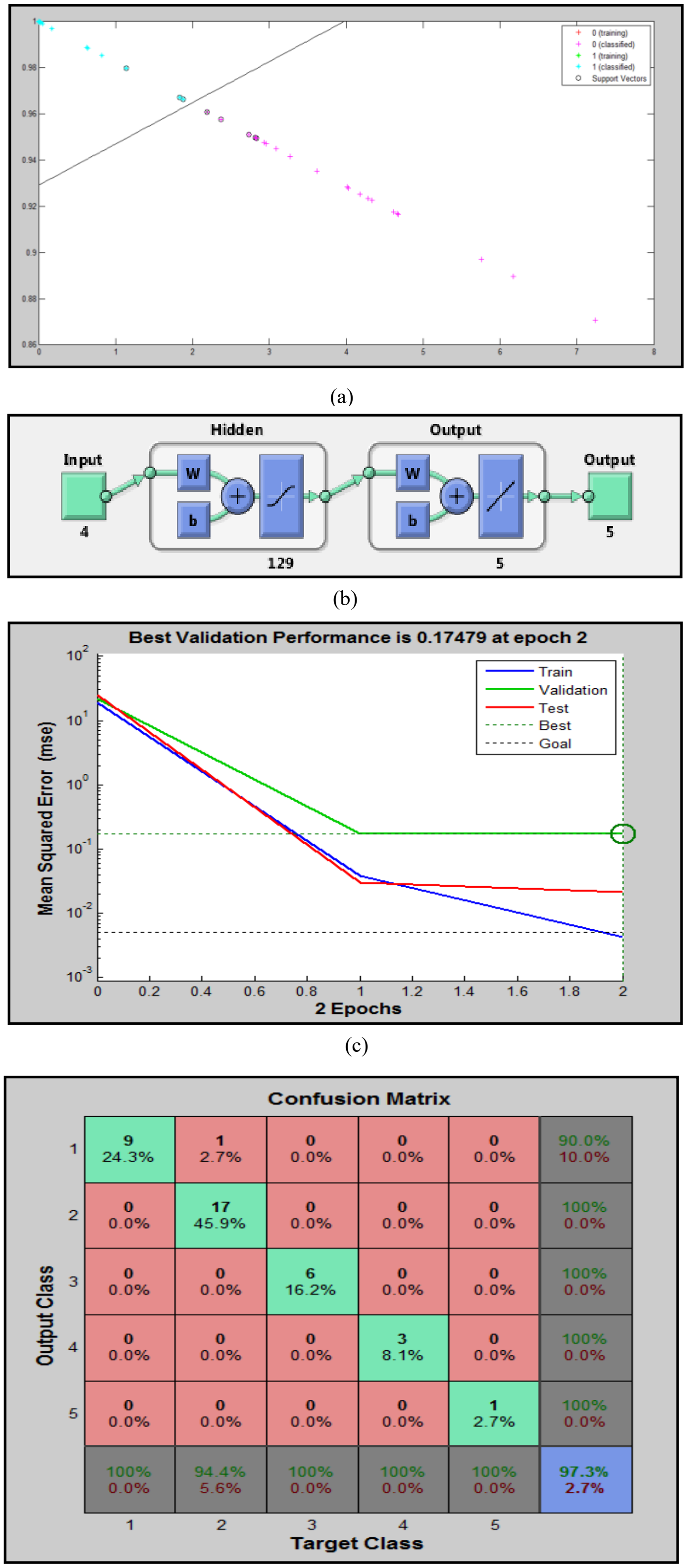
According to the previous approach the tumorous brains have to be classified into 5 defined categories. Due to implementing this classifying network, an ANN-based network was designed as the configuration given in Figure 6(b). The network consists of four input vector layers, one hidden layer with 129 neurons, and five output vector layers. The input feature vectors are evaluated as the second kind of features of MRI images which are acknowledged in this article by the region property based features.

The performances of training, validation, and testing of the proposed network are shown in Figure 6(c). The specifications of the achieved performances were 60 iterations with 0.05 increment order. Additionally, the minimum error was considered up to 0.5e-02, the gradient minima were approximately 1e-10, and the maximum validation check failure was taken 6. From the Figure 6(c), it is found that the desired performances are achieved between 8 and 9 iterations which indicate the less required time for the network compared to the 60 iterations. Best validation performance is 0.17479 at iteration 2 mentioned in Figure 6(c).

The second kind of features *i.e.*, area, eccentricity, perimeter, orientation are used to differentiate the tumor and different malignant stages of I to IV. The obtained features are denoted as input vectors of the proposed network. On this contrary, we acknowledged in this article that we have five categories of brain tumor those are I-IV malignant groups along with benign. Therefore, the results are given by the confusion matrix in Figure 6(d) are the output vectors or classified 5 groups. The network achieved 97.3% classifying accuracy given in Figure 6(d). This result is very convincing where it is found that there are 9 benign, 17 malignant I, 6 malignant II, 3 malignant III, and 1 malignant IV stages. The network misclassified 1 tumorous brain MRI images.

In Table 1, there are some feature values for classifying tumor categories as well as detecting exact tumor position in our brain lobe. The region-based features are used as input vector of ANN to classify the malignant and benign tumors. Beside one of the features mentioned as 'orientation' of the tumor defines the exact lobe position of the tumor in brain MRI images. These results of orientation by the proposed method provide the information about the position of the tumor either in left or right or center part of the brain.

Since the classification accuracy is not the only performance metric, some other performances of the classifier are necessary to be evaluated. Based on the value of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) of the proposed network, some other performances like sensitivity, specificity, beat error rate (BER) can be calculated. In this step,



(a)

(b)

(c)

Confusion Matrix

Output Class \ Target Class	1	2	3	4	5	
1	9 24.3%	1 2.7%	0 0.0%	0 0.0%	0 0.0%	90.0% 10.0%
2	0 0.0%	17 45.9%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	6 16.2%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	3 8.1%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 2.7%	100% 0.0%
	100% 0.0%	94.4% 5.6%	100% 0.0%	100% 0.0%	100% 0.0%	97.3% 2.7%

(d)

Figure 6: The relevant simulated results of proposed hybridized SVM and ANN algorithm for classifying tumor categories. 6(a): 2D plot of classified normal and tumor brain data for proposed SVM classifier, 6(b): The network architecture of proposed ANN method, 6(c): The performance curves for the proposed ANN method, and 6(d): The confusion matrix for benign and malignant stages I-IV using proposed ANN classifier for second order region based features.

Table 1: Some examples of classified tumor and identification tumor position using ANN approach

Area in mm^2	Eccentricity in mm	Perimeter in mm	Orientation in degree	Tumor type	Tumor position in lobe
9.9167	0.9754	60.275424	-84.1794	Malignant II	Left
6.472	0.9213	33.55836	-100.0502	Malignant I	Left
8.2561	0.8402	40.736784	-54.547	Malignant I	Left
9.5952	0.9414	36.863376	-57.7349	Malignant I	Left
8.6719	0.7084	31.913904	0.404	Malignant II	Centered
10.2451	0.6974	41.258448	-32.196	Malignant II	Left
21.5625	0.793	82.645728	0.6467	Malignant III	Centered
9.7787	0.6381	37.249872	-88.0731	Malignant II	Left
8.1541	0.8617	32.143056	69.189	Malignant I	Right
7.0295	0.7793	25.955424	14.0189	Malignant I	Right
7.9814	0.9886	18.599592	85.0421	Malignant I	Right
22.4633	0.6164	135.3689	60.2044	Malignant III	Right
8.0379	0.4214	28.77864	-59.6895	Malignant II	Left
8.7599	0.8547	36.848064	1.3184	Malignant I	Right
24.4937	0.8675	135.78576	-77.2517	Malignant III	Left
7.7914	0.7926	36.24984	-49.5862	Malignant I	Left
13.3835	0.8289	49.570752	12.2018	Malignant I	Right
15.8158	0.8813	75.392592	44.3799	Malignant II	Right
11.0597	0.625	89.080464	-55.4867	Malignant II	Left
36.9817	0.4032	141.77	-82.9364	Malignant IV	Left
15.7074	0.671	58.995552	89.6919	Malignant III	Right
12.3489	0.8206	58.698816	-82.3899	Malignant II	Left

Table 2: Comparison between proposed algorithm and conventional methods

Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)	BER	Computational Time
Thresholding	85	80	81.3	0.175	~3 min
Region Growing	88.46	75	86.47	0.182	~6 min
ANN	95.42	100	95.07	0.022	~8 min
FCM	86.95	85.7	86.4	0.136	~5 min
SVM	96.2	66.67	90.44	0.0234	~4 min
K-means	75	92.85	83.7	0.160	~160-170 sec
TKFCM	88.9	100	91.89	0.055	~100 sec
Fuzzy Logic Method	96.3	100	96.667	0.018	~ 120 sec
Proposed SVM+ANN Method	**98**	**100**	**97.37**	**0.0294**	**~ 2 min**

*Image size=256x256, Software=MATLAB2014a, Processor= Core2duo, RAM=2GB, windows=7

these important estimations are performed by the following relations (16)-(19) and used for comparison purposed with other conventional methods [19].

$$Sensitivity = \frac{TP}{(TP+FN)} \times 100 \quad (16)$$

$$Specificity = \frac{TN}{(TN+FP)} \times 100 \quad (17)$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100 \quad (18)$$

$$BER = \frac{1}{2}\left(\frac{(FN)}{(TP+FN)} + \frac{(FP)}{(FP+TN)}\right) \quad (19)$$

Comparisons of the proposed method with predictable techniques whose algorithms are coded to cope up with the same dataset are mentioned in Table 2. In this proposed technique, 37 brain tumor images are taken for the simulation. The sensitivity of 98%, specificity of 100%, the accuracy of 97.37%, and BER 0.0294, which are obtained by combining ANN and SVM. The results show that the proposed method is better than the conventional methods like thresholding, region growing, SVM, ANN, FCM, TKFCM, and K-means. The proposed algorithm that is actually combinations of several suitable techniques proves more effective than any other single method. Though the conventional ANN comprehends the value of accuracy of 95.07%, the computational time is so high. But the proposed technique shows less computational time with better accuracy than conventional ANN. So, it can be useful for both detecting tumors and be classifying tumor stages of brain MRI images for experts.

5. Conclusion

This work classified the images as normal and tumorous. In addition, the size of the tumor and its position in brain lobe were also identified by the proposed method from MRI image. To implement such an efficient and intelligent algorithm, a number of statistical and machine learning based algorithms like temper based *K*-means and modified Fuzzy *c*-means clustering, SVM, and ANN are combined. As a result, it is found that the previous limitations like less classifying accuracy, computational time requirement, unsatisfactory of BER, sensitivity, and specificity of conventional ideas has been overcome by this proposed algorithm. Other existing methods have some tradeoff among the performances. This method provides sensitivity 98%, specificity 100%, classifying accuracy 97.37%, BER=0.0294, and required less than 2 minutes to give the result. These results are too convincing to identify the brain tumor and its size, as well. In addition, this algorithm has provided the tumor position accurately. Considering all the performances, it can be concluded that the proposed algorithm is better than others like region growing, thresholding, and FCM compared to every parameter. On the other hand, in point of accuracy and computational time the proposed method is better than ANN, SVM, and TKFCM. Eventually, we hope that this method can be very helpful for diagnosis of brain tumor.

Conflict of Interest

So far the knowledge, the authors declare no conflict of interest regarding this article.

Acknowledgment

Authors would like to thank Prof. Dr. Md. Abdur Rafiq, Dean, Faculty of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh, for his beautiful lectures on neural network- fuzzy logic and guidelines to implement it in this research work.

References

[1] R. Ahmmed, A. S. Swakshar, M. F. Hossain, and M. A. Rafiq, "Classification of tumors and it stages in brain MRI using support vector machine and artificial neural network," International Conference on Electrical, Computer and Communication Engineering (ECCE-2017), 229-234, February 16-18, Cox's Bazar, Bangladesh. http://doi.org/ 10.1109/ECACE.2017.7912909.

[2] R. Rana, H. S. Bhadauria, and A. Singh, "Study of various methods for brain tumor segmentation from MRI images," International Journal of Emerging Technology and Advanced Engineering, **3**(2), 338-342, 2013.

[3] J. Joshi and A. Padhke, "Feature Extraction and Texture Classification in MRI," International Journal of Computer & Communication Technology (IJCCT), **2**, 130-136, 2010.

[4] A. Ahirwar, "Study of techniques used for medical image segmentation and computation of statistical test for region classification of brain MRI," International Journal on Information Technology and Computer Science, **5**, 44-53, April 2013. doi:10.5815/ijitcs.2013.05.06

[5] I. E. Kaya, A. Ç. Pehlivanli, E. G. Sekizkardes and T. Ibrikci, "PCA based clustering for brain tumor segmentation of T1W MRI Images" Computer Methods and Program in Biomedicines in Elsivier, **140**, 19-28, March, 2017. http://doi.org/ 10.1016/j.cmpb.2016.11.011

[6] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," Annual review on Biomedical Engineering, **2**, 315-337, 2000.

[7] W. Narkbuakaew, H. Nagahashi, K. Aoki, and Y. Kubota, "Integration of modified K-means clustering and morphological operations for multi-organ segmentation in CT liver-images," Recent Advances in Biomedical & Chemical Engineering and Materials Science, 34-39, March 2014.

[8] G. Deng, "A generalized unsharp masking algorithm," IEEE Transactions on Image Processing, **20(**5), 1249-1261, May 2011. http://doi.org/ 10.1109/TIP.2010.2092441

[9] S. R. Kannana, S. Ramathilagam, R. Devi, and E. Hines, "Strong fuzzy c-means in medical image data analysis," Journal of Systems and Software, **85**, 2425–2438, December 2011. http://doi.org/10.1155/2013/972970

[10] V. Harati, R. Khayati, and A. R. Farzan, "Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images," Computers in Biology and Medicine, **41**, 483–492, April 2011. http://doi.org/10.1186/s12938-016-0165-2

[11] H. R. Shally, and K. Chitharanjan, "Tumor volume calculation of brain from MRI slices," International Journal of Computer Science & Engineering Technology (IJCSET), **4**(8), 1126-1132, 2013.

[12] X. Descombes, F. Kruggel, G. Wollny, and H. J. Gertz, "An object-based approach for detecting small brain lesions: Application to Virchow-robin spaces," IEEE Transaction on Medical Imaging, **23**(2), 246–255, 2004. http://doi.org/10.1109/TMI.2003.823061

[13] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering amgorithm," Computer & Geosciences, **10(**2-3), 191-203, 1984. https://doi.org/10.1016/0098-3004(84)90020-7

[14] M. Gong, Y. Liang, J. Shi, W. Ma and J. Ma, "Fuzzy c-means clustering with local information and kernel metric for image segmentation," IEEE Transactions On Image Processing, **22**(2), 573-584, February 2013. http://doi.org/10.1109/TIP.2012.2219547

[15] (2017) PE- Brain tumor website. [Online]. Available: http:// www.mayfieldclinic.com /

[16] University of Maryland Medical Center (2017) homepage. Primary brain tumor [Online]. Available: http://umm.edu/health/medical/reports/articles/brain-tumors-primary

[17] Oncology (2017) homepage. Tumor type [Online]. Available: http://www.oncolink.org/types/article.cfm?c=52&aid=247&id=9534

[18] M. Lugina, N. D. Retno, and R. Rita, "Brain tumor detection and classification in magnetic resonance imaging (MRI) using region growing, fuzzy symmetric measure, and artificial neural network back propagation", International Journal on ICT, **1**, 20-28, December 2015.

[19] J. Selvakumar, A. Lakshmi, and T. Arivoli, "Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and fuzzy c-mean algorithm," International Conference on Advances in Engineering, Science and Management, Tamil Nadu, 186-190, 2012.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 49-57 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Interference Avoidance using Spatial Modulation based Location Aware Beamforming in Cognitive Radio IOT Systems

Jayanta Datta[*,1], Hsin-Piao Lin[2]

[1]*Department of Electrical Engineering and Computer Science, National Taipei University of Technology, Taipei City 10608, Taiwan*

[2] *Department of Electronic Engineering, National Taipei University of Technology, Taipei City 10608, Taiwan*

ARTICLE INFO

Article history:
Received: 25 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Spatial Modulation
Femto-Cell
Virtual Antenna Array

ABSTRACT

The Internet of Things (IOT) is a revolutionary communication technology which enables numerous heterogeneous objects to be inter-connected. In such a wireless system, interference management between the operating devices is an important challenge. Cognitive Radio (CR) seems to be a promising enabler transmission technology for the 5G-IOT system. The "sense-and-adapt" smart transmission strategy in CR systems can help to overcome the problem of multiple access interference (MAI) in IOT systems. In this paper, a 5G-IOT smart infrastructure system is arranged in the form of CR based virtual antenna array (VAA) system. In VAA based wireless system, knowledge of users' locations can help the transmitter to achieve interference avoidance by steering the main beam towards the intended recipient. This idea has been applied to the VAA-IOT system, where smart antenna array based location aware beamforming are applied at both transmitter and receiver cluster of smart sensors with the help of spatial modulation principle. The waveform of choice for the CR-IOT clusters is Generalized Frequency Division Multiplexing (GFDM) while corresponding waveform for the primary user (PU) cluster is conventional Orthogonal Frequency Division Multiplexing (OFDM). Computer simulation shows that under multipath fading conditions, the implemented system can reduce the interference to the primary user (PU) system, leading to better coexistence.

1. Introduction

This paper is an extension of work originally presented in 2017 International Conference of Applied System Innovation, held at Sapporo, Japan [1]. In this work, a beamforming scheme is presented which is based on transmitter activation by spatial modulation (SM) technique. Spatial Smoothing (SS) based Multiple Signal Classification (MUSIC) algorithm is applied for direction of arrival (DOA) estimation of the PU. Based on the DOA data, the combined spatial modulation-beamforming scheme is designed for avoiding interference to the PU. In [1], an overview of the methodology was presented. However, in this work, details about application of the SS-MUSIC as well as receiver side data detection are presented. Computer simulations are performed to demonstrate the effectiveness of the designed scheme for better interference avoidance in CR scenario.

Increasing demand for higher data rate and quality of mobile communications has led to the development of advanced multicarrier signaling schemes such as OFDM and GFDM among others. GFDM is a recently proposed non-orthogonal multicarrier waveform which is a potential candidate for 5G wireless technology. Its benefits are flexible carrier aggregation and low out-of-band (OOB) radiation [2-4]. Due to its flexibility in subcarrier allocation, the GFDM based user can sense the spectral bands in a CR environment and intelligently allocate the subcarriers to the bands where the interference temperature threshold can be satisfied. Apart from frequency bands, space and angle dimensions can be exploited too with the recent advances in multiple antenna technologies. In such a smart heterogeneous infra-structure, multi-antenna based GFDM system can lead to enhanced system performance and low cross-tier interference. This can be achieved with the application of beamforming technology at the Femto-Cell transceiver system. Moreover, cognitive radio technology can be combined with beamforming at

[*]Jayanta Datta, 1, Section 3, Zhongxiao Xinsheng, Taipei City 10608, Taiwan, +886-0988354648, Email: jdatta1@iit.edu

https://dx.doi.org/10.25046/aj030206

the Femto-Cell network enabling it to become a self-aware entity [5, 6]. The cognitive Femto-Cell system can use spatial modulation principle to encode the location estimates of the Macro-Cell PUs and select the Femto-Cell users closest to the Femto-Cell SU receiver system. The selected SU transmitters can perform null-steering based beamforming to avoid cross-tier interference by directing nulls towards Macro-Cell PUs. In this work, the CR-IOT [7-9] system is considered as a cognitive Femto-Cell under-laid with an existing main Macro-Cell infrastructure. This kind of cognitive interference management strategy can maximize Femto-Cell CR-IOT system performance while minimizing interference on the Macro-Cell PU system, leading to better spectral coexistence between the primary Macro-Cell user and the secondary Femto-Cell CR-IOT user.

2. System Model

It is assumed that the primary Macro-Cell and secondary Femto-Cell use the same uplink and downlink frequencies and the same bandwidth. The Macro-Cell uses OFDM for transmission, while the Femto-Cell uses GFDM as its signaling scheme. Figure 1 shows the system model.

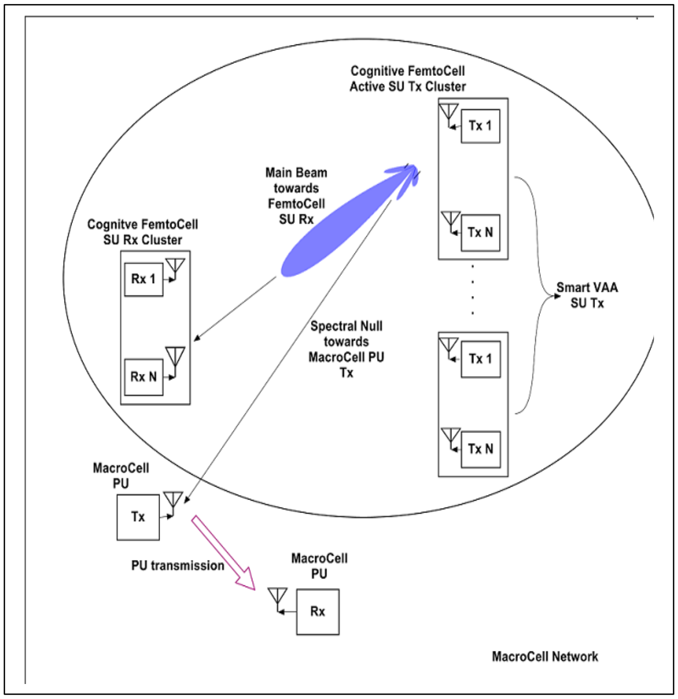


Figure 1. CR-IOT Femto-Cell VAA system under-laid with Macro-Cell PU system

The Femto-Cell based CR-IOT transmitter system is divided into multiple clusters. Each cluster contains a number of low power single antenna terminals arranged in the form of virtual antenna array (VAA) [10, 11]. The VAA based Femto-Cell CR-IOT system collects DOA information about the Macro-Cell PU and performs null-steering based beamforming to transmit its data. It uses spatial modulation principle to encode the DOA information of the PU in the form of bit blocks. This bit-block is used to identify a suitable transmit cluster which is then activated. The terminals in the active SU cluster transmit data to the SU receiver using null-steering based beamforming.

3. Transmitter Side Processing

Let there be ‘N_c’ SU transmitter clusters in total. Let there be ‘N’ terminals within each cluster. A central controller acts as the cluster head. The central controller has the role of the cognitive engine (CE). It is responsible for analyzing the received signal parameters and taking decisions on spectrum band occupancy status and estimated locations of the PUs and SUs. The transmit side processing operation can be divided into 2 phases, namely sensing phase and transmission phase.

3.1. Sensing Phase

The sensing phase operations performed at the SU transmitter side can be pictorially depicted as in Figure 2 below.

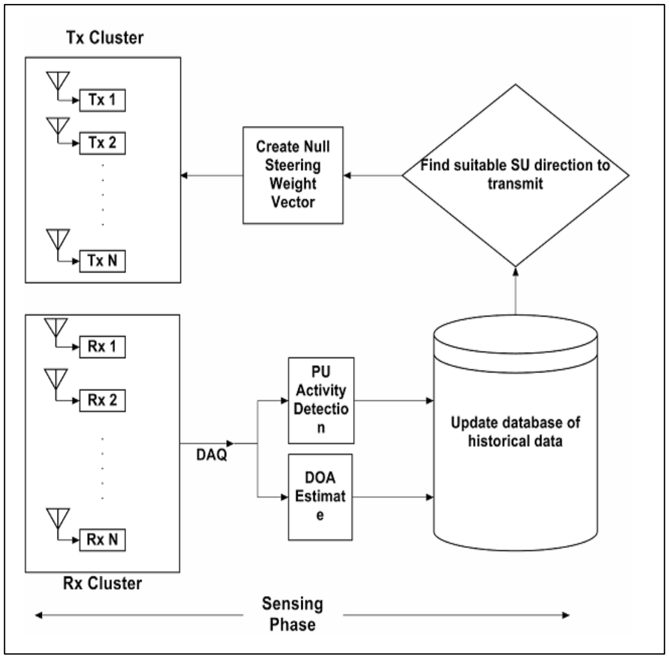

Tx Cluster
Tx 1
Tx 2
Tx N
Create Null Steering Weight Vector
Find suitable SU direction to transmit
Rx 1
Rx 2
Rx N
DAQ
PU Activity Detection
DOA Estimate
Update database of historical data
Rx Cluster
Sensing Phase

Figure 2. Sensing Phase at the Femto-Cell based CR-IOT transmitter cluster

1. PU parameters like Interference Temperature (IT), PU transmission power, available spectrum hole, PU location estimates (for e.g. DOA of received PU signal) etc. are collected and forwarded to the central controller, by the SU clusters.
2. Based on received parameters, the central controller calculates the distance of the clusters from the receiver SU. This can be achieved by localization algorithms like angle of arrival (AOA), received signal strength (RSS) estimation etc.
3. The central controller maintains a database of possible location estimates of the SU receiver clusters and PU terminals, as well as channel occupancy status within the spectrum band. The DOA estimates are matched with the database entries to verify the presence of PU activity at that location.

Based upon DOA estimation result of the received signal, the central controller creates a null-steering based beamforming weight vector. This is applied to the transmit signal during the transmission phase to steer spectral nulls towards the PU transmitter in order to avoid interference, whereas the main beam is steered towards the SU receiver cluster.

3.2. DOA Estimation by Spatial Smoothing MUSIC

SS-MUSIC [12, 13] is chosen for DOA estimation of the incoming signal under multipath propagation. Let the PU-OFDM signal be denoted by $x_{PU\_OFDM}(t)$.

$$x_{PU\_OFDM} = W^H.s \tag{1}$$

In the above equation, 'W' indicates the discrete Fourier transform (DFT) matrix. 'W^H' indicates the inverse discrete Fourier transform matrix. The term's' indicates the data subcarriers which can be *M*-ary QAM or *M*-ary PSK modulated symbols. In general for an antenna array with N terminals, the received signal at all elements of the array can be expressed by the following mathematical equation:

$$r(t) = [r_1(t)\ r_2(t) \ldots\ldots r_N(t)]^T$$

$$= \left[1\ e^{-j\omega\frac{dsinf}{c}} \ldots\ldots e^{-j\omega\frac{(N-1)dsinf}{c}}\right]^T.\left(H \otimes x_{PU_{OFDM}}(t)\right)$$

$$= a(f).y(t), where\ y(t) = (H \otimes x_{PU_{OFDM}}(t)) \tag{2}$$

Suppose the antenna array comprising 'N' terminals be divided into 'L' overlapping sub-arrays. Each sub-array contains 'M' terminals, with M>K, K being the number of source directions. Then, N=M+L-1.

Let $x_i(t)$ be the received signal vector at the i^{th} sub-array. Let $a_M(\phi)$ be the steering vector associated with each sub-array.

$$x_i(t) = e^{-j\frac{\omega(i-1)dsinf}{c}}.a_M(f).y(t)\ , i = 1,2,\ldots,M \tag{3}$$

With 'K' DOAs, the received signal '$x_i(t)$' for the *i*th sub-array can be expressed by the following mathematical equation:

$$x_i(t) = [a_M(f_1)\ldots\ldots a_M(f_K)].\begin{bmatrix} e^{-j\frac{\omega(i-1)dsinf_1}{c}} & 0 & \cdots & 0 \\ \vdots & . & \ddots & \vdots \\ 0 & . & \cdots & 0 \\ 0 & . & . & e^{-j\frac{\omega(i-1)dsinf_K}{c}} \end{bmatrix}.y(t) \tag{4}$$

$$A_M = [a_M(f_1)\ldots\ldots\ldots a_M(f_K)] \tag{5}$$

$$D(i) = \begin{bmatrix} e^{-j\frac{\omega(i-1)dsinf_1}{c}} & 0 & \cdots & 0 \\ \vdots & . & \ddots & \vdots \\ 0 & . & \cdots & 0 \\ 0 & . & . & e^{-j\frac{\omega(i-1)dsinf_K}{c}} \end{bmatrix} \tag{6}$$

$$x_i(t) = A_M.D_i.y(t) \tag{7}$$

The correlation matrix at each sub-array can be computed as follows:

$$R_{x_i} = E\{x_i(t).x_i^H(t)\}$$

$$= A_M.D_i.R_y.D_i^H.A_M^H + \sigma_0^2.I$$

$$= A_M.D_i.E\{y(t).y^H(t)\}.D_i^H.A_M^H + \sigma_0^2.I \tag{8}$$

The average of all the correlation matrices at all the sub-arrays can be expressed as follows:

$$R_{x_{avg}} = \frac{1}{L}.\sum_{i=1}^{L} E\{x_i(t).x_i^H(t)\}$$

$$= A_M.[\tfrac{1}{L}.\textstyle\sum_{i=1}^{L} D_i.R_y.D_i^H].A_M^H + \sigma_0^2.I \tag{9}$$

The conventional MUSIC [12] algorithm can be applied on the matrix 'R_{x_avg}' to obtain the DOA estimate of the incoming PU-OFDM signal. The dimension of the matrix 'A_M' is M-by-M. The dimension of the matrix within square brackets is K-by-K. Hence, it is evident that the dimension of 'R_{x_avg}' is M-by-M.

By calculating the eigen-values of R_{x_avg}, 2 disjoint eigen-spaces can be identified. One is the signal subspace, and it consists of signal eigen-vectors on which noise is overlapped. The other one is the noise subspace which comprises eigen-vectors only due to noise. The signal eigen-vectors correspond to the 'K' (with K≤M-1) greatest eigen-values. The noise eigen-vectors correspond to the other (M-K) eigen-values. The MUSIC algorithm then searches for those steering vectors which are orthonormal to the noise subspace.

Let u_i be the *i*th eigen-vector of R_{x_avg} corresponding to the eigen-value σ_i^2.

Step 1: We perform eigen-decomposition of the M-by-M matrix 'R_{x_avg}'.

$$R_{x_{avg}}.u_i = \left[A_M.R_{s_{avg}}A_M^H + \sigma_0^2.I\right].u_i = \sigma_i^2.u_i, i = 1,2,\ldots,M$$

$$where\ \sigma_i^2 > \sigma_0^2, i = 1,2,\ldots,K; \sigma_i^2 = \sigma_0^2, i = K+1,\ldots,M \tag{10}$$

Step 2: The above expression can be re-expressed as follows:

$$A_M.R_{s_{avg}}.A_M^H.u_i = \left(\sigma_i^2 - \sigma_0^2\right).u_i; i = 1,2\ldots K$$

$$= 0 \quad ; i = K+1,\ldots,M \tag{11}$$

Step 3: We partition the M-dimensional vector space into signal subspace U_s and noise subspace U_n:

$$[\boldsymbol{U_s}\,\boldsymbol{U_n}] = [\boldsymbol{u_1} \ldots\ldots \boldsymbol{u_K}\,\boldsymbol{u_{K+1}} \ldots\ldots \boldsymbol{u_M}],$$

$$\boldsymbol{where}\ \boldsymbol{U_s} = [\boldsymbol{u_1} \ldots\ldots \boldsymbol{u_K}] : \boldsymbol{\sigma_i^2}, \boldsymbol{i = 1, 2, \ldots, K}$$

$$\boldsymbol{and}\ \boldsymbol{U_n} = [\boldsymbol{u_{K+1}} \ldots\ldots \boldsymbol{u_M}] : \boldsymbol{\sigma_i^2}, \boldsymbol{i = K+1, 2, \ldots, M} \tag{12}$$

Step 4: The steering vector associated with the DOA of the PU-OFDM signal is in the signal subspace, the latter being orthogonal to the noise subspace. Hence, the MUSIC algorithm searches through all the angles 'ф' and plots the spatial spectrum given by the following equation:

$$\boldsymbol{P(f)} = \frac{\boldsymbol{1}}{\boldsymbol{a^H(f).U_n}} \tag{13}$$

Whenever the scanned angle $ф=ф_i$, the DOA of the PU-OFDM signal, the MUSIC spatial spectrum will exhibit a peak, providing an estimate of the DOA of the PU-OFDM signal.

3.3. Data Transmission Phase

This phase consists of 2 sub-phases, namely VAA cluster index modulated CR-IOT transmit signal generation and null-steering based transmit side beamforming.

i. VAA cluster index modulated CR-IOT transmit signal generation

Multiple-antenna techniques constitute a key technology for modern wireless communications, which trade-off superior error performance and higher data rates for increased system complexity and cost. The basic idea of SM-Beamforming is to map a block of information bits into information carrying units. Each information carrying unit comprises a symbol chosen from a complex constellation diagram and a unique cluster index that is chosen from a set of transmitter clusters. Each transmitter cluster contains a VAA of transmitters arranged in the form of a uniform linear array (ULA). As a consequence, we have a hybrid modulation and MIMO technique [14, 15], in which the modulated signal belongs to a tridimensional constellation diagram [15-17], which jointly combines signal and spatial information (cluster index). At the transmitter side, the bit-stream emitted from the central controller is divided into blocks containing ($\log_2 N_c + N_M.\log_2 M$) bits each, with $\log_2 N_c$ and $\log_2 M$ being the number of bits needed to identify a transmitter cluster and a transmitter symbol from the symbol constellation diagram respectively. Here, 'N_M' denotes the total number of information symbols to be transmitted for a particular time slot. Each block is then processed by a SM mapper, which splits each of them into two sub-blocks of $\log_2$ (N_c) and $N_M.\log_2$ (*M*) bits each. The spatial mapper comprises 2 separate tables, namely active cluster index (ACI) mapping table and information symbol mapping table. The ACI mapping table contains a list of the angular locations of the SU transmitter clusters, and bit combinations of size $\log_2$ (N_c) corresponding to each transmitter cluster. The incoming bit sub-block of size $\log_2$ (N_c) selects that SU transmitter cluster index which is nearest in distance to the SU receiver. It is assumed that the CE already has prior knowledge of the distances based on prior location estimation. The other bit sub-block of size $N_M.\log_2$ (*M*) selects N_M symbols according to the information symbol mapping table. The ACI mapping table is depicted in Table 1 below:

Table 1**.** Active Cluster Index (ACI) Mapping Table.

SU Tx Cluster Index	SU Tx Cluster Locations	Bit Pattern
1	30 degrees	00
2	45 degrees	01
3	60 degrees	10
4	75 degrees	11

The information symbol mapping table is shown in Table 2 below:

Table 2**.** Information Symbol Mapping Table.

Symbol	Bit Pattern
1+j	00
1-j	01
-1+j	10
-1-j	11

According to the above scheme, only one cluster of SU transmitters is activated at a particular frame instant while the other clusters remain silent. This helps to avoid the inter-cluster interference. In other SM schemes [15-18], only one transmitter antenna is active at a particular time instant, thereby transmitting only one constellation symbol per instant. However in our proposed transmission scheme, multiple information symbols are transmitted by the antenna array leading to improved spectral efficiency.

The cluster index modulated symbol stream is passed through the GFDM modulator, after which null-steering based beamforming is performed by the array of SU terminals within the activated cluster. The process can be pictorially depicted in Figure 3 below:

Details of baseband GFDM transceiver implementation can be obtained from [2].

i. *Null-Steering Transmit Side Beamforming*

The digital baseband signal snapshots $x=[x_1, x_2\ldots x_N]^T$ in a transmit beamforming cluster with N single antenna sensors (virtual antenna array) can be represented by the following equation:

$$\boldsymbol{x} = \boldsymbol{w}(\emptyset_{SU}).\boldsymbol{x}_{SU\_GFDM} \tag{14}$$

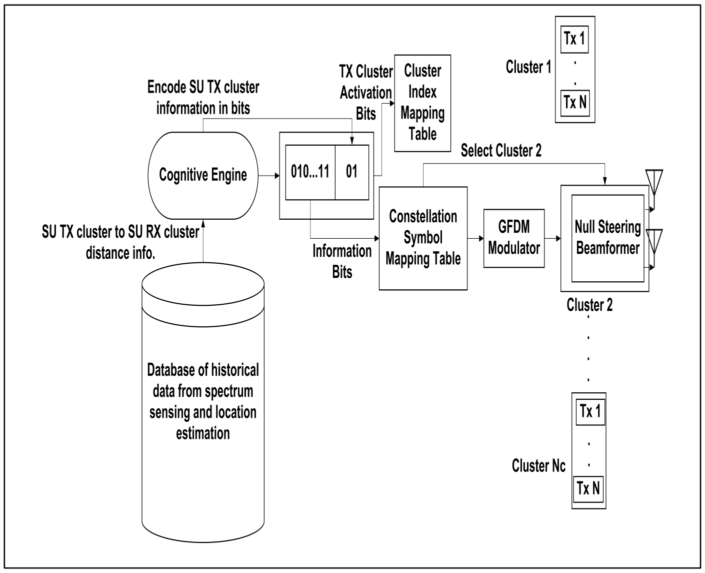

Encode SU TX cluster information in bits
TX Cluster Activation Bits
Cluster Index Mapping Table
Cluster 1
Tx 1
Tx N
Cognitive Engine
010...11
01
Select Cluster 2
SU TX cluster to SU RX cluster distance info.
Information Bits
Constellation Symbol Mapping Table
GFDM Modulator
Null Steering Beamformer
Cluster 2
Database of historical data from spectrum sensing and location estimation
Tx 1
Cluster Nc
Tx N

Figure 3. SM-Beamforming operation in CR-IOT GFDM Transmitter

In the above equation, $w(\phi_{SU})$ = $[w_1(\phi_{SU}), w_2(\phi_{SU}), \ldots, w_N(\phi_{SU})]^T$ represents the beamforming weight vector under a specific optimization criteria towards the desired direction ϕ_{SU} [19],[20]. In other words, ϕ_{SU} is the direction in which the intended SU receiver is located. 'x_{SU_GFDM}' refers to the spatially modulated symbol transmitted by the activated SU cluster. It is assumed that the sensors within each transmit cluster are closely spaced. Under this assumption, the effective channels between the different transmit antennas inside the cluster and the receiver antenna differ by phase shift included in the steering vector of the transmit cluster. The steering vector for ULA in direction 'ϕ' can be represented by the following equation [21]:

$$\boldsymbol{a}(\emptyset) = [1, e^{j.d.k.cos\emptyset}, e^{j.2d.k.cos\emptyset}, \ldots\ldots, e^{j.(N-1)d.k.cos\emptyset}]^T \quad (15)$$

In the above equation, the constant 'k'=2π/λ, where λ is the wavelength of the RF signal frequency. The distance between the elements of the cluster is expressed as'd'. The received snapshot of the signal received at a direction 'ϕ' is represented by the following equation:

$$\boldsymbol{y}(\emptyset) = \boldsymbol{a}^H(\emptyset).\boldsymbol{w}(\emptyset_{SU}).\boldsymbol{x}_{SU\_GFDM} + \boldsymbol{n} \quad (16)$$

Suppose there is 'M' number of PUs operating in the same frequency band as the SU transmitter. Let the angles of the PUs from the SU transmitter be denoted by $\phi_{PU,1}$, $\phi_{PU,2}$,…., $\phi_{PU,M}$. Let the null-steering matrix be denoted by 'A'. Then the matrix 'A' consists of the steering vectors to all the PU directions.

$$\boldsymbol{A} = [\boldsymbol{a}(\emptyset_{PU,1}), \boldsymbol{a}(\emptyset_{PU,2}), \ldots\ldots\ldots, \boldsymbol{a}(\emptyset_{PU,M})] \quad (17)$$

The N-by-N matrix, P_A, refers to the orthogonal projection matrix onto the subspace spanned by the columns of A. It is expressed mathematically as follows:

$$\boldsymbol{P}_A = \boldsymbol{A}.[\boldsymbol{A}^H.\boldsymbol{A}]^{-1}.\boldsymbol{A}^{\boldsymbol{H}} \quad (18)$$

Conventional null-steering beamforming problem can be formulated as follows [19]:

$$max_H|\boldsymbol{w}^H.\boldsymbol{a}(\emptyset_{SU})|^2 \text{ subject to } \boldsymbol{w}^H.\boldsymbol{A} = 0 \text{ and } \boldsymbol{w}^H.\boldsymbol{w} = c\,,$$
$$c: constant\ value \quad (19)$$

The solution to the above optimization problem is given as follows [19]:

$$\boldsymbol{w}(\emptyset_{SU}) = \frac{c}{||(\boldsymbol{I}-\boldsymbol{P}_A).\boldsymbol{a}(\emptyset_{SU})||}.(\boldsymbol{I}-\boldsymbol{P}_A).\boldsymbol{a}(\emptyset_{SU}) \quad (20)$$

In the above equation, 'I' refers to N-by-N identity matrix. The radiation pattern of the activated SU cluster represents the spatial response of the VAA performing the null-steering beamforming with the derived beamforming weights 'w (ϕ_{SU})'. Mathematically it can be expressed as follows:

$$\boldsymbol{R}(\emptyset) = |\boldsymbol{a}^H(\emptyset).\boldsymbol{w}(\emptyset_{SU})|^2 \quad (21)$$

4. Receiver Side Processing

4.1. Spatial Signature based Matched Filtering

In general, the wireless channel impulse response between the *i*th terminal in the *s*th SU transmitter cluster and the *j*th terminal in the receiver cluster can be represented by the following mathematical equation [22]:

$$h_{s,i,j}(t,\tau) = \sum_{l=0}^{L_i(t)} \alpha_{i,l}(t).\delta(t-\tau_{i,l}(t)) \quad (22)$$

In the above equation, '$\alpha_{i,l}(t)$' denotes the complex channel amplitude. The number of multipath components is denoted by 'L_i (t)'.The path delay of the *l*th multipath component is denoted by '$\tau_{i,l}(t)$'.Let '$H_{s,i,j}$' denote the frequency domain channel coefficients between the *i*th transmitter antenna of the *s*th SU transmitter cluster and *j*th receiver antenna.

Let '$H^{diag}{}_{s,i,j}$' denote a diagonal matrix containing the elements of '$H_{s,i,j}$' along the main diagonal.

$$H_{i,j}^{diag} = \begin{bmatrix} H_{i,j}(1) & 0 & 0 & \cdots & 0 \\ \vdots & H_{i,j}(2) & 0 & \ddots & \vdots \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & 0 & 0 & \cdots & H_{i,j}(N_{sub}) \end{bmatrix} \quad (23)$$

Let '$\mathbf{x}^i{}_{SU\_GFDM}$', i=1, 2…N_T , denote the transmitted GFDM signal from the *i*th transmitter antenna, before application of the null-steering beamforming weight vector. Its mathematical expression is same as that for baseband GFDM signal shown in [2]. Let

'$\mathbf{S}^i_{\mathrm{SU\_GFDM}}$', i=1, 2…$N_T$, denote the pre –IFFT GFDM signal as expressed below. As mentioned earlier, details of the mathematical expression for '$\mathbf{x}^i{}_{\mathrm{SU\_GFDM}}$' can be found in [2].

$$\boldsymbol{S}^{\boldsymbol{i}}_{SU_{GFDM}} = \boldsymbol{W}_{N_{sub}}.\boldsymbol{x}^{i}_{SU_{GFDM}}$$

$$= \boldsymbol{W}_{N_{sub}}.\left\{\boldsymbol{W}^{\boldsymbol{H}}_{\boldsymbol{N_{sub}}}.\sum_{k=1}^{K}\boldsymbol{P}_k\,.\boldsymbol{\Gamma}_{diag}.\boldsymbol{R}^L.\boldsymbol{W}_M.d_k\right\}$$

$$= \sum_{k=1}^{K}\boldsymbol{P}_k\,.\boldsymbol{\Gamma}_{diag}.\boldsymbol{R}^L.\boldsymbol{W}_M.d_k\,; i = 1,2,\dots,N_T \tag{24}$$

Further, '$\mathbf{S}^i_{\mathrm{SU\_GFDM}}$' can be represented by the column vector as shown below:

$$\boldsymbol{S}^i_{SU\_GFDM} = [S^i_{SU_{GFDM}}(1)\ S^i_{SU_{GFDM}}(2) \dots\dots\ S^i_{SU_{GFDM}}(N_{sub})]^T \tag{25}$$

It is assumed that the cyclic prefix (CP) is longer than the maximum delay spread of the channel. So after removal of the CP, the channel appears circular to the symbols. As such, application of FFT converts the circulant channel matrix to diagonal matrix, with frequency domain channel coefficients along the main diagonal.

The null-steering weight vector can be expressed as a column vector as follows:

$$w(\emptyset_{SU}) = [w^{(1)}(\emptyset_{SU})\ w^{(2)}(\emptyset_{SU}) \dots\dots..w^{(N_T)}(\emptyset_{SU})]^T \tag{26}$$

The received signal at the *j*th receiver antenna can be expressed as follows:

$$\boldsymbol{y}_j = \left[\boldsymbol{H}^{diag}_{1,j}\ \boldsymbol{H}^{diag}_{2,j} \dots\dots..\boldsymbol{H}^{diag}_{N_T,j}\right].\begin{bmatrix} w^{(1)}(\emptyset_{SU}).\boldsymbol{S}^1_{SU\_GFDM} \\ w^{(2)}(\emptyset_{SU}).\boldsymbol{S}^2_{SU\_GFDM} \\ \vdots \\ w^{(N_T)}(\emptyset_{SU}).\boldsymbol{S}^{N_T}_{SU\_GFDM} \end{bmatrix} + \boldsymbol{N} \tag{27}$$

The above equation can be re-expressed as:

$$y_j = \left(\boldsymbol{w}(\emptyset_{SU}) \odot \left[\boldsymbol{H}^{diag}_{1,j}\ \boldsymbol{H}^{diag}_{2,j} \dots..\boldsymbol{H}^{diag}_{N_T,j}\right]\right).\begin{bmatrix} \boldsymbol{S}^1_{SU\_GFDM} \\ \boldsymbol{S}^2_{SU\_GFDM} \\ \vdots \\ \boldsymbol{S}^{N_T}_{SU\_GFDM} \end{bmatrix} + \boldsymbol{N}$$

$$= \left(\boldsymbol{w}(\emptyset_{SU}) \odot \left[\boldsymbol{H}^{diag}_{1,j}\ \boldsymbol{H}^{diag}_{2,j} \dots..\boldsymbol{H}^{diag}_{N_T,j}\right]\right).\boldsymbol{S}^{all}_{SU\_GFDM} + \boldsymbol{N}$$

$$\boldsymbol{where}\ \boldsymbol{S}^{all}_{SU\_GFDM} = [\boldsymbol{S}^1_{SU\_GFDM}\ \boldsymbol{S}^2_{SU\_GFDM} \dots\dots \boldsymbol{S}^{N_T}_{SU\_GFDM}]^T \tag{28}$$

In (28), the sign 'Θ' indicates element-wise multiplication.

Let 'b_j' denote the spatial signature of the j^{th} receiver antenna.

$$b_j = \boldsymbol{w}(\emptyset_{SU}) \odot [\boldsymbol{H}^{diag}_{1,j}\ \boldsymbol{H}^{diag}_{2,j} \dots..\boldsymbol{H}^{diag}_{N_T,j}] \tag{29}$$

Let '**b**' denote the matrix of spatial signatures at all the receiver antennas.

$$\boldsymbol{b} = [b_1\ b_2 \dots\dots.b_{N_R}]^T \tag{30}$$

Then, the received signal at 'N_R' receiver antennas can be expressed in terms of the transmitted signal and spatial signature associated with each receiver antenna as follows:

$$\boldsymbol{y} = [y_1\ y_2 \dots\dots y_{N_R}]^T = \boldsymbol{b}.\boldsymbol{S}^{all}_{SU\_GFDM} + \boldsymbol{N} \tag{31}$$

It is assumed that the channel is temporally slowly fading and that each post-FFT sample of the received GFDM signal encounters independent and identical Rayleigh distribution. Moreover, the channel stays constant over a number of symbols and the receiver has perfect knowledge of the slowly varying channel. The receiver computes the pseudo-inverse of the spatial signature matrix '**b**' to recover the transmitted spatially modulated GFDM signal.

4.2. Active Transmitter Cluster Index Estimation

While it is assumed that the receiver has perfect knowledge of the spatial signatures, the receiver does not have knowledge which transmitter cluster is active for that particular frame instant. The receiver performs spatial signature based matched filtering on the received signal using spatial signature belonging to each SU transmitter cluster and stores the result in vector form in a buffer. Next, the receiver computes the squared absolute value of each stored vector in the buffer and compares the result to find out the maximum value. The index of the maximum value gives an estimate of the active transmitter SU cluster.

The procedure is tabulated in Table 3 below:

Table 3. Spatial Signature based Matched filtering

Inputs: Received Signal: y, Spatial Signature of s^{th} transmitter cluster: b_s ; s=1,2,……..,N_c Output: Index of Active Transmit SU cluster: s
(i) $\boldsymbol{y}^{matched}_s = (\boldsymbol{b}^H_s.\boldsymbol{b}_s)^{-1}.\boldsymbol{b}^H_s.\boldsymbol{y}, s = 1,2 \dots., N_c$

(ii) $a_s = ||\boldsymbol{y}_s^{matched}||^2 , s = 1,2, \dots , N_c$

(iii) $s = \arg max_s \, a_s, s = 1,2, \dots . , N_c$

5. Simulation Results and Analysis

In this section, simulation results are provided to evaluate the performance of the proposed system under multipath Rayleigh fading channel environment. It is assumed that the clusters are closely spaced. Inter-element distance within each cluster is d=0.5λ, where 'λ' is the wavelength. Cluster elements are arranged in the form of ULA. The center frequency of operation of both the CR-IOT Femto-Cell and primary Macro-Cell is f_c=2.3 GHz. The number of transmitter and receiver terminals within each cluster in the CR-IOT Femto-Cell network is 16. The multicarrier signaling scheme chosen for the Femto-Cell CR-IOT system is GFDM, which uses *M*-ary QAM constellation. The Macro-Cell PU system uses OFDM waveform, which also uses *M*-ary QAM constellation. The PU-OFDM transmitter is assumed to be at 45 degrees.

The following table shows the specifications of the Femto-Cell CR-IOT GFDM waveform.

Table 4. GFDM system parameters

Simulation Parameters for SM-beamformer GFDM system	
Total number of time slots	5
Active Subcarrier Indices	100:200
Pulse Shaping Filter	RRC
RRC Overlap Factor	0.1
Number of SU Tx clusters	4
Number of SU Tx terminals per cluster	16
Number of SU Rx clusters	1
Number of SU Rx terminals per cluster	16

In Figure 4 showing antenna array pattern, null is steered at 45 degrees which is the Macro-Cell PU-OFDM transmitter's location. The SS-MUSIC algorithm provides the DOA estimate of the PU-OFDM transmitter, which helps to steer null towards the PU location. As evident from Figure 4, the red dotted line at 45 degrees indicates the direction of the Macro-Cell PU-OFDM transmitter.

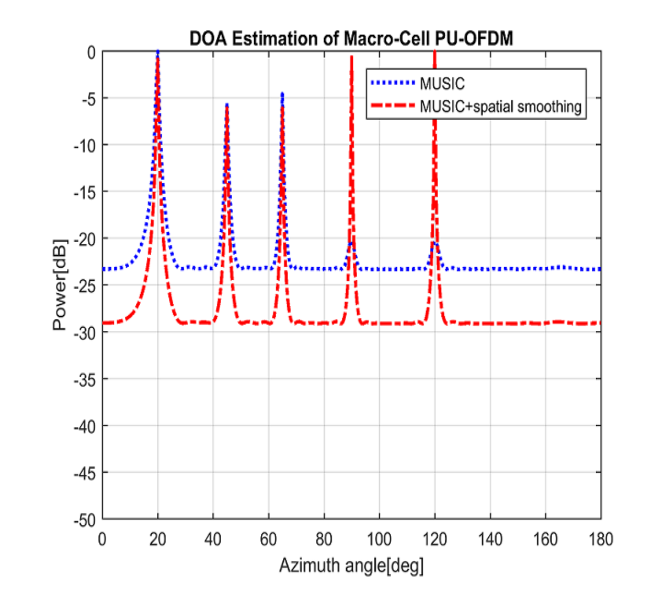


Figure 4. Cognitive Null Steering towards Macro-Cell PU-OFDM user

SS-MUSIC is preferred over conventional MUSIC primarily because of the fact that under multipath propagation, the latter does not perform well in identifying all the sources. This is because under multipath propagation environment, 2 or more DOAs might belong to the same source .i.e. for 'K' number of DOAs, actually the number of sources might be less than 'K' under multipath environment.

Figure 5 shows DOA estimation performance comparison between SS-MUSIC and conventional MUSIC for 2 sources arriving at the receiver from different directions, under multipath propagation conditions.

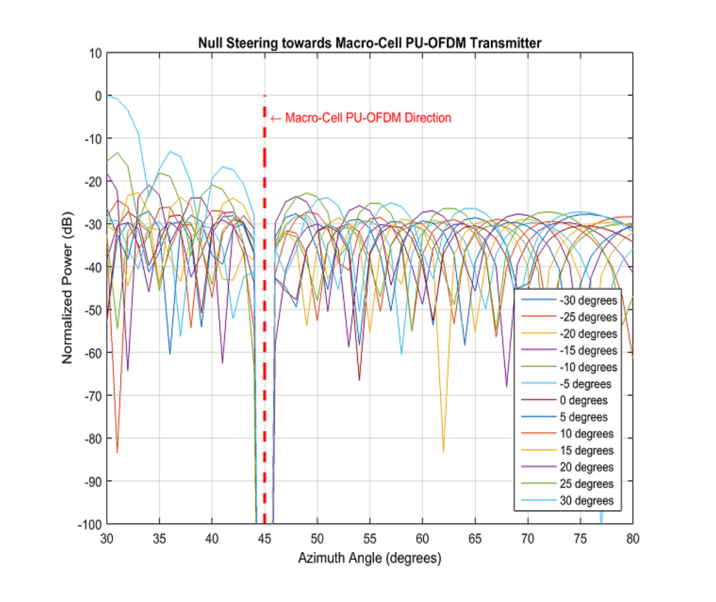


Figure 5. SS-MUSIC vs conventional MUSIC

As evident from Figure 5, for signals with multipath components too close to the main direction, the estimation capability of SS-MUSIC is better due to the application of the smoothing procedure. In the above figure, resolution of the main DOA (45°) and the multipath DOAs (20°, 65°, 90° and 120°) by the SS-MUSIC is better than that with MUSIC.

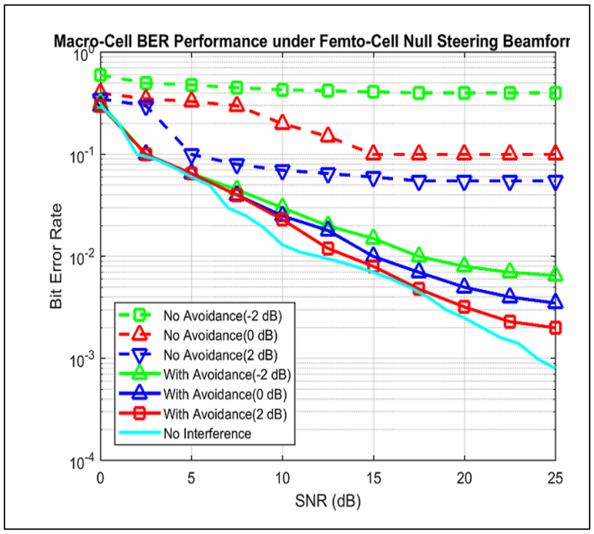


Figure 6. BER performance of Macro-Cell based PU-OFDM

Figure 6 compares the BER curves as a function of SNR for different signal-to-interference ratio (SIR) values for the Macro-Cell PU-OFDM user under conditions of no interference avoidance and with interference avoidance. As evident from the above figure, the Macro-Cell PU suffers significant performance loss under no interference avoidance scenario; however considerable performance improvement is achieved with the application of the cognitive null-steering based beamforming at the Femto-Cell CR-IOT GFDM transmitter.

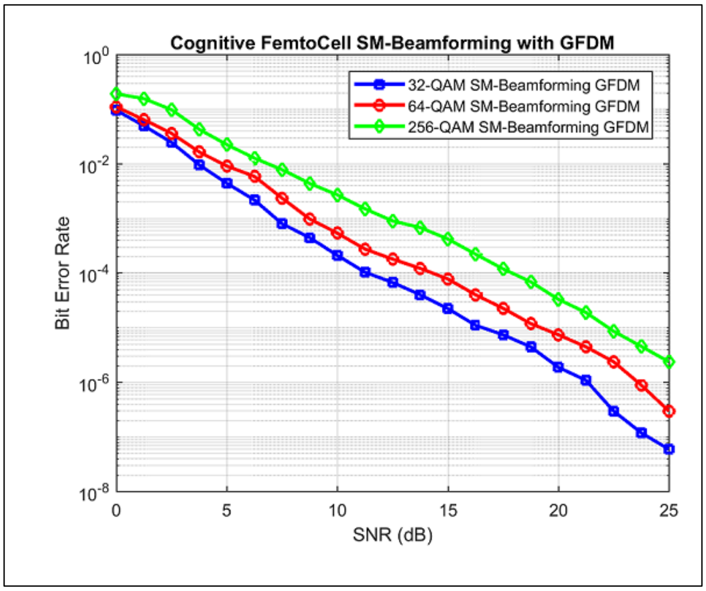


Figure 7. BER performance of Femto-Cell based CR-IOT GFDM

The proposed SM-beamforming system is very well capable of avoiding interference to the PU system for signal-to-noise ratio (SNR) less than or equal to 15 dB. At higher SNR values, it is observed that there is presence of residual interference. This happens due to imperfect steering of nulls at higher SNRs, for different SIR values.

Figure 7 shows the BER performance of the Femto-Cell based CR-IOT GFDM employing SM-beamforming scheme. As expected, BER for 256-QAM system is higher than that of 32-QAM and 64-QAM system under multipath Rayleigh fading channel conditions.

6. Conclusion

This paper introduces the application of SM based null-steering beamforming in VAA based CR-IOT system, under multipath fading channel conditions. Computer simulation results show that the proposed scheme can lead to good BER performance of the PU system by appropriate design of the null-steering beam-former. Moreover, the application of beamforming at the spatial modulation based SU transmitter also provides good BER performance of the SU system under multipath channel conditions. Spatial Smoothing based MUSIC is selected for DOA estimation due to its superiority over conventional MUSIC under multipath propagation. However, performance analysis of other DOA estimation algorithms needs to be evaluated for multicarrier signaling case. As a future work, transmit-receive beamforming schemes with DOA estimation will be investigated under full duplex spatial modulation based transmission conditions.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by National Taipei University of Technology, Taiwan.

References

[1] J. Datta, H.P. Lin, D.B.Lin, "Spatial modulation based location aware beam-forming in GFDM modulated cognitive radio systems", in International Conference on Applied System Innovation, Sapporo Japan, 2017. https://doi.org/ 10.1109/ICASI.2017.7988357

[2] I. Gaspar, N. Michailow, A. Navarro, E. Ohlmer, S. Krone, G. Fettweis, "Low Complexity GFDM Receiver Based on Sparse Frequency Domain Processing" in Proceedings of the IEEE 77th Vehicular Technology Conference, Dresden Germany, 2013. https://doi.org/ 10.1109/VTCSpring.2013.6692619

[3] N. Michailow, M. Matthé, I.S. Gaspar, A. Navarro, L.L. Mendes, A. Festag, "Generalized Frequency Division Multiplexing for 5th Generation Cellular Networks" IEEE Trans. Comm.., **62(9)**, 3045-3061, 2014. https://doi.org/10.1109/TCOMM.2014.2345566

[4] N. Michailow, I. Gaspar, S. Krone, M. Lentmaier, G. Fettweis, "Generalized frequency division multiplexing: Analysis of an alternative multi-carrier technique for next generation cellular systems" in Proceedings of the IEEE International Symposium on Wireless Communication Systems, Paris France, 2012. https://doi.org/ 10.1109/ISWCS.2012.6328352

[5] G. Gur, S. Bayhan, F. Alagoz, "Cognitive femtocell networks: an overlay architecture for localized dynamic spectrum access." IEEE Mag. Wireless Comm, **17**(4), 62-70, 2010. https://doi.org/ 10.1109/MWC.2010.5547923

[6] M.Z. Shakir, R. Atat, M.S. Alouini, "On the interference suppression capabilities of cognitive enabled femtocellular networks", Proceedings of the IEEE International Conference on Communications and Information

Technology, Hammamet Tunisia, 2012. https://doi.org/ 10.1109/ICCITechnol.2012.6285835

[7] P. Rawat, K.D. Singh, J.M. Bonnin, “Cognitive radio for M2M and Internet of Things: A survey”, Comp Comm., **94**, 1-29, 2016. https://doi.org/10.1016/j.comcom.2016.07.012

[8] K. Katzis, H. Ahmadi, “Challenges Implementing Internet of Things (IoT) Using Cognitive Radio Capabilities in 5G Mobile Networks”, Internet of Things (IoT) in 5G Mobile Technologies, **8**, 55-76, 2016.

[9] R. Fantacci, D. Marabissi, “Cognitive Spectrum Sharing: An Enabling Wireless Communication Technology for a Wide Use of Smart Systems”, Future Internet, **8**, 23, 2016. https:// doi:10.3390/fi8020023

[10] M. Dohler, E. Lefranc, H. Aghvami, “Space-time block codes for virtual antenna arrays”, Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Lisboa, Portugal, 2002. https://doi.org/ 10.1109/PIMRC.2002.1046733

[11] M. Dohler, E. Lefranc, H. Aghvami, “Link capacity analysis for virtual antenna arrays”, Proceedings of the IEEE 56th Vehicular Technology Conference, Vancouver, Canada, 2002. https://doi.org/ 10.1109/VETECF.2002.1040381

[12] B.D. Rao, K.V.S. Hari, “MUSIC and spatial smoothing: A statistical performance analysis”, Proceedings of the Twenty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 1989. https://doi.org/ 10.1109/ACSSC.1989.1201035

[13] A. Paulraj, V.U. Reddy, T.J. Shan, T. Kailath, “Performance Analysis of the Music Algorithm with Spatial Smoothing in the Presence of Coherent Sources”, Proceedings of the IEEE Military Communications Conference, Monterey, USA, 1986. https://doi.org/ 10.1109/MILCOM.1986.4805849

[14] R, Schmidt, “Multiple emitter location and signal parameter estimation”, IEEE Trans. Antennas and Propagation, **34** (3), 276-280, 1986. 10.1109/TAP.1986.1143830

[15] M. D. Renzo, H. Haas, P.M. Grant, “Spatial modulation for multiple-antenna wireless systems: a survey”, IEEE Communications Mag., **49** (12), 182-191, **2011**. https://doi.org/10.1109/MCOM.2011.6094024

[16] M.D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, L. Hanzo, “Spatial Modulation for Generalized MIMO: Challenges, Opportunities, and Implementation”, Proceedings of the IEEE, **102** (1), 56-103, 2014. https://doi.org/ 10.1109/JPROC.2013.2287851

[17] R. Mesleh, H. Haas, C.W. Ahn, S. Yun, “Spatial Modulation-A low Complexity Spectral Efficiency Enhancing Technique”, in Proceedings of First International Conference on Communications and Networking in China, Beijing, China, 2006. https://doi.org/ 10.1109/CHINACOM.2006.344658

[18] J. Jeganathan, A. Ghrayeb, L. Szczecinski, “Spatial Modulation: Optimal Detection and Performance Analysis”, IEEE Communications Letters, **12**(8), 545-547, 2008. https://doi.org/ 10.1109/LCOMM.2008.080739

[19] K. Zarifi, S. Affes, A. Ghrayeb, “Collaborative Null-Steering Beamforming for Uniformly Distributed Wireless Sensor Networks”, IEEE Trans. Signal Process. **58** (3), 1889-1903, 2010. https://doi.org/ 10.1109/TSP.2009.2036476

[20] B. Friedlander, B. Porat, “Performance analysis of a null-steering algorithm based on direction-of-arrival estimation”, IEEE Trans. Acoust., Speech, Signal Process. **37** (4), 461 –466, 1989. https://doi.org/ 10.1109/29.17526

[21] L. Godara, “Application of antenna arrays to mobile communications. II. beam-forming and direction-of-arrival considerations”, Proc. IEEE **85** (8), 1195 –1245, 1997. https://doi.org/ 10.1109/5.622504

[22] R.B. Ertel, P. Cardieri, K.W. Sowerby, T.S. Rappaport, J.H. Reed, “ Overview of spatial channel models for antenna array communication systems”, IEEE Pers. Comm., **5** (1), 10-22, 1998. https://doi.org/ 10.1109/98.656151

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 58-68 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Enhanced Outdoor to Indoor Propagation Models and Impact of Different Ray Tracing Approaches at Higher Frequencies

Muhammad Usman Sheikh[*,1], Kimmo Hiltunen[2], Jukka Lempiainen[3]

[1]Tampere University of Technology, Department of Electronics and Communications Engineering, Finland.

[2]Ericsson Research, Helsinki, Finland.

[3]Tampere University of Technology, Department of Electronics and Communications Engineering, Finland.

ARTICLE INFO

Article history:
Received: 26 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Outdoor to Indoor
Propagation
Angular loss
Wall loss model
Building penetration loss
Ray tracing

ABSTRACT

The main target of this article is to study the provision of indoor service (coverage) using outdoor base station at higher frequencies i.e. 10 GHz, 30 GHz and 60 GHz. In an outdoor to indoor propagation, an angular wall loss model is used in the General Building Penetration (GBP) model for estimating the additional loss at the intercept point of the building exterior wall. A novel angular wall loss model based on a separate incidence angle in azimuth and elevation plane is proposed in this paper. In the second part of this study, an Extended Building Penetration (EBP) model is proposed, and the performance of EBP model is compared with the GBP model. In EBP model, the additional fifth path known as the "Direct path" is proposed to be included in the GBP model. Based on the evaluation results, the impact of the direct path is found significant for the indoor users having the same or closed by height as that of the height of the transmitter. For the indoor users located far away from the exterior wall of building, a modified and enhanced approach of ray tracing type is proposed in this article. In the light of acquired simulation results, the impact of a modified ray tracing approach is emphasized.

1. Introduction

This article is an extension of research work originally presented at International Wireless Communication and Mobile Computing (IWCMC'17) conference [1]. In reference [1], studies were made at 10 GHz only; whereas in this article the impact of different propagation models is also analyzed at 30 GHz and 60 GHz. Additionally, two different ray tracing approaches are also analyzed in this article.

Outdoor to indoor propagation in a small cell environment involves Line of Sight (LOS) and Non-LOS (NLOS) propagation. It includes path loss computation, determination of reflection and diffraction loss, penetration loss, and other indoor losses. The penetration loss can be divided into four major categories i.e. building exterior wall penetration loss also known as building penetration loss, floor (ceiling) penetration loss, indoor propagation (room penetration) loss, and angular wall loss [2]. Several propagation models are given in literature e.g. General Building Penetration (GBP) model provides and evaluates the candidate paths, Berg's recursive method for micro cell path loss [3], and the linear attenuation model for indoor propagation. The frequency dependent penetration loss models are presented at [4].

The angular wall loss model presented in [2] and used in [4] depends on a single three-dimensional incidence angle. This paper presents a new angular wall loss model based on a separate incidence angle in azimuth and elevation plane. An Extended Building Penetration (EBP) model is proposed in this article. Path gain is used as a metric to compare the performance of different angular wall loss models and building penetration loss models. Section II provides the details about the GBP model, frequency

[*] Muhammad Usman Sheikh, Tampere University of Technology
Email: muhammad.sheikh@tut.fi

https://dx.doi.org/10.25046/aj030207

dependent penetration models, angular wall loss model, and ray tracing. Whereas Section III explains the proposed angular wall loss model and extended building penetration model. Section IV gives the description of the simulation environment and provides the details about the assumptions and simulation parameters. Section IV discusses the results in detail. Finally, Section V concludes the paper.

2. Background Theory

2.1. General Building Penetration (GBP) Model

The General Building Penetration (GBP) model is based on the COST231 building penetration model presented at [2]. The model considers a path from each exterior wall of the building. A top view of a single building scenario with a Transmitter (TX) located in an outdoor environment is shown in Figure 1. For each Receiver (RX) point located inside the building, there exist four candidate paths. A path coming from the front face of the building without any diffraction is called the "LOS path", and the paths which are reaching the receiver point after diffracting from the corners of the building are known as "NLOS paths". For the NLOS paths, the Berg's recursive model is used to capture the additional path loss due to the diffraction around the building, before entering the building, whereas the outdoor to indoor and other indoor losses are determined from the COST231 building penetration model as done in [4]. Each multipath has two propagation parts; the first part comprises free space propagation between the transmitter and the building's exterior wall intercept point. The second part is the propagation of a path in an indoor environment after penetrating through the exterior wall to the receiver point.

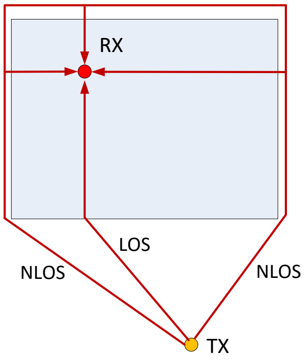


Figure 1. Illustration of propagation paths from outdoor base station to indoor location (top view).

A well-known Friis free space path loss model is used to find the path loss in the first free space propagation part. For the second part of the propagation path, along with the free space path loss model, the building penetration loss model, an indoor wall loss model, and an angular wall loss model is used to compute the additional loss due to indoor propagation. The received signal strength at the receiver location inside the building is the sum of four paths as shown in Figure 1.

Figure 2 shows the side view of the paths entering the building for the two receiver points RX1 and RX2 located on the Floor 1 and Floor 4, respectively. The signal path intercepts the building wall at the same height as that of a receiver height; therefore the ceiling penetration loss is not taken into account. It is important here to mention that in GBP model both the LOS path and NLOS paths do not penetrate through the ceiling.

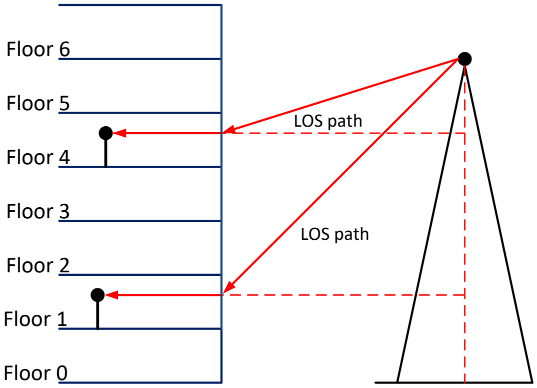


Figure 2. Illustration of LOS paths entering the building (side view).

2.2. Building Penetration Loss (BPL)

The signal experiences a penetration loss while penetrating from the outdoor environment to the building. Outdoor to indoor penetration loss is generally termed as Building Penetration Loss (BPL). The building penetration loss relies heavily on the frequency and on the material characteristics of the building; therefore the BPL can be significantly different for different material types at different frequencies. Generally, the old houses are composed of plane standard glass windows and concrete wall, while the Infrared Reflective (IRR) glass windows are commonly used in the new modern energy saving houses. In reference [4], the old buildings are assumed to have 30 % of the standard glass windows and 70 % of the concrete wall. Similarly, the assumption for new modern building type corresponds to the 70 % of the IRR glass windows and 30 % of the concrete wall.

A simple model structure has been proposed in [4] to model a single material frequency dependent penetration loss. The penetration loss for different material types is provided at references [5-9].

$$\mathrm{L_{Single\ glass,dB}} = 0.1 * \mathrm{Frequency_{GHz}} + 1, \tag{1}$$

$$\mathrm{L_{Double\ glass,dB}} = 0.2 * \mathrm{Frequency_{GHz}} + 2, \tag{2}$$

$$\mathrm{L_{IRR\ glass,dB}} = 0.3 * \mathrm{Frequency_{GHz}} + 23, \tag{3}$$

The penetration loss for the concrete wall is modeled as

$$\mathrm{L_{Concrete,dB}} = 4 * \mathrm{Frequency_{GHz}} + 5, \tag{4}$$

As the buildings are composite of windows and concrete wall, the building penetration loss for old buildings and new buildings is modeled as shown in (5) and (6), respectively [10].

$$\mathrm{L_{Old\ building,dB}} = -10\mathrm{Log_{10}}\left[0.3 * 10^{\frac{-\mathrm{L_{Double\ glass,dB}}}{10}} + 0.7 * 10^{\frac{-\mathrm{L_{Concrete,dB}}}{10}}\right], \tag{5}$$

$$\mathrm{L_{New\ building,dB}} = -10\mathrm{Log_{10}}\left[0.7 * 10^{\frac{-\mathrm{L_{IRR\ glass,dB}}}{10}} + 0.3 * 10^{\frac{-\mathrm{L_{Concrete,dB}}}{10}}\right], \quad (6)$$

The building penetration loss as a function of frequency for different types of building is shown in Figure 3.

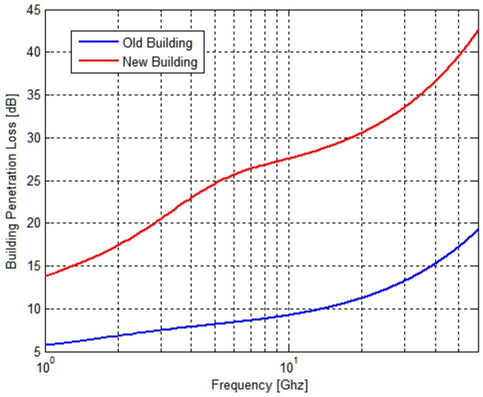


Figure 3. Building penetration loss as a function of frequency.

2.3. Indoor Propagation Loss

In an indoor environment, generally the indoor walls are made up of standard glass alternatively plaster. In [4], two different indoor wall loss models are presented as a function of the frequency assuming an average wall distance of 4 m. The Indoor Loss Model 1 assumes an indoor wall of standard glass, whereas Indoor Loss Model 2 is based on the measurements performed in [5]. Two indoor wall loss models are modeled as shown in (7) and (8).

$$\mathrm{L^{(1)}_{Wall\ loss,dB/m}} = \mathrm{L_{Single\ glass,dB}}, \quad (7)$$

$$\mathrm{L^{(2)}_{Wall\ loss,dB/m}} = 0.2 * \mathrm{Frequency_{GHz}} + 1.7, \quad (8)$$

Indoor loss as a function of frequency for two different indoor wall loss models, expressed as db/m is shown in Figure 4.

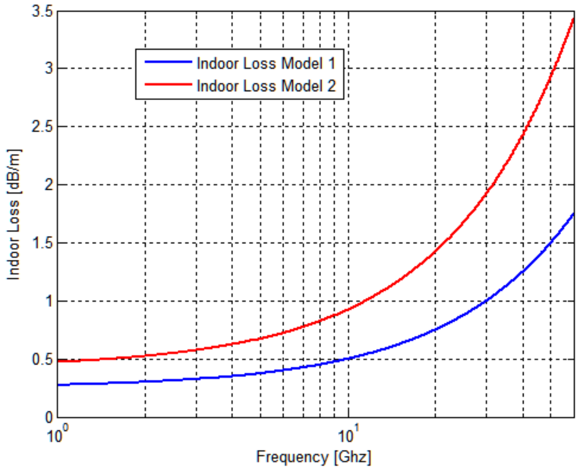


Figure 4. Indoor loss models as a function of frequency.

2.4. Body Loss as Function of Frequency

In reference [10], the frequency dependency of the body loss is modeled as given in (9).

$$\mathrm{L_{Body,dB}} = \frac{\mathrm{Frequency_{GHz}}}{60} + 3, \quad (9)$$

Equation (9) shows that frequency has a negligible impact on a considered body loss model at 10 GHz or lower frequencies, as traditionally the body loss is assumed to be 3 dB. However, equation (9) gives additional 0.5 dB and 1 dB body loss at 30 GHz and 60 GHz, respectively.

2.5. Angular Wall Loss Model Based on a Single Three-dimensional Incidence Angle

In addition to the building penetration loss and indoor wall loss, there exists an angular wall loss. An angular wall loss model presented at [2, 4] is used to include the angular loss that can be experienced at the building's exterior wall intercept point. The angular wall loss model is given by (10),

$$\mathrm{L_{angular,dB}} = 20 * [1 - \mathrm{Cos}(\theta_i)]^2, \quad (10)$$

where θ_i is the single three-dimensional incidence angle.

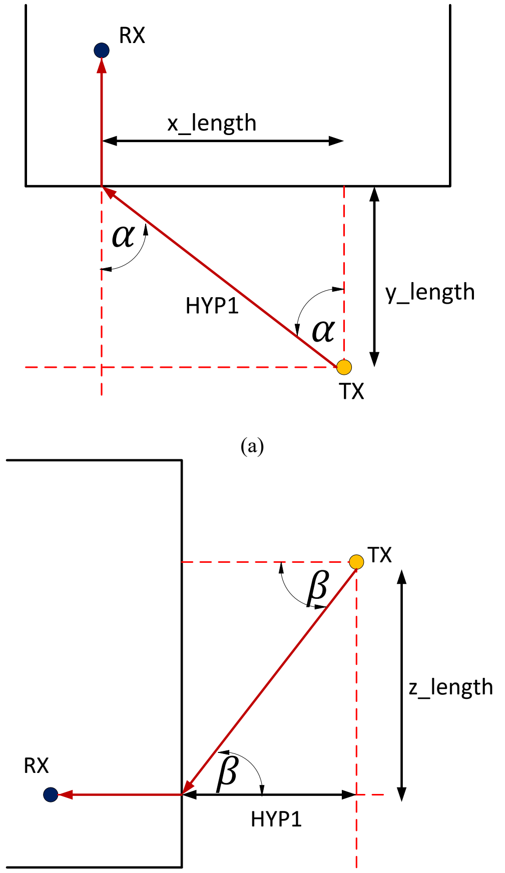


Figure 5. Illustration of incidence angles at intercept point on wall in, a) Azimuth (horizontal) plane, and b) Elevation (vertical plane).

To understand the geometry and the computation of θ_i, consider an example scenario in an azimuth plane (top view) as shown in Figure 5(a). The receiver point is located inside the building and the transmitter is located in an outdoor environment. The distance between the transmitter and the receiver point along the x-axis and the y-axis is given by x_length and y_length, respectively. HYP1 is the distance between the transmitter and the incidence point on the wall, and is given in (11).

$$\mathrm{HYP1} = \sqrt{\mathrm{x\_length}^2 + \mathrm{y\_length}^2}, \qquad (11)$$

In Figure 5(b), z_length is the difference of height between the transmitter and the receiver point, and HYP2 is the three dimensional distance between the transmitter and the wall intercept point. HYP2 is calculated as follows:

$$\mathrm{HYP2} = \sqrt{\mathrm{HYP1}^2 + \mathrm{z\_length}^2}, \qquad (12)$$

The single three-dimensional incidence angle θ_i can be computed as

$$\theta_{\mathrm{i}} = \cos^{-1}\left(\frac{\mathrm{y\_length}}{\mathrm{HYP2}}\right), \qquad (13)$$

Considering the geometry shown in Figure 5, the equation given in (13) can also be re-written in a simplified form as

$$\mathrm{L}_{\mathrm{angular,dB}} = 20 * \left[1 - \frac{\mathrm{y\_length}}{\mathrm{HYP2}}\right]^2, \qquad (14)$$

2.6. Ray Tracing (RT)

Ray tracing techniques can be used for estimating the received signal level and for the characterization of the radio propagation environment. Finding the multipath components between the transmitter and receiver is the first step towards the computation of the received electric field or power at the receiver point. By using Image Theory (IT) algorithm, all multipaths with the given finite number of reflections and diffractions can theoretically be found between the transmitter and receiver. An Image theory algorithm shows a high level of accuracy and precision. Determination of multipath components by image based ray tracing technique may require large computation time. The complexity and the computational time of the ray tracing algorithm increases with the increase in number of supported reflections and diffractions [11-13]. Reflection losses are determined by reflection coefficients. The reflection coefficient depends upon the polarization and on the material permittivity. For perpendicular and parallel polarization, the reflection coefficients are given in (15) and (16), respectively. β is the angle between the incident ray and the reflected surface, and ε_r is the material permittivity of the reflecting surface.

$$|\Gamma_{\perp}| = \frac{\mathrm{Sin}(\beta) - \sqrt{\varepsilon_r - \mathrm{Cos}^2(\beta)}}{\mathrm{Sin}(\beta) + \sqrt{\varepsilon_r - \mathrm{Cos}^2(\beta)}} \qquad (15)$$

$$|\Gamma_{\parallel}| = \frac{-\varepsilon_r\,\mathrm{Sin}(\beta) + \sqrt{\varepsilon_r - \mathrm{Cos}^2(\beta)}}{\varepsilon_r\,\mathrm{Sin}(\beta) + \sqrt{\varepsilon_r - \mathrm{Cos}^2(\beta)}} \qquad (16)$$

There are several ways for computing the diffraction loss; however a recursive method proposed in [14] is used in this article to compute the diffraction loss.

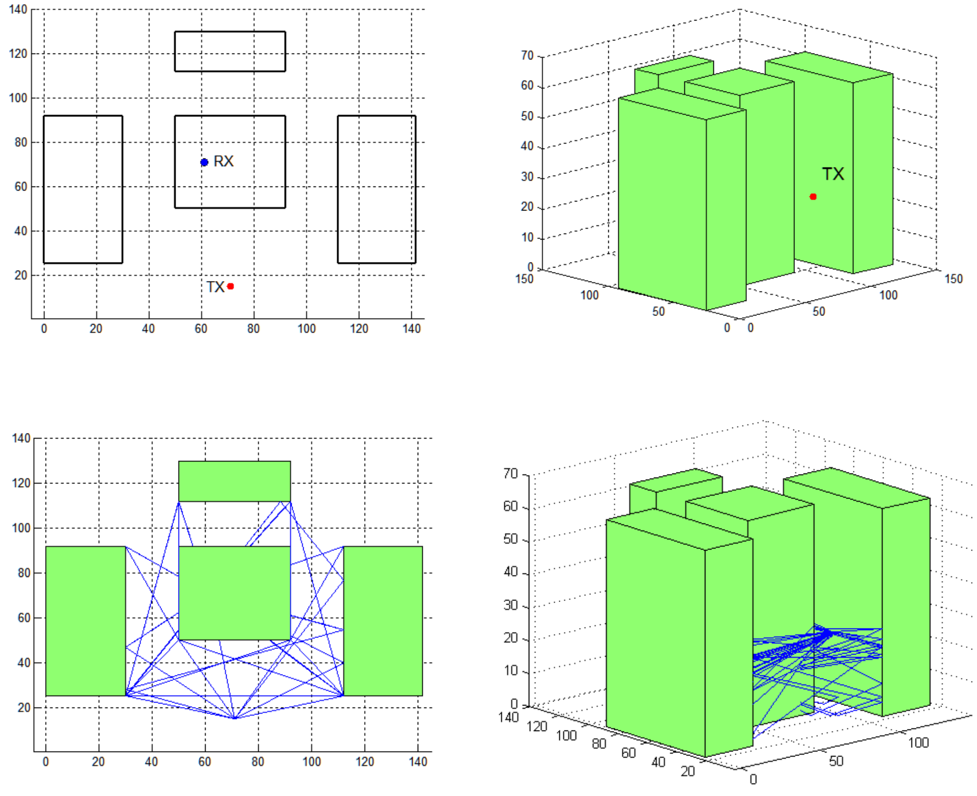


Figure 6. Illustration of ray tracing, (a) Two dimensional view of sample building scenario, (b) Three dimensional view of sample building scenario, (c) Two dimensional view of ray tracing, and (d) Three dimensional view of ray tracing.

Figure 6(a) shows the sample building scenario in two-dimensional environment with transmitter and receiver locations marked with TX and RX, respectively. Figure 6(b) shows the three-dimensional view of a sample scenario with TX height at 31.5 meter. The receiver point is located inside the central building at the height of 10.5 m. Figure 6(c) illustrates the ray tracing in two-dimensional view, whereas Figure 6(d) shows the three-dimensional view of ray tracing. Blue lines show different propagation paths with finite number of reflections and diffractions, and LOS path.

3. Proposed Models

3.1. Angular Wall Loss Model based on Separate Incidence Angle in Azimuth and Elevation Plane

In this article, a novel angular wall loss model based on a separate incidence angle in azimuth and elevation plane is proposed. It is proposed to compute the total angular loss as a sum of the angular loss in an azimuth and elevation plane. Unlike the single three-dimensional angle, in this proposed approach the incidence angle in the horizontal plane and vertical plane should be computed separately. The azimuth angular loss and elevation angular loss is given in (17) and (18), respectively.

$$L_{\text{Azimuth ang,dB}} = 10 * [1 - \text{Cos}(\alpha)]^2, \qquad (17)$$

$$L_{\text{Elevation ang,dB}} = 10 * [1 - \text{Cos}(\beta)]^2 \qquad (18)$$

The angles α and β used in (17) and (18), respectively, are shown in Figure 5. The angles α and β are the incidence angles computed separately in an azimuth and elevation plane. Now, the total angular loss is given by (19).

$$L_{\text{angular,dB}} = L_{\text{Azimuth ang,dB}} + L_{\text{Elevation ang,dB}} \qquad (19)$$

While considering a fixed x_length of 50 m for a geometry shown in Figure 5, the Figure 7 shows the angular loss based on a single three-dimensional incidence angle as a function of y_length and z_length.

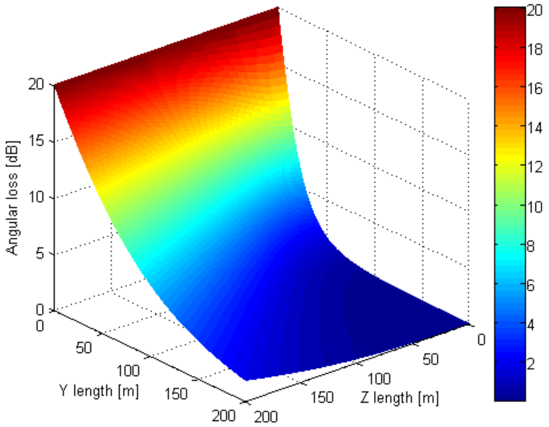


Figure 7. The angular loss based on single 3D incidence angle.

Figure 8 shows the angular loss based on a separate incidence angle in azimuth and elevation plane. The first approach of the angular wall loss model based on a single three-dimensional incidence angle gives higher angular loss at the lower values of y_length and z_length. The angular wall loss model based on a separate incidence angle in azimuth and elevation plane considers the change of angle of an incident ray in both domains separately, and therefore it can be seen in Figure 8 that the angular loss increases with the increase in z_length which is not the case in Figure 7. The impact of change of z_length especially at the lower values of y_length, is minimal on the angular loss based on a single three-dimensional approach, whereas Figure 8 shows that dual-angle model has greater dependency on the change of angle in the elevation plane compared with the single three-dimensional angle approach.

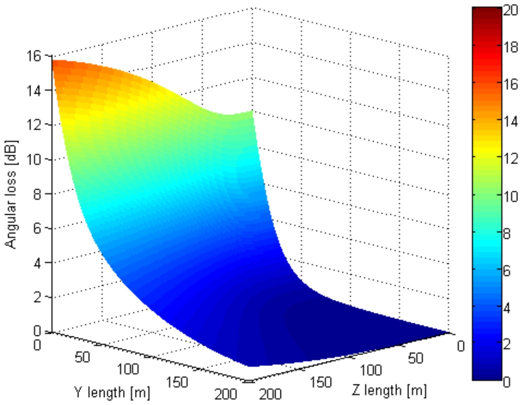


Figure 8. The angular loss based on separate incidence angle in an azimuth (horizontal) and elevation (vertical) plane.

3.2. Extended Building Penetration (EBP) Model

The extended building penetration model is the extension of the general building penetration model. In this model, the additional fifth path known as the "Direct path" is proposed to include as shown in Figure 9. In 3D environment, the direct path is the shortest path between the transmitter and the receiver and it can penetrate through the building wall and through the ceilings of the floors. Figure 10 shows two receiver points RX1 and RX2 located inside the building at Floor 1 and Floor 4, respectively.

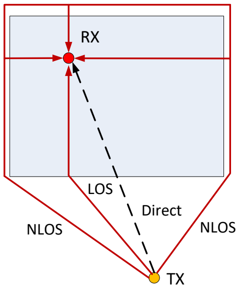


Figure 9. Illustration of direct path along with other paths (top view).

In Figure 10, the LOS paths are shown with the red arrows and the direct paths are shown with the black arrows. It can be seen that the direct ray follows the shortest path between the transmitter and the receiver, and intercepts the building wall at different height compared with that of a receiver height. In Figure 10, it can also be seen that the direct path to RX1 penetrates through the single ceiling, whereas the direct path to RX2 penetrates through the ceilings of two floors. Therefore, the ceiling penetration loss needs to be taken in to account for the direct path.

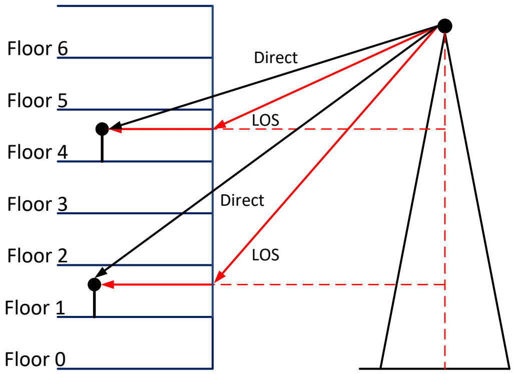


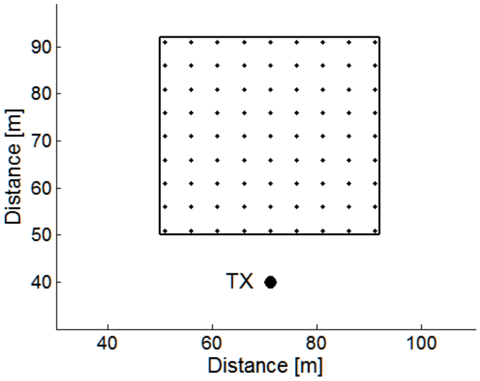
Figure 10. Ray path entering the building while penetrating through the ceiling (side view).

Generally, the ceilings are made up of concrete. Therefore, the ceiling penetration loss model is assumed to follow the concrete penetration loss model as given by (20). The given ceiling penetration loss model provides the ceiling penetration loss of 45 dB, 85 dB, and 125 dB at 10 GHz, 30 GHz and 60 GHz, respectively.

$$L_{\text{Ceiling,dB}} = 4 * \text{Frequency}_{\text{GHz}} + 5, \qquad (20)$$

4. Simulation Environment and Simulation Results

For the first and second part of research work, a single twenty-one stories building with an average floor height of 3 m is considered for simulation as shown in Figure 11. An outdoor base station is located at a distance of 10 m away from the building as shown in Figure 11.

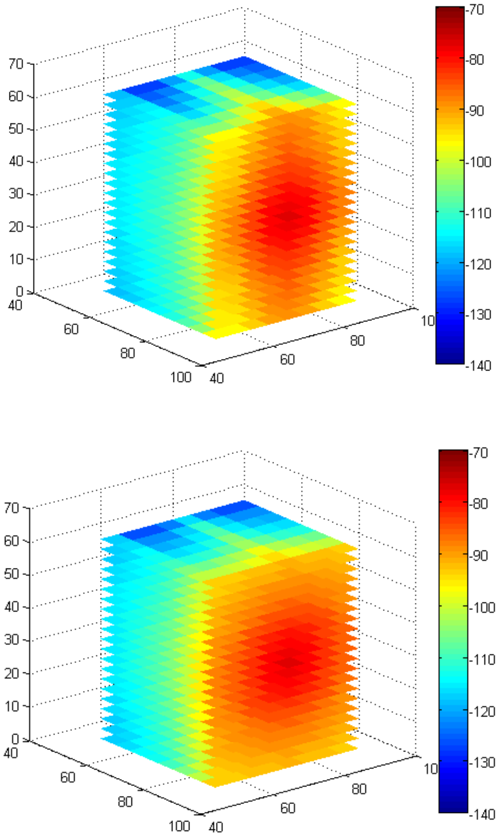


(a)

Figure 11. Two-dimensional map of a single building scenario with transmitter 10 m away from the building.

The transmission power and the height of the base station antenna are 33 dBm and 31.5 m, respectively. The simulations are performed at the frequency of 10 GHz. The indoor location points are placed on each floor with the separation of 5 m among them. An old building type was assumed with indoor wall loss model 1.

In the first part of this research work, the impact of angular wall loss models is analyzed. In case of an angular wall loss model based on a single three-dimensional incidence angle, the incidence angle θ_i was set to 60° for the NLOS paths. Similarly, in case of an angular wall loss model based on a separate incidence angle in azimuth and elevation plane, the incidence angle in an azimuth plane 'α' and the incidence angle in an elevation plane 'β' were also set to 60° for NLOS paths. Therefore, both models give similar results for NLOS paths.



(b)

Figure 12. Path gain, a) Angular wall loss model based on single three-dimensional incidence angle, and (b) Angular wall loss model based on separate incidence angle in azimuth and elevation.

Figure 12 shows the heat map of a path gain achieved with two angular wall loss models. The difference between the two angular wall loss models is evident and visible. In Figure 12(b) as we move away from the centre point of the building wall, the path gain starts to fade (deteriorate) in a perfect circular pattern. However, in Figure 12(a) the change of the path gain with the change in an azimuth and elevation angle is not in a perfectly circular way.

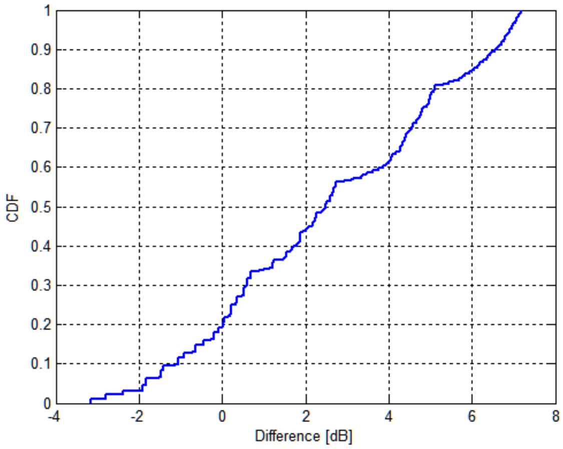
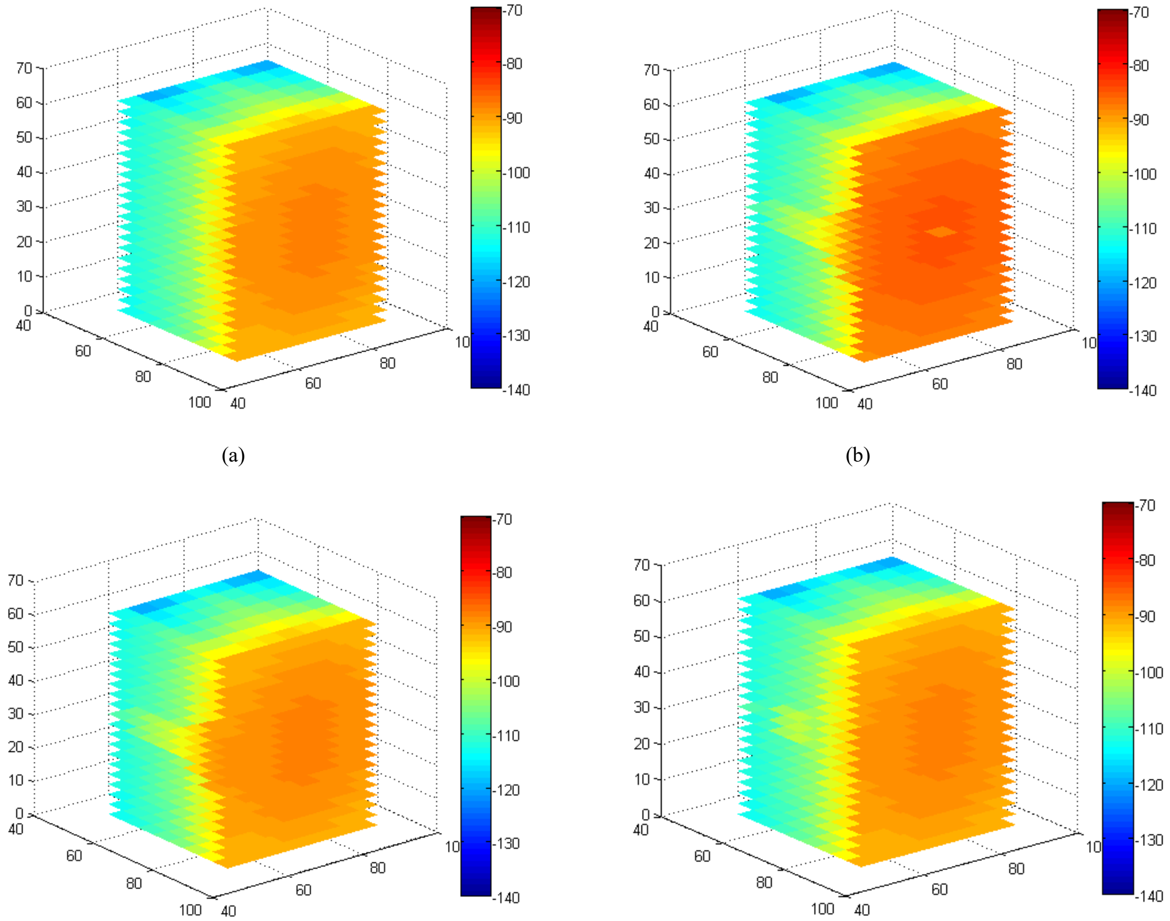


Figure 13. Difference of angular wall loss models.

The difference between the two angular wall loss models is computed by subtracting the received signal power at each point using angular loss model based on separate incidence angle in azimuth and elevation plane from the received signal power calculated using angular wall loss model based on a single three-dimensional incidence angle. The difference was computed assuming only LOS path. The NLOS paths were neglected for calculating the difference, as both models show similar results for NLOS paths. Figure 13 shows the CDF plot of difference between the two angular wall loss models. It can be seen in Fig. 10 that the difference between the two angular wall loss models has values up to 7 dB. The CDF curve with the large number of positive values shows that the angular wall loss model based on a single three-dimensional angle provides aggressive angular loss compared with the dual-angle based angular loss model.



(a)

(b)

(c)

(d)

Figure 14. Path gain at 10 GHz, a) GBP model, (b) EBP model without any filter, (c) EBP model with 3 dB filter, and (d) EBP model with 7 dB filter.

In the second part of this research work, the impact of the direct path in the extended building penetration model is analyzed, and the performance of the general building penetration model is compared with the extended building penetration model. The angular wall loss model based on a separate incidence angle in horizontal and vertical plane is assumed, and the separation between the building wall and transmitter is set to 35 m.

Figure 14 shows a heat map of path gain for different cases. Figure 14(a) shows a path gain for GBP model, and while going from front to the back of the building an almost similar path gain is achieved on all the floors of the building. In Figure 14(b), a direct path is included. Now, due to an additional path a higher path gain is obtained. It is interesting to see that an additional direct path improves the received signal level of the users located on the floors which are close to the height of the outdoor transmitter

It can be seen in Figure 14(b) that due to higher ceiling penetration loss at higher frequency, the impact of a direct path diminishes as the difference between the outdoor antenna height and the user height increases. However, the impact of a direct path is clearly evident to the users located deep inside the building at the same floor height as an outdoor antenna, or on one floor above and below. On the other hand, it can also be observed that in Figure 14(b) the path gain is overestimated at the front face of a building, as the “LOS path” and the “Direct path” are almost identical, and they are both included in the modeling. Therefore, a filtering threshold of 3 dB is used in Figure 14(c) to filter out the direct path if the difference between the LOS path and the direct path is less than or equal to 3 dB. An impact of simple 3 dB filter can be seen in Figure 14(c) as the front face of a building is showing now almost similar results as in Figure 14(a). Similarly, a filtering threshold of 7 dB was used in Figure 14(d). The direct path has quite significant impact on the floors close to the outdoor antenna height even with 3 dB filtering threshold. Therefore, it will improve the model of outdoor to indoor propagation by including a direct path along with the certain filtering threshold, as we have used in this study.

For ray tracing simulations in order to acquire reflected and diffracted multipaths a multiple building scenario is considered as shown in Figure 15.

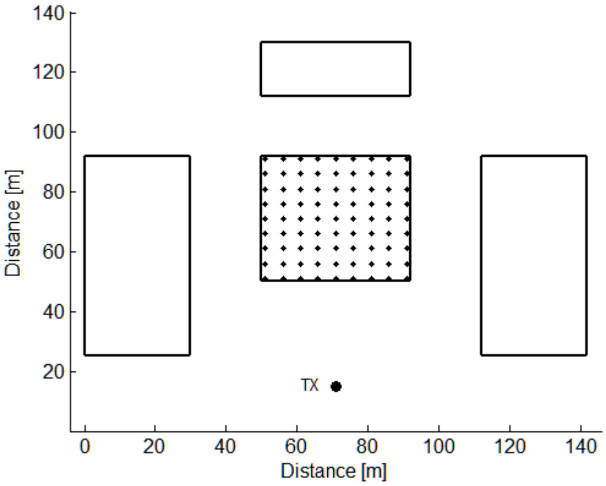


Figure 15. Two-dimensional map of a multiple building scenario with transmitter 35 m away from the building.

In this scenario, a single twenty stories building surrounded by four other buildings of the same height are considered. However, the focus area is the central building, and therefore indoor location test points are distributed in the central building only. An outdoor base station is located at a distance of 35 m away from the central building as in the case of a single building case.

In the third part of this research work, two different ray tracing approaches are considered for the indoor users:

Ray tracing type 1 (RT type 1): In this case, an Image Theory (IT) based ray tracing technique is used to find the propagation paths between the outdoor transmitter and the indoor receiver point. A smooth building surface is assumed, which acts as a perfect reflecting surface. For the transmitter and the receiver points at different heights, the incidence ray path does not intercept the building wall at receiver height. Therefore, for the paths which pass through the ceiling of a floor as shown in Figure 10, the ceiling penetration loss model is used to include the ceiling penetration loss.

Ray tracing type 2 (RT type 2): It is the modified approach of ray tracing type 1. In this case, the propagation paths between the transmitter and receiver in an azimuth plane are found by using the image theory. However, in an elevation plane for the transmitter and receiver at different heights the incidence ray path "always" intercepts the building wall at receiver height. It means that the ray path would never penetrate through the ceiling of a floor. Therefore, the ceiling penetration loss model is not used in this case.

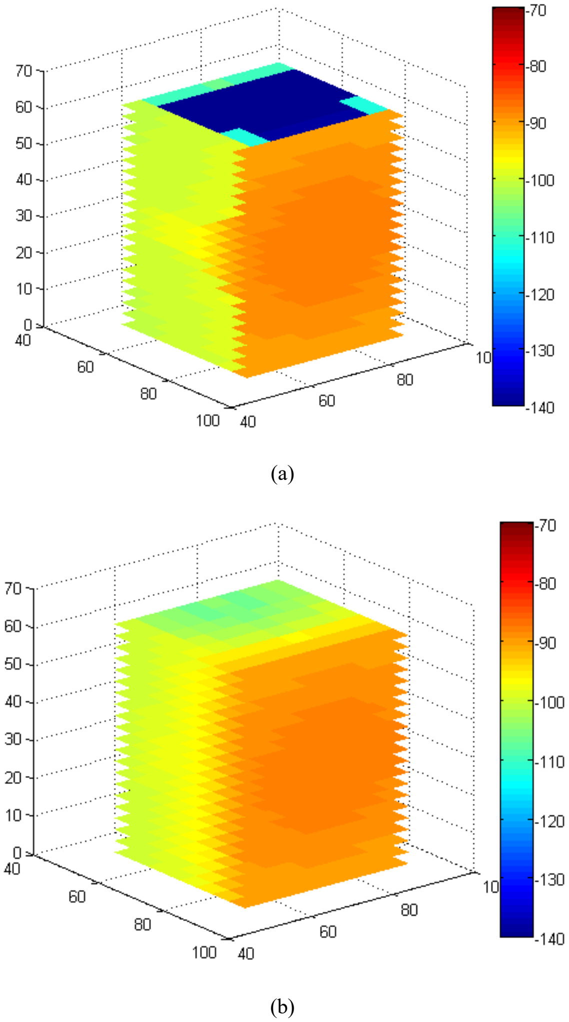


Figure 16. Path gain at 10 GHz, a) Ray tracing type 1, (b) Ray tracing type 2.

Figure 16 shows a path gain map for different cases. It is interesting to see the results obtained by ray tracing type 1. In Figure 16(a), the acquired results from ray tracing type 1 for the indoor users located 6 m or more deep inside the building are quite pessimistic in comparison with the results obtained with ray tracing type 1. As explained earlier, that in case of ray tracing type 1 approach, the ray path entering the building intercepts the building wall at different height compared with that of a receiver height. For the users located deep inside the building the most of the ray paths reach at the receiver point after penetrating through a single or through the multiple ceilings of the floor. Due to high ceiling penetration loss, the signal power attenuates significantly after passing through the ceilings.

Therefore, the ray tracing type 1 is found not suitable for the indoor users located far away from the exterior wall of the building. However, the results obtained with ray tracing type 2 seem more optimistic and realistic in comparison with the ray tracing type 1. It is critical here to mention that in ray tracing type 2, the incident ray always intercepts the building wall at the same height as that of a receiver height, and therefore the ceiling penetration loss is ignored in ray tracing type 2.

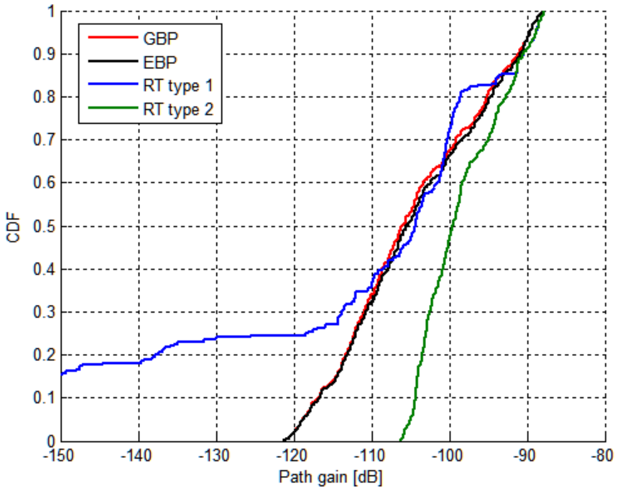
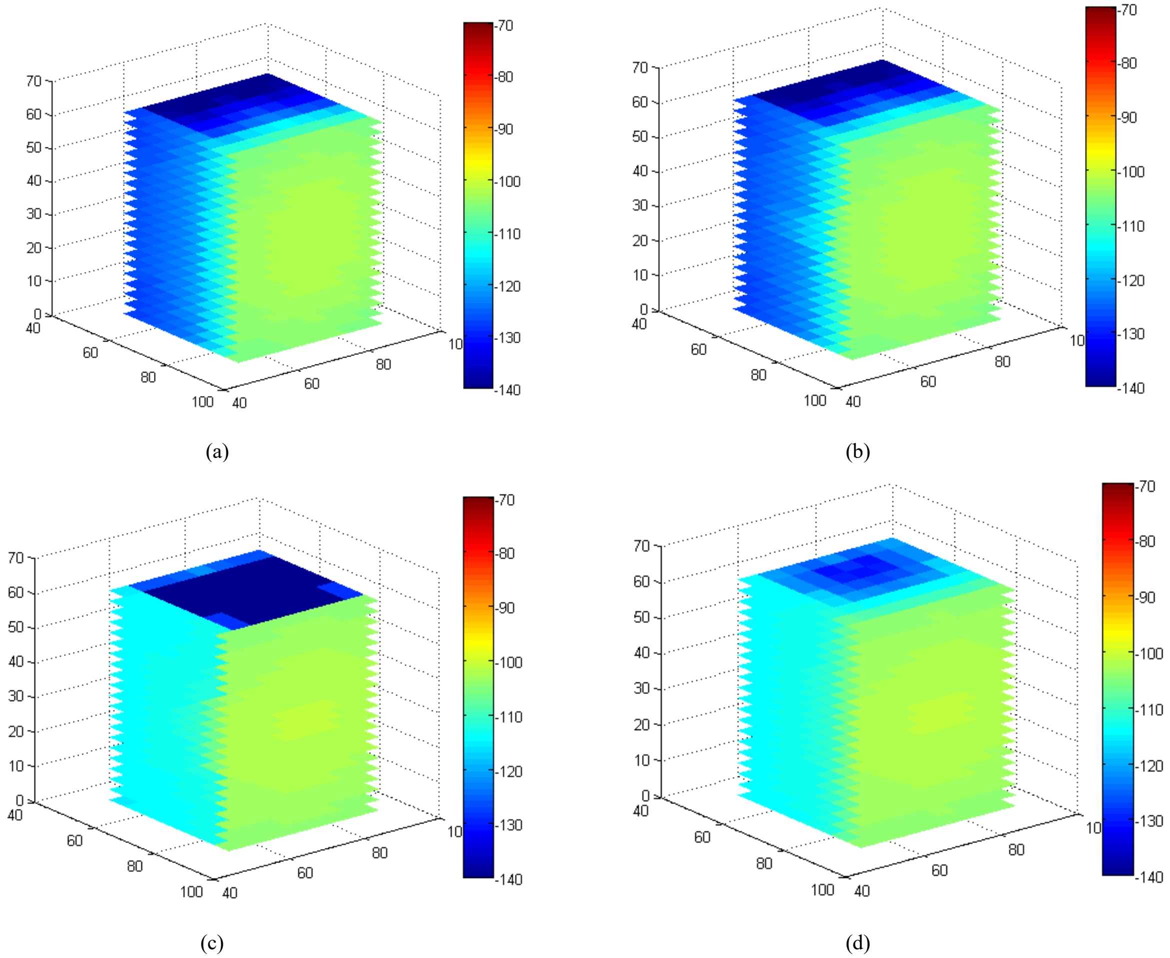


Figure 17. CDF plot of path gain for different propagation models at 10 GHz.

Figure 17 shows the CDF plot of path gain for different considered cases. For an extended building penetration model, a 3 dB filtering threshold was used. For the top nearly 14.5 % of the samples, the ray tracing type 1 shows an almost identical result as ray tracing type 2. Those 14.5 % of the samples mainly represent the receiver locations from the front row and few from the second row i.e. receiver points at the front face of the building. Again, for the same top 14.5 % of the samples the general building penetration model provides an almost similar result like that of an extended building penetration model and other ray tracing approaches. Then, the path gain starts to drop significantly for the receiver points located deep inside the building. A large portion of samples have the extremely pessimistic values of the path gain with a ray tracing type 1. For the considered cases, the ray tracing type 2 was found the most optimistic approach. For ray tracing simulations, all the possible ray paths with two reflections and two diffractions were considered.

Therefore, there is a large number of considered ray paths in case of ray tracing compared with the GBP and EBP. The results provided in Figure 14 and Figure 16 reveal that ray tracing type 2 is a better and realistic approach compared with ray tracing type 1 for estimating the received signal level, especially for the indoor users located deep inside the building.



(a)

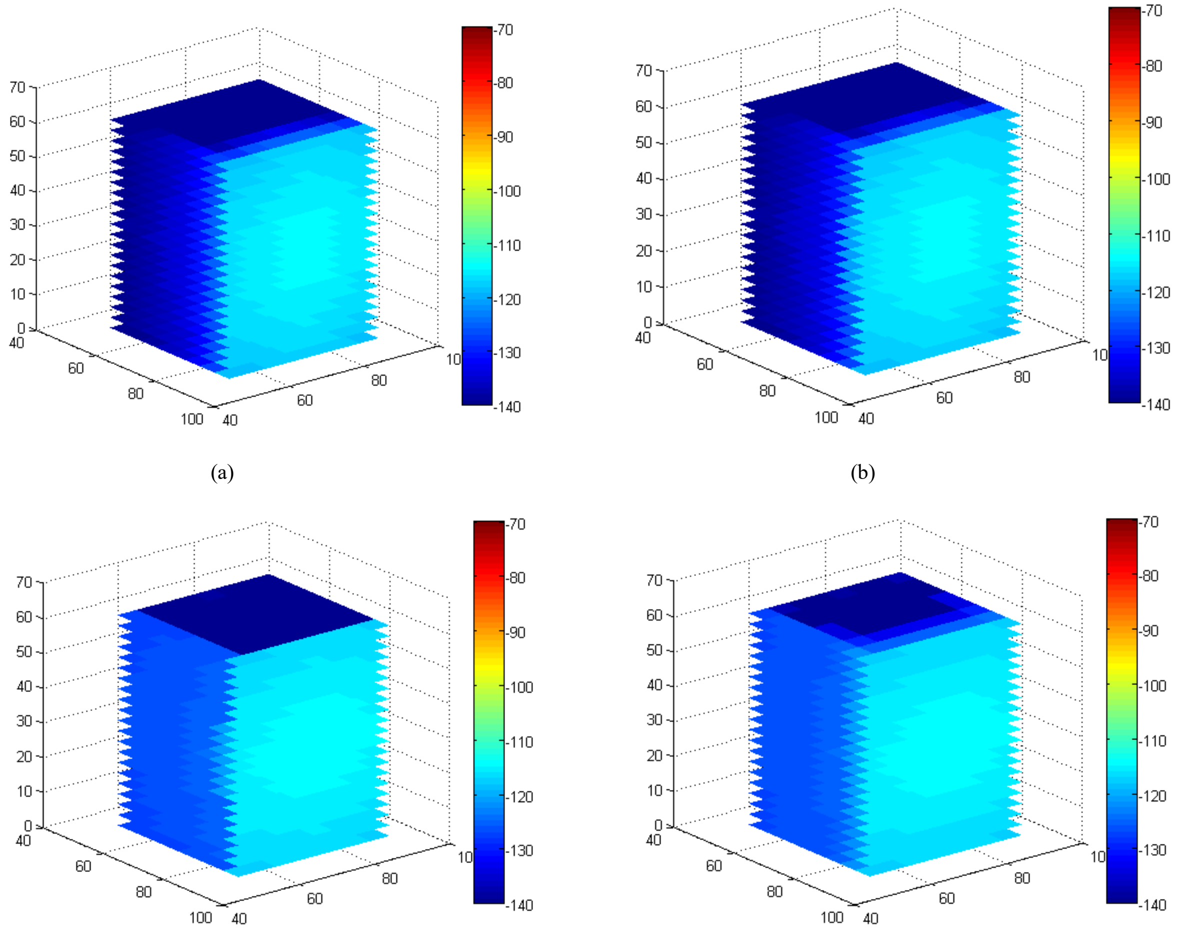
(b)

(c)

(d)

Figure 18. Path gain at 30 GHz, a) GBP model, (b) EBP model with 3 dB filtering threshold, c) Ray tracing type 1, and (d) Ray tracing type 2.

Figure 18 shows a heat map of path gain for different models assuming old building type at 30 GHz and indoor wall loss model 1. Path gain is the function of frequency and a significant impact of frequency of operation can be seen between the results presented in Figure 16 (at 10 GHz) and Figure 18 (at 30 GHz). In Figure 18(b), the path gain for EBP model with 3 dB filtering threshold is shown, and the impact of the direct path is even



(a)

(b)

(c)

(d)

Figure 19. Path gain at 60 GHz, a) GBP model, (b) EBP model with 3 dB filtering threshold, c) Ray tracing type 1, and (d) Ray tracing type 2.

visible at 30 GHz frequency for the center floors. The impact of a direct path is not prominent in overall results. Again, the ray tracing type 1 is found as an inadequate approach for estimating the path gain, and it provides the extremely low values of path gain at 30 GHz. However, it is interesting to see the results acquired with ray tracing type 2. It is fascinating to compare the results of two ray tracing approaches, as ray tracing type 1 clearly shows lack of indoor coverage with the outdoor base station except the front wall of the building. On the other hand, the simulation results of ray tracing type 2 show that the low indoor coverage can be provided with the outdoor base station. Similarly, GBP and EBP models show lack of indoor coverage at 30 GHz. These results leave an open discussion that the results acquired from the ray tracing type 2 are more realistic or the results acquired with ray tracing type 1 are more realistic. As we don't have any measured reference data therefore it is hard to make a clear statement.

Figure 19 shows a heat map of path gain for old building type at 60 GHz assuming indoor wall loss model 1. At 60 GHz, all of the considered propagation models show that due to high propagation and penetration losses the indoor coverage cannot be provided with the outdoor base station. Higher antenna gain at higher frequencies can compensate the high penetration and propagation losses. Otherwise, an indoor coverage with the outdoor base station at 60 GHz can be provided if new buildings are constructed with the material having better signal penetration properties.

5. Conclusion

In this paper, the outdoor to indoor propagation at higher frequency is studied. Furthermore, a novel angular wall loss model based on a separate incidence angle in azimuth and elevation plane is proposed. In the proposed angular wall loss model, the total angular loss is the sum of loss due to the change of angle of an incident ray in both azimuth and elevation plane. The simulation results show that the angular wall loss model based on a single three-dimensional angle provides aggressive angular loss compared with the proposed dual-angle angular wall loss model. In the second part of this study, an Extended Building Penetration (EBP) model is proposed, and the performance of EBP model is compared with the COST231 building penetration model. In an extended building penetration model a fifth path known as a "Direct path" is added. Through the simulation results, it is shown that the direct path has a significant impact on the indoor receiver points located close to the height of the transmitter. Therefore, the extended building penetration model can be considered as a better approach for modelling the outdoor to indoor propagation at higher frequencies.

The ray tracing is a promising and precise technique for estimating the received signal level and for channel modeling. In this study, a performance comparison between the general building penetration model and the ray tracing model is done. It was found that the traditional ray tracing technique provides good approximation for the users located close to the wall of the building. However, the traditional ray tracing approach was found in-efficient for modeling the outdoor to indoor propagation especially for the indoor users located far away from the exterior wall of the building. A new approach of ray tracing is also proposed in this study. The simulation results show that ray tracing performance is significantly improved by the proposed recommendation. With the proposed ray tracing approach the results of a path gain were fairly better compared with the other models. Later, the performance of different propagation models was compared at 10 GHz, 30 GHz and 60 GHz frequency of operation. It was found that for the considered simulation scenario the adequate coverage can be provided for the indoor users with the outdoor base station at 10 GHz. However at 30 GHz, the general penetration model clearly showed the lack of indoor coverage. Whereas the proposed ray tracing type 2 approach showed a fair indoor coverage at 30 GHz. The propagation and building penetration losses were quite high at 60 GHz; therefore the simulation results show no indoor service at 60 GHz.

Acknowledgment

Authors would like to thank European Communications Engineering (ECE) Ltd, TAKE-5 project, and Oy L M Ericsson Ab for supporting this research work.

References

[1] M. U. Sheikh, K. Hiltunen and J. Lempiäinen, "Angular wall loss model and Extended Building Penetration model for outdoor to indoor propagation," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, 2017, pp. 1291-1296.

[2] E. Damosso and L. M. Correia, Eds., Digital mobile radio towards future generation systems, final report ed. European Commission, 1999.

[3] J. E. Berg, "A recursive method for street microcell path loss calculations," Proceedings of 6th International Symposium on Personal, Indoor and Mobile Radio Communications, Toronto, Canada, September 1995, pp. 140−143 vol. 1.

[4] E. Semaan, F. Harrysson, A. Furuskär and H. Asplund, "Outdoor-to-indoor coverage in high frequency bands", 2014 IEEE Globecom Workshops (GC Wkshps), Austin, TX, 2014, pp. 393−398.

[5] C. Larsson, F. Harrysson, B.-E. Olsson, and J. E. Berg, "An outdoor-to-indoor propagation scenario at 28 GHz," in 8th European Conference on Antennas and Propagation (EuCAP 2014), The Hague,The Netherlands, April 2014, pp. 3301−3304.

[6] W. C. Stone, "Electromagnetic signal attenuation in construction materials, "NIST Building and Fire Research Laboratory, Gaithersburg, Maryland, NISTIR 6055 Report No. 3 6055, Oct. 1997.

[7] L. M. Frazier, "Radar surveillance through solid materials," in Proceedings of the SPIE - The International Society for Optical Engineering, vol. 2938, Hughes Missile Syst. Co., Rancho Cucamonga, CA, USA, 1997, pp. 139−146.

[8] R. Wilson, "Propagation losses through common building materials 2.4 GHz vs 5 GHz," University of Southern California, CA, Tech. Rep. E10589, Aug. 2002.

[9] C. A. Remley, G. H. Koepke, C. L. Holloway, C. A. Grosvenor, D. G. Camell, J. M. Ladbury, R. Johnk, D. R. Novotny, W. F. Young, G. Hough, M. McKinley, Y. Becquet, and J. Korsnes, "Measurements to support modulated-signal radio transmissions for the public-safety sector," NIST, Boulder, CO, Tech. Rep. Tech. Note 1546, Apr. 2008.

[10] White paper on "5G Channel Model for bands upto 100 GHz".

[11] H.W. Son, and N.H. Myung, "A deterministic ray tube method for microcellular wave propagation prediction model," Antennas and Propagation, IEEE Transactions on , vol.47, no.8, pp.1344-1350, Aug 1999.

[12] D.N. Schettino, F.J.S. Moreira, C.G. Rego, "Efficient Ray Tracing for Radio Channel Characterization of Urban Scenarios," Magnetics, IEEE Transactions on , vol.43, no.4, pp.1305-1308, April 2007

[13] S. Soni, and A. Bhattacharya, "An efficient two-dimensional ray-tracing algorithm for modeling or urban microcellular environment", International Journal of Electronics and Communications (AEU), volume 66, issue 6, pp. 439-447, June 2012.

[14] J. E. Berg, "A recursive method for street microcell path loss calculation", 6th PIMRC' 95, Toronto, Canada, September 1995, pp. 140-143.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 69-73 (2018)

www.astesj.com

Special issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

TPMTM: Topic Modeling over Papers' Abstract

Than Than Wai[*], *Sint Sint Aung*

Web Mining, University of Computer Studies, Mandalay, ZIP Code 05071, Myanmar

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 05 February, 2018
Online: 08 March, 2018

Keywords:
Frequent pattern mining
Latent Dirichlet Allocation
Entropy

ABSTRACT

Probabilities topic models are active research area in text mining, machine learning, information retrieval, etc. Most of the current statistical topic modeling methods, such as Probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA). They are used to build models from unstructured text and produce a term-based representation to describe a topic by choosing single words from multinomial word distribution. There are two main weaknesses. First, popular or common words are different topics, often causing ambiguity for understanding the topics; Second, lack of consistent semantics for single words to be represented correctly. To address these problems, this paper proposes a model (A Two-Phase Method for Constructing Topic Model, TPMTM) that combines statistical modeling (LDA) with frequent pattern mining and produces better presentations of rich topics and semantics. Empirical evaluation shows that the results of the proposed model are better than LDA.

1. Introduction

Topic models are Bayesian statistical models that are structured in accordance with a hidden theme, usually called unstructured data in a set of textual documents, topics with multiple distributions of words. Due to a collection of unstructured text documents, the topic model assumes that the collection of documents (corpus) has a certain number of hidden topics and that each document contains more than one topic in different sizes. Researchers have developed several topic models such as Latent Semantic Indexing (LSA) [1], Probability Latent Semantic Analysis (PLSA) [2] and Latent Dirichlet Allocation (LDA) [3]. Topic modeling automatically selects topics from the text and identifies topics over time [4], explores the connection between topics[5], supervised the topics [6], recommendation [7], and so on.

LDA or unsupervised generation probabilistic methods for modeling the document collection (corpus), is the most commonly used topic modeling method. The LDA, each document can be described as a probabilistic distribution for latent topics, and that the topic distribution of all documents is distributed a common Dirichlet prior. Within each topic in the LDA model is described as a probabilistic distribution over-represented as a probabilistic distribution of words and words distributions of topics distribute the same Dirichlet prior. Each latent topic in the LDA model is also distributions of topics distribute a common Dirichlet prior as well. A corpus D consists of M documents, with document d having N_d words (d ∈{1,..., M}), LDA models D as stated in the following process[8].

LDA concludes the following generative process for each document w in a D corpus:

1. Choose N ~ Poisson (ξ).

2. Choose θ ~Dir (α).

3. For each of the N words w_n:

(a) Choose a topic z_n~ Multinomial (θ).

(b) Choose a word w_n from $p(w_n|z_n.\beta)$, a multinomial probability conditioned on the topic Z_n.

The discovered variables are words in documents although others are hidden variables (θ and ϕ) and hyperparameters (β and α).To provide hyperparameters and hidden variables, the probability of discovered data D and maximized as follows:

$$p\,(D|\alpha,\beta)= \prod_{d=1}^{M}\int p(\theta_d\,|\alpha)\big(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p(\omega_{dn}|z_{dn},\beta)\big)d\theta_d \quad (1)$$

The LDA model has three levels of the representation. In corpus level, there are two parameters (α and β) that are involved

[*] Corresponding Author: Than Than Wai, Web mining, University of Computer Studies, Mandalay, ZIP Code 05071, Myanmar | Email: thanwai85@gmail.com

https://dx.doi.org/10.25046/aj030208

in the process of building a corpus. Document-level variables are the variables θ_d, which are sampled once per document. Finally, word-level variables $z_{d,n}$ and $w_{d,n}$ are the variables that are collected for each word in each document. The current statistical topic modeling techniques make multinomial distributions in words to represent topics in a given collection of texts. For example, Table 1 displays a sample of multinomial distributions used to describe the three themes of a scientific collection of publications.

Table 1: A Sample of Topic Presentation on AAAI Dataset

TopicId	Words
7	problem, result,order ,solver, present
12	algorithm,show, state,find,result
18	behavior ,system ,agent, develop ,result

From the above results in Table1, a sample of word distributions used to present three themes of a scientific paper collection of AAAI dataset. The term "result" is a general term and very general term in showing research papers in all different fields. The general words cause ambiguous to the topic presentation. So, a new model is required to solve these problems. The new method should take higher special representations and explore latent associations under multinomial word distributions.

The LDA and other topic models are portions of the better field of probabilistic modeling. Generative probabilistic topic modeling is a method for unsupervised classification of documents, by modeling each document as a mix of themes and each theme as a mix of words. But there exist the problems of word uncertainty and semantic integrity [8].

Text mining is a technique that supports users' assets effective information from a variety of digital documents. Most text mining methods are keyword-based strategies that need single words to show the documents. Based on the theory that the phrase may have more linguistic meaning than the keyword, strategies for using phrases instead of keywords are also suggested. However, surveys have shown that phrase-based techniques are not always better than keyword-based techniques [9, 10]. Many strategies in the field of data mining are used in patterns to mine useful documents that yield encouraging results [11, 12].

Topic modeling provides a convenient way to analyze large classified text collections while extracting interesting features to express collections in text mining. Thus, it advances the proposed model to improve the accuracy and relevance of the topic's representations by using the techniques of text mining, especially frequent itemset mining techniques.

In this model, Latent Dirichlet Allocation (LDA) is integrated into data mining techniques and achieves successful accuracy for the collection documents (corpus). The proposed model is composed of two phases: 1) LDA is applied to accomplish first topic models and 2) the frequent itemset (pattern) mining method is applied to obtain further particular patterns to produce topics of the document collections. Furthermore, the frequent itemsets (patterns) often explain information about the structure of the relationship between words that provide topics that are understandable, relevant and broad.

2. Related Work

Probabilistic topic modeling is expanded to capture more interesting features [13], but they show topics through the distribution of multinomial words. The papers [14, 15] are a widespread way to express the linguistic meaning of the topics as mentioned in the introduction. The authors [16] show a way to calculate the similarities between given themes and a known hierarchy then select the most grant labels to show the topics. But, the weakness of existing methods is that they are strictly limited to resource candidates and are limited to linguistic coverage. The proposed model is a work extension originally presented at 16th IEEE/ACIS International Conference on Computer and Information Science, [17]. This paper [18] discusses the topic-related model phrase by Markov dependencies in word order based on LDA structure, related to this paper. The results provided on [19, 20, 21] show that the topics described by the phrases are easier to interpret than their LDA. But phrases may contribute low-level events in documents, which cannot be accomplished with efficient retrieval performance.

3. Phase1: Topic Presentation Propagation Using LDA

The Latent Dirichlet Allocation is an algorithm that automatically detects the themes that are present in the collection of documents. At the LDA, each document can be viewed as a mix of different topics. The LDA provides the topic using word distribution and representation of the document using the topic distribution. The description of the topic means which words are important to which the topic and representation of the document in which topics are important in the documents [22-26].

Let D = {d_1, d_2, ..., d_M} be a corpus. Each document is considered as a mixture of themes and each theme can be defined as a distribution over frozen vocabulary words composed of documents using (1). In general, the proposed model has $\theta = \emptyset_1, \emptyset_2, \ldots, \emptyset_V$ for all topics.

Example results of LDA, the topic presentation is shown in Table 2. At document level, table 3 shows LDA's example results, document representation and Table 4 also shows word-topic assignments results of LDA's example.

Table 2: Topic representations – probability distribution over words

Topic	θ
$\emptyset_7$	problem:$\frac{1}{3}$, result:$\frac{4}{12}$,order: $\frac{2}{12}$,solver$\frac{2}{12}$, present : $\frac{2}{12}$
$\emptyset_{12}$	algorithm$\frac{2}{15}$,show : $\frac{4}{15}$, state: $\frac{1}{3}$,find: $\frac{2}{15}$,result **:** $\frac{2}{15}$
$\emptyset_{18}$	behavior : $\frac{4}{15}$,system : $\frac{2}{15}$,agent : $\frac{1}{15}$, develop: $\frac{2}{15}$, result: $\frac{2}{15}$

LDA contributions are the representation of the topic that uses the word distribution that the words are important to what the topic matter is, and the representation of the document by distributing the topic which themes are important for a particular document. These representations are used for obtaining information, document classification, text mining, machine learning, etc.

Table 3: Document representation-probability distribution over topics

Document	z_7	z_{12}	z_{18}
d_7	0.385	0.123	0.108
d_{12}	0.464	0.118	0.073
d_{18}	0.305	0.11	0.098

Then, specified topics also indicate which words are important in which topics, similar to the representation on the subject. The proposed model performs word-topic emphasis on the LDA for more precise or more specific topic presentation for a given corpus.

4. Phase2: Topic Presentation Expansion

Words with high probability in topic distributions are selected to represent topics in most LDA based applications. For example, the top 5 words for the 3 topics, as shown in Table 2, are: problem, result, order ,solver, present for topic 7, algorithm, show, state, find, result for topic 12 and behavior ,system ,agent, develop ,result for topic 18. So, they are likely to represent the general concepts or common concepts of the three topics and can not describe the three topics that are noteworthy. Furthermore, the words in representations on the topics formed by the LDA are individual single words. Single words provide very little information about relationships between words and very limited language definitions to make the topic matter clear.

In this section, we propose a method based on frequent patterns (itemset) mining techniques, detailed in the following sub-sections, aimed at reducing the above-mentioned problems.

4.1. Frequent Itemsets based on LDA

The frequent itemset mining is the method of mining data in a set of items or words in large data sets. These patterns are usually described in various forms such as frequent itemsets, sequential patterns, or substructures.

In this paper, the proposed model uses frequent itemsets. Typical itemset generally indicates that a set of items often happens together in a transaction dataset. The methods of using frequent itemsets are categorized into three basic skills: horizontal data format, vertical data format, and expected database strategy.

We use the vertical data format in the mine frequent itemsets in this paper and believe that frequent pattern mining based representation can be more meaningful and more accurate to describe topics. In addition, frequent pattern based representations contain structural information that shows the relationship between the words.

Create Transactional Dataset: The aim of the proposed frequent pattern-based method is to discover related words (i.e., frequent itemsets) from the words assigned by LDA to the topics. From this purpose, we build a set of words from each word-topic assignment instead of using the order of words, because, for frequent pattern mining, the frequency of a word within a transaction is less important. A topical document transaction is a set of words without any duplicates. Let D={d_1,d_2,…,d_M}be the original document collection, the transactional dataset for topic Z_j is defined as T_j .

For the topics in D, we can develop V transactional datasets. An example of the transactional datasets is described in Table 5, made from the example in Table 4.

Generate Frequent Pattern-based Representation: The basic idea of the proposed method is to reduce frequent itemsets generated from each transactional dataset T_j to represent Z_j. Here the frequent items in each transactional dataset are taken. Then documents associated with each item are converted into the vertical format and a number of documents containing each item are counted. Items are sorted according to their number of hits (i.e. number of documents of each item). Removing noise items that the number of hits is greater than the user provided threshold. If document-set is a subset of another documents-sets in each transactional dataset then items are merged into enhanced frequent itemsets. For example, w_1:d_1d_2:3, w_2:d_1d_3:2, w_3:d_1d_4:3, w_1:d_2d_3:2, w_8:$d_1d_2d_3$:2, w_7:$d_1d_2d_4$:3 can be compressed as the enhanced frequent itemsets w_1,w_2,w_8:$d_1d_2d_3$:2, w_1,w_3, w_7:d_1d_2d4:3 (format - "items" : "document" : "topic (transactional dataset)"). Such meaningless itemsets may harm document filtering tasks using frequent itemsets because it has duplicates the same frequent itemsets. So, we can regard the proposed method as lossless compression because we can cover all the removed frequent itemsets with the exact topic and it can effectively filter out the redundant itemsets. Finally, frequent itemsets in each transactional dataset are reconstructed. ("itemset" and "pattern" are interchangeable in this thesis). Frequent itemsets are the most widely used patterns created from transactional datasets to illustrate useful or interesting patterns. The main idea of the proposed frequent itemset method is to use of frequent patterns generated from each transactional dataset to represent topic Zj. For a given minimal support threshold δ, and itemset p in T_j is frequent if supp (p) >= δ where supp(p) is the support of p where the number of transactions containing p. Take T_7 as an example, which is the transactional dataset for topic Z_7. For a minimal support threshold δ = 2, all the frequent itemsets generated from are given in Table 6. Patterns represent words related to specific and recognizable meanings.

Table 6: The Frequent Itemsets Explored from T_7

Frequent Patterns	minimal support threshold δ
<result>	3
<algorithm,show,state,find>	2

5. Evaluation

We made experiments to evaluate the performance of the proposed method. In this section, we show the results of the evaluation.

5.1. Datasets

Two datasets are used in experiments, containing abstracts of papers published in the AAAI from 2013 to 2014 and NSF from 1990 to 2003. The two datasets contain 548 and 129000 abstracts. The abstracts are obtained from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml) and by using Porter's stemmer in Java (http://www.tartarus.org/~martin/PorterStemmer).

Table 4: Word-topic Assignments

Document	$\mathbf{z_7}$		$\mathbf{z_{12}}$		$\mathbf{z_{18}}$	
	Proportion of Topic	***Terms***	***Proportion of Topic***	***Terms***	***Proportion of Topic***	***Terms***
d_7	0.385	problem, result, order ,solver, present	0.123	algorithm ,state, find ,algorithm	0.108	behavior , system ,agent
d_{12}	0.464	algorithm,show, state,find,result	0.118	spars ,distribute ,fin d	0.073	class, complex
d_{18}	0.305	algorithm,show, state,find,result	0.11	algorithm, state	0.098	behavior, system ,system

Table 5: Transactional datasets

Document	$\mathbf{z_7}$(*T₇*)		$\mathbf{z_{12}}$(*T₁₂*)		$\mathbf{z_{18}}$(*T₁₈*)	
	Proportion of Topic	***Terms***	***Proportion of Topic***	***Terms***	***Proportion of Topic***	***Terms***
d_7	0.385	problem, result, order ,solver, present	0.123	algorithm,state, find	0.108	behavior , system ,agent
d_{12}	0.464	algorithm,show, state,find,result	0.118	spars,distribute, find	0.073	class, complex
d_{18}	0.305	algorithm,show, state,find,result	0.11	algorithm,state	0.098	behavior, system

5.2. Settings

Firstly, we are preparing datasets from the UCI Machine Learning Repository and preprocessing of all documents by removing stop and stemming words.

Secondly, we apply the LDA model to construct a topic model with V = 20 topics for each data collection, using the MALLET topic modeling toolkit (http://mallet.cs.umass.edu//index.php). Our experiments show that an inadequate number of topics will mainly lead to abundant expanded patterns in the topic model. We run Gibbs sampling for 1000 iterations, the LDA hyperparameters are α = 50 / V and β = 0.01.

Thirdly, topical transaction datasets for optimizing topic representations are developed.

Finally, frequent itemsets based on topic representations using the proposed method presented in Section 4 are developed. We used 10% of the documents for testing purpose and trained the resting model at 90%.

5.3. Baseline Model

In order to compare the suggested method, the LDA chose the baseline model in the experiments. Examples of the results of the two models (the LDA model and the frequent itemset based on model) are shown in Table 7. The top 10 terms or frequently itemsets in each of the topic presentations produced by two models are shown in Table 7.

Table 7: Topic Presentations of All Models Using the AAAI Dataset

Topic8		**Topic4**	
LDA	***Frequent itemsets***	***LDA***	***Frequent itemsets***
change	algorithm,method	parallel	markov
system	gener	semant	agent
data	differ	complex	popular,demonstre
unsupervise	propos	algorithm	analyze,learn
input	mani	mathemat	design
compact	consid	condit	relationship,topic
benefit	effici	represent	tempor
superior	sever,train	tool	recommend,exploit
state	imag	regular	adapt
trust	target,problem	improve	time

5.4. Results

The objective of the proposed approach and other current topic modeling methodologies is to show the topics of a collection of documents as specific potential. For existing topic modeling methods and the proposed methodology, the topic representations are word distributions or patterns with probabilities. The more specific selected words or patterns are in the representation of the topic, the more precise representation of the topic matter becomes. The performance of the proposed method is evaluated by the use of information entropy. The higher the entropy, the more the proposed model is disordered.

Table 8: Comparison of All Models in Information Entropy Using a collection of documents of AAAI and NSF

Datasets	**Latent Dirichlet Allocation**	**Frequent Itemsets (patterns)**
AAAI	4.86159	2.68171
NSF	28.3747	20.7532

From the above results in Table 8, the suggested method based on frequent itemsets has lower entropy rates than the baseline model. Thus, it can make more exact presentations of the topics of a corpus.

6. Conclusion

This paper proposes a model to produce more discriminative and semantic rich representations for modeling topics in a given collection of documents. The main contribution of this paper is the novel approach of incorporating the pattern mining method and topic modeling method (Latent Dirichlet Allocation) to produce representations based on the pattern for modeling the topics. The test results show that representations based on patterns are more specific than representations developed by the Latent Dirichlet Allocation. In the future, we will examine the structure of the patterns and discover the relationship between words that represent topics at a more granular level.

References

[1] Blei, David M. "Probabilistic topic models." Communications of the ACM 55, no. 4 (2012): 77-84.

[2] Blei, David M., and D. Jon. "McAuliffe. supervised topic models." Advances in Neural Information Processing Systems 20 (2007): 121128.

[3] Blei, David M., and John D. Lafferty. "A correlated topic model of science." The Annals of Applied Statistics (2007): 17-35.

[4] Blei, David M., and John D. Lafferty. "Dynamic topic models." In Proceedings of the 23rd international conference on Machine learning, pp. 113-120. ACM, 2006.

[5] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3, no. Jan (2003): 993-1022.

[6] Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. "Indexing by latent semantic analysis." Journal of the American society for information science 41, no. 6 (1990): 391.

[7] Fürnkranz, Johannes. "A study using n-gram features for text categorization." Austrian Research Institute for Artifical Intelligence 3, no. 1998 (1998): 1-10.

[8] Gao, Yang, Yue Xu, Yuefeng Li, and Bin Liu. "A two-stage approach for generating topic models." In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 221-232. Springer, Berlin, Heidelberg, 2013.

[9] Gupta, Shivani D., and B. P. Vasgi. "Effective Pattern Discovery and Retrieving Relevant Document for Text Mining."

[10] Hofmann, Thomas. "Unsupervised learning by probabilistic latent semantic analysis." Machine learning 42, no. 1 (2001): 177-196.

[11] Lau, Jey Han, David Newman, Sarvnaz Karimi, and Timothy Baldwin. "Best topic word selection for topic labelling." In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 605-613. Association for Computational Linguistics, 2010.

[12] Lau, Jey Han, Karl Grieser, David Newman, and Timothy Baldwin. "Automatic labelling of topic models." In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 1536-1545. Association for Computational Linguistics, 2011.

[13] Magatti, Davide, Silvia Calegari, Davide Ciucci, and Fabio Stella. "Automatic labeling of topics." In Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on, pp. 1227-1232. IEEE, 2009.

[14] Moran, Kelly, Byron C. Wallace, and Carla E. Brodley. "Discovering Better AAAI Keywords via Clustering with Community-Sourced Constraints." In AAAI, pp. 1265-1271. 2014.

[15] Náther, Peter. "N-gram based Text Categorization." Lomonosov Moscow State Univ (2005).

[16] Sebastiani, Fabrizio. "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34, no. 1 (2002): 1-47.

[17] Wai, Than Than, and Sint Sint Aung. "Enhanced frequent itemsets based on topic modeling in information filtering." In Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on, pp. 155-160. IEEE, 2017.

[18] Wang, Chong, and David M. Blei. "Collaborative topic modeling for recommending scientific articles." In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 448-456. ACM, 2011.

[19] Wang, Xuerui, Andrew McCallum, and Xing Wei. "Topical n-grams: Phrase and topic discovery, with an application to information retrieval." In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on, pp. 697-702. IEEE, 2007.

[20] Westergaard, D., Stærfeldt, H.H., Tønsberg, C., Jensen, L.J. and Brunak, S., 2017. "Text mining of 15 million full-text scientific articles." bioRxiv, p.162099.

[21] Wu, Sheng-Tang, Yuefeng Li, and Yue Xu. "Deploying approaches for pattern refinement in text mining." In Data Mining, 2006. ICDM'06. Sixth International Conference on, pp. 1157-1161. IEEE, 2006.

[22] Zeng, J., 2012. "A topic modeling toolbox using belief propagation." Journal of Machine Learning Research, 13(Jul), pp.2233-2236.

[23] Zhang, W., Ma, D. and Yao, W., 2014. "Medical Diagnosis Data Mining Based on Improved Apriori Algorithm." JNW, 9(5), pp.1339-1345.

[24] Zhao, Wayne Xin, Jing Jiang, Jing He, Yang Song, Palakorn Achananuparp, Ee-Peng Lim, and Xiaoming Li. "Topical keyphrase extraction from twitter." In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, pp. 379-388. Association for Computational Linguistics, 2011.

[25] Zhu, X., Ming, Z., Hao, Y. and Zhu, X., 2015, June. "Tackling Data Sparseness in Recommendation using Social Media based Topic Hierarchy Modeling." In IJCAI (pp. 2415-2423).

[26] Zhu, J., Wang, K., Wu, Y., Hu, Z. and Wang, H., 2016. "Mining User-Aware Rare Sequential Topic Patterns in Document Streams." IEEE Transactions on Knowledge and Data Engineering, 28(7), pp.1790-1804.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 74-84 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

MPC-based energy efficiency improvement in a pusher type billets reheating furnace

Silvia Maria Zanoli*,1, Francesco Cocchioni2, Crescenzo Pepe2

1Università Politecnica delle Marche, Dipartimento di Ingegneria dell'Informazione, Ancona (AN), 60131, Italy

2i.Process S.r.l., Falconara Marittima (AN), 60015, Italy

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 21 December, 2017
Online: 08 March, 2018

Keywords:
Model Predictive Control
Energy efficiency
Billets reheating furnace

ABSTRACT

The research reported in this paper proposes an Advanced Process Control system, denoted "i.Process | Steel – RHF", oriented to energy efficiency improvement in a pusher type billets reheating furnace located in an Italian steel plant. A tailored control method based on a two-layer Model Predictive Control strategy has been created that involves cooperating modules. Different types of linear models have been combined and an overall furnace global linear model has been developed and included in the controller formulation. The developed controller allows handling all furnace conditions, guaranteeing the fulfillment of the defined specifications. The reliability of the proposed approach has been tested through significant simulation scenarios. The controller has been installed on the considered real plant, replacing local standalone controllers manually conducted by plant operators. Very satisfactory field results have been achieved, both on process control and energy efficiency improvement. Optimized trade-offs between energy saving, environmental impact decreasing, product quality improvement and production maximization have been guaranteed. Consequently, Italian energy efficiency certificates have been obtained. The formulated steel industry reheating furnaces control method has been patented.

1. Introduction

Process industries were recently observing significant changes, due to the need to increase the automation features. This requirement is strictly related to the increasingly stringent environmental and pollution standards that are connected to the need to improve the energy efficiency level [1]. For this purposes, Advanced Process Control (APC) solutions are receiving the attention of process engineers [2]. Significant benefits can be obtained with respect to standalone controllers, both from process control and energy efficiency achievement and improvement point of view [3,4]. The magnitude of these results is directly tied to the amount of energy required for the considered processes.

Steel industry represents a high-energy intensive process industry: it is characterized by different complex phases that have to be carefully managed. Figure 1 briefly depicts the production chain of a steel industry [5]. Initially, raw materials (e.g. waste materials) are processed (Figure 1, *Raw Materials Processing Phase*) in order to obtain steel bars at an intermediate stage of manufacture, e.g. *billets*. Billets are then introduced in a reheating furnace in order to be suitably reheated (Figure 1, *Reheating Phase*). There are different typologies of reheating furnaces, distinguished by the furnace movement methodology: for example, in a *pusher type* reheating furnace, billets are moved along the furnace through the action of pushers. Billets can enter the furnace at different temperatures. In the case study that will be proposed in the present paper, i.e. a pusher type billets reheating furnace located in an Italian steel plant, the billets inlet temperature range is 20 [*°C*] – 700 [*°C*]. Through the combustion reactions triggered by some burners, i.e. air/fuel burners, billets are exposed to heating reactions during their path along the furnace. The billets must be reheated in order to fulfill the specifications required for the subsequent plastic deformation phase (Figure 1, *Rolling Phase*). In the proposed case study, a range example for billets discharge (furnace outlet) temperature is 1000 [*°C*] – 1100 [*°C*]. From the final phase, the finished products, e.g. iron rods or tube rounds, are obtained.

*Corresponding Author: Silvia Maria Zanoli, Università Politecnica delle Marche, Ancona (AN), 60131, Italy. Email: s.zanoli@univpm.it

https://dx.doi.org/10.25046/aj030209

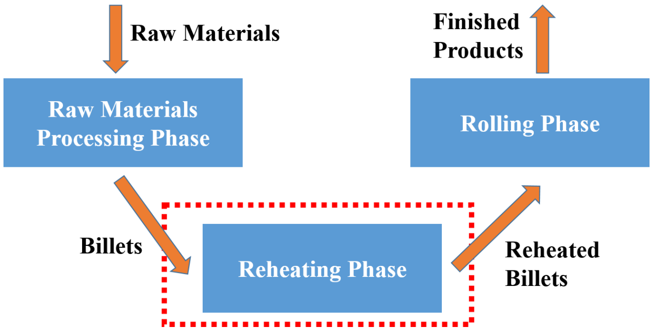


Figure 1. General production chain of a steel industry.

In a steel industry, the previously described *Reheating Phase* represents a crucial phase from an energy efficiency and products quality point of view [6]. In order to guarantee the desired trade-offs between the conflicting challenges of the steel industry *Reheating Phase*, i.e. energy saving and environmental impact decreasing versus product quality and production maximization, different approaches have been proposed in the control literature, ranging from classic to innovative techniques. In [7], a nonlinear optimization problem is formulated through a genetic algorithms approach; the minimization of fuel cost and the satisfaction of a desired discharge temperature represent the control objectives. In [8], two models and two related tracking control systems are proposed. In [9], a mixed neural network/heat transfer model approach is exploited, achieving an integrated intelligent control method. In [10], a nonlinear predictive approach is exploited for controlling a steel slabs continuous reheating furnace. A first principles mathematical model is developed that allows the controller to define the appropriate local furnace temperatures required for the achievement of the desired slabs discharge temperatures. The enormous potential of model-based control and optimization for steel reheating furnaces is described in [11], where one of the proposed control approaches is based on a transient nonlinear furnace model.

This paper is an extension of work originally presented in the 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP) [12]. In [12], the authors have presented an APC framework for the optimization of a pusher type billets reheating furnace (*Reheating Phase*) located in an Italian steel plant. Some details about the process features, the controller formulation and field results have been provided, highlighting the achieved benefits with respect to the previous control system, based on local temperature controllers managed by plant operators. In this paper, in addition to the previously reported literature analysis, additional aspects about the process modellization are provided. Furthermore, more details on simulation results are given through specific scenarios. Finally, results related to the installation of the developed APC system, denoted "i.Process | Steel – RHF", on the Italian industrial plant are depicted and discussed, analyzing process control and energy efficiency aspects.

The paper organization is the following: Section 2 details the main process features, focusing on the control specifications and on the formulated process model. Details on the obtained relationships between the main process variables are provided. Section 3 depicts the control architecture, detailing the formulation of the two MPC modules. A significant plant scenario is reported in Section 4, where a typical condition is simulated and managed with the activation of the developed APC system. In Section 5, results related to the installation on the Italian steel industry are reported, focusing on process control performances and on energy efficiency evaluation with respect to the computed baseline. Section 6 reports the conclusions.

2. Process description and modellization

This section reports the description of the considered case study, i.e. a pusher type billets reheating furnace located in an Italian steel plant. Furthermore, the formulated process model is described.

2.1. Pusher Type Reheating Furnace

As previously described, a pusher type billets reheating furnace located in an Italian steel plant represents the case study that is proposed in this paper for the illustration of the main peculiarities of "i.Process | Steel – RHF" APC system. Figure 2 reports a synoptic of the developed Graphical User Interface (GUI) where a schematic representation of the considered reheating furnace is given. Note the three furnace areas: *Preheating Area* (green rectangle), *Heating Area* (yellow rectangle), *Soaking Area* (red rectangle). The main features of the furnace areas have been reported in Table 1. The furnace billets capacity is 136 (m_b=136); the billets are moved along the furnace based on the defined furnace production rate (up to 120 [t/h]). Billets enter the furnace through the *Preheating Area* (Figure 2, left side). In their path along this furnace area, they cross a unique zone (*tunnel*, 4.733 [m] length) that is not equipped with an own burners set. Subsequently, billets are moved towards the *Heating Area* that is constituted by zone 6 (3.477 [m] length), zone 5 (6.4 [m] length), and zone 4 (3.2 [m] length). All these furnace zones have air/fuel burners (Figure 2, yellow circles). Finally, billets enter the last furnace area, i.e. *Soaking Area* (Figure 2, right side). Three zones constitute this furnace area, characterized by an own burners set (Figure 2, red circles): zone 3 (4.546 [m] length), zone 2 (3.2 [m] length), and zone 1 (3.2 [m] length). Zone 2 and zone 1 are vertically disposed. Table 2 contains some billets features related disposed. Table 2 contains some billets features related to the considered case study.

Flowmeters detect air and fuel (natural gas) flow rates. The furnace zones temperatures are measured through thermocouples. Furnace and air pressures are measured by manometers placed near the furnace inlet. Billets temperature is measured by optical pyrometers only at the furnace inlet and after the billets furnace discharge (in the rolling mill area). Before the installation of "i.Process | Steel – RHF" APC system, no information was available about billets heating profile within the furnace. Plant operators managed the local Proportional Integral Derivative (PID) temperature controllers; they try to ensure the desired billets discharge temperature exploiting their experience and skills. However, due to the multivariable, nonlinear and time- varying characteristics of the considered process, the observed billets furnace discharge temperature was often higher than the minimum required temperature.

Through deepened process preliminary studies, significant energy efficiency margins (fuel specific consumption minimization) have been identified. For this reason, an APC

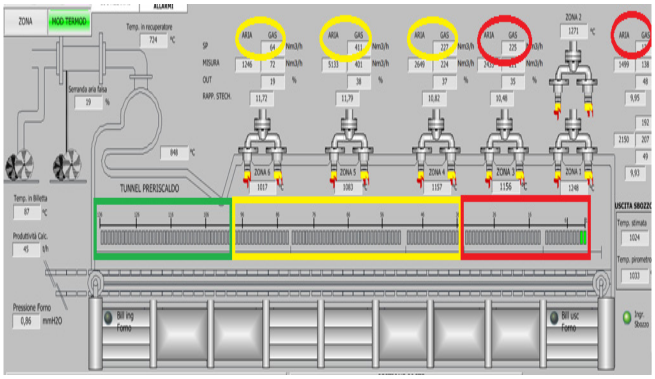


Figure 2. Pusher type reheating furnace synoptic.

Table 1: Details of furnace areas

Area	Billets Number / Area Length	Temperature	Acronym [*Units*]
Preheating	38 / 4.733 [*m*]	Tunnel Temp.	*Tun* [*°C*]
Heating	64 / 13.077 [*m*]	Zone 6 Temp.	$Temp_6$ [*°C*]
		Zone 5 Temp.	$Temp_5$ [*°C*]
		Zone 4 Temp.	$Temp_4$ [*°C*]
Soaking	34 / 7.746 [*m*]	Zone 3 Temp.	$Temp_3$ [*°C*]
		Zone 2 Temp.	$Temp_2$ [*°C*]
		Zone 1 Temp.	$Temp_1$ [*°C*]

Table 2: Billets details

Billets Feature	Case Study Billets Feature Value
Mass	2.2815 [*t*]
Length	9 [*m*]
Section	0.2 [*m*] × 0.16 [*m*]
Inlet Temperature	20 [*°C*] – 700 [*°C*]
Discharge Temperature	1000 [*°C*] – 1100 [*°C*]

system customized for the analyzed process has been developed. In particular, as it will be described in Section 3, a Model Predictive Control (MPC) strategy based on linear models has been formulated [13-15].

2.2. Process Modellization

In order to design an APC system, the availability of billets temperature estimations during their path within the furnace has been evaluated as a crucial milestone to satisfy. For this purpose, a virtual sensor has been developed [16]. The virtual sensor implements, for each billet that at each control instant lies in the furnace, a first principles adaptive nonlinear model. In this model, the involved heat phenomena and the billets movement information have been included. The inputs of the model are represented by the temperatures of the five furnace zones that are closer to the furnace inlet (tunnel, zone 6, zone 5, zone 4, and zone 3) and by the mean ($TempM_{21}$ [*°C*]) of the temperatures related to the two furnace zones that are closer to the furnace outlet (zone 2 and zone 1). The developed model has been based on the conduction model:

$$\dot{Q}_{cond} = -\lambda A \frac{dT}{dx} \quad [W] \tag{1}$$

In (1), A [m^2] represents the area related to the billet section that is normal to the heat transfer direction, λ [$W/(m{\cdot}K)$] is the billet thermal conductivity and dT/dx [K/m] is the temperature variation along the considered layer direction. The model reported in (1) can be customized with the needed number of billet layers. The convection and radiation phenomena have been considered through the following equations [17]:

$$\dot{Q}_{conv} = hA(T_{bill} - T_{env}) \quad [W] \tag{2}$$

$$\dot{Q}_{rad} = \varepsilon\sigma A(T_{bill}^4 - T_{env}^4) \quad [W] \tag{3}$$

In (2)-(3), A [m^2] represents the area related to the exposed surface, h [$W/(m^2{\cdot}K)$] is the convection heat transfer coefficient, T_{bill} [K] is the billet temperature and T_{env} [K] is the environment temperature of the fluid around the billet. σ is the Stefan-Boltzmann constant and ε is the emissivity coefficient [17].

Through the equations reported in (1)-(3), a discretized model has been formulated for each billet that, at each considered sampling instant, lies in the reheating furnace. An important remark is the lack of an exact knowledge of the involved heat transfer coefficients. Online adaptation procedures for them have been included, based on customized constrained optimization problems that have been formulated exploiting feedback information. Figures 3-5 show an example of the virtual sensor field results related to October 2016. In Figure 3, blue stars indicate the measurements provided by the optical pyrometer in the rolling mill area; green stars represents the related virtual sensor temperature estimations. In Figures 4-5, trends related to the inputs of the virtual sensor model and related to the furnace production rate have been reported, respectively. Using the Root Mean Square Error of Prediction (RMSEP) as a performance indicator for the virtual sensor estimation, an RMSEP less than 10 [*°C*] has been detected (about 1 [*%*] of an optical pyrometer measurement range). In Figure 4, examples of zone temperatures ranges can be observed. Note the increasing monotonicity of the temperatures according to the proximity to the furnace outlet.

The virtual sensor nonlinear model has then been linearized, in order to include billets temperatures estimations within an MPC strategy based on linear models. A Linear Parameter-Varying (LPV) model has been obtained [18]. The billets temperatures have been included in an ad hoc group of Controlled Variables (CVs), denoted as bCVs (*b*) group. In the considered case study, the bCVs group is composed by 136 elements (m_b=136). All terms involved in the inputs vector related to the bCVs model, i.e. the temperatures of all furnace zones, have been included in an own group, called as *zones* Controlled Variables (zCVs, *y*) group. This group also includes temperature differences between adjacent furnace zones, smoke-exchanger temperature (T_{SE}, [*°C*]), and fuel valves opening position [12]. As typical in industrial APC applications, other two categories of measured input process variables have been defined: Manipulated Variables (MVs, *u*) and Disturbance Variables (DVs, *d*). The MVs group includes the six fuel flow rates ($Fuel_i$ (i=1-6), [Nm^3/h]) and the six stoichiometric ratios (R_i (i=1-6), []) related to the furnace zones equipped with an own burners set. In DVs group,

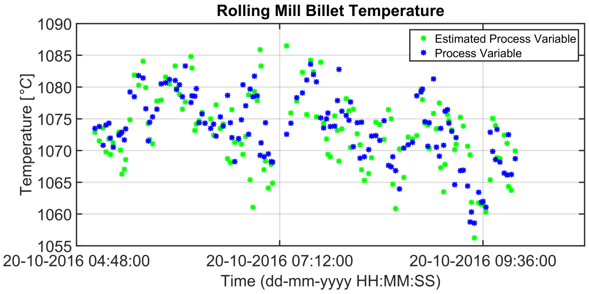


Figure 3. Virtual sensor results: rolling mill billets temperature.

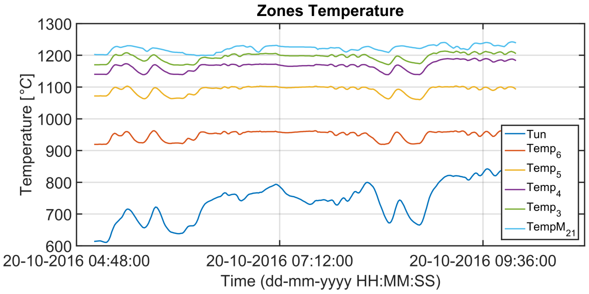


Figure 4. Virtual sensor results: inputs.

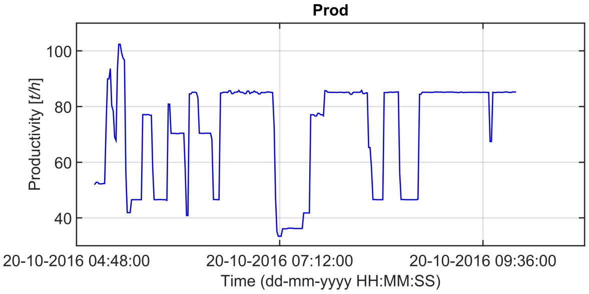


Figure 5. Virtual sensor results: furnace production rate.

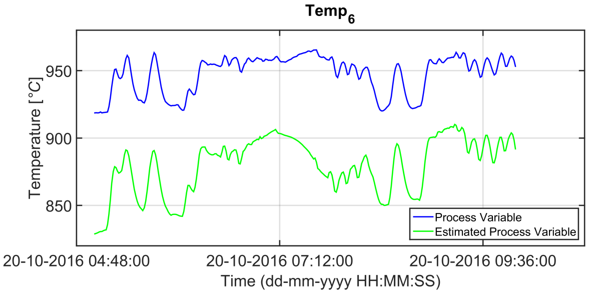


Figure 6. Zone 6 temperature model performances.

the furnace production rate (*Prod*, [*t/h*]), the furnace pressure (*FurnPress*, [*mmH2O*]) and the air pressure (*FurnPress*, [*mbar*]) have been included. The zCVs-MVs/DVs relationships have been modelled through asymptotically stable linear time invariant models without delays on the input-output channels. Steady-state gain signs related to some zCVs-MVs/DVs models have been reported in Tables 3-4. The empty boxes indicate the absence of a relationship.

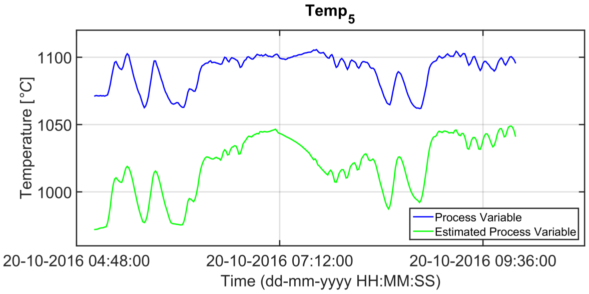


Figure 7. Zone 5 temperature model performances.

Table 3: zCVs-MVs models: steady-state gains signs

Acronym	$Fuel_6$	$Fuel_5$	$Fuel_4$	$Fuel_3$	$Fuel_2$	$Fuel_1$
Tun	+	+	+	+	+	+
$Temp_6$	+	+	+	+	+	+
$Temp_5$		+	+	+	+	+
$Temp_4$			+	+	+	+
$Temp_3$				+	+	+
$Temp_2$					+	+
$Temp_1$					+	+
T_{SE}	+	+	+	+	+	+

Table 4: zCVs-DVs models: steady-state gains signs

Acronym	*FurnPress*	*Prod*	*AirPress*
Tun	+	-	-
$Temp_6$	+	-	-
$Temp_5$	+	-	-
$Temp_4$	+	-	-
$Temp_3$	+	-	-
$Temp_2$	+	-	-
$Temp_1$	+	-	-
T_{SE}	+	-	-

Note, in Table 3, that the temperatures of the furnace zones located closer to the furnace inlet are influenced by the fuel flow rates of the downstream zones. Figures 6-7 report examples of the performances of the models related to zone 6 and zone 5 temperatures. Green lines represent the model prediction while blue lines represent the field process variables. The depicted trends refer to the same period taken into account in the scenario proposed in Figures 3-5.

3. “i.Process | Steel – RHF” APC system

In this section, “i.Process | Steel – RHF” APC system technology is described, focusing on the optimization problems solved by the two MPC modules and on tuning procedures.

3.1. APC Architecture Description

Figure 8 schematically describes “i.Process | Steel – RHF” APC system architecture. At each control instant *k*, a Supervisory Control and Data Acquisition (*SCADA*) system provides updated data; note the location of the developed billets temperature estimation virtual sensor. A *Data Conditioning & Decoupling Selector* (DC&DS) block receives updated process variables values (Figure 8, right side, *u(k-1), d(k), y(k), b(k)*). Furthermore, DC&DS block exploits other information, i.e. local control loops conditions (Figure 8, *Plant Signals & Parameters*), status information related to the selected process variables (Figure 8, right side, *u-d-y-b Status*), and control requirements related to MVs manipulation (Figure 8, right side, *Decoupling Matrix*) [19,20]. The status information related to the selected process variables defines which process variables have to be included in the MPC problem at each control instant. The *Decoupling Matrix* defines which MVs have to be moved by MPC strategy for the satisfaction of the specifications related to the zCVs. Table 5 shows the initial *Decoupling Matrix* provided to DC&DS block. For example, considering the zone 6 temperature, this zCVs is tied to all fuel flow rates (see Table 3); for this reason, without any expedients, all fuel flow rates could be moved by MPC system for the satisfaction of the zone 6 temperature specifications. According to some additional specifications that have been defined for the considered case study, only zone 6 fuel flow rate should be moved for the cited zCV. With regard to this aspect, observe the second row of Table 5 (related to zone 6 temperature): the “0” value indicates the MVs to be inhibited for its control [20].

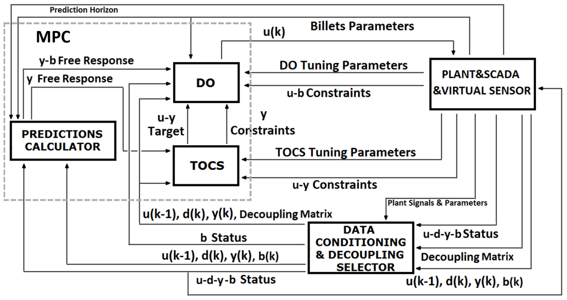


Figure 8. “i.Process | Steel – RHF” architecture.

Table 5: zCVs-MVs models: initial decoupling matrix

Acronym	$Fuel_6$	$Fuel_5$	$Fuel_4$	$Fuel_3$	$Fuel_2$	$Fuel_1$
Tun	1	1	0	0	0	0
$Temp_6$	1	0	0	0	0	0
$Temp_5$	1	1	0	0	0	0
$Temp_4$	1	1	1	0	0	0
$Temp_3$	1	1	1	1	0	0
$Temp_2$	1	1	1	1	1	0
$Temp_1$	1	1	1	1	0	1
T_{SE}	1	1	0	0	0	0

Based on the received information, DC&DS block performs different operations and checks, e.g. bad detection and data conditioning. From this phase, the possibly modified process variables values (Figure 8, left side, *u(k-1), d(k), y(k), b(k)*), the final status values for process variables (Figure 8, left side, *u-d-y-b Status*) and the final *Decoupling Matrix* (Figure 8, left side, *Decoupling Matrix*) resulted. The final *Decoupling Matrix* contains information about process variables final status values and about MVs action inhibition specifications [20].

MPC block, based on all the detailed parameters and information, computes the current optimal input to be applied to the plant (Figure 8, *u(k)*). *MPC* block has been based on a two-layer scheme formulated on linear models, constituted by an upper layer module called as Targets Optimizing and Constraints Softening (Figure 8, *TOCS*) and by a lower layer module called as Dynamic Optimizer (Figure 8, *DO*). A *Predictions Calculator* module cooperates with the two-layer scheme.

3.2. Two-Layer MPC Scheme

The MPC scheme exploits process variables predictions on a prediction horizon H_p. The zCVs and bCVs *free response* ([13]) is computed by *Predictions Calculator* module (Figure 8, *y Free Response* and *y-b Free Response*).

At the upper layer of the proposed MPC scheme, *TOCS* module solves a Linear Programming (LP) problem. A linear cost function is minimized, subject to linear constraints:

$$V_{TOCS}(k) = c_u{}^T \cdot \Delta\hat{u}_{TOCS}(k) + \rho_{y\_TOCS}{}^T \cdot \varepsilon_{y\_TOCS}(k) \qquad (4)$$

subject to

$$\begin{aligned} &\text{i. } lb_{du\_TOCS} \le \Delta\hat{u}_{TOCS}(k) \le ub_{du\_TOCS} \\ &\text{ii. } lb_{u\_TOCS} \le \hat{u}_{TOCS}(k) \le ub_{u\_TOCS} \\ &\text{iii. } lb_{y\_TOCS} - \gamma_{lby\_TOCS} \cdot \varepsilon_{y\_TOCS}(k) \le \hat{y}_{TOCS}(k) \le \\ &\qquad \le ub_{y\_TOCS} + \gamma_{uby\_TOCS} \cdot \varepsilon_{y\_TOCS}(k) \\ &\text{iv. } \varepsilon_{y\_TOCS}(k) \ge 0 \end{aligned} \qquad (5)$$

In the LP problem represented by (4)-(5), minimization or maximization directions for MVs can be preferred through c_u term that multiplies the MVs steady-state move $\Delta\hat{u}_{TOCS}(k)$. $\Delta\hat{u}_{TOCS}(k)$ is constrained by lb_{du_TOCS} and ub_{du_TOCS}; with regard to MVs, the steady-state value $\hat{u}_{TOCS}(k)$ is constrained by lb_{u_TOCS} and ub_{u_TOCS}. zCVs steady state value $\hat{y}_{TOCS}(k)$ is constrained by lb_{y_TOCS} and ub_{y_TOCS}. *TOCS* MVs constraints have been considered as *hard* constraints: they can never be violated and their feasibility has been suitably imposed. On the other hand, *TOCS* zCVs constraints have been considered as *soft* constraints: they can be violated thanks to the slack variables contained in $\varepsilon_{y\_TOCS}(k)$ term. This term contains two nonnegative slack variables for each zCV; it has been introduced in (4) through ρ_{y_TOCS} term and in (5) through γ_{lby_TOCS} and γ_{uby_TOCS} terms.

In Figure 8, lb_{du_TOCS}, ub_{du_TOCS}, lb_{u_TOCS}, ub_{u_TOCS} are among *u-y Constraints*. c_u, ρ_{y_TOCS},γ_{lby_TOCS} and γ_{uby_TOCS} terms are among *TOCS Tuning Parameters*.

TOCS module formulation exploits zCVs-MVs/DVs models and predictions at the end of the prediction horizon H_p. Solving its LP problem, *TOCS* module computes MVs and CVs steady-state targets (Figure 8, *u- y Target*) and zCVs constraints (Figure 8, *y Constraints*); these terms are provided to *DO* module [19,20].

At the lower layer of the proposed MPC scheme, *DO* module computes the H_u (H_u is denoted as control horizon [13]) optimal MVs moves, solving a Quadratic Programming (QP) problem. These moves are included in a $\Delta\hat{u}(k+M_i|k)$ vector ($i=1,\dots,H_u$). M_i represent the MVs movement instant ($M_1=0$; $M_i<H_p$) [12]. A quadratic cost function is minimized, subject to linear constraints:

$$\begin{aligned} V_{DO}(k) = {} & \textstyle\sum_{i=0}^{H_p-1}\|\hat{u}(k+i|k)-u_t(k+i|k)\|^2_{\mathcal{S}(i)} + \\ & +\textstyle\sum_{i=1}^{H_p}\|\hat{y}(k+i|k)-y_t(k+i|k)\|^2_{Q(i)} + \\ & +\textstyle\sum_{i=1}^{H_u}\|\Delta\hat{u}(k+M_i|k)\|^2_{\mathcal{R}(i)} + \|\varepsilon_y(k)\|^2_{\rho_y} + \\ & +\textstyle\sum_{j=1}^{m_b}\left\|\hat{b}_j(k+e_j|k)-lb_{b\_DO_j}\right\|^2_{T_j} + \|\varepsilon_b(k)\|^2_{\rho_b} \end{aligned} \tag{6}$$

$$\begin{aligned} &\text{subject to} \\ &\text{i. } lb_{du\_DO}(i) \le \Delta\hat{u}(k+M_i|k) \le ub_{du\_DO}(i), i=1,\dots,H_u \\ &\text{ii. } lb_{u\_DO}(i) \le \hat{u}(k+M_i|k) \le ub_{u\_DO}(i), i=1,\dots,H_u \\ &\text{iii. } lb_{y\_DO}(i) - \gamma_{lby\_DO}(i)\cdot\varepsilon_y(k) \le \hat{y}(k+i|k) \le \\ &\qquad \le ub_{y\_DO}(i) + \gamma_{uby\_DO}(i)\cdot\varepsilon_y(k), i=1,\dots,H_p \\ &\text{iv. } lb_{b\_DO_j} - \gamma_{lbb\_DO_j}\cdot\varepsilon_{b_j}(k) \le \hat{b}_j(k+e_j|k) \le \\ &\qquad \le ub_{b\_DO_j} + \gamma_{ubb\_DO_j}\cdot\varepsilon_{b_j}(k), j=1,\dots,m_b \\ &\text{v. } \varepsilon_y(k)\ge 0;\ \varepsilon_b(k)\ge 0 \end{aligned} \tag{7}$$

In the QP problem represented by (6)-(7), tracking errors over H_p between MVs and zCVs reference trajectories ($u_t(k+i|k)$ and $y_t(k+i|k)$) and the related predicted values ($\hat{u}(k+i|k)$ and $\hat{y}(k+i|k)$) are weighted by $\mathcal{S}(i)$ and $Q(i)$ positive semidefinite matrices. $\hat{u}(k+i|k)$ and $\hat{y}(k+i|k)$ are constrained by lb_{u_DO}, ub_{u_DO}, lb_{y_DO} and ub_{y_DO}. *DO* MVs moves are weighted in (6) by positive definite matrices $\mathcal{R}(i)$ and they are constrained in (7) by lb_{du_DO} and ub_{du_DO}. *DO* MVs constraints have been considered as *hard* constraints: they can never be violated and their feasibility has been suitably imposed. On the other hand, *DO* zCVs constraints have been considered as *soft* constraints: they can be violated thanks to the slack variables contained in $\varepsilon_y(k)$ term. This term contains a nonnegative slack variable for each zCV; it has been introduced in (6) through ρ_y term and in (7) through γ_{lby_DO} and γ_{uby_DO} terms.

In Figure 8, lb_{du_DO}, ub_{du_DO}, lb_{u_DO}, ub_{u_DO} are among *u-b Constraints*. lb_{y_DO} and ub_{y_DO} terms are among *y Constraints*. $\mathcal{S}(i)$, $Q(i)$, $\mathcal{R}(i)$, ρ_y, γ_{lby_DO} and γ_{uby_DO} terms are among *DO Tuning Parameters*.

Taking into account the just described terms related to MVs and zCVs, and exploiting zCVs-MVs/DVs models, a first “i.Process | Steel – RHF” APC system control mode has been formulated, denoted *zones* APC mode.

Including terms related to bCVs in the *DO* QP problem represented by (6)-(7), a second control mode for “i.Process | Steel – RHF” APC system has been obtained, denoted *adaptive* APC mode. It constitutes the main control mode and it exploits, besides zCVs-MVs/DVs models, also first principles bCVs LPV model and billets virtual sensor information. In this way, an adaptive two-layer MPC strategy has been formulated. In (6)-(7), for the generic billet that lies on the j ($j=1\dots m_b$) furnace place at the current control instant k, its predicted furnace discharge instant is computed (e_j) taking into account the furnace production rate. The billets temperature predictions ($\hat{b}_j(k+e_j|k)$) at the related furnace discharge instants e_j ($j=1\dots m_b$) are constrained in (7) by $lb_{b\_DO_j}$ and $ub_{b\_DO_j}$ (Figure 8, *u-b Constraints*). These constraints have been considered as *soft* constraints: they can be violated thanks to the slack variables contained in $\varepsilon_b(k)$ term. This term contains a nonnegative slack variable for each bCV; it has been introduced in (6) through ρ_b term and in (7) through γ_{lbb_DO} and γ_{ubb_DO} terms (Figure 8, *DO Tuning Parameters*). Tracking option of desired values ($lb_{b\_DO_j}$) for bCVs has been included in *DO* cost function (6), exploiting T_j nonnegative scalars.

3.3. Tuning Details

Tailored tuning methods have been developed for optimizing the controller performances in the two defined control modes. The control moves are computed by the APC system once a minute, according to the formulated process model.

With regard to the choice of the prediction horizon H_p, it varies based on the control mode that has to be exploited. Consequently, also the control horizon H_u and the MVs movement instants M_i are adapted. For example, in the simulation example that will be proposed in Section 4, the *adaptive* APC mode will be activated. In this case, a furnace movement time equal to about 95 [*s*] is assumed, which corresponds, for the present case study, to a furnace production rate equal to about 85 [*t/h*]. Accordingly, in order to guarantee the predicted reaching of the furnace outlet to the billet closer to the furnace inlet, a prediction horizon H_p of 216 [*min*] is set. In a parametric way, the control horizon H_u is set equal to 44 moves suitably spaced over the prediction horizon H_p.

In *TOCS* module formulation, the elements of c_u have been set as positive, in order to prefer, within the process variables defined constraints, minimization directions for fuel flow rates and stoichiometric ratios. Furthermore, ρ_{y_TOCS}, γ_{lby_TOCS} and γ_{uby_TOCS} terms have been set in order to guarantee the desired priority on constraints satisfaction. For example, constraints related to smoke-exchanger temperature are more important than those related to zones temperature: the related ρ_{y_TOCS},γ_{lby_TOCS} and γ_{uby_TOCS} terms have been set accordingly to this specification, taking into account also the magnitude of the involved process variables.

In *DO* module formulation, common tuning aspects between the two formulated control modes have been proposed. The priority of the soft constraints terms (if these constraints are included in the controller setup) ρ_y, γ_{lby_DO} and γ_{uby_DO} is guaranteed; furthermore, zCVs constraints are always present within *DO* module formulation. The option of tracking desired reference trajectories related to MVs is another similar tuning aspect, together with the need to take into account the magnitude of MVs moves. In *DO* module formulation related to the *adaptive* APC mode, billets (tracking and/or constraints satisfaction) specifications can be considered; optimal trade-offs between control and energy efficiency specifications must be ensured.

4. Simulation Results

This section proposes some simulation results related to "i.Process | Steel – RHF" APC system. In particular, a simulation scenario where the *adaptive* APC mode is activated is described.

4.1. A Simulated Scenario

The *adaptive* APC mode performances are shown through a simulated scenario: the zCVs-MVs/DVs plant model exploits the identified zCVs-MVs/DVs model and no measurement noise is assumed. The plant model for simulating the relationships between billets temperature and furnace zones temperature is based on the developed billets temperature nonlinear model that is exploited by the virtual sensor.

At the initial control instant of the proposed simulation, the *adaptive* APC mode is requested to be activated. The virtual sensor estimation gives reliable results and bCVs (billets temperature) can be included in the control problem. The zCVs reported in Table 1, the temperature differences between the bCVs model inputs and all fuel flow rates represent the other process variables that are considered in the simulation. The other MVs and all DVs are considered constant, so not influencing the proposed simulation scenario. Constraints related to zCVs and fuel flow rates have been reported in Tables 6-7. The temperature differences between the bCVs model inputs are constrained so as to ensure an increasing monotonicity of the temperatures along the furnace. As mentioned in Subsection 3.3, the furnace production rate is equal to about 85 [*t/h*]; the temperature of the 136 billets that initially lie within the furnace is in the range 20 [°*C*] - 1140 [°*C*], while the billets that will enter the furnace during the simulation are characterized by a temperature of about 550 [°*C*]. The *Rolling Phase* specifications require that the billets temperature in the rolling area must be in the range 1030 [°*C*] - 1045 [°*C*]. This specification can be converted in a temperature range of about 1051 [°*C*] - 1066 [°*C*] at the furnace outlet. The "i.Process | Steel – RHF" *adaptive* APC mode ensures that the billets temperature detected by the optical pyrometer in the rolling mill area converges towards the minimum required temperature (1030 [°*C*]), as can be observed in Figure 9.

Table 6: Simulation scenario: zCVs constraints

Acronym	Upper Constraint	Lower Constraint
Tun	950 [°*C*]	550 [°*C*]
$Temp_6$	1150 [°*C*]	800 [°*C*]
$Temp_5$	1150 [°*C*]	800 [°*C*]
$Temp_4$	1200 [°*C*]	800 [°*C*]
$Temp_3$	1250 [°*C*]	1000 [°*C*]
$Temp_2$	1250 [°*C*]	1000 [°*C*]
$Temp_1$	1250 [°*C*]	1000 [°*C*]

Table 7: Simulation scenario: MVs constraints

Acronym	Upper Constraint	Lower Constraint
$Fuel_6$	800 [Nm^3/h]	0 [Nm^3/h]
$Fuel_5$	1600 [Nm^3/h]	0 [Nm^3/h]
$Fuel_4$	650 [Nm^3/h]	0 [Nm^3/h]
$Fuel_3$	650 [Nm^3/h]	0 [Nm^3/h]
$Fuel_2$	250 [Nm^3/h]	0 [Nm^3/h]
$Fuel_1$	250 [Nm^3/h]	0 [Nm^3/h]

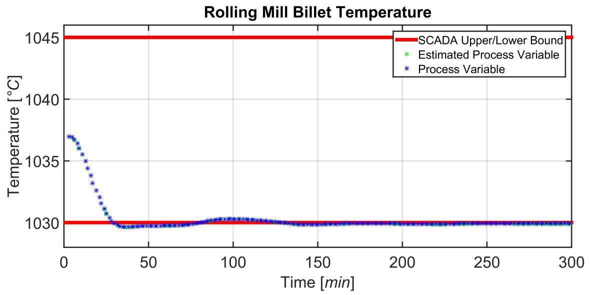


Figure 9. Simulation results: bCVs trends.

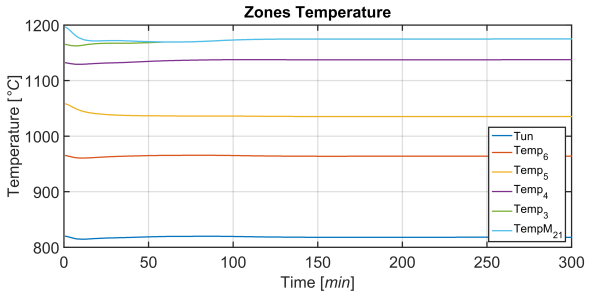


Figure 10. Simulation results: bCVs model inputs trends.

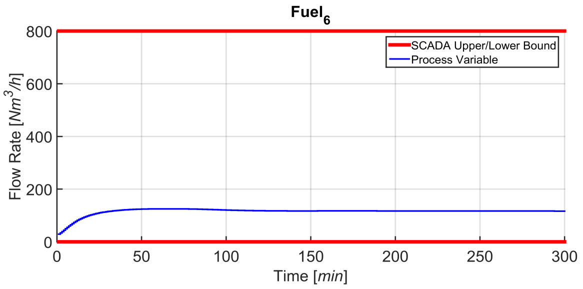


Figure 11. Simulation results: $Fuel_6$ trends.

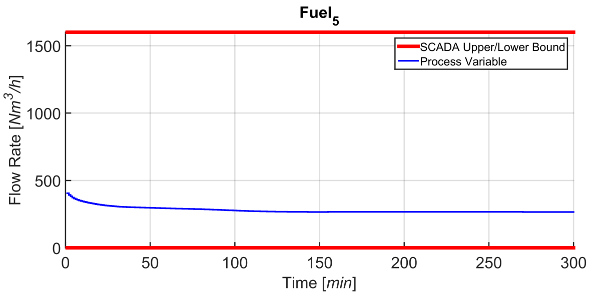


Figure 12. Simulation results: $Fuel_5$ trends.

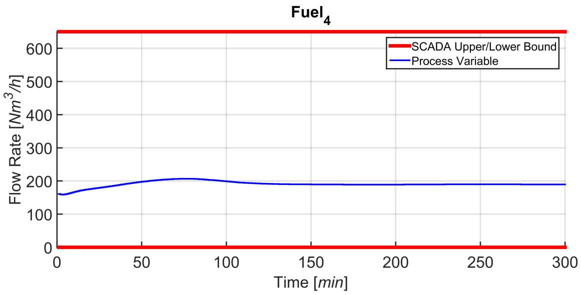


Figure 13. Simulation results: *Fuel$_4$* trends.

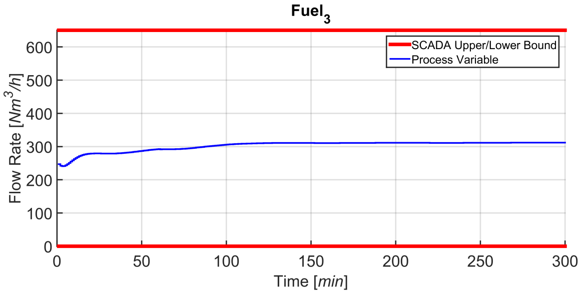


Figure 14. Simulation results: *Fuel$_3$* trends.

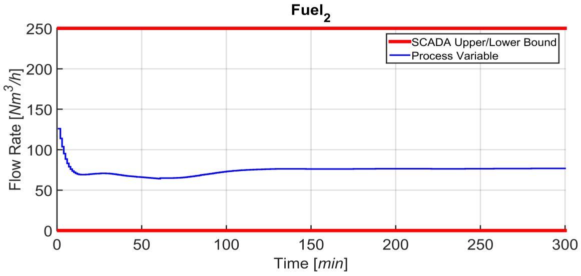


Figure 15. Simulation results: *Fuel$_2$* trends.

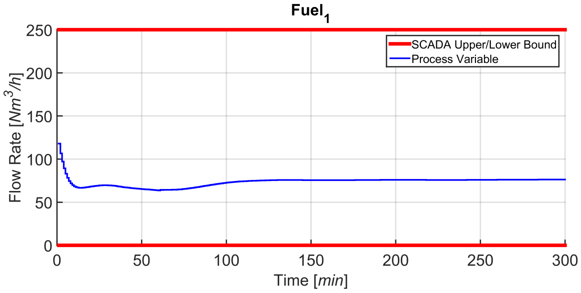


Figure 16. Simulation results: *Fuel$_1$* trends.

The cooperative action between *TOCS* and *DO* modules, that exploits virtual sensor information and billets temperature LPV model ensures a coordinated management of the furnace zones temperature (Figure 10) that is directly tied to the manipulation of the fuel flow rates (Figures 11-16). Consequently, a more profitable plant configuration is guaranteed that, at the same time, respect all the defined specifications. For example, in Figure 10, note the increasing monotonicity of the temperatures along the furnace.

5. Field Results

The study and design phases of the project related to the considered process began in January 2015 and ended in May 2015. The APC system has been installed on the considered Italian steel plant in early June 2015, substituting the local PID temperature controllers managed by plant operators. This section shows a plant scenario under the control of "i.Process | Steel – RHF" APC system and some results about the obtained fuel specific consumption.

5.1. A Plant Scenario

Figures 17-26 show a field scenario where "i.Process | Steel – RHF" APC system is active. A four hours period is taken into account. Figure 17 shows the bCVs trends: virtual sensor estimation and optical pyrometer measurements are depicted, together with the defined constraints in the rolling mill area (1125 [*°C*] - 1065 [*°C*]). The furnace production rate is shown in Figure 18, while the billets furnace inlet temperature is shown in Figure 19. The inputs related to the bCVs model are depicted in Figure 20 and Figures 21-26 show the fuel flow rates. Note the fuel flow rates trends (Figures 21-26, blue line) and the defined constraints (Figures 21-26, red lines). All MVs and DVs are considered in the control problem, together with some zCVs (furnace zones temperatures, temperature differences between adjacent furnace zones, smoke-exchanger temperature) and the bCVs (some process variables have not been shown for brevity). Examples of the constraints defined for the furnace zones and smoke-exchanger temperatures have been reported in Table 8. The 136 billets that are already present in the furnace at the beginning of the considered plant scenario are characterized by temperatures in the range 55 [*°C*] – 1100 [*°C*] and the inlet temperature of the billets that will enter the furnace is in the range 40 [*°C*] – 105 [*°C*] (Figure 19). Besides the inlet temperature of the billets, also the furnace production rate is not constant (Figure 18): it assumes a

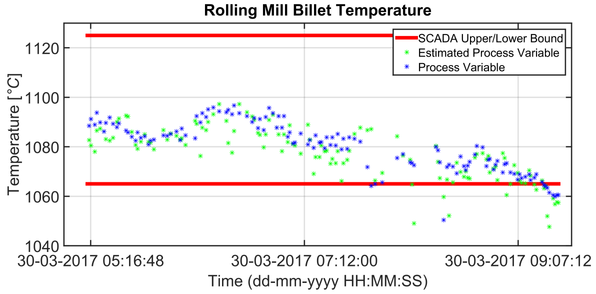


Figure 17. Field results: bCVs trends.

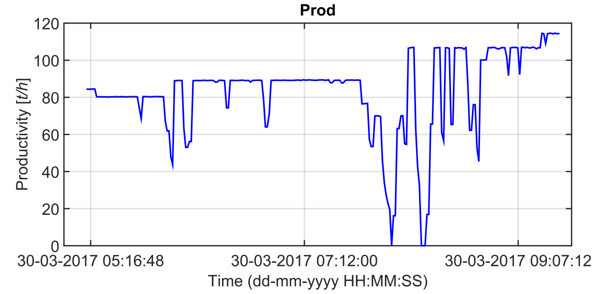


Figure 18. Field results: furnace production rate trends.

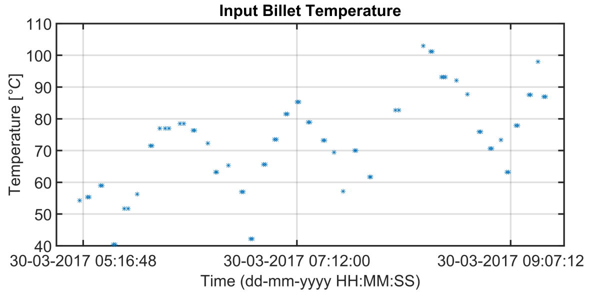


Figure 19. Field results: billets furnace inlet temperatures trends.

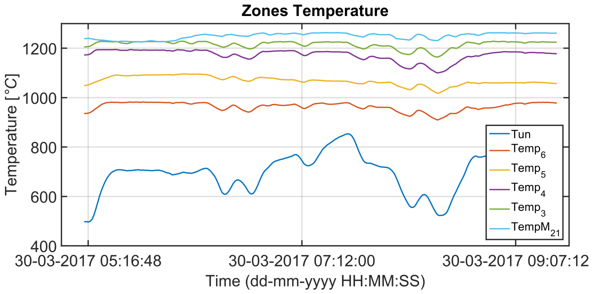


Figure 20. Field results: bCVs model inputs trends.

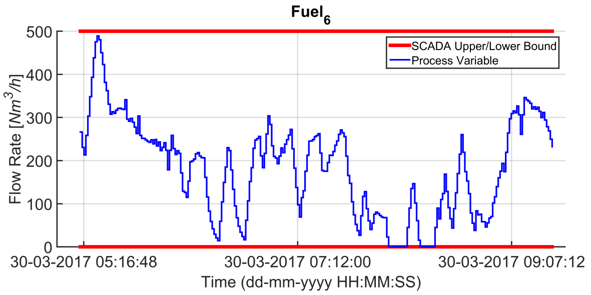


Figure 21. Field results: *Fuel*$_6$ trends.

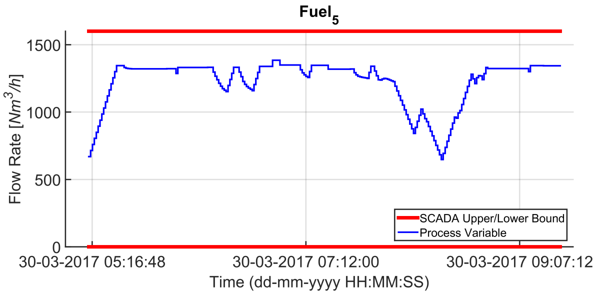


Figure 22. Field results: *Fuel*$_5$ trends.

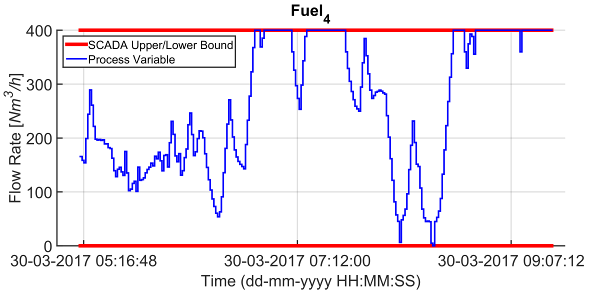


Figure 23. Field results: *Fuel*$_4$ trends.

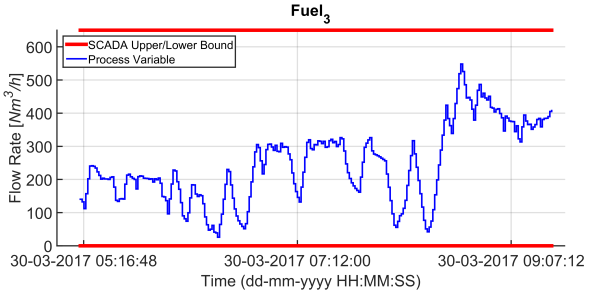


Figure 24. Field results: *Fuel*$_3$ trends.

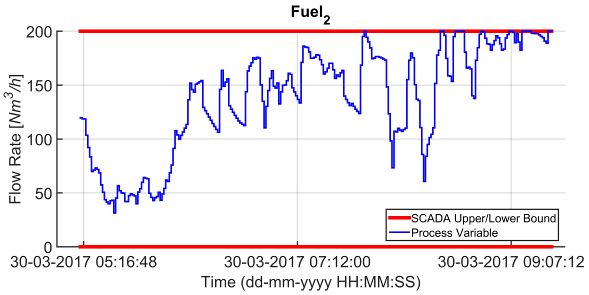


Figure 25. Field results: *Fuel*$_2$ trends.

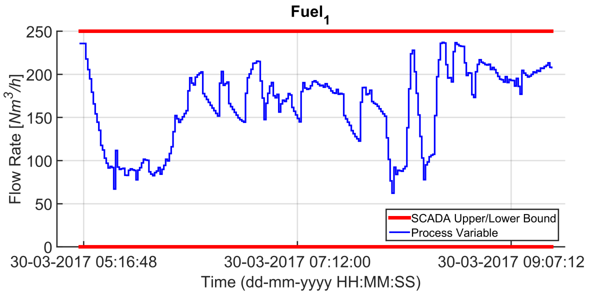


Figure 26. Field results: *Fuel*$_1$ trends.

maximum value that is equal to about 120 [*t/h*]. The furnace pressure and the air pressure (DVs) are about 0.9 [*mmH2O*] and 84 [*mbar*], respectively.

As can be observed in Figure 17, "i.Process | Steel – RHF" APC system ensures that the billets temperature detected by the optical pyrometer in the rolling mill area converges towards the minimum required temperature (1065 [*°C*]). The *TOCS-DO* cooperative action, exploiting virtual sensor information and billets temperature LPV model, ensures a profitable management of the furnace zones temperature (Figure 20) that is directly tied to the manipulation of the fuel flow rates (Figures 21-26). All the imposed constraints and specifications are fulfilled, despite a not constant furnace production rate (Figure 18) and a not constant billets furnace inlet temperature (Figure 19). The benefits deriving from the proposed multivariable predictive approach led the controller to conduct the plant to very profitable operating regions, but, at the same time, all the control specifications are satisfied. As it will be shown in the next subsection, an energy efficiency

Table 8: Field scenario: zCVs constraints

Acronym	Upper Constraint	Lower Constraint
Tun	950 [°C]	400 [°C]
$Temp_6$	980 [°C]	900 [°C]
$Temp_5$	1100 [°C]	1050 [°C]
$Temp_4$	1190 [°C]	1140 [°C]
$Temp_3$	1220 [°C]	1190 [°C]
$Temp_2$	1260 [°C]	1220 [°C]
$Temp_1$	1260 [°C]	1220 [°C]
T_{SE}	730 [°C]	300 [°C]

improvement has been obtained with the developed APC system, with respect to the previous furnace conduction.

5.2. Fuel Specific Consumption Results

The installation of the developed controller on the real industrial plant has guaranteed an improvement on the process control that has directly influenced the fuel specific consumption. The fuel specific consumption, that takes into account the fuel usage and the furnace production rate, represents a very significant indicator for the evaluation of the energy efficiency performances of "i.Process | Steel – RHF". A project baseline for the fuel specific consumption has been computed, that varies with the furnace hot charge.

Figure 27 shows a subpart of a synoptic of the developed GUI: the daily fuel specific consumption ([Sm^3/t]) is depicted. The "i.Process | Steel – RHF" APC system daily specific consumption related to July 2017 is represented though a blue line, while the defined daily project baseline is shown through a red line. This page can be online monitored by plant operators; in this way, they can practically evaluate the controller performances from an energy efficiency point of view.

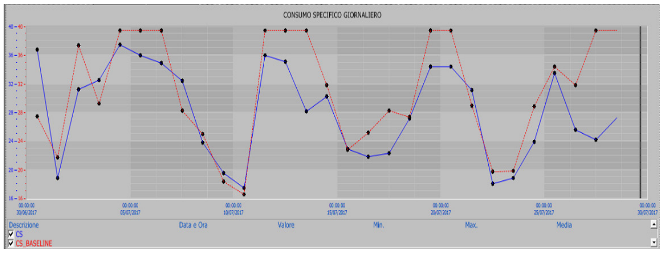


Figure 27. Field results: comparison between daily baseline and daily official specific consumption.

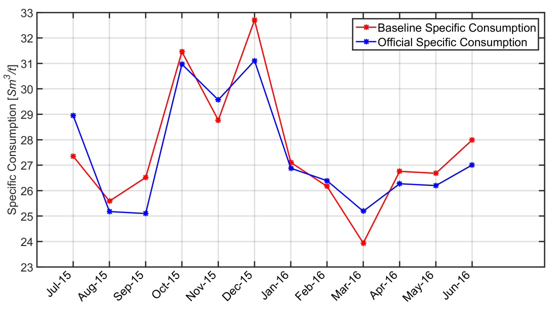


Figure 28. Field results: comparison between monthly baseline and monthly official specific consumption.

Figure 28 shows the monthly fuel specific consumption ([Sm^3/t]) related to the first year of "i.Process | Steel – RHF" APC system performances. The specific consumption is represented though a blue line, while the defined project baseline is shown through a red line. After about two years from the installation of "i.Process | Steel – RHF" APC system on the described pusher type billets reheating furnace, about 2 [*%*] reduction of the fuel specific consumption with respect to the defined project baseline has been achieved. A controller service factor about equal to 95 [*%*] has been observed.

6. Conclusions

In this paper, an Advanced Process Control system aimed at optimizing a pusher type billets reheating furnace located in an Italian steel plant has been proposed. The control system, denoted "i.Process | Steel – RHF", has been based on two-layer Model Predictive Control strategy formulated with linear models. The two-layer predictive controller also interacts with additional functional blocks.

Simulation and field results have demonstrated significant performances improvements guaranteed by the developed controller with respect to the previous control system based on Proportional Integral Derivative (PID) temperature controllers managed by plant operators. Thanks to the multivariable predictive approach, "i.Process | Steel – RHF" recognizes efficient operating zones and allows the process reaching them; in this way, process control and energy efficiency improvements have been guaranteed. Specifications related to the billets reheating are fulfilled and, at the same time, optimal configurations of the manipulated variables are reached. The developed control method has been patented [16].

After about two years from the installation of "i.Process | Steel – RHF" APC system on the described pusher type billets reheating furnace, about 2 [*%*] reduction of the fuel specific consumption with respect to the defined project baseline has been obtained, together with a controller service factor about equal to 95 [*%*].

References

[1] www.enea.it

[2] P. L. Latour, J. H. Sharpe, M. C. Delaney, "Estimating Benefits from Advanced Control" ISA Trans., **25**(4), 13–21, 1986.

[3] M. Bauer, I. K. Craig, "Economic assessment of advanced process control – A survey and framework" J. Proc. Contr., **18**(1), 2–18, 2008. https://doi.org/10.1016/j.jprocont.2007.05.007

[4] W. M. Canney, "Are you getting the full benefit from your advanced process control system?" Hydroc. Process., **84**(6), 55–58, 2005.

[5] W. Trinks, M. H. Mawhinney, R. A. Shannon, R. J. Reed, J. R. Garvey, Industrial Furnaces, John Wiley & Sons, 2004.

[6] A. Martensson, "Energy efficiency improvement by measurement and control: a case study of reheating furnaces in the steel industry" in 14th National Industrial Energy Technology Conference, Houston Texas USA, 1992. http://hdl.handle.net/1969.1/92210

[7] H. S. O. Santos, P. E. M. Almeida, R. T. N. Cardoso, "Fuel Costs Minimization on a Steel Billet Reheating Furnace Using Genetic Algorithms" Modelling and Simulation in Engineering, **2017**, Article ID 2731902, 2017. https://doi.org/10.1155/2017/2731902

[8] Z. Yi, Z. Su, G. Li, Q. Yang, W. Zhang, "Development of a double model slab tracking control system for the continuous reheating furnace" International Journal of Heat and Mass Transfer, **113**, 861–874, 2017. https://doi.org/10.1016/j.ijheatmasstransfer.2017.05.093

[9] Y. X. Liao, J. H. She, M. Wu, "Integrated Hybrid-PSO and Fuzzy-NN Decoupling Control for Temperature of Reheating Furnace" IEEE Trans. Ind. Electr., **56**(7), 2704–2714, 2009. https://doi.org/10.1109/TIE.2009.2019753

[10] A. Steinboeck, D. Wild, A. Kugi, "Nonlinear model predictive control of a continuous slab reheating furnace" Control Eng. Pract., **21**(4), 495–508, 2013. https://doi.org/10.1016/j.conengprac.2012.11.012

[11] A. Steinboeck, Model-based Control and Optimization of a Continuous Slab Reheating Furnace, Shaker Verlag GmbH, 2011.

[12] G. Astolfi, L. Barboni, F. Cocchioni, C. Pepe, S. M. Zanoli, "Optimization of a pusher type reheating furnace: an adaptive Model Predictive Control approach" in 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP), Taipei Taiwan, 2017. https://doi.org/10.1109/ADCONIP.2017.7983749

[13] J. Maciejowski, Predictive Control with Constraints, Prentice-Hall, 2002.

[14] E. F. Camacho, C. Bordons, Model Predictive Control, Springer-Verlag, 2007.

[15] J. B. Rawlings, D. Q. Mayne, Model Predictive Control: Theory and Design, Nob Hill Publishing, 2013.

[16] G. Astolfi, L. Barboni, F. Cocchioni, C. Pepe, "Metodo per il controllo di forni di riscaldo," Italian Patent n. 0001424136 awarded by Ufficio Italiano Brevetti e Marchi (UIBM), 2016. http://www.uibm.gov.it/uibm/dati/Avanzata.aspx?load=info_list_uno&id=2253683&table=Invention&

[17] Y. A. Cengel, Introduction to Thermodynamics and Heat Transfer, McGraw-Hill Companies, 2008.

[18] J. Mohammadpour, C. W. Scherer, Control of Linear Parameter Varying Systems with Applications, Springer, 2012.

[19] C. Pepe, "Model Predictive Control aimed at energy efficiency improvement in process industries," Ph.D. Thesis, Università Politecnica delle Marche, 2017.

[20] S. M. Zanoli, C. Pepe, "Two-Layer Linear MPC Approach Aimed at Walking Beam Billets Reheating Furnace Optimization" Journal of Control Science and Engineering, **2017**, Article ID 5401616, 2017. https://doi.org/10.1155/2017/5401616

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 85-89 (2018)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Estimating short time interval densities in a CTM-KF model

Arlinda Alimehaj Rrecaj*,[1], Marija Malenkovska Todorova[2]

[1]*University of Prishtina, Mechanical Engineering Faculty, Traffic and Transportation Department, 10000 Prishtina, Kosovo*

[2]*University of St.Klement Ohrid, Faculty of Technical Sciences, Transport Engineering, 7000 Bitola, Republic of Macedonia*

ARTICLE INFO

Article history:
Received: 27 January, 2018
Accepted: 21 February, 2018
Online: 08 March, 2018

Keywords:
Cell
Density
Kalman Filter

ABSTRACT

On-ramping is being widely used as e method to increase the freeway operational efficiency. The main traffic parameter that must be taken in consideration for the implementation of the feedback control strategies for the on-ramp metering is density on the main section of road. In this paper is given discretized model of traffic which is then improved by a recursive technique called Kalman-Filter with the aid of which is possible to predict the density, by only having the traffic flow measured on the start and end road section. Kalman Filter is based on linear relationship of flow and density. By minimizing the square of error between of the measurements and the estimated values of flows, a gain is derived which then is applied to the densities of the model in order to obtain the greatest accuracy of these values.

1. Introduction

The increasing demand of motorized vehicles is becoming one of the major problems that face the developed world places in nowadays. Based on some statistics in Great Britain, 90% of the journeys are by road, during the last decade, and for more the road distances travelled, have increased by over 1000% on the last sixty years. [1]

On-ramp metering [2] is being widely used as e method to increase the freeway operational efficiency by regulating the traffic interruption by minor roads, while maintaining the right of way on the major section until it reaches me critical densities, which assure the maximal flow [3]. Measure of the density on the major road is difficult to measure. In this paper is proposed a discretized traffic model (CTM) to obtain traffic densities which then will be accurate through Kalman Filter [4, 5 and 6].

2. CTM model of a Highway with three cells

If we denote with ρ_i (k) the density of a cell (uniform or non-uniform length), instead of the number of vehicles n_i on the a unit length cell, then we can bring equation *(1.1)*[6].

for density $\rho_i(k+1)$ of the cell i updated time step in *(k+1)*, where T_s is the discrete time unit in seconds.

$$\rho_i(k+1) = \rho_i(k) + \frac{T_s}{L}(q_i(k) - \ q_{i+1}(k)) \tag{1.1}$$

Analyzing a highway partitioned in three cells (for the sake of simply illustration) with an on ramp and an off ramp, by assuming that the belonging cells can be in the free flow either in congested mode the densities on each cell can be written as in equations (*1.2, 1.3, 1.4*).

The densities on each cell are:

$$\begin{gathered}\rho_{i-1}(k+1) = \rho_{i-1}(k) + \frac{T_s}{L}(q_{i-1}(k) - q_i(k) + r(k)) \text{ or} \\ \rho_1(k+1) = \rho_1(k) + \frac{T_s}{L}(q_1(k) - q_2(k) + r(k))\end{gathered} \tag{1.2}$$

$$\begin{gathered}\rho_i(k+1) = \rho_i(k) + \frac{T_s}{L}(q_i(k) - q_{i+1}(k)) \text{ or} \\ \rho_2(k+1) = \rho_2(k) + \frac{T_s}{L}(q_2(k) - q_3(k))\end{gathered} \tag{1.3}$$

$$\begin{gathered}\rho_{i+1}(k+1) = \rho_{i+1}(k) + \frac{T_s}{L}(q_{i+1}(k) - q_4(k) - f(k)) \text{ or} \\ \rho_3(k+1) = \rho_3(k) + \frac{T_s}{L}(q_3(k) - q_4(k) - f(k))\end{gathered} \tag{1.4}$$

With the elaboration of the inter-cell flow law can be defined the expressions for the inter cell flows q_1, q_2 and q_3 in the above equations (*1.1, 1.2 and 1.3)*. Before the inter cell flows elaboration is given, a reasonable description of the congestion must be given further, since as we assumed above, the cells can be in either free

*Arlinda Alimehaj Rrecaj, Email: arlinda.alimehaj@uni-pr.edu

https://dx.doi.org/10.25046/aj030210

or congested mode. Congestion is defined as the state of the traffic with high density rates, or with other words the density of that part of the highway expressed in cell is equal or higher than the critical density based on the fundamental diagram of relationship of flow and density. Referred to the mentioned diagram, can be noticed that the congested flow belongs to higher values of the density, above the critical density values where the flow drops down. That can be described with enormous number of vehicles travelling at low speeds and with short distance spaces between each other.

The common modes, used in analysis of researchers are the fully congested mode when the three cells are congested, denoted by *CCC* mode, and free flow mode when the three cells are in free flow mode, denoted by *FFF* mode. The other middle modes that are out of the scope of this paper are those with last one and two cells in congested mode, written by *FCC* and *FFC*, respectively. To emphasize the modes, the congested cells are highlighted further.

Now, for the *FFF* mode, the densities of the cells are lower than the critical density and the inter cell flows are as follows:

$$q_i(k) = \min\left(v_{fi-1}.\rho_{i-1},\ Q_{i-1}\ w_i(\rho_J - \rho_i)\right) or$$

$$q_2(k) = \min(v_{f1}.\rho_1,\ Q_1\ w_2(\rho_J - \rho_2)\ ) = v_{f1}.\rho_1 \tag{1.5}$$

$$q_{i+1}(k) = \min(v_{fi}.\rho_i,\ Q_i\ w_{i+1}(\rho_J - \rho_{i+1}),)$$

$$q_3(k) = \min(v_{f2}.\rho_2,\ Q_2\ w_3(\rho_J - \rho_3)\ ) = v_{f2}.\rho_2 \tag{1.6}$$

In *CCC* mode, the densities of the cells are higher that the critical density, and the inter cell flows are:

$$q_i(k) = \min\left(v_{fi-1}.\rho_{i-1},\ Q_{i-1}\ w_i(\rho_J - \rho_i)\right) or$$

$$q_2(k) = \min(v_{f1}.\rho_1,\ Q_1\ w_2(\rho_J - \rho_2)\ ) = w_2(\rho_J - \rho_2) \tag{1.7}$$

$$q_{i+1}(k) = \min(v_{fi}.\rho_i,\ Q_i\ w_{i+1}(\rho_J - \rho_{i+1}),)$$

$$q_3(k) = \min(v_{f2}.\rho_2,\ Q_2\ w_3(\rho_J - \rho_3)\ ) = w_3(\rho_J - \rho_3) \tag{1.8}$$

After subtracting the expressions for inter cell flows (1.*5 and 1.6*) in the equations of densities *(1.2, 1.3* and *1.4),* for the *FFF* mode, we have:

$$\rho_1(k+1) = \rho_1(k) + \frac{T_s}{L}(q_1(k) - v_{f1}.\rho_1(k) + r(k)) \tag{1.9}$$

$$\rho_2(k+1) = \rho_2(k) + \frac{T_s}{L}(v_{f1}.\rho_1(k) - v_{f2}.\rho_2(k)) \tag{1.10}$$

$$\rho_3(k+1) = \rho_3(k) + \frac{T_s}{L}(v_{f2}.\rho_2(k) - q_4(k) - f(k)) \tag{1.11}$$

And after subtracting the expressions for inter cell flows (*1.7* and *1.8*) in the equations of densities (*1.2, 1.3* and *1.4*), for the *CCC* mode, we have:

$$\rho_1(k+1) = \rho_1(k) + \frac{T_s}{L}\left(q_1(k) - w_2\left(\rho_J - \rho_2(k)\right) + r(k)\right) \tag{1.12}$$

$$\rho_2(k+1) = \rho_2(k) + \frac{T_s}{L}\left(w_2(\rho_J - \rho_2(k)) - w_3(\rho_J - \rho_3(k))\right) \tag{1.13}$$

$$\rho_3(k+1) = \rho_3(k) + \frac{T_s}{L}\left(w_3(\rho_J - \rho_3(k)) - q_4(k) - f(k)\right) \tag{1.14}$$

3. State-Space presentation of CTM Model

The state space presentation (particularly the state space, eq.*1.17*) of *CTM* based traffic densities of a highway segment in *FFF* mode differs from that of *CCC* mode. [7].

What it characterizes the state space model of the traffic density based on *CTM* model, is implication of some other extension parts of the state space, B_q which is the input matrix of upstream and downstream flows q_1 and q_4 respectively, B_r the input matrix for on ramp and off ramp flows, *r* and *f* respectively that are applicable on the *FFF* mode and input matrices, B_w which takes into consideration the backward waves w_2 and w_3 and B_J the input matrix of the jam density that are applicable on the state space model of the *CCC* mode. (1.17) and (1.18)

$$\rho_{(k+1)FFF} = A\rho_{(k)} + B_u q_{(k)} + B_{r,f}(r,f)_{(k)} + Bw_{(k)} \tag{1.17}$$

$$\rho_{(k+1)CCC} = A\rho_{(k)} + B_u u_{(k)} + B_{r,f}(r,f)_{(k)} + B_J\rho_{J(k)} + Bw_{(k)} \tag{1.18}$$

Where: *x (k+1)* is the system state vector and in this paper, according to the *CTM* model, corresponds to the density in cell of cell *i.*
A is the state matrix*, B* is the input matrix*, u(k)* is the input or control and w_k represents the process noise.
They are assumed to be independent (of each other), white, and with normal probability distributions (Gaussian) as:
$p(w)\tilde{}\ N(0,Q)\ and\ p(v)\tilde{}\ N(0,R)$.
From the system of equations in (*3.12, 3.13* and *3.14*) can be drawn (after some regulations finding partial derivatives of the functions of densities to the parameters previous densities *ρ(k),* flows q_1 and q_4 and backward waves w_2 and w_3 which provide the elements of the respective matrices of the i^{th} row that correspond to density of i^{th} cell) the belonging matrices of the system matrices.
After bringing up together (eq. in *1.19* and *1.20* can be expanded to the form of state space: For *FFF* mode:

$$\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}_{FFF}(k+1) = \begin{pmatrix}1 - v_{f1}\cdot\frac{T_s}{l_1} & 0 & 0\\ v_{f1}\cdot\frac{T_s}{l_2} & 1 - v_{f2}\cdot\frac{T_s}{l_2} & 0\\ 0 & v_{f2}\cdot\frac{T_s}{l_3} & 1\end{pmatrix}\cdot\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}(k) + \begin{pmatrix}\frac{T_s}{l_1} & 0 & 0\\ 0 & \frac{T_s}{l_2} & 0\\ 0 & 0 & \frac{T_s}{l_3}\end{pmatrix}\cdot\begin{pmatrix}q_1\\ 0\\ -q_4\end{pmatrix}(k) + \begin{pmatrix}\frac{T_s}{l_1} & 0 & 0\\ 0 & \frac{T_s}{l_2} & 0\\ 0 & 0 & \frac{T_s}{l_3}\end{pmatrix}\cdot\begin{pmatrix}r\\ 0\\ 0\end{pmatrix}(k) \tag{1.19}$$

By recalling the standard state space models in (*1.19*) and *(1.20*), for completely free flow mode FFF and congested mode CCC respectively, there can be derived other variants, by changing the outflow from which is considered to have congestion the density formula, that is dictated by the backward speed and jam density of the downstream cell.

For CCC mode:

$$\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}_{CCC}(k+1)=\begin{pmatrix}1 & w_2\cdot\frac{T_s}{l_1} & 0 & 0\\ & 0 & 1-w_2\frac{T_s}{l_2} & w_3\cdot\frac{T_s}{l_2}\\ & 0 & 0 & 1-w_3\cdot\frac{T_s}{l_3}\end{pmatrix}\cdot$$

$$\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}(k)+\begin{pmatrix}\frac{T_s}{l_1} & 0 & 0\\ 0 & \frac{T_s}{l_2} & 0\\ 0 & 0 & -\frac{T_s}{l_3}\end{pmatrix}\cdot\begin{pmatrix}q_1\\ 0\\ q4\end{pmatrix}(k)+\begin{pmatrix}\frac{T_s}{l_1} & 0 & 0\\ 0 & \frac{T_s}{l_2} & 0\\ 0 & 0 & \frac{T_s}{l_3}\end{pmatrix}\cdot\begin{pmatrix}r\\ 0\\ 0\end{pmatrix}(k)+$$

$$\begin{pmatrix}0 & -w_2\cdot\frac{T_s}{l_1} & 0\\ 0 & w_2\cdot\frac{T_s}{l_2}0 & 0\\ 0 & 0 & w_3\cdot\frac{T_s}{l_3}\end{pmatrix}\cdot\begin{pmatrix}\rho_{1J}\\ \rho_{2J}\\ \rho_{3J}\end{pmatrix}(k) \qquad (1.20)$$

4. A numerical example of the CTM model-Traffic density

For the purpose of the demonstration of the CTM model, in this paper is performed a numerical example which is described below. For the sake of simplicity, are chosen the approximately the same freeway segment partitioning characteristics as that in earlier sections in order to do an interconnection with the laid state space model of traffic density. The system of performance measurements of the traffic road networks of the Californian state *PeMs* is used for traffic collection data and is considered a freeway link for on the street "Broadway Avenue", Stockton/San Francisco. The freeway is consisted from three cells with different lengths with one on-ramp on the first cell. (*fig.1*).

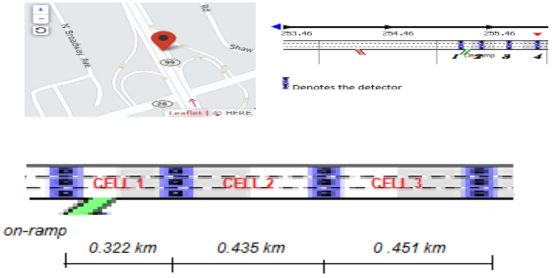


Fig. 1. Freeway segment with three cells

Calibrated parameters are given below [8].

Table 1. Summary of traffic parameters for three cells

	Q_M	vf	ρcr	ρJ	w	FF/ /CC
Cell 1	5580	84.8	65.8	248	30.6	FF
Cell 2	4176	96.8	43.1	248	20.3	FF
Cell 3	4268	106.7	40.0	248	20.5	FF

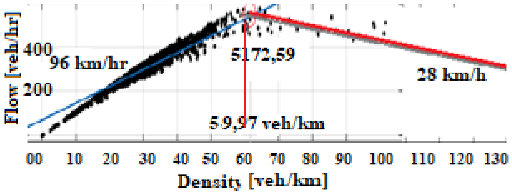


Fig.2. Calibrated parameters

5. Kalman Filter

Kalman filter-*KF* (*Kalman, 1960; Welch and Bishop 2001)* is a recursive data processing algorithm that uses only the previous time-step's prediction with the current measurement in order to make an estimate for the current state [4]. This means the *KF* does not require previous data to be stored or reprocessed with new measurements.

1.1. The building structure of the KF

The Kalman Filter consists of a set of mathematical equations that provides an efficient recursive computation to estimate the state of a process by minimizing the mean of the squared error. [4] The KF estimates the value of the variable x at any time (k+1), represented by a linear stochastic equation.

$$x_{k+1} = Ax_k + BX_k + w_k \qquad (1.22)$$

Where: A (k) is matrix which relates the state a time interval *k* with the state at current time interval *k+1. B (k)* is matrix which relates the current state to the control inputX_k.

The random variable *w* represents noise in modelling process. It is assumed to be within normal probability distributions with zero mean and variance Q (Gaussian) as: $p(w)\tilde{}\ N(0,Q)$

The system measurement equation describes the relationship between system states and measurements. Acknowledging that measurements inevitably contain noise, the output equation is expressed as follows:

$$\mathrm{Z(k+1) = H\cdot x(k),} \qquad (1.23)$$

$\tilde{Z}(k)$ is the measurement variable (outflow of vehicles from cell 3-measured by loop detector 2) *H (k)* is the output matrix, and *v (k)* is the measurement noise variable. The errors in estimating *a priori* and *a posteriori* states are defined as follows:

$$P_k^- = x_k - \hat{x}_k^- \qquad (1.24)$$

$$\mathrm{P_k = x_k - \hat{x}_k} \qquad (1.25)$$

The *a priori* and a posteriori estimate covariance is given by:

$$P_k^- = E[e_k^-, e_k^{-T}] = AP_{k-1}\cdot A^T + Q \qquad (1.26)$$

$$P_k = E[e_k, e_k^{\,T}) = (I - K_kH)\cdot P_k^- \qquad (1.27)$$

The KF estimates a posteriori state of the process using a linear combination of a priori state and a weighted difference between the actual measurement and the model measurement of the state.

$$\hat{x}_k = \hat{x}^-{}_k + K(z_k - H\cdot\hat{x}_k^-) \qquad (1.28)$$

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \qquad (1.29)$$

Based on the above equation, especially on (1.29), the *KF* process can be divided in two steps ore phases. The first step is the prediction step and the second step is the correction step.

6. Estimation with Kalman Filter

In this section are described in detail the applied matrices to the algorithm of the *CTM-KF* model. It is necessary to recall the equations state space of *CTM* model (Section V.1) first and then to do an interconnection of it with the *KF* algorithm equations. Since in our model, we are estimating the traffic densities of the three cells of the mentioned link, by the usage of the inputs values of the inflow q_1, output values q_4 and the flow from on ramp, then the state space vector of our algorithm are the densities $x=[x_1,x_2,x_3]=[\rho_1, \rho_2, \rho_3]^T$,the input vectors are

$$x_u=[q_1,0,q_4]^T;\ x_r=[qr,0,0]^T$$

$$x_{k+1} = Ax_k + BX_{Uk} + w_k$$

$$\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix} = A\cdot\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix} + Bu\cdot\begin{pmatrix}q_1\\ 0\\ -q_4\end{pmatrix} Br\cdot\begin{pmatrix}r\\ 0\\ 0\end{pmatrix} \qquad (1.30)$$

On this paper we are going to use the measurement of the outflow from the cell three, that corresponds to the flow q_4 in the figure (3.1). Based on the fundamental diagram we model the traffic flow measurement through the densities on the last cell (ρ_3) and the free flow speed on cell 3 v_{f3} we will have $q_{out} = v_{f3}\cdot\rho_3$ that is consistent with the equation (6.2)

$$Z(k+1) = H\cdot x(k) \qquad (1.31)$$

$$H = [0\ 0\ v_{f3}]\ \ and\ x = \begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}, q_{out} - [0\ 0\ v_{f3}]\begin{pmatrix}\rho_1\\ \rho_2\\ \rho_3\end{pmatrix}$$

Where*:* Q -the model error covariance matrix which elements standard deviations of the density variables. The off diagonal elements are equal to zero while R is the measurement or output error covariance. In this seminar paper, the matrices Q and R are assumed to be constant.[9]

$$\mathrm{Q} = \begin{vmatrix}\mathrm{w} & 0 & 0\\ 0 & \mathrm{w} & 0\\ 0 & 0 & \mathrm{w}\end{vmatrix};\ \mathrm{R} = \begin{vmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & \mathrm{v}\end{vmatrix}$$

7. Results and Conclusion

Evaluation of the density values of cell is performed with discrete time intervals of $T_s=10$ seconds, where the initial values of the densities $\rho_0 = [\rho_{10}, \rho_{20}, \rho_{30}]^T$ and estimated covariance matrix P_o are assumed. T_s is chosen to be 10 second in order to fill the conditions $T<L/v_f$, for proper work with system matrices, otherwise there will be obtained negative values of density parameters. For the purpose of the results evaluation, measured traffic densities for five minute intervals are used for comparison with the estimated densities with *CTM* model. The performance of the model was quantified by calculating the Mean Absolute Percentage Error (*MAPE)* given in (*1.32)*.

$$MAPE = [\tfrac{1}{n}\cdot\textstyle\sum_{k=1}^{n}\left|100\frac{\rho_{mod}(k)-\rho_{meas}(k)}{\rho_{meas}(k)}]\cdot 100\right| \qquad (1.32)$$

The MAPE results for CTM model for modes FFF -KF,

for Cell 1, Cell 2 and Cell3, 2 %, 0.6 % and 1 % respectively, nd for CCC-KF, 14% on the three cells. The results of the estimated values by CTM model against the measured values of traffic density are also graphically presented on the below figure (3-5).

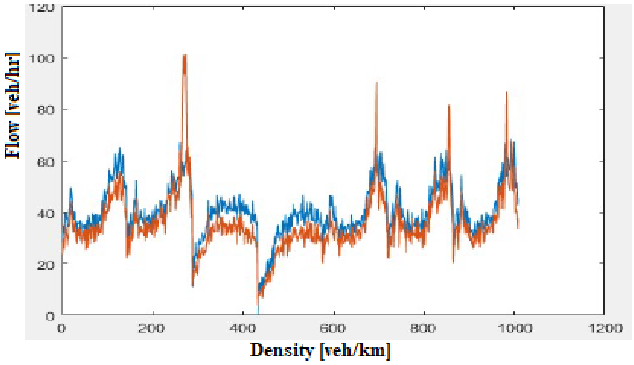


Fig.3. Graphic results densities of KF FFF and measurements-Cell 1

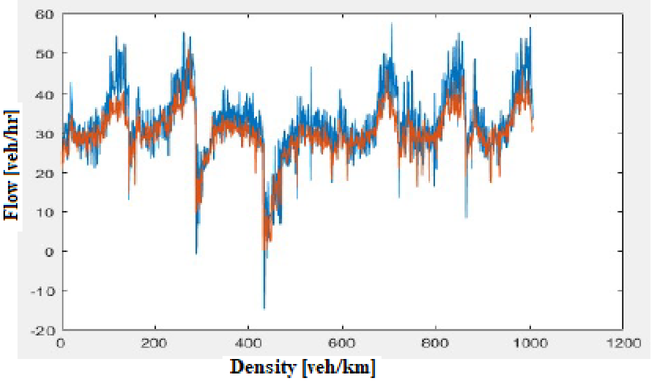


Fig.4. Graphic results densities of KF FFF and measurements-Cell 2

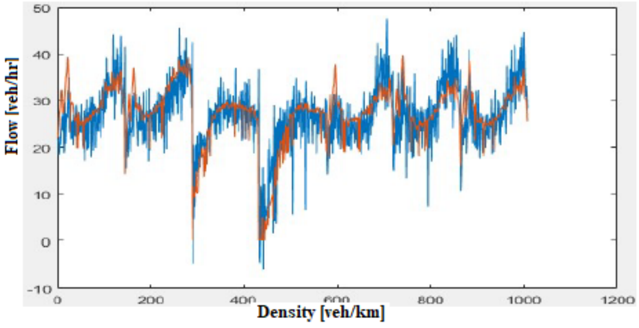


Fig.5. Graphic results densities of KF FFF and measurements- Cell3

References

[1] Nicola George - Kathryn Kershaw, Road Use Statistics, Great Britain, 2016

[2] Ramp Metering: A Review of the Literature, E.D. Arnold ,Jr (1998)

[3] Chen X.M, Li, L; Stochastic Evolutions of Dynamic Traffic Flow Modeling and Applications, Chapter 2, http://www.springer.com/978-3-662-44571-6 Springer (2015)

[4] M.J .Lighthill ; G.B. Whitham "On Kinematic Waves, II: A theory of traffic flow on long crowded roads, Proceedings of the Royal Society of London, Series A-Mathematical and Physical Science.

[5] X. Sun, L. Munoz. R. Horowitz, A Mixture Kalman Filter Highway Congestion Mode and Vehicle Density Estimator and its Application (2004)

[6] Carlos F. Daganzo, "The Cell Transmission Model: Network Traffic", (1996)

[7] L. Munoz, X. Sun, R. Horowitz, L. Alvarez; "Traffic Density Estimation with the Cell Transmission Model1" (2003)

[8] G.B. Witham "Linear and Nonlinear Waves", Pure and Applied Mathematics, John Wiley Sons, New York City, USA, (1974)

[9] L. Munoz, X. Sun, R. Horowitz, D. Sun, G. Gomes; "Methodological calibration of the cell transmission model" Proceedings of the American Control Conference, Massachusetts 2004.

[10] A. Katsivalis, M. Papa Georgiou, 'The importance of Traffic Flow Modeling for Motorway Traffic Control', 2001

Advances in Science, Technology and Engineering Systems Journal
Vol. 3, No. 2, 90-98 (2018)
www.astesj.com
Special issue on Advancement in Engineering Technology

ASTES Journal
ISSN: 2415-6698

Agent Based Fault Detection System for Chemical Processes using Negative Selection Algorithm

Naoki Kimura*, Yuya Takeda, Yoshifumi Tsuge

Department of Chemical Engineering, Faculty of Engineering, Kyushu University, 819-0395, Japan

ARTICLEINFO

Article history:
Received: 15 November, 2017
Accepted: 07 January, 2018
Online: 12 March, 2018

Keywords:
Negative selection algorithm
Plant alarm system
Fault detection

ABSTRACT

Recently, the number of industrial accidents of chemical plants has been increasing in Japan. The fault detection system is required to keep chemical plant safely. In this study, a fault detection system for a chemical plants using agent framework and negative selection algorithm was proposed. The negative selection algorithm is one of artificial immune systems. The artificial immune system is an imitative mechanism of vital actions to discriminate self/nonself to protect itself. The method was implemented and applied to a complicated chemical plant—which is a boiler plant virtually operated using a dynamic plant simulator. The simulations of fault detection were carried out. And also the results of simulations are presented in this paper.

1 Introduction

This paper is an extension of work originally presented in 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP) [1].

Recently, it has been increasing the accidents of chemical plants. Figure 1 shows the annual numbers of the industrial accidents in Japan (the data is based on the summary of accidents in specified business facilities inside petrochemical complex in Japan, published by Fire and Disaster Management Agency, Ministry of Internal Affairs and Communications, Japan, posted on their official web site on May 2017, article in Japanese). The numbers of the industrial accidents in Japan—except accidents caused by earthquakes or tsunami—has been increasing from 45 in 1993 to over 250 in 2016. It is said that the remote causes of the rise of accidents in Japan are mass retirements of skilled engineers, insufficient technical tradition, labor-savings in production lines or plant operations, aged deterioration of productive facilities, or maintenance cost reduction in assertive ways. Therefore, effective fault detection system for chemical plants is required. In the general chemical plant operations, *plant alarm system* has been used to notify the process deviance to operators via warning lights or buzzers in the operation rooms, where the upper and lower thresholds of the measured values or the thresholds of their amounts of changes have been set to the sensors in the chemical plants. However, it is so difficult to determine the adequate values of thresholds (that is 'alarm setpoint') that if the alarm setpoints are too small, the alarm floods will be caused, if the setpoints are too large, missed detection of deviation will be caused. And also it is difficult to detect if the plant has normal and regular load fluctuations under both the normal and the abnormal situations. Therefore, a method is required that observes the relationship among several variables to detect faults in a complicated system. We focus on the Artificial Immune System.

Artificial Immune Systems—which are imitative mechanisms of vital actions of discrimination between self and nonself—have been proposed since 1990s. And a lot of methods have been proposed using various parts of artificial immune systems, for example, pattern recognition by B-cells for fault detection in gas lift oil well by Aguilar [2, 3], Natural Killer (NK) immune cells mechanisms by Laurentys [4], clonal selection algorithm for maintenance scheduling of power generators by El-sharkh [5], dendritic cell for failure detection of aircraft by Azzawi [6]. Also lots of applications have been proposed—Wada et al. [7] proposed an fault mode detection method for automotive exhaust gas treatment system, Inomo et al. [8] proposed an failure diagnosis method for water supply network by using immune system.

*Corresponding Author, nkimura@chem-eng.kyushu-u.ac.jp

https://dx.doi.org/10.25046/aj030211

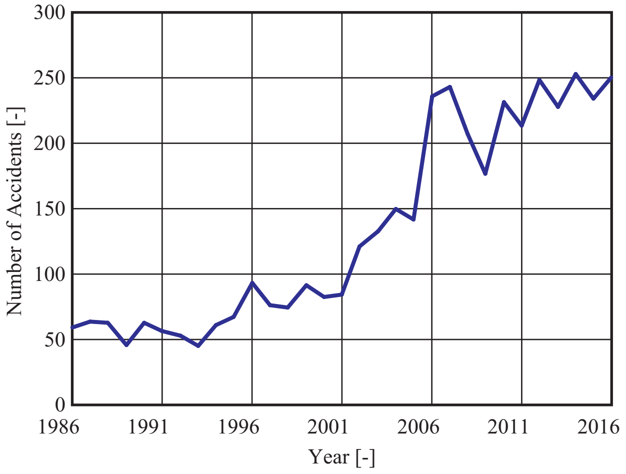


Figure 1: The annual numbers of the industrial accidents in Japan, data from FDMA, Japan.

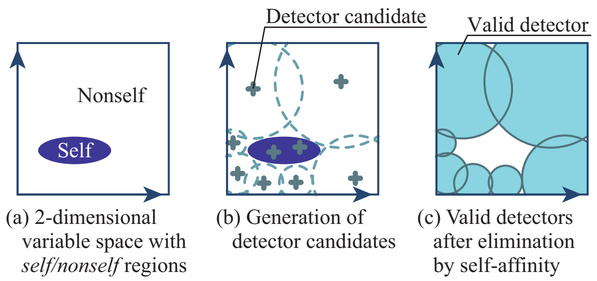
In this study, we adopt a negative selection algorithm to detect faults in a chemical plant. Negative selection algorithm is an imitative method of mechanisms of differentiation and maturation, and discrimination of normal/abnormal. The algorithm was proposed by Forrest in 1994 [13] for the detection of computer virus. The negative selection algorithm has been applied to various domains—Dasgupta et al. [9] applied to aircraft fault detection, Gao et .al [10] applied to motor fault detection, Xiong et al. [11] and Prasad et al. [12] applied to fault detection in the Tennessee Eastman process.

In this study, we introduce detectors to detect faults based on the negative selection algorithm. The adopted method of the negative selection algorithm is mentioned in section 2. In order to utilize negative selection algorithm, we designed and implemented an agent based framework of fault detection system, mentioned in section 3. The target chemical process, the simulation conditions and the results are mentioned in section 4. And the conclusion in section 5.

2 Negative Selection Algorithm

Negative selection algorithm is one of the methods of the artificial immune systems inspired by the vital immune systems. Negative selection algorithm borrowed from the mechanism of T-cell generation in thymus. On T-cell generation in vital system, immature T-cells are randomly generated with various immunological types. And then some of T-cells are eliminated if they have high affinity with self-antigen to avoid response to "self". T-cells which are not self-affinitive can be matured to react with foreign antigen to protect itself.

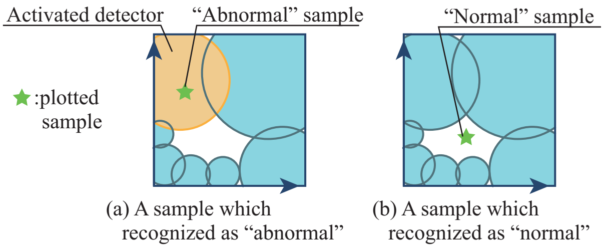
In our system, detectors—correspond to mature T-cells in vital—can detect faults by recognizing the normal operational data of chemical processes which are assumed as *self* and the abnormal operational data which are assumed as *nonself*. In Figure 2(a)–(c), the steps of the generation phase of detectors are illustrated. Figure 2(a) illustrates that there are *self* regions in a 2-dimensional process variables space. And the rest area of the variables space are *nonself* regions. Figure 2(b) illustrates that detector candidates which indicated by plus(+) signs are generated with various position, where the radius is set as the minimum distance to the *self* region. In this study, we implement two steps of candidate generation. The first step is a grid-based generation. Place the detector candidates at every 0.1 of each axis—axis is normalized between [0.0, 1.0]—, therefore 121 (each axis divided into eleven) grid-based candidates are generated. The second step is a semi-randomly generation. Place a detector candidate at random position. Then check whether the new candidate is overlapped on the previously generated candidates. If it overlapped, the new candidate is eliminated in this step, if not, the new candidate is added to the candidates. And then, check the coverage of the detectors. If the coverage is over 90%, candidate generation will be terminated. From Figure 2(b) to (c), the elimination of some candidates which have affinity with "self" is carried out. Figure 2(c) illustrates that seven valid detectors are remained. Then the generation phase is finished.



(a) 2-dimensional variable space with *self/nonself* regions (b) Generation of detector candidates (c) Valid detectors after elimination by self-affinity

Figure 2: A schematic diagram of the detectors generation.

In the detection phase, a sample consists of the values of the current process variables—which are to be examined— is plotted into the variable space indicated by a 'star'. If the plotted sample is inside the detection area of at least one detectors, the sample is recognized as "abnormal" (Figure 3(a)). And if the sample is out of any detection area, the sample is recognized as "normal" (Figure 3(b)).



(a) A sample which recognized as "abnormal" (b) A sample which recognized as "normal"

Figure 3: A schematic diagram of the detection by detectors.

3 Fault Detection System

A multiagent framework is adopted to implement a fault detection system using negative selection algorithm. In our system, there are *Detector Manager*, *Detector Leader(s)* and *Database Agent* illustrated in Figure 4. *Detector Manager* is an interface between human operator and the system. *Detector Leader(s)* have their own variable spaces to detect faults. These variable spaces represent relationships between two certain process variables. And there are a lot of detectors under the dominion of a *Detector Leader*. In order to avoid missing detection, some *Detector Leaders* with variety of combinations of the process variables are required. *Database Agent* have a database which stores operational database of the target process. All these agents can communicate with other agents via TCP/IP network connection.

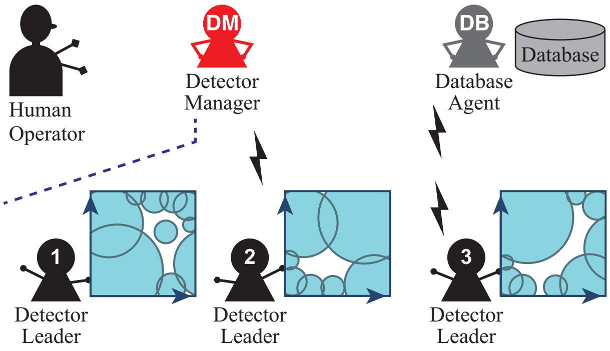


Figure 4: A schematic diagram of the detection by detectors.

4 Simulations and Results

4.1 Target Process

Figure 5 shows a boiler plant which is the target process of this study. It is an utility plant which supplies three steam headers—high pressure steam (HP), middle pressure steam (MP), and low pressure steam (LP) to the nearby plants by boiling pure water in a furnace. Due to the fluctuation of the steam demands from user plants, the boiler plant is always under unsteady operation. Therefore, it is difficult to detect faults by setting up the constant thresholds to some process variables of the boiler plant.

In this study, nine variables were selected from among 120 measured variables in the boiler plant. The selected nine variables are listed in Table 1 and also illustrated in Fig.5. The operational data was obtained by using dynamic plant simulator "Visual Modeler" (Omega Simulation Co., LTD).

4.2 Normal Operational Data

Data of a normal operation—in which the steam demands were stepwise changed without any malfunction—were obtained using the dynamic simulator. The data contain 7200 samples of the above-mentioned nine process variables and its sampling interval is one second. The trend graphs of PI1422 and TI1422 were indicated in Figure 6. These data were normal operational data which should be recognized as *self* in the artificial immune system even though the steam demands had increased at time 600 second.

4.3 Abnormal Operational Data

Abnormal operational data with three kinds of assumed malfunctions were obtained using the dynamic simulator. Table 2 shows the list of three assumed malfunctions. And Figures 8–10 illustrate the trend graphs of PI1422 and TI1422 when the one of malfunctions was occurred at time 600 second, but without any steam demand change during these 7200 seconds.

Table 1: Selected nine process variables of the boiler plant.

Variable name	Description
FI1422	Flow rate of the second extraction steam from the turbine
PI1422	Pressure of the second extraction steam from the turbine
TI1422	Temperature of the second extraction steam from the turbine
TC1423	Temperature of the low pressure steam from the desuperheater
PI1315	Pressure inside the furnace (upper)
GB401.pos	Valve position of the combustion air
PI1311	Pressure of the outlet of the forced draft fan
TI1310	Temperature of the exhaust gas at the gas air heater outlet
TI1308	Temperature of the combustion air at the gas air heater outlet

4.4 Generation of Detectors

Detectors were generated by using normal operational data for each of 36 combinations of 2-dimensional variable spaces—all the 2-dimensional variable spaces consists of two different process variables from the nine variables. Figure 7 illustrates the detectors in the variable space consisting of TI1422 vs. PI1422. The axes are normalized to the range of the normal operation occupies [0.05, 0.95] in the normalized axis respectively. In this figure, there are 261 detectors (118 grid detectors and 143 randomized detectors) which indicated as sky blue circles. In other words, it is recognized that there exist the normal operational data—which correspond "self" region—at

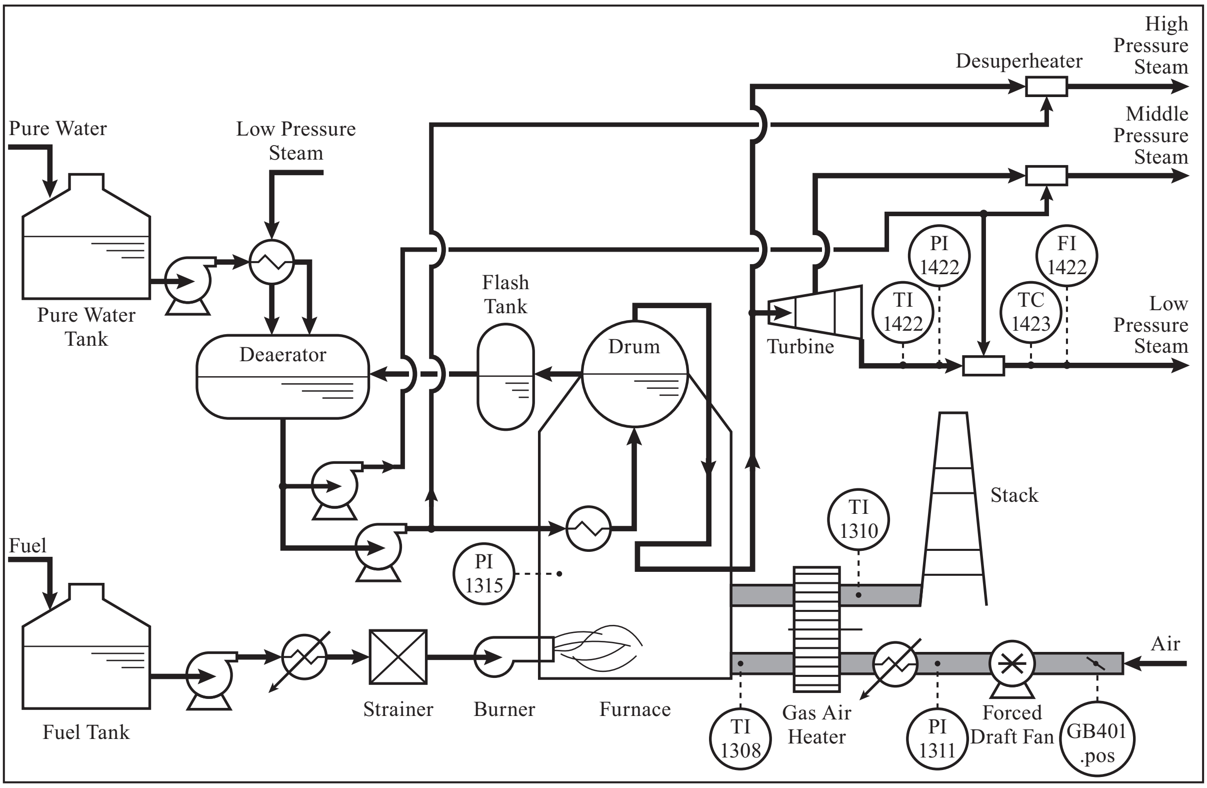


Figure 5: A schematic diagram of the boiler plant.

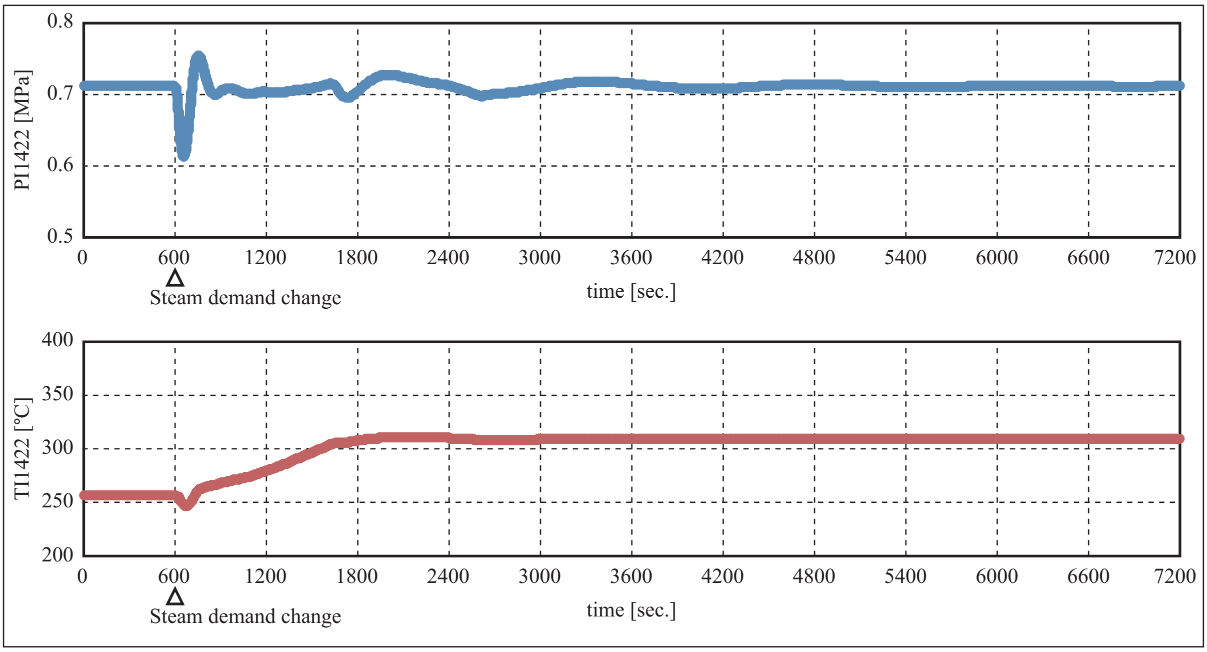


Figure 6: The trend graphs of PI1422 and TI1422 during normal operation with steam demands change at time 600 second.

the unpainted parts in Figure 7, and not exist the normal operational data—which correspond "nonself" region—at the painted sky blue parts. The sets of detectors were generated for the rest of variable spaces in the same way. The numbers of the detectors in each variable space were from 121 to 316 including both the grid detectors and randomized detectors.

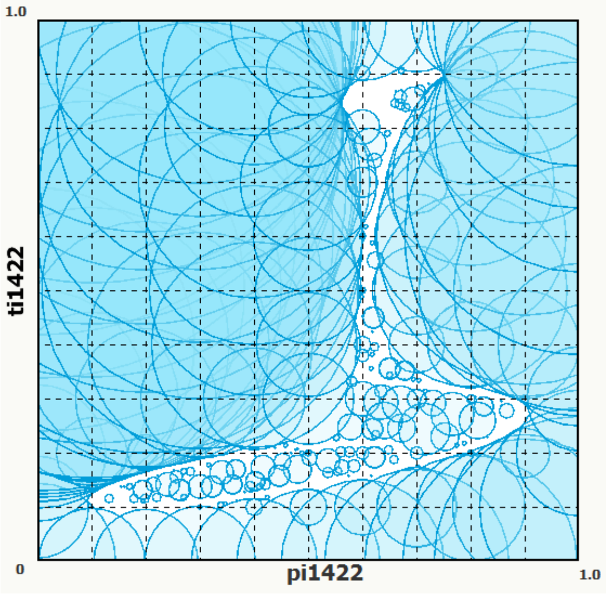


Figure 7: Generated detectors for a 2-dimensional variable space (TI1422 vs. PI1422).

The bold dashed lines in the trend graphs are the maximum and minimum values under the normal operation (Figure 6). We can find that it is impossible to detect fault using upper/lower thresholds of variable PI1422 and/or TI1422 when the malfunction BFO or GAH occurred, because the values of PI1422 and TI1422 have not exceeded the ranges under the normal operation. On the other hand, it may possible to detect at time 671 second by lower threshold of TI1422 when the malfunction FDFclose occured at time 600 second. TI1422 also exceeds upper threshold after time 1340 second, shown in Figure 10.

Table 2: Three assumed malfunctions in the boiler plant.

Malfunction ID	Description
BFO	Burner frame out
GAH	Gas air heater rotation failure
FDFclose	Forced draft fan inlet vane closure

4.5 Fault Detection using Detectors

Fault detections for the abnormal operational data were carried out. Figure 11 shows an outline of the fault detection in a 2-dimensional variable space consisting of TI1422 and PI1422 when malfunction BFO occurred. The axes and the detectors—indicated by sky blue circles— are the same as Figure 7. The sampling data were plotted by green dots on the 2-dimensional variable space after the normalization for every second serially. The blue dots are the normal operational data and the green dots are the sampling data to be examined by the detectors. These dots are moving in the variable space as time proceeds. If the green dots were placed on the unpainted region, they are recognized as "self"—where the values are similar to the normal operational data. If the green dots were placed over the sky blue region, they are recognized as "nonself"—in other words, a fault was detected in this variable space by detector(s).

In Figure 11, the painted orange circles indicate activated detectors—which have detected fault. Figure 12 shows the detection status through time. If the value is '1', at least one detector in this variable space detects fault, and if the value is '0', no detector detects fault at that time. The figure shows that the first detection was at time 630 second—which is 30 second after malfunction occurred—, and there are missing between time 897 and 904 second in this variable space. The detections using the sets of detectors in the rest of variable spaces were simultaneously carried out in the same way. Figure 13 shows the number of variable spaces whose detection status is '1'. The figure shows that there are 15 variable spaces detected at time 601 second, 30 variable spaces—which corresponds to 80% of 36 variable spaces—detected at time 604 second, and the number does not fall below the 80% after that time. Therefore it can be said that this system can detect malfunction BFO successfully.

Figure 14 shows the outline of the fault detection, Figure 12 shows the detection status through time when malfunction GAH occurred. The figures show that the fault was detected at time 666 second by the detectors. On the other hand, it have not been detected after time 1047 second in this variable space TI1422 vs. PI1422. However 16 shows that 21 variable spaces—which corresponds to 58% of 36 variable spaces—detected at time 602 second, and the number does not fall below 58% after that time. It can be said that this system can detect malfunction GAH successfully, although the variable space TI1422 vs. PI1422 could not detect after time 1047 second.

Figure 17 also shows the outline of the fault detection, Figure 18 shows the detection status through time when malfunction FDFclose occurred. The figures show that the fault was detected at time 630 second by the detectors. Although the detection status fluctuate between time 897–1102 second in the variable space TI1422 vs. PI1422, over 30 variable spaces detect after time 604 second and the number does not fall below the 80% of the 36 variable spaces after time 604 second. It can be also said that this system can detect malfunction FDFclose successfully.

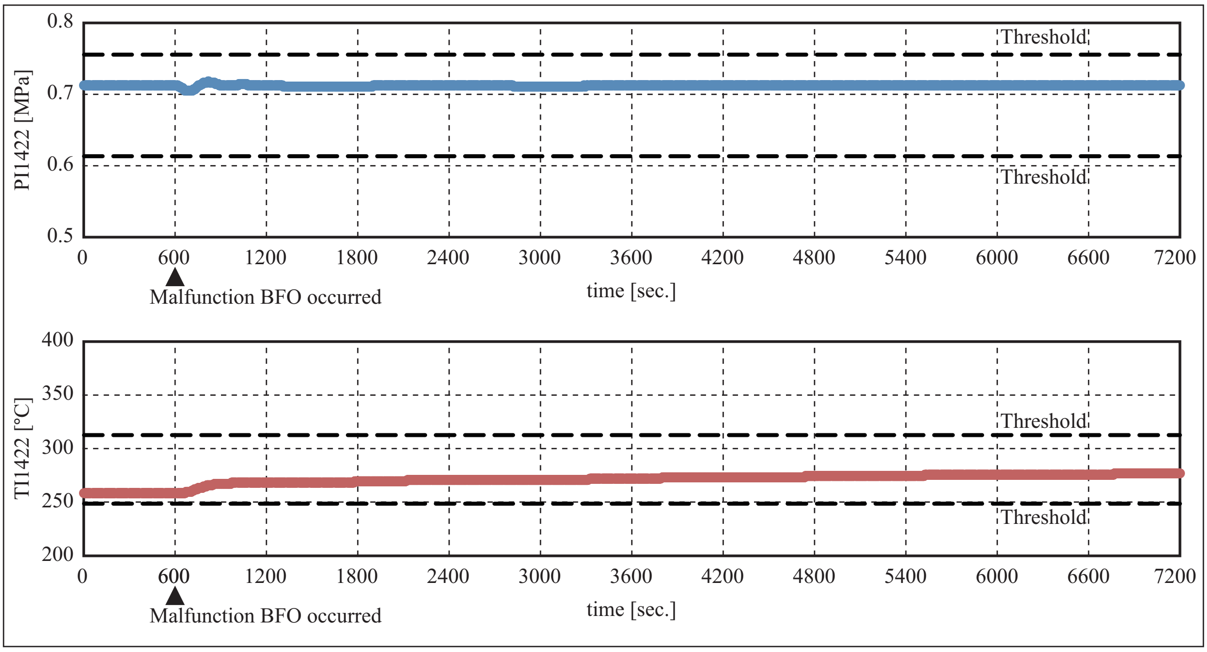


Figure 8: The trend graphs of PI1422 and TI1422 when malfunction BFO was occurred at time 600 second without steam demands change.

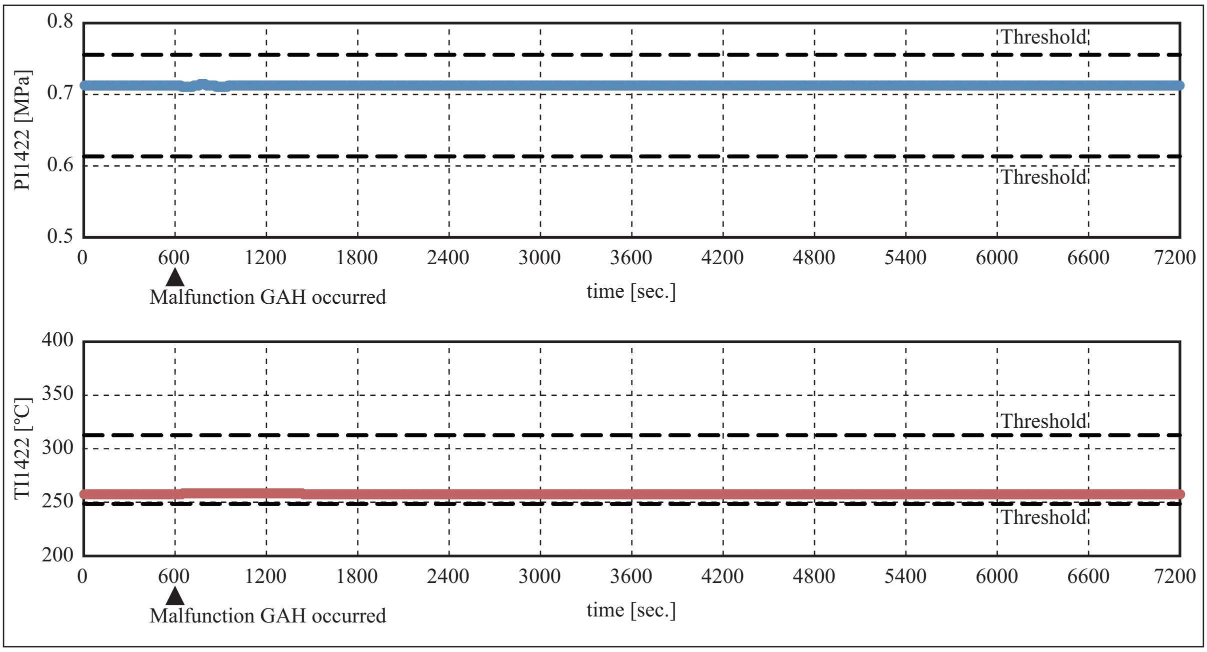


Figure 9: The trend graphs of PI1422 and TI1422 when malfunction GAH was occurred at time 600 second without steam demands change.

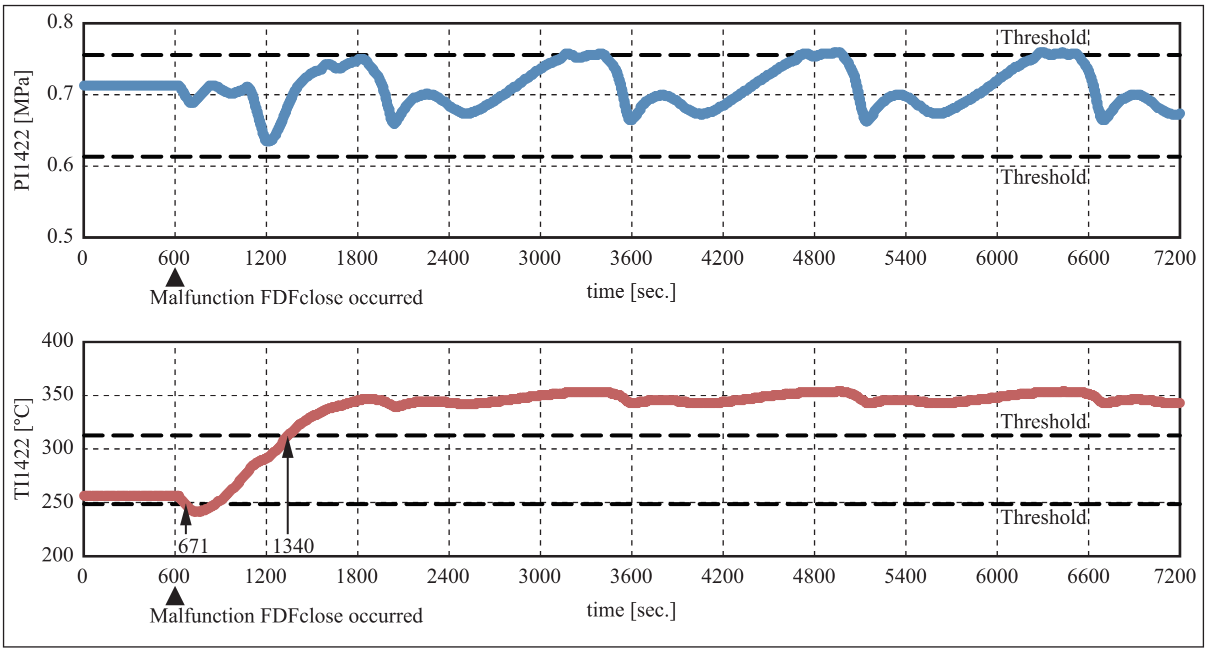


Figure 10: The trend graphs of PI1422 and TI1422 when malfunction FDFclose was occurred at time 600 second without steam demands change.

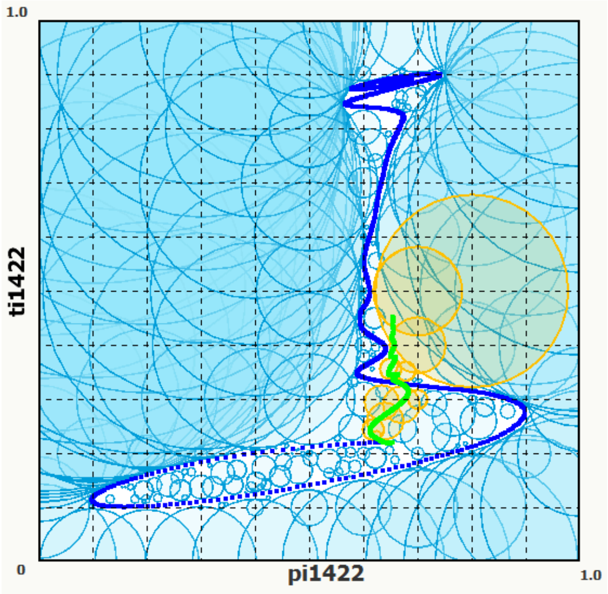


Figure 11: Fault detection by detectors in a 2-dimensional variable space when malfunction BFO was occurred.

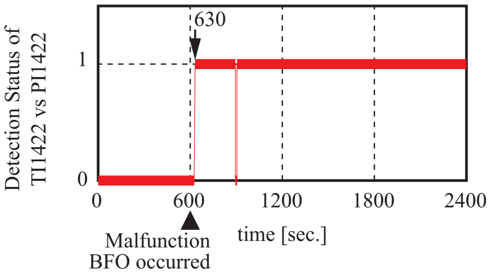


Figure 12: Detection status of the TI1422 vs. PI1422 variable space when malfunction BFO was occurred.

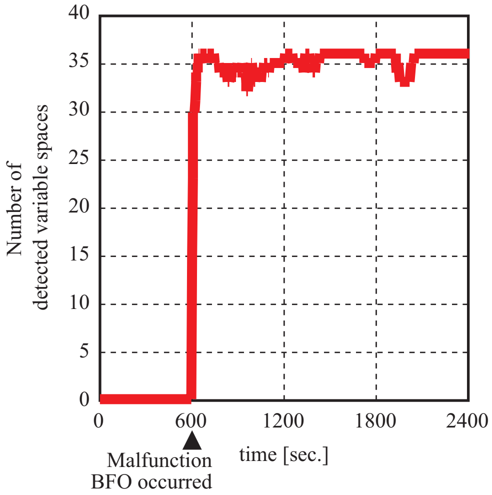


Figure 13: The number of detected variable spaces when malfunction BFO was occurred.

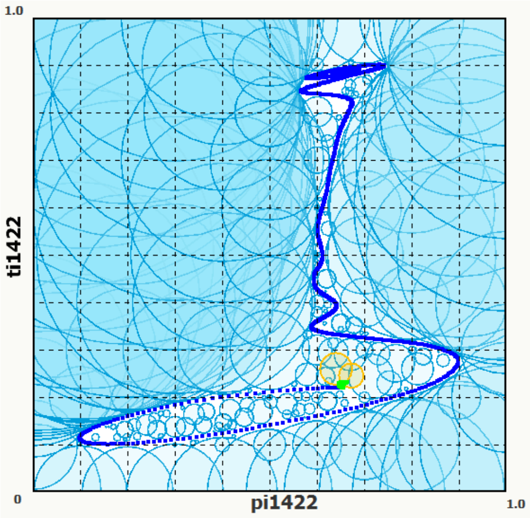


Figure 14: Fault detection by detectors in a 2-dimensional variable space when malfunction GAH was occurred.

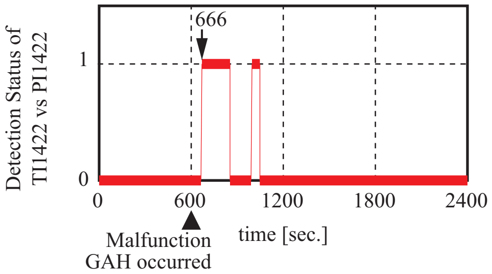


Figure 15: Detection status of the TI1422 vs. PI1422 variable space when malfunction GAH was occurred.

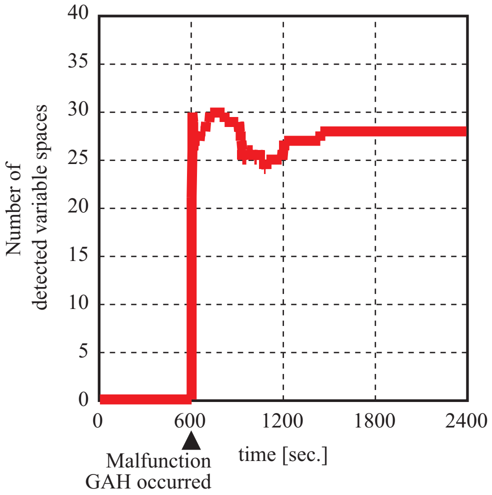


Figure 16: The number of detected variable spaces when malfunction GAH was occurred.

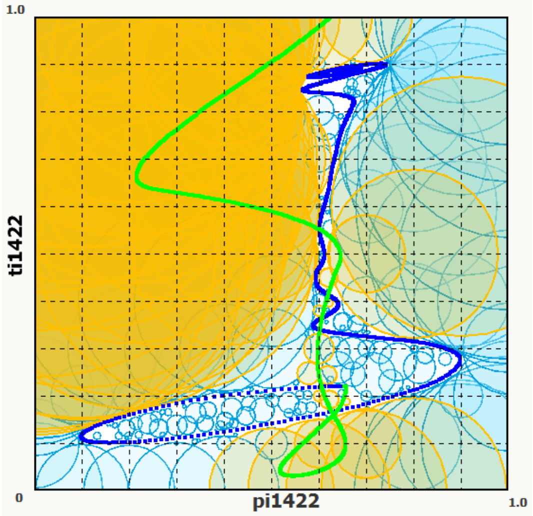


Figure 17: Fault detection by detectors in a 2-dimensional variable space when malfunction FDF-close was occurred.

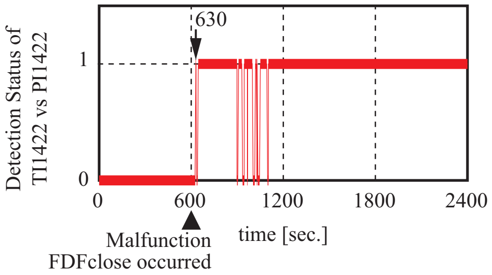


Figure 18: Detection status of the TI1422 vs. PI1422 variable space when malfunction FDFclose was occurred.

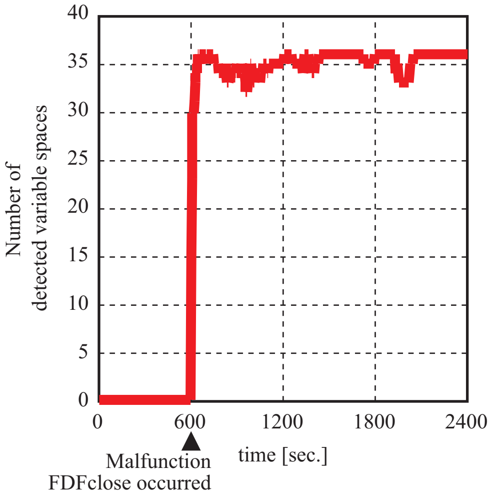


Figure 19: The number of detected variable spaces when malfunction FDFclose was occurred.

5 Conclusion

We built up a fault detection system using negative selection algorithms which can focus on the relationships between process variables.

References

1. N. Kimura, Y. Takeda, T. Hasegawa, Y. Tsuge, “Agent based fault detection using negative selection algorithm for chemical processes” in 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP), Taipei, Taiwan, 2017. https://doi.org/10.1109/ADCONIP.2017.7983822
2. J. Aguilar, “An artificial immune system for fault detection” Innovations in Applied Artificial Intelligence (17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2004, Lecture Note in Artificial Intelligence), **3029**, 219–228, 2004. https://doi.org/10.1007/978-3-540-24677-0_24
3. M. Araujo, J. Aguilar, H. Aponte, “Fault detection system in gas lift well based on artificial immune systems” in the International Joint Conference on Neural Networks, 2003 , 1673–1677, 2003. https://doi.org/10.1109/IJCNN.2003.1223658
4. C.A. Laurentys R.M. Palhares, W.M. Caminhas, “A novel artificial immune system for fault behavior detection” Expert Systems with Applications, **38**, 6957–6966, 2011. https://doi.org/10.1016/j.eswa.2010.12.019
5. M.Y. El-Sharkh, “Clonal selection algorithm for power generators maintenance scheduling” International Journal of Electrical Power & Energy Systems, **57**, 73–78, 2014. https://doi.org/10.1016/j.ijepes.2013.11.051
6. Dia Al Azzawi, Mario G. Perhinschi, Hever Moncayo, Andres Perez, “A dendritic cell mechanism for detection, identification, and evaluation of aircraft failures” Control Engineering Practice, **41**, 134–148, 2015. https://doi.org/10.1016/j.conengprac.2015.04.010
7. K. Wada, T. Toriu, H. Hama, “Improving the efficiency of known fault mode detection for immunity-based diagnosis” Transactions of the Institute of Systems, Control and Information Engineers, **27**(2), 59–66, 2014 (article in Japanese with English abstract). https://doi.org/10.5687/iscie.27.59
8. H. Inomo, W. Shiraki, Y. Imai, H. Kanamaru, “Failure diagnosis of water supply network by immune system” J. Soc. of Mat. Sci., Japan, **52**(1), 40–45, 2003. (article in Japanese with English abstract). https://doi.org/10.2472/jsms.52.40

9. D. Dasgupta, K. KrishnaKumar, D. Wong, M. Berry, “Negative selection algorithm for aircraft fault detection” Artificial Immune Systems (Third International Conference, ICARIS 2004, Lecture Note in Computer Science), **3239**, 1–13, 2004. https://doi.org/10.1007/978-3-540-30220-9_1

10. X.Z. Gao, H. Xu, X. Wang, K. Zenger, “A study of negative selection algorithm-based motor fault detection and diagnosis” International Journal of Innovative Computing, Information and Control, **9**(2), 875–901, 2013.

11. C. Xiong, Y. Zhao, W. Liu, “Fault detection method based on artificial immune system for complicated process”, Computational Intelligence (International Conference on Intelligent Computing, ICIC 2006, Lecture Note in Artificial Intelligence), **4114**, 625–630, 2006. https://doi.org/10.1007/11816171_77

12. J. V. Prasad, K. Ghosh, “Negative selection algorithm for monitoring processes with large number of variables”, in 2014 IEEE Conference on Control Applications (CCA), 778–783, Antibes, France, 2014. https://doi.org/10.1109/CCA.2014.6981435

13. S. Forrest, A. S. Perelson, L. Allen, R. Cherukuri, “Self-nonself discrimination in a computer” in 1994 IEEE Computer Society Symposium on Research in Security and Privacy, 202–212, 1994 https://doi.org/10.1109/RISP.1994.296580

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 99-105(2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Detection of ExoMars launcher during its passage over Europe with Space Surveillance radar breadboard

Stéphane Saillant*, Marc Flécheux, Yann Mourot

Department of Electromagnetism and Radar, ONERA, Palaiseau 91123, France

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 01 February, 2018
Online: 12 March, 2018

Keywords:
Radar
Bistatic
Kinematic Analysis
Space surveillance
ExoMars
Space launcher

ABSTRACT

A bistatic radar breadboard for space surveillance has been developed by ONERA for the European Space Agency. This item was operated during the launch of the ExoMars mission on March 14th 2016. The spacecraft, attached to the Proton launcher, was well detected in real-time during its passage over South Europe. This paper presents the setting up of an experiment to detect this particular type of targets with the radar breadboard. The results of its operation as space surveillance system as well as a specific kinematic analysis of the ExoMars spacecraft as viewed from the radar.

1. Introduction

This paper is an extension of works originally presented in 2017 IEEE Radar Conference (RadarConf17) [1].

ABISS project (Antenna BIstatic for Space Surveillance) is relative to the breadboarding of a ground-based surveillance radar demonstrator for European Space Agency (ESA) [2]. This paper presents the setting up of an experiment to detect certain types of targets with this Space Surveillance radar demonstrator

The first section presents a description of the radar demonstrator and its principle of operation for Space Surveillance. Results obtained during the different campaigns of measurements are presented.

The second section is relative to the experiments which have been done during ExoMars launch in March 2016. A kinematic analysis of the radar parameters has been realized from the measurements of detection of this particular target.

2. Radar demonstrator

The European Space Surveillance radar breadboard is a bistatic system that is located on two separate sites in the northern part of France around Paris as shown in Figure 1.

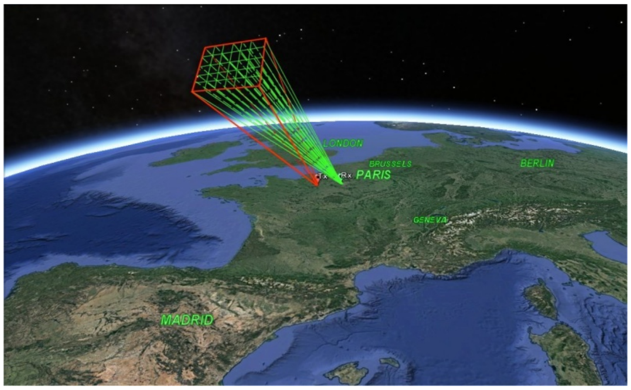


Fig.1. Bistatic radar configuration (*Tx*: transmitting site in red, *Rx*: receiving site in green)

This radar operates in L-band for the detection of space objects over a steerable Field of Regard (FoR) defined as 30° in azimuth by 25° in elevation oriented preferably to the South.

One of the advantages of such a system is its operation with a continuous wave (CW) signal, which is easier for spectral management as it is strictly limited to the carrier frequency.

[*] Stéphane Saillant, Email: stephane.saillant@onera.fr

https://dx.doi.org/10.25046/aj030212

The breadboard has been designed on the basis of simulations of performances, as a radar which should have a reference detection range at 500 km for a reference target of 0 dBm² as Radar Cross Section (RCS) for a revisit time of 10 seconds, meaning that the beam of the radar should revisit the same azimuth and elevation position within 10 seconds.

These are the reference values that have been used in the power budget computation for the design of the system. The principle of the detection for such radar is only based on Doppler measurement; there is no evaluation of the distance of the target.

The transmitting system (*Tx*, presented in figure 2 and figure 3) is constituted of one antenna array of surface area 1.5 m² with 49 radiating elements, and 49 CW solid state power amplifiers, controlled in both phase and gain, associated to each radiating element.

All electronics equipments for transmitting and control are in a shelter. The antenna array is set up under a radome on the roof oriented to the South. The positioner which hosts the antennas array can be steered in +/-90° in azimuth around a central position oriented in South direction, and 0° to 90° in elevation.

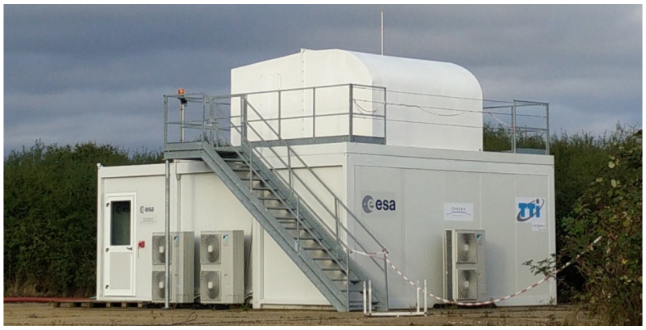


Fig.2. The transmitting system the shelters with the antenna on the roof under a radome)

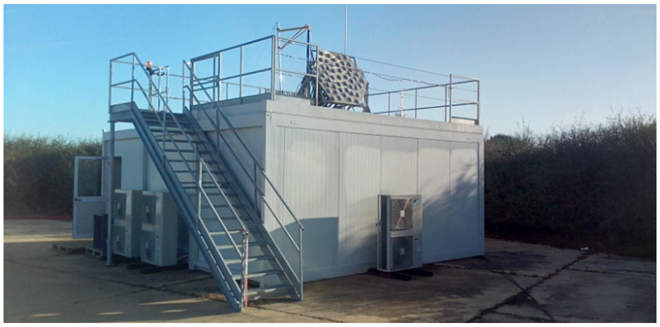


Fig.3. Transmitting system (the shelters with the antenna on the roof without the radome)

The receiving site (*Rx*) is within ONERA centre in Palaiseau. The antenna panel, with a surface area of 10.5 m² and 64 receiving antennas, is installed on the roof of an office building as shown on Figure 4.

The antennas array can be steered in +/-90° in azimuth around a central position oriented in South direction, and it can be inclined from 0° to 90° in elevation. The receiving system is constituted of 64 analog front-edge stages in one cabinet, 4 digital receivers of 16 channels and a cluster of 10 PC servers to process the data.

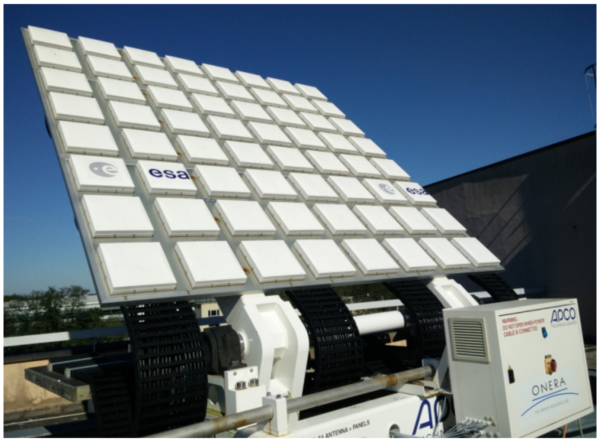


Fig.4. The receiving antenna panel in Palaiseau (ONERA centre)

In order to achieve the simulated detection performances, the whole Field of Regard of 30° in azimuth by 25° in elevation as shown in Fig.5, should be revisited every 10 seconds.

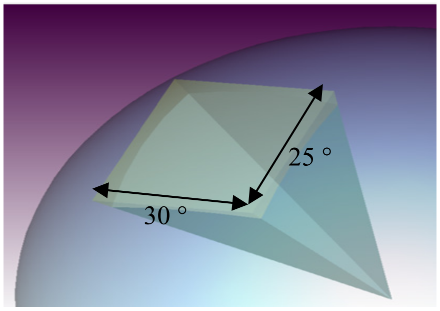


Fig.5. The Field of Regard (FoR) of the radar

According to simulations taking into account the geographical constraints and the maximization of objects detection, the difference of pointing between the two beams oriented South is few degrees between the two boresight directions.

The aperture of the transmitting beam plotted on Figure 6 is 11° azimuth by 11° elevation. Thus, the entire Field of Regard is covered by 10 scanning positions of this beam as presented on Fig.7.

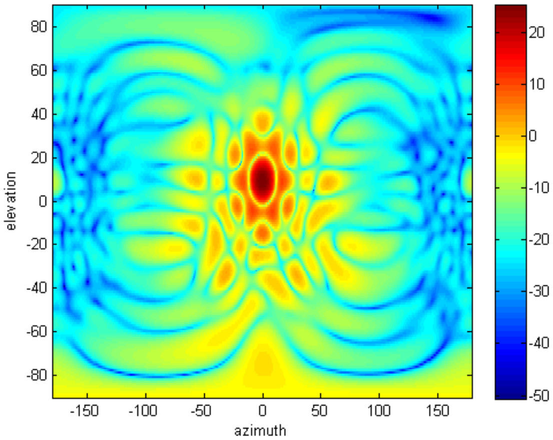


Fig.6. Pattern of the transmitting beam formed with transmitting array (11° aperture in azimuth and elevation).

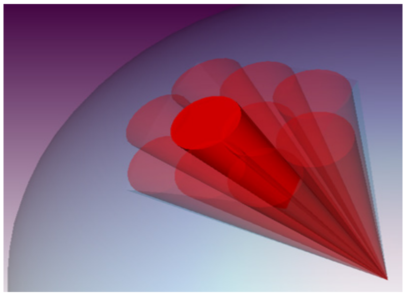


Fig.7. The 10 beams positions necessary to cover the Field of Regard

The Field of Regard is completely covered with an average gain greater than 22 dB as presented on Fig.8.

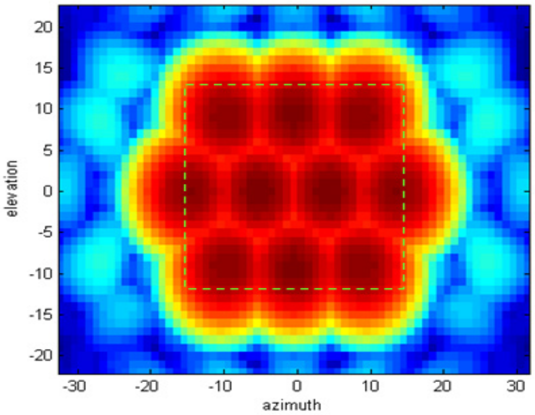


Fig.8. Optimization of the coverage of the Field of Regard for detection requirement

An evaluation of the number of receiving beams to form to cover each transmitting beam is done with an overlapping factor of 0.8 with a beam width of 4° (the aperture of receiving beam). The Figure 9 shows the results of the evaluation of receiving beams to form for extreme positions of the transmitting beam in the four corners and in the middle of the Field of Regard.

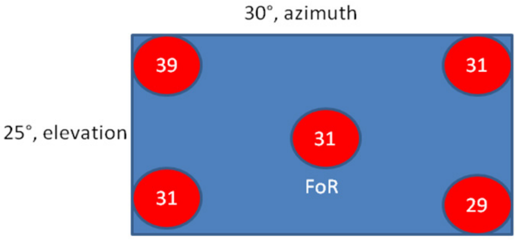


Fig.9. Number of *Rx* beams required in different locations of transmitting beam (Red areas) in the Field of Regard (Blue area: 30° azimuth x 25° elevation).

The most critical position which requires most of receiving beams to compute is when transmitting beam is focalized in the position corresponding to the upper left corner of the Field of Regard. Maximum number of beams to form for this steering direction of the transmitting beam is around 39, according to the following result represented on Figure 10 with a drawing which takes into account the distortion of the illuminated area.

The full receiving antenna array (including coupling and losses of materials) has been simulated. The global gain in the central boresight of the panel is 31.8 dB as shown in the 3 dimension display of the Figure 11.

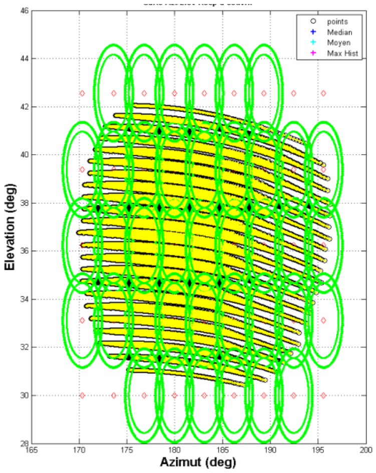


Fig.10. Transmitting beam deformed (yellow) and viewed from receiving site with the coverage of associated receiving beams (green)digitally formed

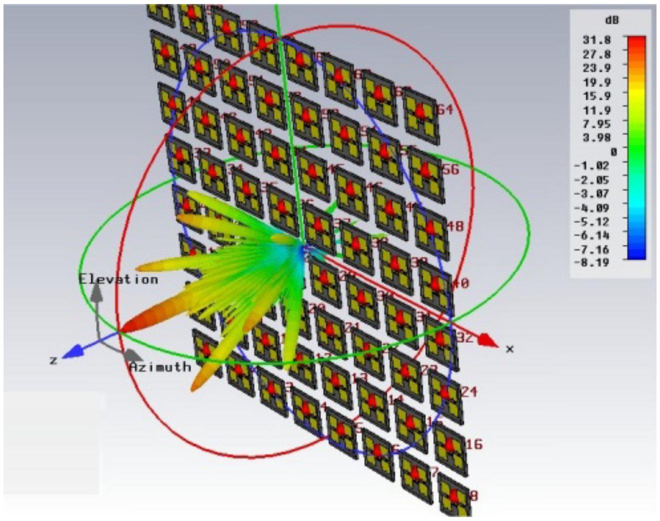


Fig.11. 3D simulation of the full receiving antenna panel (normal direction pointing)

The average gain has been evaluated over the whole Field of Regard of the radar. It is greater than 29.9 dB by computing all the beams formed for each of the 10 positions of the transmitting beam.

The signal processing implemented in the radar consists in Doppler/acceleration measurements (*i.e.*, speed estimation), detection and monopulse compression for precise direction finding. As the Field of Regard is covered by 10 positions of the transmitting beam in 10 seconds, one second of signal is analyzed at each position. Consequently, the coherent integration time is 1 second.

Thus, detections of objects that pass through the Field of Regard are characterized by Doppler traces versus time. The passage of one space object is signaled by a series of closely spaced radar plots as it can be observed on Fig.12.

Each extracted plot contains all measurements parameters, namely azimuth, elevation, Doppler (speed), radial acceleration and time of detection.

Tracking of a previously detected target is realized with consecutive plots using the Track-While-Scan (TWS) method, which combines both search and track functions. The orbital trajectory can be built from the track from latitude and longitude positions of the transmitting site and receiving site of the radar.

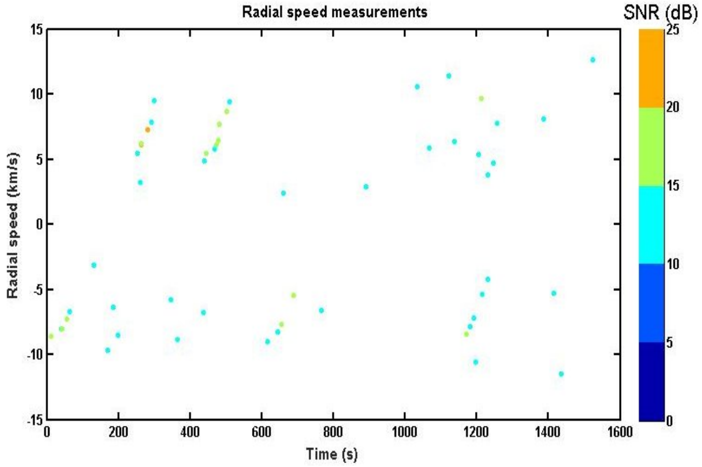


Fig.12. Doppler traces (speed) versus time (many traces of detected space objects are characterized by alignments of closed detected plots at 50 s, 250 s, 430 s, 620 s and 1200 s)

The breadboard radar has been regularly operating for more than one and half year.

Around 700 space objects have been detected and their associated tracks have been generated. The tracked objects are plotted as red dots over the Space track catalogue display on the Figure 13. More than 60% of detections are related to targets at least 3 meters in diameter at ranges less than 700 km. Objects tracked over 800 km to 1000 km are generally bigger (more or less 5 meters in diameter).

More than 40% of detections are identified as rocket debris.

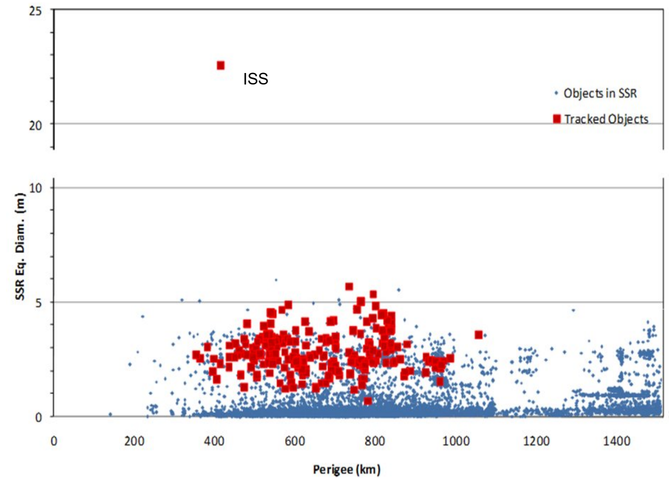


Fig.13. Estimated size of the detected objects with ABISS radar (Red) - Equivalent diameter (m) versus perigee (km) given by Space Track catalogue (Blue dots)

3. Operation of ABISS radar during ExoMars mission

The first ExoMars mission started with the launch of the spacecraft in collaboration with Russian agency Roscosmos. The launch is operated with a Proton rocket from Baikonur cosmodrome, Russian launch center.

Before separating from ExoMars, the Proton launcher vehicle needs three steps to place the spacecraft on its required Earth escape trajectory as shown on Figure 14:

- One first complete revolution around the Earth on a low parking orbit after the lift-off concluded by a boost to change its altitude,
- A second complete revolution on an intermediate eccentric orbit with a new burn to place the launcher on its transfer orbit just above North-West of Spain (in term of ground track),
- A third quasi-complete revolution with a final maneuver to eject the spacecraft on the way to Mars, above Central Africa.

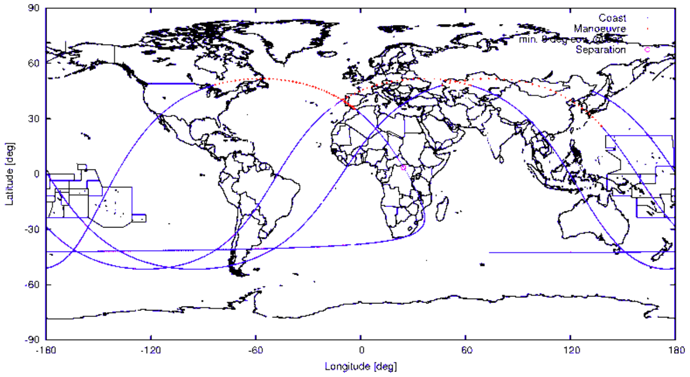


Fig.14. Estimate trajectory of the launcher (Maneuver sections in red, separation from ExoMars above Africa at magenta dot) (Sources: blogs.esa.int, http://www.russianspaceweb.com)

The maneuver to move from intermediate eccentric orbit to its transfer orbit during the second phase of the flight is exactly realized in the area where the Field of Regard of ABISS radar can be steered. The radar was set up to steer its Field of Regard so as to cover the area where this maneuver would take place as it can be seen on Figure 15.

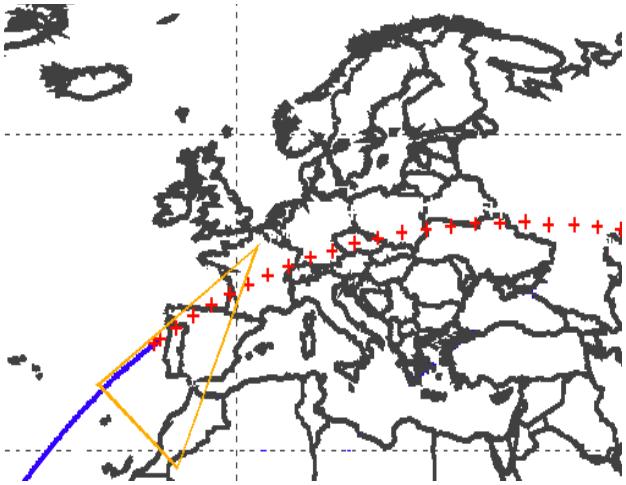


Fig.15. Trajectory of the launcher in the Field of Regard of ABISS radar (orange) during the second passage

The spacecraft was launched on March14th 2016 at 10:31 CET. The burn for this change of orbit was operated around 5 hours after lift-off.

The last stage of the launcher, with the ExoMars spacecraft, was detected by the radar approximately at the expected time when it passed through the Field of Regard [3]. At the beginning of the detection the object was just coming off the perigee of its second revolution and a new burn was operated to go on its transfer orbit.

Few plots were directly extracted from the signal processing but a corrected trajectory was calculated after a correlation with a refined trajectory. As the ExoMars launcher is a specific object in terms of kinematic behavior, we decided to have a detailed analysis of radar parameters concerning this detection [1].

Other space objects were detected during the measurements campaign for ExoMars observation. One of them is particularly useful because it appeared almost at the same time as the passage of the launcher as it can be observed on Fig.16. This space object has been identified by data processing as the rocket-debris named SL-14 R/B with a well-known actualized TLE (Two lines Elements).

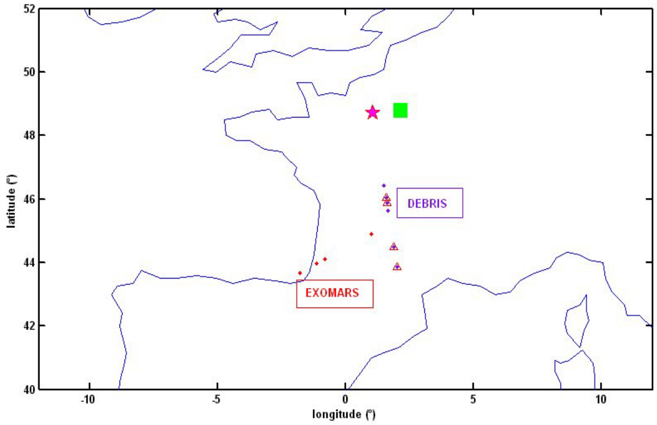


Fig.16. ExoMars launcher and rocket-debris SL-14 R/B detected simultaneously – 14th march 2016 – 10:31 CET – Triangles represent the confirmed positions of detected plots (dots)

Rocket-debris SL-14 R/B, which was detected just before ExoMars passage, is used as a calibration source for speed and acceleration measurements. The visualization of its trajectory in radar axis on Figure 17 shows the good correlation between the plots positions and the TLE of the object.

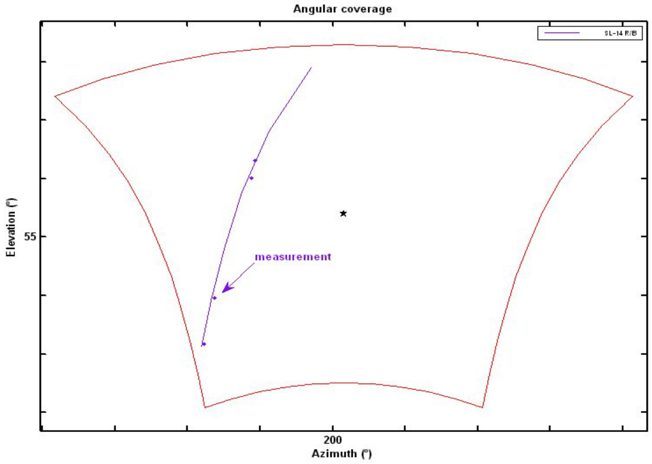


Fig.17. Rocket-debris SL-14 R/B passage in radar axis in UTC reference

As the passage of this object through the Field of Regard was confirmed by the successful tracking and approved by the identification process, the calibration of the radar was realized by evaluating the shift observed on radial speed ΔV (as shown in Fig.18) and radial acceleration parameters$\Delta\gamma$ (as shown in Fig.19) for each detected plot compared to TLE.

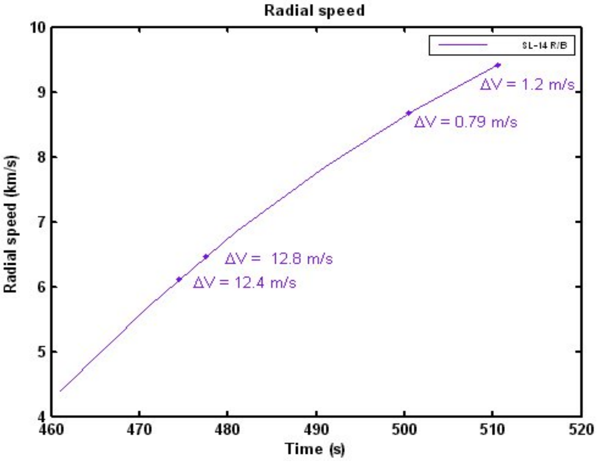


Fig.18. Rocket-debris SL-14 R/B radial speed and error margins

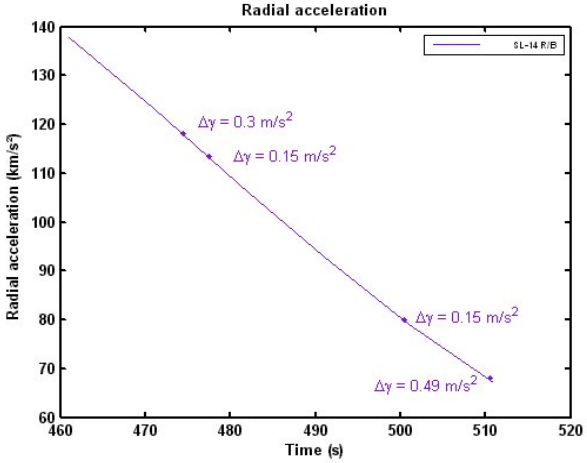


Fig.19. Rocket-debris SL-14 R/B radial acceleration and error margins

Measurements are in accordance with the parameters of movement of the target, no major bias can be observed in comparison with the values deduced from the TLE. The good correlation of acceleration versus speed confirms the synchronization of detections as presented in Fig.20. This could be used to estimate the time shift to apply for recalibration of ExoMars detections. Indeed, biases are expected to be observed along the trajectory of ExoMars because its orbit is not keplerian especially during the thrust phase for the change of altitude of the spacecraft.

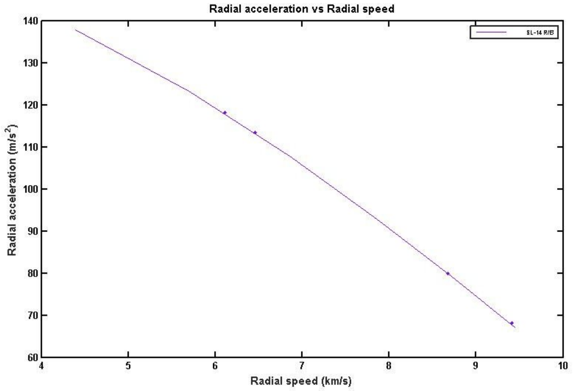


Fig.20. Radial acceleration versus radial speed of SL-14 R/B

Four detected plots of the ExoMars launcher appeared during the time window of the passage through the Field of Regard as spotted on Figure 21.

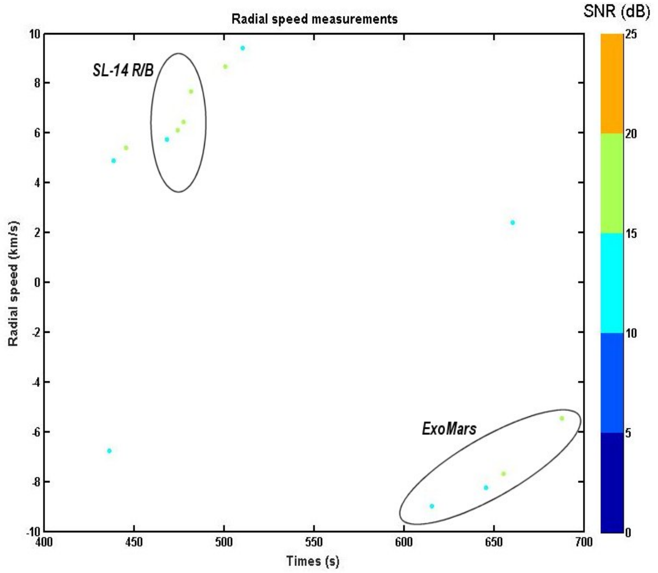


Fig.21. Detected plots around the time of ExoMars passage in the Field of Regard - Radial speed versus time (The two targets of interest have been circled - 4 plots characterize the passage of ExoMars)

The synchronization of these detections was obtained by shifting the reference time by +6 seconds (UTC + 6 s). Only two of them were matched to the estimated trajectory, especially for the beginning of the passage when the object is still in an eccentric orbit at its perigee around 600 km as shown with an estimated trajectory of classical orbit on Figure 22.

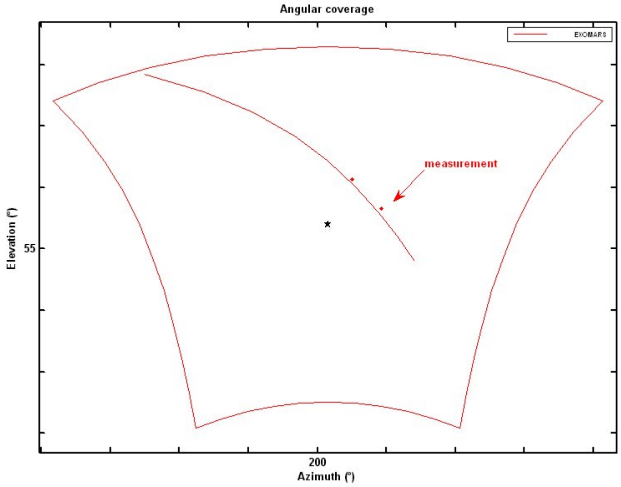


Fig.22. ExoMars launcher passage in the radar axis in UTC + 6 sec

The other two plots are not visible due to the thrust phase of the launcher (change from the second orbit to the third). The figure of radial acceleration ($\Delta\gamma$) versus radial speed (ΔV) for the two matched plots shows an important shift between measurements and predicted parameters from the TLE as observed on Figure 23.

These two biases measured on the parameters of radial speed and radial acceleration, are presented respectively on Figure 24 and Fig.25.

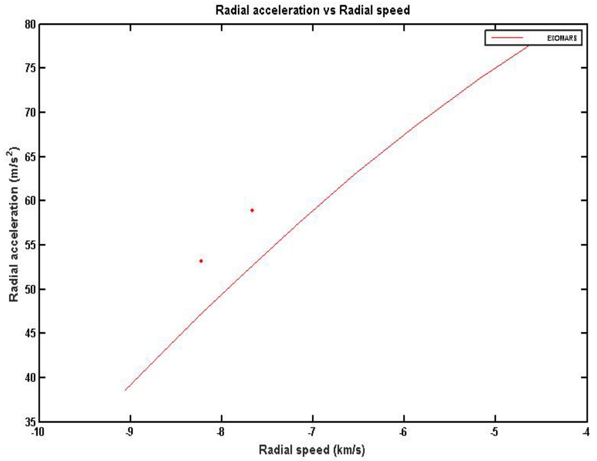


Fig.23. Radial acceleration ($\Delta\gamma$) versus radial speed (ΔV) of ExoMars

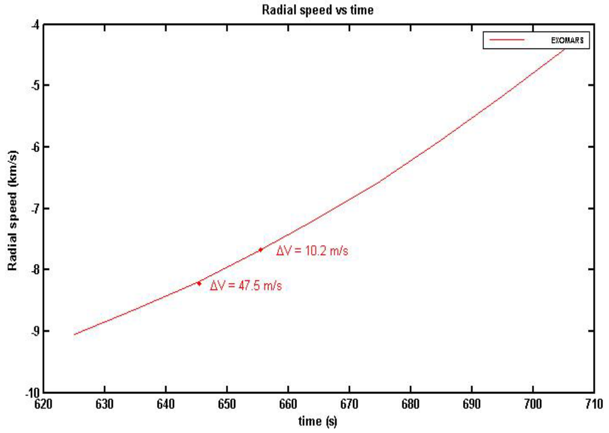


Fig.24. Radial speed of ExoMars for the first 2 detected plots

It can be noticed that the observation of the launcher was excellent during the thrust phase for orbit change, in accordance to the predicted trajectory. As shown on Fig.25, the two last plots of detection of the launcher are outside the limits of the coherence domain with a classical orbit consideration. This is due to the increase of the acceleration for the thrust.).

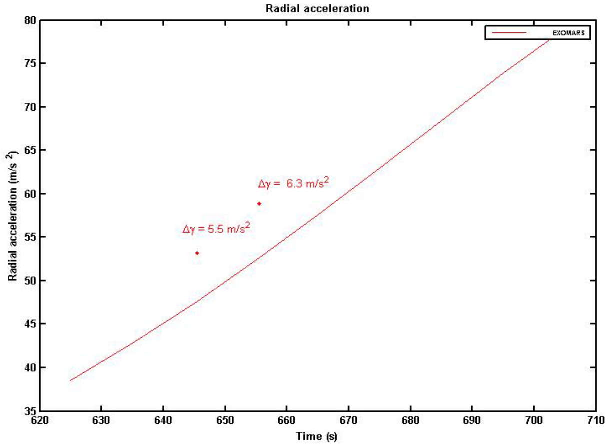


Fig.25. Radial acceleration of ExoMars for the last 2 detected plots

4. Conclusion

The challenge was to detect the ExoMars spacecraft still attached to the launcher during its passage of over South Europe with the demonstrator of Space surveillance radar ABISS. The results obtained during this campaign of experiments have shown that such a system has the capacity to detect as well as to match a specific target in the radar area of surveillance during its thrust phase as long as it stays within the domain of coherence of a classical orbit consideration in terms of kinematic.

Conflict of Interest

The authors declare no conflict of interest.

References

[1.] Saillant, S., " Kinematic analysis of the ExoMars launcher change of orbit as detected during its passage over Europe ", Proceedings of IEEE Radar Conference (RadarConf17), Seattle (2017).

[2.] Saillant, S., " Bistatic L-Band Radar To Monitor Space", Proceedings of International Radar Conference, Lille (2014).

[3.] Saillant, S., " ExoMars spacecraft detection with European Space Surveillance bistatic radar", Proceedings of CIE International Conference On Radar, Guangzhou (2016).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 106-118 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Performance of Location and Positioning Systems: a 3D-Ultrasonic System Case

Khaoula Mannay[*,1,2], Jesus Urena[1], Álvaro Hernández[1], Mohsen Machhout[3]

[1] *Department of Electronics, University of Alcala, 28805, Alcalá de Henares, Madrid, Spain*

[2] *EμE Lab Faculty of Sciences of Monastir, National Engineer School of Tunis, University of Tunis EI Manar, 1002, Tunis, Tunisia*

[3] *EμE Lab, Faculty of Sciences of Monastir, University of Monastir, 5019, Monastir, Tunisia*

ARTICLE INFO

Article history:
Received: 01 November, 2017
Accepted: 21 February, 2018
Online: 12 March, 2018

Keywords :
Indoor Ultrasonic LPS
LPS distribution
3D positioning
PDOP
Positioning error
Accuracy
LPS fusion

ABSTRACT

The necessity of navigation in people and mobile robots (MR) through specific environments (indoors or outdoors) has become more and more relevant nowadays. For indoors, generally speaking, the positioning systems can be divided into 2D (two dimensions) or 3D (three dimensions) approaches, where Ultrasonic Local Positioning Systems (ULPS) are often a common solution for MRs in 2D. This work proposes the extension of an already developed 2D ULPS to a 3D ULPS, where the compact design and the suitable performance of the initial 2D ULPS have been maintained. The ultrasonic beacons have been re-arranged to avoid co-planarity, then improving the third coordinate estimation. Furthermore, this work proposes the use of up to four ULPSs together to cover the 3D region of interest. Two configurations have actually been considered, one involving three ULPSs and another based on four. A heuristic Position Dilution of Precision (PDOP) estimation has been carried out, by taking into account two ways of obtaining the 3D-position: a) all beacons from the three different ULPSs are processed simultaneously, so all measurements are considered in the same set of positioning equations; and b) every ULPS is detected and considered separately and, later, the different estimated positions are merged. The second option is more likely to happen in a real scenario and, furthermore, the fusion of the independent positions obtained from each one of the ULPS improves the final position accuracy.

1. Introduction

The necessity for positioning and navigating objects and people indoors and outdoors is expanding more and more every day. For outdoor environments, the Global Positioning System (GPS) is widely used. Nevertheless, whether there is a lack of RF signals, indoors or in constrained outdoor environments, other approaches have already emerged as a supporting and/or alternative technology, thus providing the so-called Local Positioning Systems (LPS). The demand for these LPSs, and in general terms for indoor positioning, has also become more relevant due to the worldwide spreading of smart devices, and their corresponding location applications.

Different sensory technologies have already been used in the development of LPS, such as Wi-Fi, infrareds (IR), ultrasounds (US) or radio-frequency (RF). The final decision on one technology or another is mainly related to the type of application, the environment, the required accuracy, as well as other secondary parameters [1] [2] [3]. Among them, ultrasounds-based approach is often considered whether the required accuracy is in the range of centimetres, not only for 2D but also for 3D deployments.

This work has been initially based on the previous LOCATE-US LPS (ULPS), firstly developed by the GEINTRA-US/RF Research Group from the University of Alcalá for 2D positioning [4]. A performance analysis of an ULPS is presented here when it is adapted for 3D positioning in a certain region. The proposal consists of using several ULPSs, which cover the area under scanning from different points of view in order to achieve an accurate estimation of the receiver's position in 3D. The main goal of the paper is the study of different configurations of these ULPSs (number and location in the environment) and compare them in terms of the accuracy obtained in a grid of test points that cover all the 3D region of interest. As the positioning accuracy depends on

[*] Khaoula Mannay, Email: khaoula.mannay@gmail.com

https://dx.doi.org/10.25046/aj030213

the position itself to be determined, for every particular configuration it is necessary to study the positioning errors in all the points for all the cases. The rest of the manuscript is organized as follows: Section 2 presents a general background about LPSs, their main characteristics and technologies; Section 3 describes the previous LOCATE-US LPS, whereas the proposed 3-D ultrasonic LPS (ULPS) is presented in Section 4; Section 5 explains the 3D positioning algorithm implemented for the ULPS; Sections 6, 7 and 8 shows some simulation results; Section 9 presents a short discussion; and, finally, conclusions are considered in Section 10.

2. LPS Background

Local Positioning Systems (LPS) try to position one or multiple mobile objects accurately in an indoor environment. They are often used in applications, such as environment monitoring, people tracking, robot localization, resource management, or location-based services. The moving object is usually equipped with a small receiver (or transmitter depending on each LPS design), which acquires the emissions coming from the beacons forming the LPS, in order to estimate its own position. The nature of the transmitted signals can be different, depending on the technology involved: IR, RF, US, etc. [5]. According to this, Fig. 1 shows a general scheme for a LPS, where the beacons are fixed in the environment and the receiver is moving and computing its own position.

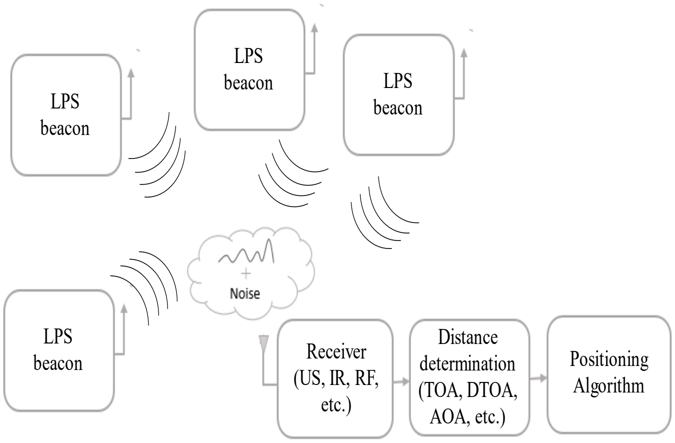

LPS beacon
LPS beacon
LPS beacon
LPS beacon
Noise
Receiver (US, IR, RF, etc.)
Distance determination (TOA, DTOA, AOA, etc.)
Positioning Algorithm

Fig. 1. General scheme of a LPS based on transmitting beacons and a moving receiver.

Furthermore, LPSs can also be classified into two groups: absolute for those where the mobile object is able to estimate its own position with respect to a reference point at any time; relative, when the mobile object can only know its position in a relative way to other receivers (or nodes) existing in the same coverage area [5].

Common LPS applications are those dedicated to 2D positioning and navigation of mobile robots and/or vehicles, as well as those focused on 3D positioning of smart devices and vehicles, such as smart phones, drones, etc. [7] [8]. Furthermore, Location-Based Services (LBS) are also potentially commercial applications for the market, not only to deliver context-dependent information accessible with a mobile device, but also to obtain information or navigate in the corresponding indoor environment. Finally, LPSs become particularly interesting in some emergency scenarios, such as positioning medical staff or equipment in hospitals, assisting rescue services in critical situations, or those proposals applied to intelligent transportation systems and/or industry manufacturing [9].

The final accuracy of a LPS in the position determination often depends on the system configuration, the sensory technology and the type and difficulty of the coverage area. Related to this, proposed LPSs are typically parametrized by the geometric dilution of precision (GDOP), where the distribution of the distance error between the estimated position and the true position is computed. Another key parameter is the range of coverage, where common values are between 5 meters and 50 meters. In this sense, for large coverage areas, scalability is key to guarantee an average positioning performance, as the positioning estimation degrades with the distance between the transmitters and receiver [10] [11] [12].

As far as the positioning algorithms are concerned, all of them are based on the determination of a variable from the beacon-receiver transmission. Some previous works based on ultra-wide band (UWB) have been recently proposed for indoor positioning by using the time differences of arrival (TDOA) from the RF transmissions between a reference beacon and the others. Typical accuracy values for these approaches are in the range of metres or even centimetres, although they present significant drawbacks, such as complexity or multi-path effects on the achieved accuracy [13] [14].

On the other hand, wireless local area networks (WLAN) have also been applied to indoor positioning [15]. Although they can also be based on TDOA, most cases deal with the received signal strength (RSS), reaching accuracies from 3m to 30m approximately [16] [17]. A similar case is the use of radio-frequency identification (RFID) [18]. In the case of infrareds, they often require a direct line-of-sight (LOS) communication between devices along a very short distance [19], where proximity, differential phase shift, and angle of arrival (AoA) are predominant.

Finally, ultrasounds-based systems also deal with TDOA [20], when there is no synchronization between the beacons and the receiver, and, afterwards, a hyperbolic positioning algorithm allows the position of the receiver to be estimated. The global accuracy of ultrasonic LPS (ULPS) is high and the structure is simple, but they often present drawbacks, such as multipath effects, Doppler, etc. [21] [22] [23].

3. General view of LOCATE_US system

The 2D LOCATE-US ULPS, designed by the GEINTRA-US/RF Research Group from the University of Alcala [4] [24], is a compact, light and portable ultrasonic beacon architecture [5]. The ULPS is formed of five coplanar ultrasonic beacons (B_i, where i=1,2,…5), placed at the four corners of a 0.707m x 0.707m square and at the centre, as is shown in Fig. 2. The five beacons have the same orientation (usually the ULPS is placed in the ceiling) and they cover roughly the same area on the floor. As the emission pattern of transducers is 120º [5], the ULPS can cover an area of 40m² roughly for a height of 3.5m. Any receiver inside the coverage area (attached to smartphones, a mobile robot (MR), a drone, ...) can compute its position in an independent and autonomous way [4] [24].

In order to cover wide indoor areas, several ULPSs can be easily deployed. Particular calibration techniques have been proposed in [27] to facilitate this deployment.

The emitters in the ULPS use a code division multiple access (CDMA) and a time division multiple access (TDMA) protocols. Every emitter is encoded with a different code, with good auto-correlation properties and low mutual interference properties with the others. The ultrasonic transmissions of the different emitters can be configured in terms of sampling frequency, modulation schemes and code patterns to be emitted. The obtained accuracy for the measurement of distances is in the centimetre range (we assume a typical deviation of 1cm). The distances measured can be up to about 20m. All that is enough for MR applications in 2D spaces, and if extensive areas must be covered we can use a set of single ULPSs [4].

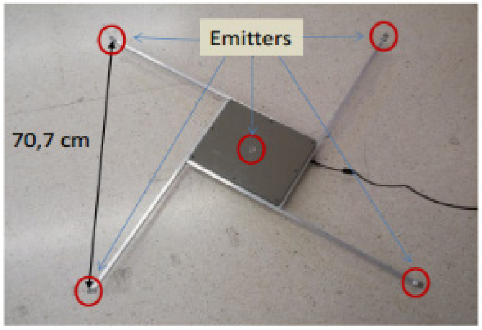


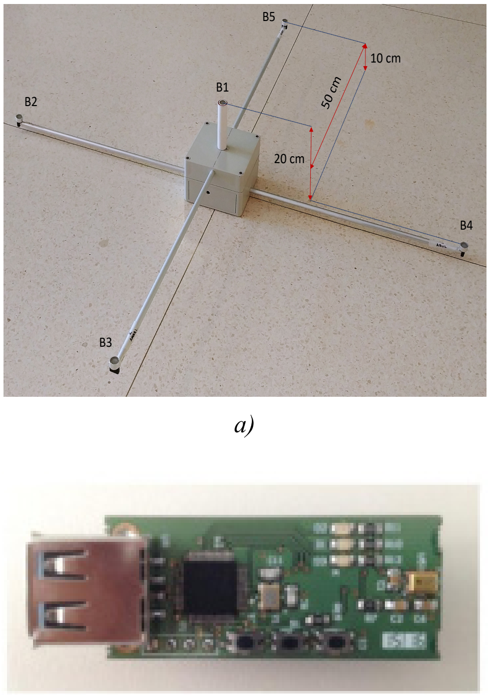
Fig. 2. General view of the 2D LOCATE-US LPS

4. Proposed 3D ULPS

The 3D ULPS described hereinafter is an extension of the 2D LOCATE-US, so it is also formed by five ultrasonic beacons B_i placed at the four vertices of a 0.707m x 0.70 m square and at the centre, as is shown in Fig. 3.a). All of them present slight variations in the z coordinate to improve this coordinate estimation and to avoid co-planarity; but still keeping the same properties of the previous 2D ULPS, such as the common orientation and coverage area. The new beacon distribution is: B2 and B4 are in the base plane, B3 and B5 are 10cm high from the base plane, and finally B1 is placed at 20cm high from the base plane [24].

The ultrasonic beacons are wired-synchronized to enable simultaneous and periodic emission [4] [24]. The ultrasonic transmission are encoded with orthogonal 1023-bit Kasami sequences, in order to mitigate any effect coming from multiple access interference (MAI) as much as possible [25]. These codes have been selected due to their suitable auto-correlation and cross-correlation properties. For their transmission, a binary phase shift-keying (BPSK) modulation has been carried out, with a carrier placed at the central frequency of the transducer bandwidth, f_c=41.67kHz. Two carrier cycles per modulation symbol have been applied, and ultrasonic transmissions are carried out periodically every 50ms to reduce multipath effects.

The involved ultrasonic transducer is the Prowave 328ST160 [26], together with an ad-hoc front-end designed for this application. The five beacons B_i are managed by a FPGA-based system, in charge of controlling the global operation of the ULPS through an Ethernet link [24].The type of binary sequence and its length, the number of carrier periods, the type of carrier, and the time interval between different emissions can be configured by the PC. A digital pass-band filter with a 40 kHz central frequency and a 10 kHz bandwidth has been included to constrain information to the ultrasonic transducer bandwidth [26].



a)

b)

Fig. 3. a) General view of the 3D LOCATE-US LPS; b) Designed ultrasonic receiver.

With regard to the reception, the moving device estimates its position asynchronously by hyperbolic trilateration from the TDOA measurements between a reference beacon and the others. To detect the TDOAs, the mobile device demodulates the received signal and correlates it with the corresponding emitted Kasami sequences, based on a generalized cross-correlation (GCC) of the received signal. A main lobe appears at the arrival instant for every Kasami sequence, so it is possible to calculate the associated TDOAs. Finally, a Gauss-Newton minimization method computes the position. Fig. 3.b) shows the general aspect of the used ultrasonic receiver [24].

5. 3D Positioning

In positioning systems, it is important to know the accuracy of the estimated position, which depends on the quality of measurements (ranging distances, signal strength, etc.), and on the geometry of the positioning system (beacons' geometry) with respect to the mobile node. The Position Dilution of Precision (PDOP) includes such dependencies and can be obtained empirically using (1).

$$PDOP \approx \frac{\sqrt{\sigma_x^2+\sigma_y^2+\sigma_z^2}}{\sigma_m} \qquad (1)$$

Where σ_x^2, σ_y^2 and σ_z^2 are the position variances in the three axes X, Y and Z, respectively; and σ_m is the standard deviation in the distance measurements (assumed to be 1cm hereinafter). As has been already mentioned, the technique used to position objects in 3D is based on the TDOAs, thus requiring synchronization between beacons (emitters), but not with receivers.

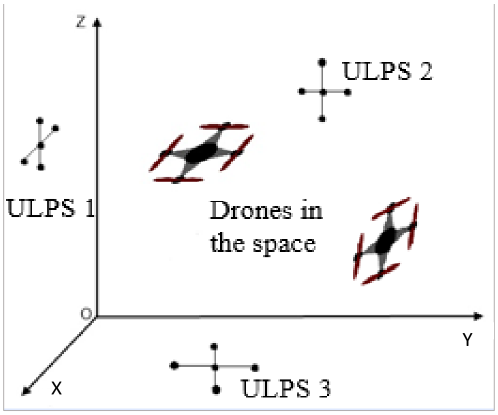


Fig. 4. Example of 3D configuration using three ULPSs and two receivers (drones).

In order to position a mobile target in a 3D space (see Fig. 4), a single ULPS is not enough to cover all the space with enough accuracy, assuming an 8x8x8m^3 volume. The setup with a single ULPS implies that the performance significantly degrades with the height variation, due to the poor Vertical Dilution of Precision (VDOP), which determines the performance by only taking into account the typical deviation in the *z* coordinate.

Thereby, the use of only one ULPS is not a suitable solution. As an example of that, Fig. 6 shows the case in which one ULPS is placed at the centre of the floor at coordinates (4m, 4m, 0m), whereas Fig. 8 depicts the case of one ULPS placed at the lower corner with coordinates (0m, 0m, 0m). Both Figs. represent the cloud of obtained position points, assuming the receiver in the X-Y plane (with steps of 1m in both axes), for three different heights: *z*=6m in the half upper volume of the room; *z*=4m in the middle of the room; and *z*=2m in the half lower volume of the room.

Additionally, Figs. 7 and 9 show the different PDOP values for the same X-Y planes and heights (*z*=2m, 4m, 6m). Note that the PDOP values are high in general terms (above 100). The contour map is a representation of the PDOP in planes at different height. The values of the PDOPs have been calculated for every point in the grid (according to the cloud of points with the estimated positions after the simulations).

Each color represent a particular value of PDOP; and the greater the PDOP the greater the positioning error in this point (even in the case that all the distances has been measured always with a typical deviation of 1 cm). With this representation one can have an idea about the error we can wait in each region of the environment in a real situation with a particular ULPS arrangement.

On the other hand, the receiver has been placed at every point in a 9x9x9 grid (8x8x8m^3 volume using 1m interval). At each point, a hundred simulations have been run by using an hyperbolic trilateration with the Gauss-Newton Positioning Method (GNPM) [10]; the standard deviation in the distance measurements is σ_m =1cm, which is consistent with ultrasonic measurements of distances.

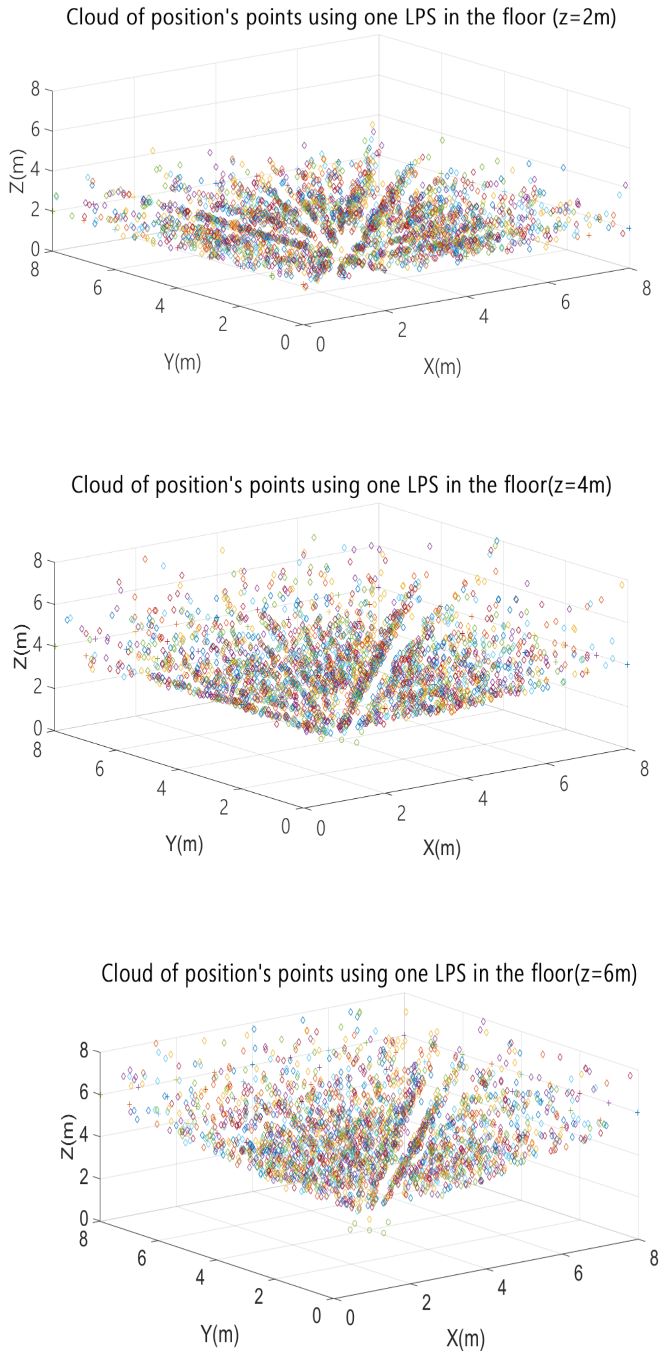


Fig. 6. Cloud of position points for one ULPS placed at the centre of the floor (4m,4m,0m) for X-Y planes at different heights (z=2m, z=4m, z=6m).

As can be observed in Fig. 7, the PDOP values differ from one plane to another, but, in general, it is smaller in the centre of every plan, and then it increases towards the sides of the room. For the first plane *z*=2m (lower half of the room), the PDOP varies from 100 to 900.

Whereas these values increase in the middle of the room (*z*=4m), where the PDOP is between 250 and 800; and, finally, they range from 450 to 850 in the third plane *z*=6m (upper half of the room).

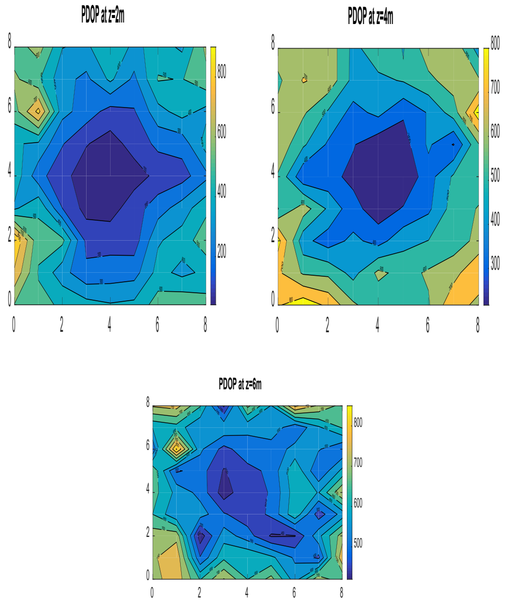


Fig. 7. Colour map of PDOPs for an ULPS placed at the centre of the floor (4m, 4m, 0m) for different X-Y planes (z=2m, z=4m, z=6m).

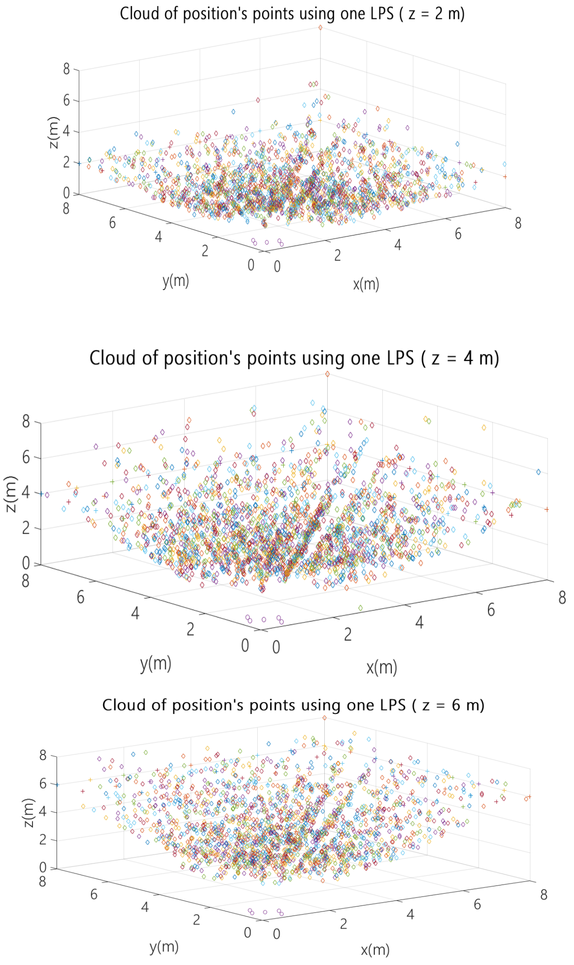


Fig. 8. Cloud of position points using an ULPS placed at (0m, 0m, 0m) corner for X-Y planes z=2m, z=4m and z=6m.

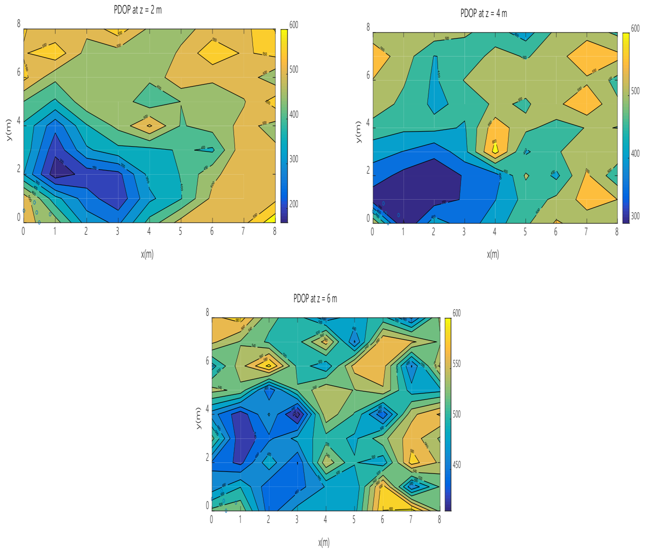


Fig. 9. Colour map of PDOPs for an ULPS placed at (0m, 0m, 0m) corner for X-Y planes (z=2m, z=4m, z=6m).

Similar conclusions can be derived from Fig. 9 with the LPS placed at corner (0m, 0m, 0m) and pointing at the centre of the room. For the first plane *z*=2m, the PDOP values are between 200 and 600; for *z*=4m, they are between 300 and 600; and between 420 and 600 for *z*=6m.

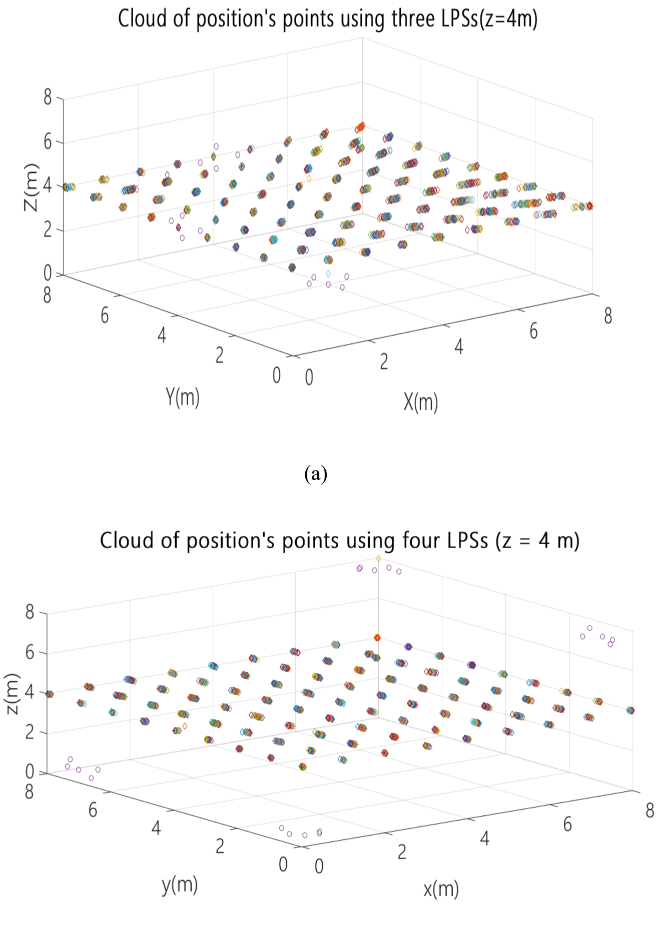
In order to improve these results, as well as to enhance the coverage area, the use of several ULPSs placed at different points of the room has been considered. These tests have been carried out by using two different configurations:

- Configuration A: three ULPSs placed at the centres of X-Y plane *z*=0m, Y-Z plane *x*=0m and X-Z plane *y*=8m, respectively. All these ULPS are emitting perpendicularly to the plane in which they are placed.
- Configuration B: four ULPSs placed at corners (0m, 0m, 0m), (8m, 8m, 8m), (0m, 8m, 0m) and (8m, 0m, 8m), respectively. Each LPS is emitting in the direction of the cube diagonal corresponding to the corner at which it is placed.

In practical situations, the 3D position can be computed with one or an array of microphones to cover all the incoming signals. In this way, two options have been considered here:

- Simultaneous measurements: all the distances (derived from TDOAs), from all the ULPSs that must be synchronized, are obtained at the same time, so the positioning algorithm involves as many equations as measured distances. For the three ULPSs configuration, the positioning algorithm requires fifteen equations, whereas, in the four ULPSs configuration, twenty equations.
- Independent measurements for each ULPS: whether five distances from one single ULPS are obtained at the receiver, it obtains a 3D position. In parallel, several 3D positions can be computed (one per every detected ULPS). To combine all these 3D positions, the Maximum Likelihood Estimation (MLE) is applied. In the case of having available three independent measurements q_1, q_2 and q_3 for a certain position q, provided that the positioning error may be modelled as $p(q_i|q)=N(q, \sigma_i)$, the

merged estimate q_{MLE} for position q can be obtained by following the procedure (2):



(a)

(b)

Fig. 10. Cloud of position points for the X-Y plane z=4m using: (a) configuration A with three ULPSs; (b) configuration B with four ULPSs.

$$q_{MLE} = \frac{\sigma_1^{-2}\cdot q_1+\sigma_2^{-2}\cdot q_2+q_3^{-2}\cdot z_3}{\sigma_1^{-2}+\sigma_2^{-2}+\sigma_3^{-2}} \qquad (2)$$

Where q_1, q_2 and q_3 are three independent measurements for position q in the case of detecting three different ULPSs. Since the statistical information is additive, the new standard deviation σ will be (3):

$$\sigma^{-2} = \sigma_1^{-2} + \sigma_2^{-2} + \sigma_3^{-2} \qquad (3)$$

Where σ_1, σ_2 and σ_3 are the standard deviations for the corresponding position measurements q_1, q_2 and q_3.

6. Results for configuration A (three ULPSs on the walls)

In order to compare the performance of each ULPS arrangement and the total number of distance measurements available, for both configurations A and B, one hundred simulations have been conducted at every point in the 3D grid (9x9x9m^3) with a step in x, y and z of 1m. The standard deviation considered in the distance measurements is σ_m=1cm.

6.1. Simultaneous measurements from all ULPSs

In this case, the 3D receiver position is estimated by the GNPM at each point, assuming there are available simultaneous measurements of TDOAs from all the ULPSs at the receiver. For configuration A, the space is covered by three ULPSs, as shown in Fig. 10. Two cases have been studied: only two of them are available; all the three are available.

6.1.1. Two ULPSs available

Firstly, in addition to the ULPS placed at the position (4m, 4m, 0m), a second ULPS has been placed at the centre of the wall *y*=8m, at position (4m, 8m, 4m). Figs. 11 and 12 show the cloud of positions obtained, as well as the PDOP for planes *z*=2m, *z*=4m and *z*=6m.

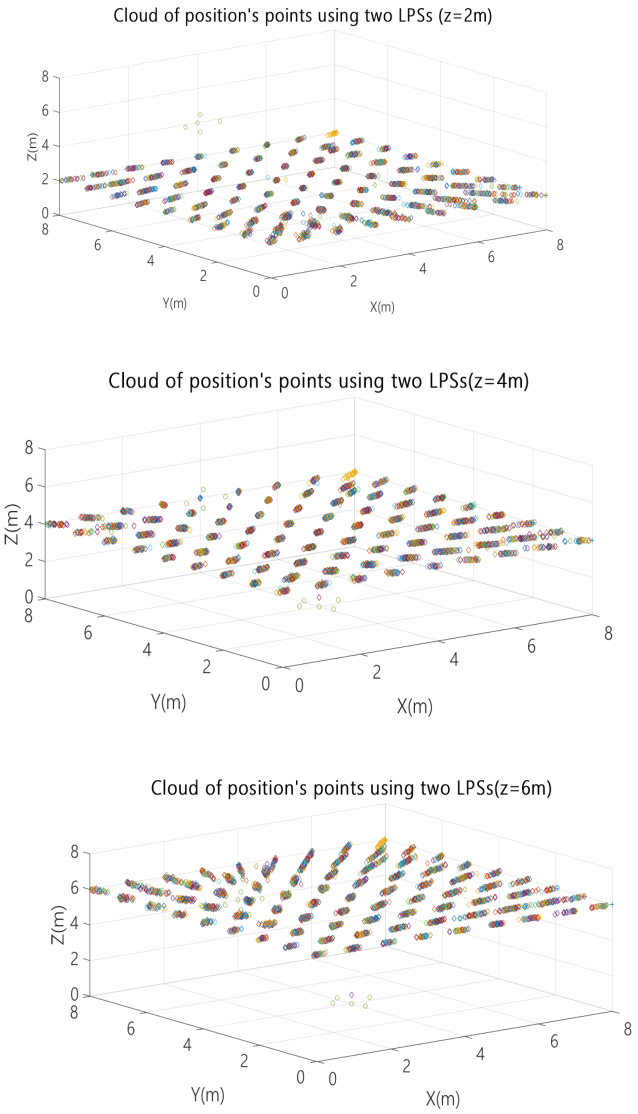


Fig. 11. Cloud of position points for two ULPSs: one placed at (4m, 4m, 0m) and the other at (4m, 8m, 4m), for different X-Y planes (z=2m, z=4m, z=6m).

Comparing with the previous results achieved with only one ULPS (Figs. 6 to 9), the improvement is clear: now the PDOP varies from 5 to 100, providing larger volumes with lower values. The lowest PDOP values are around the centre of the room.

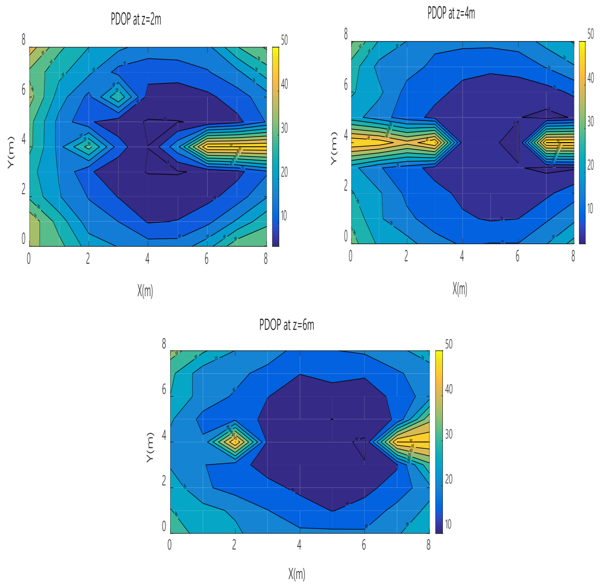


Fig. 12. Colour map of PDOP values for two ULPSs placed at (4m, 4m, 0m) and the other at (4m, 8m, 4m), for different X-Y planes (z=2m, z=4m, z=6m).

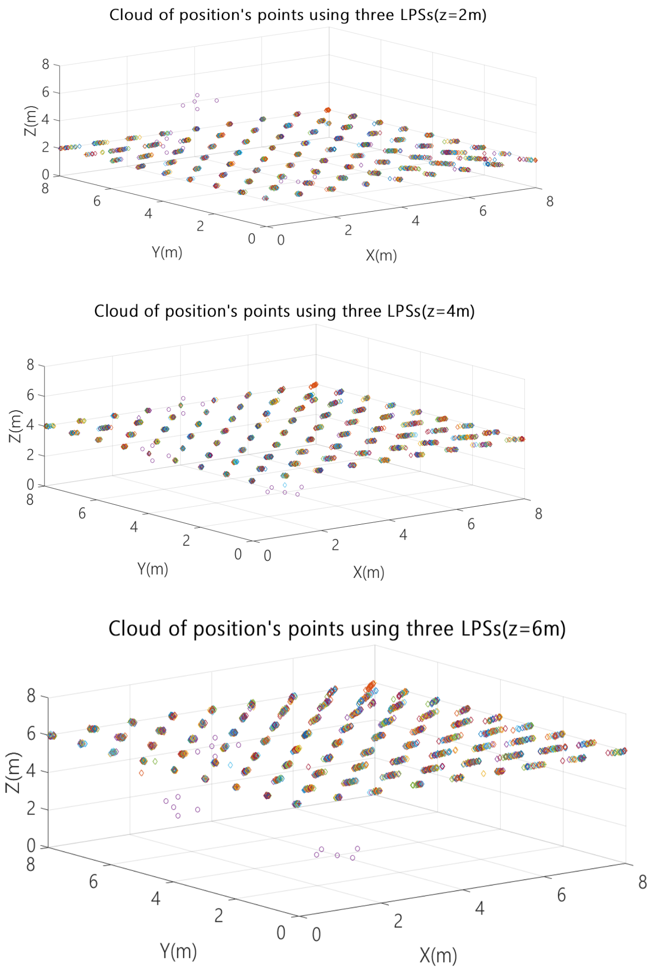


Fig. 13. Cloud of position points for three ULPSs placed at (4m, 4m, 0m), (4m, 8m, 4m) and (0m, 4m, 4m) for different X-Y planes (z=2m, z=4m and z=6m).

6.1.2. Three ULPSs available

A third ULPS has been added at coordinates (0m, 4m, 4m). Figs. 13 and 14 show the same clouds of points and PDOPs as before, for X-Y planes *z*=2m, *z*=4m and *z*=6m.

By adding the third ULPSs, the cloud of the 100 position points simulated is more accurate around the real positions considered in the grid. The PDOP values decrease below 30 in general terms, and below 15 in almost all the space, independently of the height.

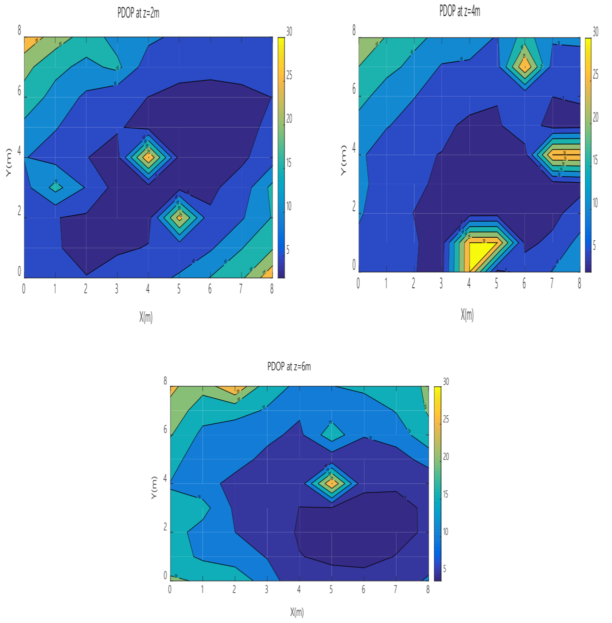


Fig. 14. Colour map of PDOP values for three ULPSs placed at (4m, 4m, 0m), (4m, 8m, 4m) and (0m, 4m, 4m) for different X-Y planes (z=2m, z=4m, z=6m).

6.2. *Independent measurements for each ULPs*

6.2.1. Fusion of two ULPSs

The configuration analysed here is the same as that one in Section 5.1.1. Now, the 3D positions from the two different ULPSs are obtained separately, and, afterwards, these positions are combined by using a MLE to estimate a final position. The process has been repeated a hundred times at each point in the aforementioned grid. The final PDOP values are shown in Fig. 15, varying from 5 and 100, also providing large areas with low values. It is worth noting that these results are similar to those in Fig. 12 for the case of simultaneous measurements.

6.2.1. Fusion of three ULPSs

When having a third ULPS placed at (4m, 0m, 4m), the PDOP values have been calculated for a hundred simulations at each point in grid, and they are shown in Fig. 16. Again, the distribution of the PDOP values with the MLE merging method are very close to the case when using simultaneous measurements (see Fig. 14).

In order to summarize the results for the case analysed in this section, for the whole volume under analysis the Cumulative Distribution Function (CDF) of the position error has been obtained, by taking into account all the points in the grid (100 simulations per each position). Fig. 17 shows the CDF for simultaneous (a) and independent (b) measurement approaches.

According to Fig. 17, for simultaneous measurements, in the 90% of the cases the error is below 0.7m when using only one ULPS, below 0.2m with two ULPSs, and below 0.1m for three ULPSs. On the other hand, for independent measurements, in the 90% of the cases the error is below 0.9m for one ULPS, 0.8m for the fusion of two LPSs, and 0.7m for the fusion of three ULPSs.

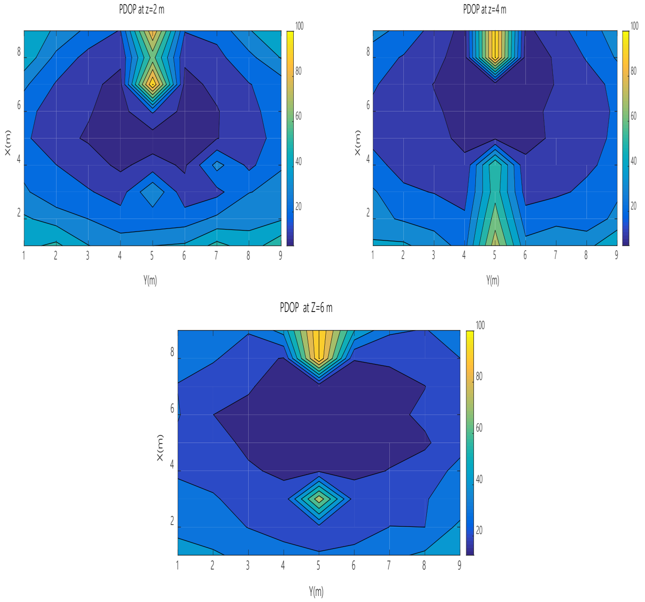


Fig. 15. Colour map of the PDOP values obtained by the MLE when using two ULPSs at (4m, 4m, 0m) and at (4m, 8m, 4m) for X-Y different planes (z=2m, z=4m, z=6m).

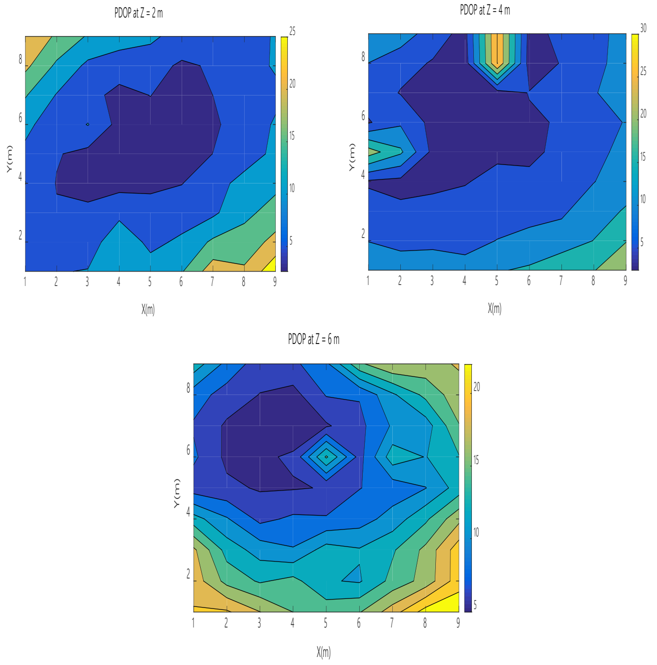
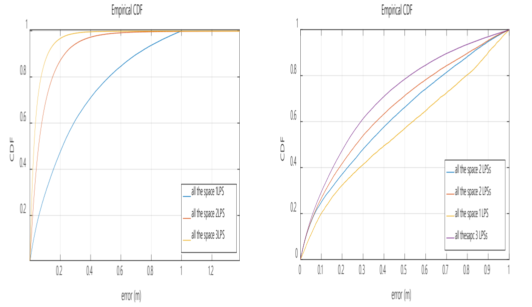


Fig. 16. Colour map of the PDOP values obtained by the MLE when using three ULPSs at (4m, 4m, 0m), (4m, 8m, 4m) and (0m, 4m, 4m) for X-Y different planes (z=2m, z=4m, z=6m).



(a) (b)

Fig. 17. CDF for the position error in the whole volume: a) using simultaneous measurements; and b) using independent measurements (and fusion).

7. Results for configuration B (with four ULPSs)

As has been already explained in Section 4, the whole volume is covered by four ULPSs in the configuration B. The goal is to improve the results presented in Figs. 8 and 9, where only one ULPS located at position (0m, 0m, 0m) was used.

7.1. *Simultaneous measurements*

7.1.1. Two ULPSs

Together with the ULPS located at (0m, 0m, 0m), a second one is added at the opposite corner, that is, located at (8m, 8m, 8m). The resulting clouds of points and PDOP after 100 simulations at each grid point are shown in Figs. 18 and 19, respectively, for planes *z*=2m, *z*=4m and *z*=6m.

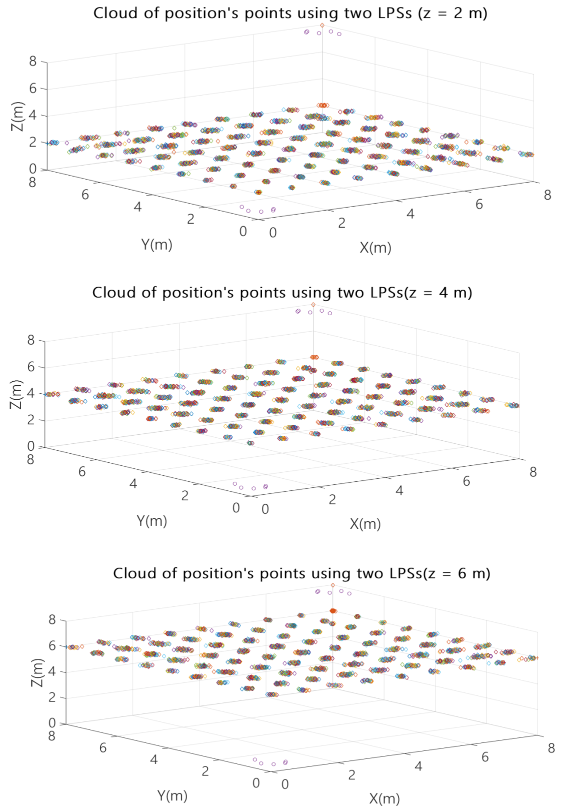


Fig. 18. Cloud of position points for two ULPSs at corners (0m, 0m, 0m) and (8m, 8m, 8m) for planes (z=2m, z=4m, z=6m).

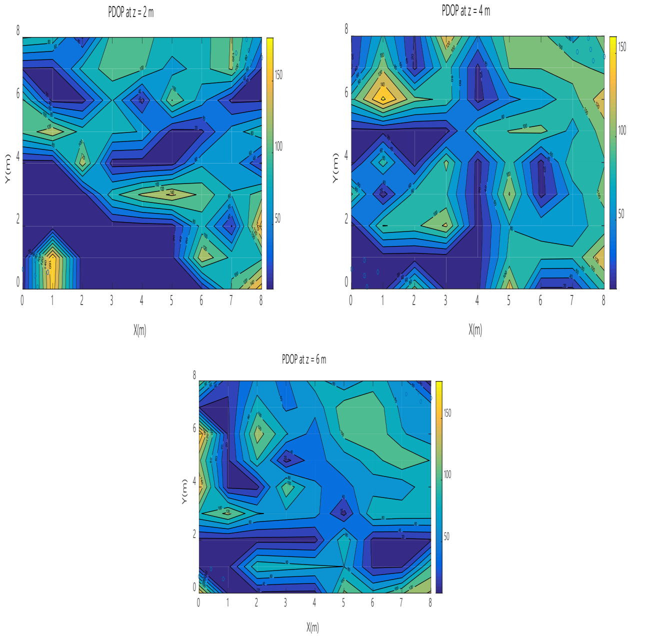


Fig. 19. Colour map of PDOP values for two ULPSs at (0m, 0m, 0m) and at (8m, 8m, 8m) for X-Y planes (z=2m, z=4m, z=6m).

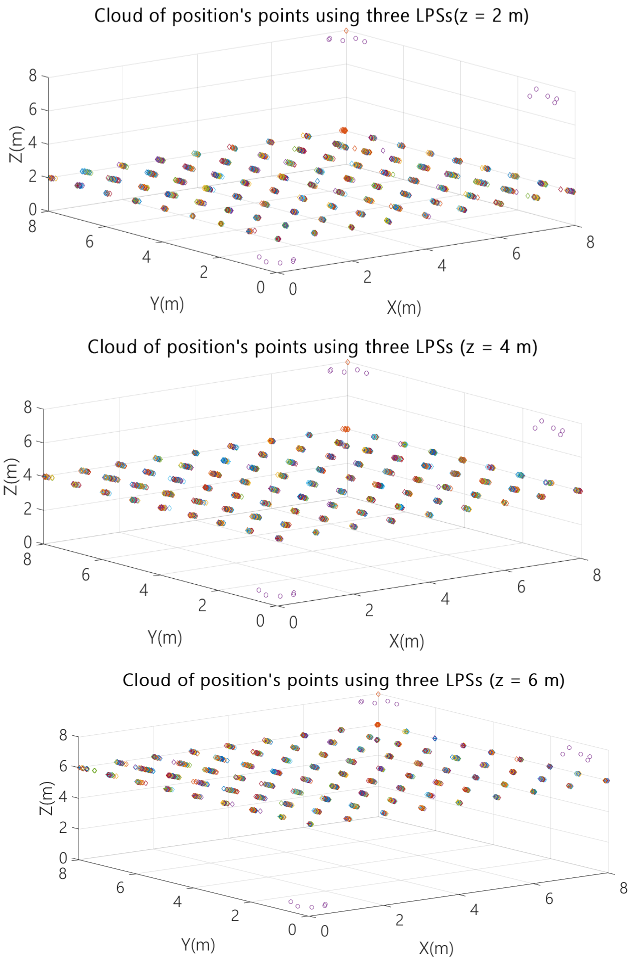


Fig. 20. Cloud of position points for three ULPSs, placed at (0m, 0m, 0m), (8m, 8m, 8m) and (8m, 0m, 8m) for X-Y planes z=2m, z=4m and z=6m.

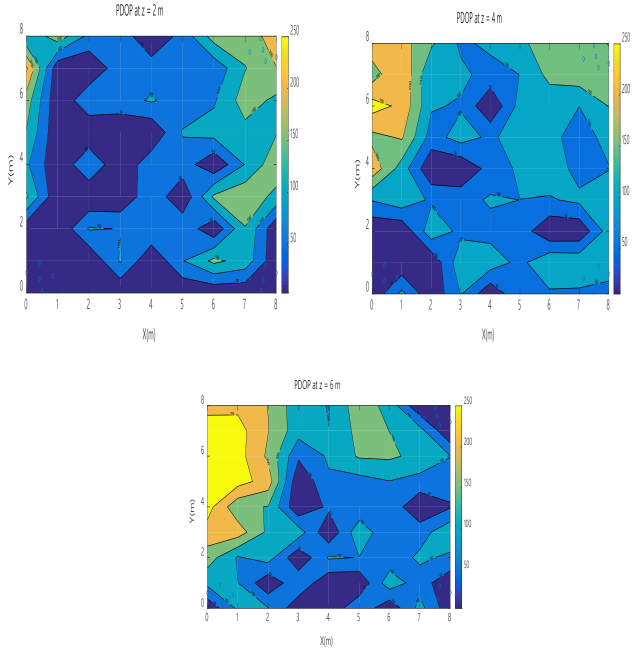


Fig. 21. Colour map of PDOPs for three ULPSs at (0m, 0m, 0m), (8m, 8m, 8m) and (8m, 0m, 8m) for X-Y planes (z=2m, z=4m,z=6m).

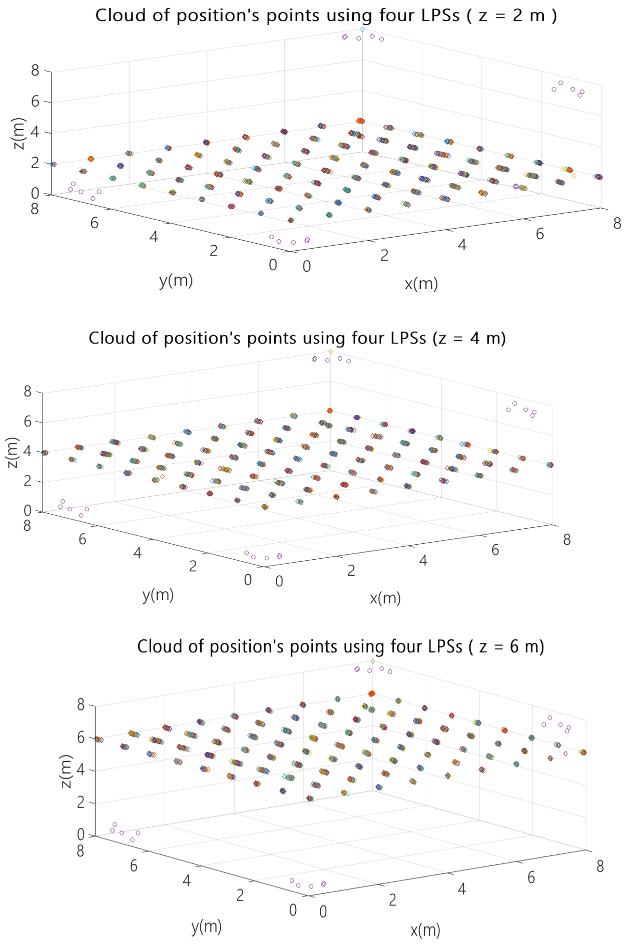


Fig. 22. Cloud of position points for four ULPSs at (0m, 0m, 0m), (0m, 8m, 0m), (8m, 0m, 8m) and (8m, 8m, 8m) for X-Y planes (z=2m, z=4m, z=6m).

It can be observed that there is an improvement in the error for all the grid points, compared to the case of only one ULPS, but still the PDOP values are in the interval from 10 to 180.

7.1.1. Three ULPSs

With a third ULPS placed at the corner (0m, 8m, 0m), the obtained results are shown in Figs. 20 and 21, also for the three planes *z*=2m, *z*=4m and *z*=6m.

The 3D positions are now more accurate, since the errors decrease especially in the neighbourhood of the ULPSs. The PDOP values are below 50 in large areas of the analysed planes.

7.1.2. Four ULPSs

Finally, a fourth ULPS is inserted at (0m, 8m, 0m), in addition to the previous three ones. The corresponding results are plotted in Figs. 20 and 21.

The errors have been considerably reduced with PDOPs between 5 and100, including large areas below 20.

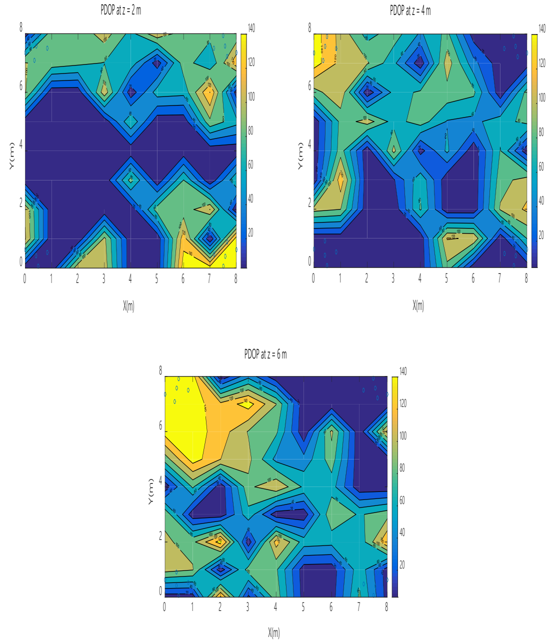


Fig. 23. Colour map of PDOP values for four ULPSs at (0m, 0m, 0m), (0m, 8m, 0m), (8m, 0m, 8m) and (8m, 8m, 8m) for X-Y planes (z=2m, z=4m, z=6m).

7.2. Independent measurements

In the second case analysed with the configuration B, each ULPS is considered independently. Figs. 24, 25 and 26 represent the PDOP values for two, three or four ULPSs,

respectively, after the fusion of the independent calculations for the positions.

Note that for two ULPSs the PDOP values are still as high as 420 in relatively large areas around the centre of the space. These areas are greatly reduced with three ULPSs (see Fig. 25) and even more with four ULPSs (see Fig. 26).

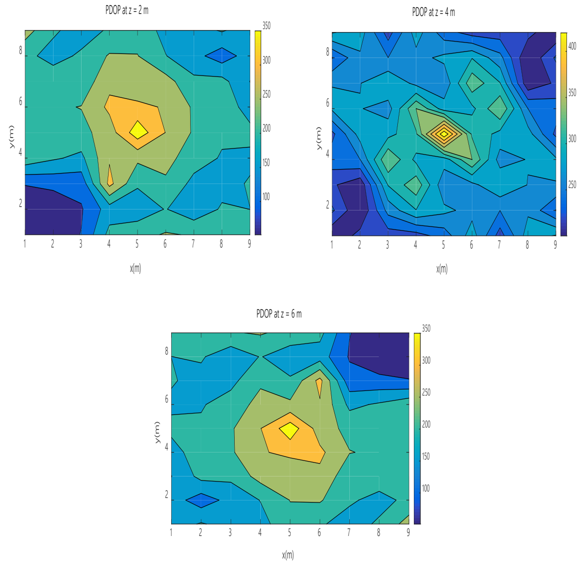


Figure 24. Colour map of the PDOP values obtained by the MLE when using two ULPSs placed at (0m,0m,0m) and (8m,8m,8m) for different X-Y planes (z=2m, z=4m, z=6m).

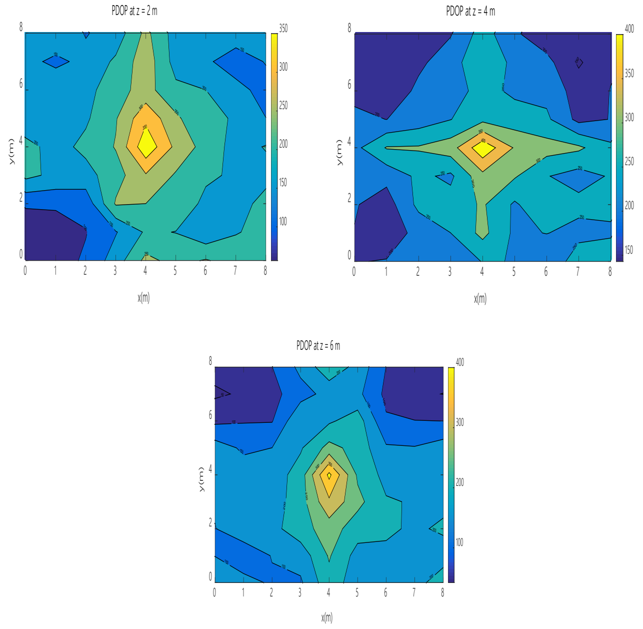


Figure 25. Colour map of the PDOP values obtained by the MLE when using three ULPSs placed at (0m,0m,0m), (8m,8m,8m) and (0m,8m,0m) for different planes (z=2m, z=4m and z=6m).

7.2.1. Error CDF

The error CDF, again obtained in all the points (and for a hundred simulations at each) for these distributions of ULPSs previously described, can be seen in Fig. 27. The upper plot considers simultaneous measurements, whereas the second one involves the fusion of independent measurements from each ULPS. In both, different CDF plots are shown for one, two, three and four ULPSs.

For the simultaneous emission, the errors for the 90% of the cases are below 0.75m using one ULPS, 0.15m for two ULPSs, 0.1m for three ULPSs and, finally, 0.07m for four LPSs (see Fig.

27.a). In case of fusion of data from independent ULPSs, these errors are below 0.85m for one ULPS and below 0.7m for four ULPSs (see Fig. 27.b).

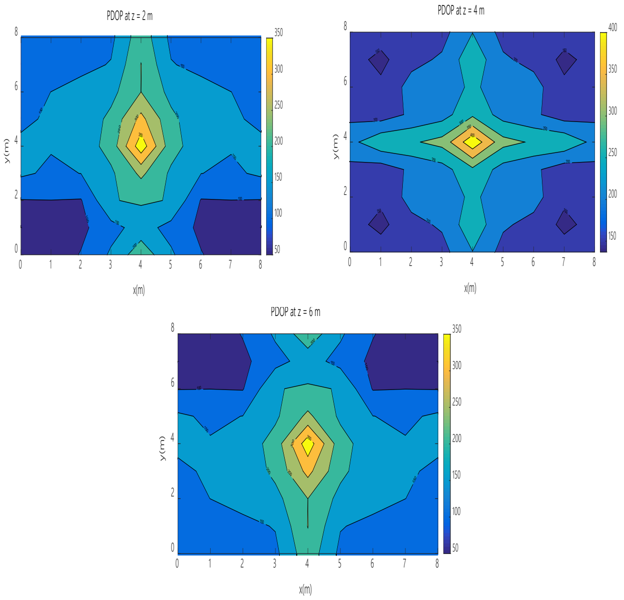
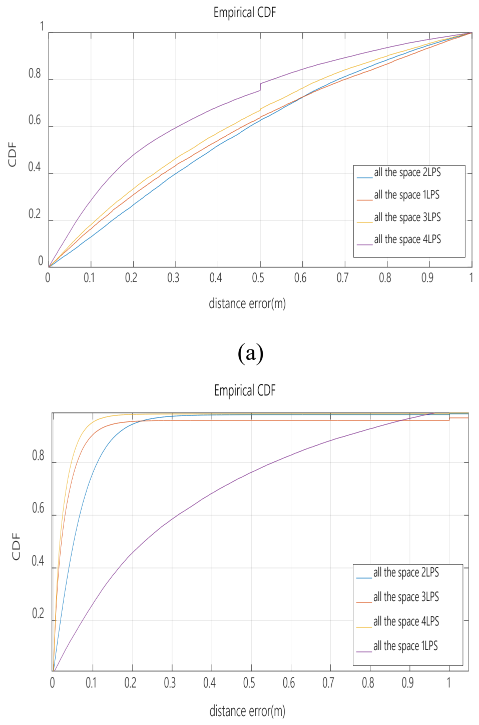


Figure 26. Colour map of the PDOP values obtained by the MLE when using four ULPSs placed at (0m,0m,0m), (8m,8m,8m), (8m,0m,8m) and (0m,8m,0m) for different X-Y planes (z=2m, z=4m, z=6m).



(a)

(b)

Figure 27. Error CDF in all the space, for 1, 2, 3 or 4 ULPSs and considering: a) simultaneous measurements with all the ULPSs; and b) independent measurements from each ULPS.

Finally, for the best case (simultaneous measurements with four ULPSs), the CDF has been split and represented in Fig. 28 for the points in every analysed plane *z*=0m, 1m, 2m …8m. Note that the error is quite similar in all the planes and, for the 90% of the cases, it is below 6-8 cm. That is, the performance of the system is similar in the whole volume.

8. Experiment results

In order to validate the proposal with real data, firstly, configuration A presented in Section 5 has been considered by using one ULPS. The tests have been carried out by positioning a receiver at each one of the 16 specific positions in a grid on the floor, as can be observed in Fig. 29. The ULPS is placed in the ceiling at a height of 3.5m. This workspace is not very complex and only the points of the corners of the grid have some problems due to signal attenuation and multipath.

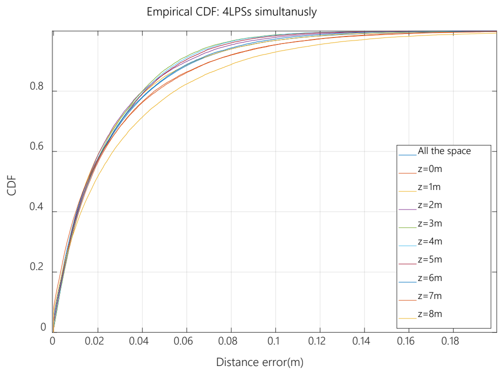


Fig. 28. Error CDF in every X-Y plane (z=1,2,…,8) for 4 ULPSs with simultaneous measurements.

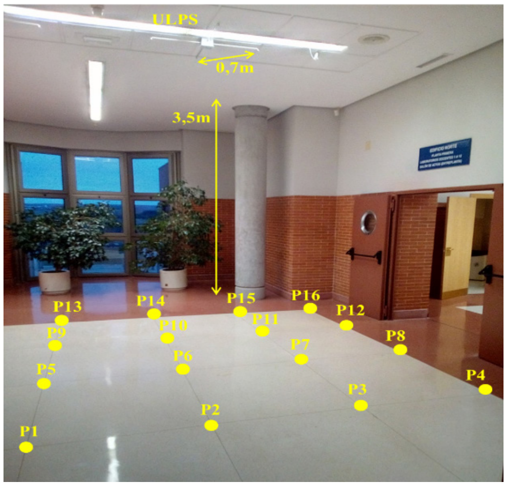


Fig. 29. Experimental setup formed by one ULPS placed in the ceiling at a height of 3.5m and the receiver placed at 16 points in a 4x4m² grid on the floor.

A set of 100 measurements have been carried out for every position, where the final estimated positions are represented by blue crosses in Fig. 30. On the other hand, the red circles are the real positions. An error ellipse containing the 95% of estimates has also been represented for every cloud of points. The green diamonds are the beacons' projections on the floor. It can be observed that the estimated positions at the test points below the beacons have a lower error than those further away. Nevertheless, all the positions are reasonably estimated.

Fig. 31 shows the error CDF obtained by simulating only one ULPS for the grid of points at z=3.5m. It can be observed that, for the 90% of cases, the positioning error is less than 1 cm. On the other hand, Fig. 32 represents a similar CDF for experimental data, considering all the measurements or grouping them into two subsets: points P6, P7, P10 and P11 (placed at the center of the grid, just below the ULPS) and the rest of points. The differences between simulated and experimental data can be due to several factors: inaccuracies in the position of the receiver when tests were performed, and propagation effects not considered in simulations (attenuation, multipath, non-Gaussian noise, etc.).

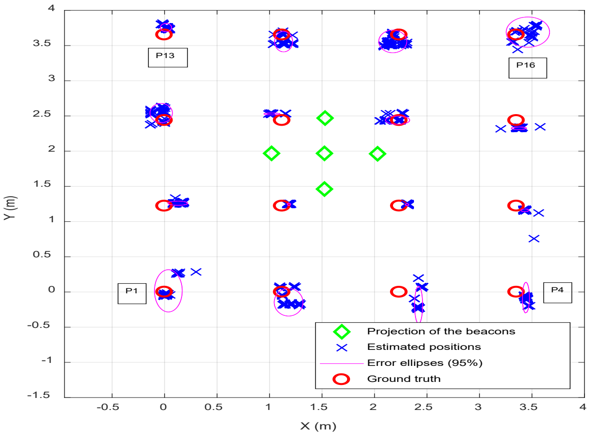


Fig. 30. Experimental test positions (red circles), estimated positions (blue crosses), and beacons' projections (green diamonds)

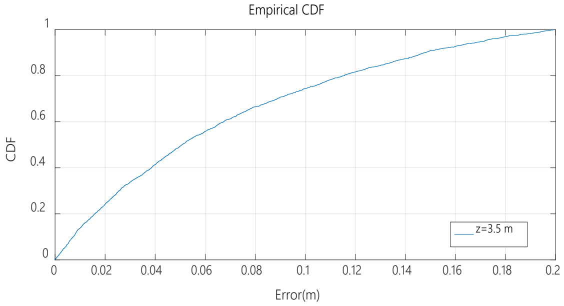


Fig. 31. Error CDF obtained by simulation for the X-Y plane (z=3.5m) using only one ULPS.

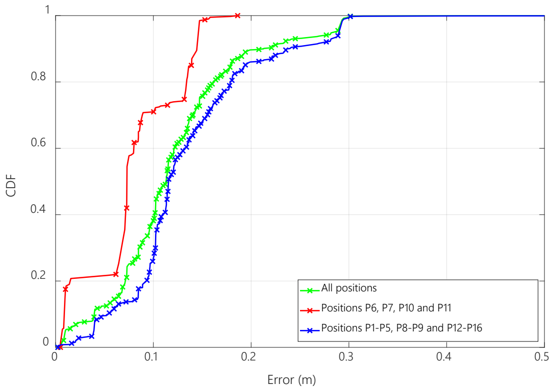


Fig. 32. Error CDF for the experimentally estimated positions in the X-Y plan (z=3.5m) using one ULPS.

Note that, according to the different CDFs in Fig. 32, the error is less than 20cm in the 90% of cases when considering all the test positions. This error is smaller (about 15cm) for those points just below the ULPS, whereas larger (in the range of 25cm) for the test positions that are far away from the ULPS' projection.

9. Discussion

When an ULPS is used in 2D, its final position can be fixed at the ceiling with all the beacons emitting downwards and the targets with receivers pointing upwards. In that case, the coverage area on the floor is a circle about 40m², assuming the ULPS is placed 3.5m high. For an extended 2D workspace, it is needed to add more ULPSs at the ceiling at different positions to cover a larger area on the floor. Anyway, these different ULPSs are always placed in the ceiling, they work independently and the obtained accuracy is similar for all the 2D floor positions considered.

Nevertheless, for a 3D positioning it is necessary to cover all the space and not only the floor surface. This is why a set of ULPSs must be placed at different positions and with different orientations (see Fig. 4) to cover all the space, thus having enough accuracy in all the volume of interest.

Two different configurations for 3D positioning in a cubical region of 8x8x8m^3 have been studied in this work. The first is the configuration A with three ULPSs pointing towards the centre of the cube and placed at the centres of the X-Y plane (z=0m), the Y-Z plane (x=0m), and the X-Z plane (y=8m). The second is the configuration B with four ULPSs pointing towards the centre of the cube and placed at corners (0m, 0m, 0m), (8m, 8m, 8m), (0m, 8m, 0m) and (8m, 0m, 8m).

Furthermore, two alternatives have been considered for each configuration:

- To have measurements from all the ULPSs simultaneously (with all the emitters synchronized). In this case, all the measured distances can be provided to the positioning algorithms.
- To have measurements independently for each ULPS (avoiding the need of synchronization among the different ULPSs). In this case, a position is estimated for every ULPS and afterwards all the results are merged.

The use of simultaneous measurements from different ULPSs is better in terms of the obtained accuracy, but it requires synchronization among ULPSs, as well as a wide receiver coverage that can be difficult in practical systems. On the other hand, with independent measurements for each ULPS, these issues can be avoided at the cost of reducing accuracy (that can still be enough for many applications).

10. Conclusions

The performance of a 3D ultrasonic positioning system (ULPS), firstly developed with a structure and characteristics for 2D positioning, has been studied and adapted to be deployed for 3D positioning. The study has covered the use of different ULPSs configurations, considering simultaneous or independent measurements. The comparison has been carried out and based on the estimation of the PDOP, by using a grid that covers all the space of interest (a cube of 8x8x8 m^3).

The achieved accuracy varies in the two cases analysed (case A and case B), depending on the number of ULPSs involved (1 to 4 ULPSs) and their positions/orientations (at the corners, in the ceiling, in walls....). Some preliminary experimental tests have also shown the performance for a single ULPS. The positioning error is less than 20cm for the 90% of considered measurements. This centimetre accuracy is in accordance with other positioning systems based on ultrasounds as can be observed in [5], which also compares this technology with others also applied to this type of systems.

Acknowledgment

This work has been possible thanks to the Spanish Ministerio de Economía, Industria y Competitividad (TARSIUS, ref. TIN2015-71564-c4-1-R, and SOC-PLC, ref. TEC2015-64835-C3-2-R) and the University of Alcala (LOCATE-US, ref. CCG2016/EXP-078). Authors thank Dr. David Gualda for his support with experimental measurements.

References

[1] Z. Li, L. Feng, A. Yang, "Fusion Based on Visible Light Positioning and Inertial Navigation Using Extended Kalman Filters", Sensors (Basel). 2017 May; 17(5): 1093. doi: 10.3390/s17051093

[2] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[3] K. Curran, E. Furey, T. Lunney, J. Santos, "An Evaluation of Indoor Location Determination Technologies", Journal of Location Based Services, vol. 5, no. 2, pp. 61-78, 2011.

[4] D. Gualda, J. Ureña, J. C. García and A. Lindo, "Locally-Referenced Ultrasonic-LPS for Localization and Navigation", Sensors, vol. 14(11), pp. 21750-21769, 2014.

[5] R. Mautz, Indoor Positioning Technologies, Habilitation Thesis, ETH Zurich, 2012.

[6] A. Alarifi, A Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, H. S. Al-Khalifa, "Ultra-Wideband Indoor Positioning Technologies: Analysis and Recent advances", Journal of Sensors 2016

[7] N. Parnian, *Integration of Local Positioning System & Strapdown Inertial Navigation System for Hand-Held Tool Tracking*, Habilitation Thesis, University of Waterloo, 2008.

[8] S. Sameshima, E. P. Katz, "Experiences with Cricket/Ultrasound Technology for 3-Dimensional Locationing within an Indoor Smart Environment Harry", Technical Report, Carnegie Mellon University, 2009.

[9] K. Pahlavan, X. Li, and J. P. Makela, "Indoor geolocation science and technology", IEEE Communications Magazine, vol. 40, no. 2, pp. 112-118, 2002.

[10] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, J. Sievanen, "A probabilistic ap-proach to WLAN user location estimation" International Journal of Wireless Information Networks, July 2002, Volume 9, Issue 3, pp 155–164

[11] N. Bodhi, M, Goraczko, *The Cricket Indoor Location System*, PhD Thesis, Massachusetts Institute of Technology, 2005.

[12] K. Mannay, J. Ureña, Á. Hernández, D. Gualda, J. M. Villadangos," Analysis of performance of Ultrasonic Local Positioning Systems for 3D Spaces", 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN). pp. 1-4. (ISBN 978-4-86049-074-4).

[13] Y. Zhou, C. L. Law, Y. L. Guan and F. Chin, "Indoor Elliptical Localization Based on Asynchronous UWB Range Measurement", IEEE Trans. on Instrumentation and Measurement, vol. 60, no. 1, pp. 248-257, 2011.

[14] E. García, P. Poudereux, A. Hernández, J. J. García and J. Ureña. "DS-UWB indoor positioning system implementation based on FPGAs", Sensors and Actuators A: Physical, vol. 201, pp. 172-181, 2013.

[15] J. Prieto, S. Mazuelas, A. Bahillo, P. Fernandez, R. M. Lorenzo and E. J. Abril, "Adaptive data fusion for wireless localization in harsh environments", IEEE Trans. on Signal Processing, vol. 60(4), pp. 1585-1596, 2012.

[16] M. Ocaña, M. A. Sotelo, L. M. Bergasa, R. Flores, "Low level navigation system for a POMDP based on WiFi and ultrasound observations", Proc. of the 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation, (CIRA'05), pp. 335–340, 2005.

[17] Y. Álvarez, M. E. De Cos Gómez, J. Lorenzo, F. Las-Heras, "Evaluation of an RSS-based indoor location system", Sensors and Actuators, A: Physical, vol. 167 (1), pp. 110–116, 2011.

[18] A. D. Koutsou, F. Seco, A. R. Jimenez, J. O. Roa, J. L. Ealo, C. Prieto, J. Guevara, "Preliminary localization results with an RFID based indoor guiding system", Proc. of the 2007 IEEE International Symposium on Intelligent Signal Processing (WISP'07), pp. 1–6, 2007.

[19] E. M. Gorostiza, J. L. Lázaro Galilea, F. J. Meca Meca, D. Salido Monzú, F. Espinosa Zapata and L. Pallarés Puerto, "Infrared sensor system for mobile-robot positioning in intelligent spaces", Sensors, vol. 11(5), pp. 5416-5438, 2011.

[20] D. Ruíz, J. Ureña, J. C. García, C. Pérez, J. M. Villadangos, E. García, "Efficient trilateration algorithm using time differences of arrival", Sensors and Actuators A: Physical, vol. 193, pp. 220-232, 2013.

[21] J. M. Villadangos, J. Ureña, J. J. García, M. Mazo, A. Hernández, A. Jiménez, D. Ruíz, C. De Marziani, "Measuring Time-of-Flight in an Ultrasonic LPS System Using General Generalized Cross-Correlation", Sensors 2011, vol. 11, pp. 10326-10342, 2011.

[22] D. Gualda, Mª C. Pérez, J. Ureña, J. C. García, D. Ruiz, E. García and A. Lindo, "Ultrasonic LPS Adaptation for Smartphones", International Conference on Indoor Positioning and Indoor Navigation (IPIN'13), pp. 1-6, 2013.

[23] A. Lindo, MC. Perez, J. Urena, D. Gualda, E. Garcıa and J. Manuel Villadangos, "Ultrasonic signal acquisition module for smartphone indoor positioning", Proc. of 2014 IEEE Conference on Emerging Technology and Factory Automation (ETFA), pp. 1-4, 2014.

[24] A. Hernández, E. García, D. Gualda, J. M. Villadangos, F. Nombela, J. Ureña, "FPGA-based Architecture for Managing Ultrasonic Beacons in a Local Positioning System", IEEE Trans. on Instrumentation and Measurement, vol. 66, no. 8, pp. 1954-1964, 2017.

[25] T. Kasami, "Weight distribution formula for some class of cyclic codes," Technical reportR-285, Coordinated Science Lab. University of Illinois, April 1968.

[26] Pro-Wave Electronics Corporation, *Air Ultrasonic Ceramic Transducers 328ST/R160*, Product Specification, 2014.

[27] D. Gualda, J. Ureña, J.C. García, J. Alcalá, A.N. Miyadaira. "Calibration of Beacons for Indoor Environments based on a Map-Matching Technique". *2016 International Conference on Indoor Positioning Systems.* pp. 1-7, Alcalá de Henares, Spain, October 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 119-127 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Automated Text Annotation for Social Media Data during Natural Disasters

Si Si Mar Win[*,1], Than Nwe Aung[2]

[1]*University of Computer Studies, Mandalay, Web Data Mining Lab, 05071, Myanmar*

[2]*University of Computer Studies, Mandalay, Faculty of Computer Science, 05071, Myanmar*

ARTICLE INFO

Article history:
Received: 21 November, 2017
Accepted: 27 February, 2018
Online: 12 March, 2018

Keywords:
LIBLINEAR
Twitter
Annotation
POS
Ngram

ABSTRACT

Nowadays, text annotation plays an important role within real-time social media mining. Social media analysis provides actionable information to its users in times of natural disasters. This paper presents an approach to a real-time two layer text annotation system for social media stream to the domain of natural disasters. The proposed system annotates raw tweets from Twitter into two types such as Informative or Not Informative as first layer. And then it annotates again five information types based on Informative tweets only as second layer. Based on the first and second layer annotation results, this system provides the tweets with user desired informative type in real time. In this system, annotation is done at tweet level by using word and phrase level features with LIBLINEAR classifier. All features are in the form of Ngram nature based on part of speech (POS) tag, Sentiment Lexicon and especially created Disaster Lexicon. The validation of this system is performed based on different disaster related datasets and new Myanmar_Earthquake_2016 dataset derived from Twitter. The annotated datasets generated from this work can also be used by interested research communities to study the social media natural disaster related research.

1. Introduction

Today, online social networking sites like Twitter, YouTube Facebook and Weibo play the important news sources during mass emergencies. Among them, Twitter, the most popular social networking site, provides a wealth of information during a natural disaster. It is often the first medium to break important disaster events such as earthquakes often in a matter of seconds after they occur and more importantly. Recent observation proofs that some events and news emerge and spread first using this media channel rather than other the traditional media like online news sites, blogs or even television and radio breaking news.

People also used social media to share advice, opinions, news, moods, concerns, facts, rumors, and everything else imaginable. Corporations use social media to make announcements of products, services, events, and news media companies use social media to publish near real-time information about breaking news. However, due to questionable source, uncontrollable broadcasting, and small amount of informational tweets among large number of non-informational tweets, Twitter is hardly an actionable source of breaking news.

Tweets from Twitter are highly vary in terms of subject and content and the influx of tweets particularly in the event of a disaster may be overwhelming. It is impractical to generation of efficient features vector based on uniform vocabulary. Therefore, effective feature extraction is first challenge. It is infeasible to automatically classify these varied tweets by using particular annotated corpora for specific messages of every disaster events. Cross event classification is a major challenge.

Another challenge is occurred by the development of supervised learning based systems trained on a single corpus and able to achieve a good performance over a broad range of different events. The annotation of corpus of messages for every disaster by human annotators is obviously time-expensive and practically infeasible on real time manner. Therefore annotation of tweets corpora by human is additional challenge.

In summary, the proposed system is aimed to address these issues by using three main functions: 1. Create annotated disasters corpus of tweets with five labels for Informative tweets on real

[*] S. S. M. Win, U.C.S.M, Email: sisimarwin@gmai.com

https://dx.doi.org/10.25046/aj030214

time manner. 2. Competitive, easily implementable feature extraction method that act as a benchmark for automated accurate classification approaches for natural disaster related datasets by using natural disaster lexicon. 3. Creation of extended natural disasters lexicon based on publicly available annotated datasets and newly annotated corpus. This paper is an extension of the work originally presented in IEEE/ACIS 16th International Conference on In Computer and Information Science (ICIS) [1]. In the previous work, we identified the tweets into only three labels such as Informative, Not Informative and Other Information as single layer annotation. Therefore, we continue to identify the Informative tweets into more specific information types. Our annotation model in this paper is based on a more relevant and small set of features than our previous work.

The rest of the paper is organized as follow: Section 2 presents the overview of closely related work to this paper. Section 3 explains the methodology that we used in collecting, preprocessing, feature extraction, disaster lexicon creation, and classification scheme used for annotating the tweets. Section 4 describes the architecture of the proposed system. Section 5 expresses the datasets details, experiments and analysis performed. Section 6 summarizes the results from our analysis and highlights the implications of our results. In this section, we also describe the future work of the proposed system.

2. Related Work

This section presents the current state-of-the-art systems, algorithms and methodologies to access the social media data analysis. Social media allows users to exchange small digital content such as short texts, links, images, or videos. Although it is a relatively new communication medium compared with traditional media, microblogging has gained increased attention among users, organizations, and research scholars in different disciplines. There are several researches on social media mining for text based data for classification and prediction of informational posts in different domains.

Among them, the authors in [2] proposed Artificial Intelligence for Disaster Response (AIDR) system to annotate the posts from Twitter into a set of user defined categories such as damage, needs etc. by using hybrid unigram and bigram features. In AIDR system, the tweets were identified into Informative and Non-Informative types during Pakistan earthquake in 2013.

The authors in follow up study automatically and provided human annotated Twitter corpora for 19 different events that occurred between 2013 and 2015. They also experimented their corpora by using the similar features set [3].

The other authors also presented the Tweedr, twitter-mining tool, to retrieve actionable information from Twitter. They applied several different types of features for their CRF clustering. For each token in a tweet, they extracted capitalization, pluralization, whether it is numeric or includes a number, WordNet hypernyms, Ngrams, and part of speech tags to provide specific information about different classes of infrastructure damage, damage types, and casualties [4].

The authors in [5] used six types of features such as Tweet Meta-data Features, Tweet Content Features, User based Features, Network Features, Linguistic Features and External Resource Features for credibility analysis.

They also developed a real time web application, TweetCred, to provide the one of the seven credibility scores of user generated content on Twitter by using 45 features. They tested their application within three weeks period. Their result showed that high credibility tweets were 8% [6].

Moreover, hashtags have been effectively utilized as critical features for various tasks of text or social media analysis, including tweet classification system [7].

In [8], the authors studied the linguistic method to analyze the importance of linguistic and behavioral annotations. They applied the datasets of four crisis events such as Hurricane Gustav in 2008, the 2009 Oklahoma Fires, the 2009 and 2010 Red River Floods, and the 2010 Haiti Earthquake. They observed that the usage of specific vocabulary to convey tactical information on Twitter can achieve higher accuracy than the usage of bag of words (BOW) model for classification of context-aware tweets.

The classification of tweets into Credible or Not Credible was presented in [9]. However, most of the recent research focuses on the information extraction and detection of situational awareness during natural disasters, it is still needed to provide a cohesive pipeline that takes into consideration all of the facets of data extraction.

This system focuses on the content based features set such as Linguistic features, Disaster Lexicon based features, twitter specific features (hashtags and URLs), unigram POS tag features and other salient features from tweet content.

3. Methodology

At the core function of this system is the capability of annotating tweets into predefined information types in real time. We propose, implement and evaluate the approach for determining and assigning a label for each tweet, taking into account terms from the tweet itself and from disaster lexicon. For this study, we first collect the tweets from Twitter. And then we extracted content based features from the collected tweets.

3.1. Data Collection

This function works for tweets collection. It collects messages from Twitter for training and testing using the Twitter streaming API. At first, it collected different annotated datasets published by using annotated tweet_id from Imran et al. [3].

The new data collection process focuses on the exact matching of keywords to acquire tweets and build the query using user defined keywords or hashtags. Using the relevant keywords or hashtags for queries are the best way to extract the most relevant tweets during crisis or disasters. For example, #MyanmarEarthquake hashtag is applied to acquire the news of earthquake that struck in Myanmar.

3.2. Preprocessing

Firstly, this task removes the tweets which already contains the same text in the previous preprocessed tweets to reduce the redundancy and noise by using the cosine similarity.

Secondly, stop-words from tweets are removed to reduce dimensionality of the dataset and thus terms left in the tweets can be identified more easily by the feature extraction process. Stop-words are common and high frequency words such as "a", "the", "of", "and", "an" "in" etc. [10]. User mention and URLs are also eliminated.

Finally, we used lemmatization instead of stemming to convert all the inflected words present in the text into a base form called a lemma. For the purpose of lemmatization, the proposed system uses Stanford Core NLP.

3.3. Feature Extraction

The most important step in text analysis using supervised learning techniques is generating feature vectors from the text data or documents. This work is intended to build a real time system based on tweets from Twitter, feature extraction is therefore concerned with altering tweet contents into a simple numeric vector representation.

In our previous work, we used hybrid unigram and bigram, unigram Brown cluster, unigram part of speech (POS) tags, number of hashtags, number of URLs and two lexicon based features such as NRC hashtags lexicon and our disaster lexicon as our features set. We found these features outperform the neural word embeddings and only hybrid unigram and bigram features.

According to the constant vocabulary, hybrid unigram and bigram features outperforms the classification of same events (i.e. the training and test datasets are equal). However we need to annotate the unknown disaster events and the contents described for different events may have different vocabulary. Even the same type disaster events may contains the different language style.

The analysis of social media data is heavily rely on the ability to analyze text data. However, there are some unique considerations in the analysis of social media data that make it different than a normal text mining analysis. To overcome the informal social media data to be formal consideration, text in tweets are tokenized using ARK Tweet NLP [11]. This process receives the tweets from preprocessing step, it extracts the features by using ARK POS tagger and different lexicons. The features used in this work are only extracted from tweet contents.

To derive the most relevant feature, this work investigated the three types of feature extraction methods. The first one is BOW model with unigram and bigram based features used in AIDR. The second is the neural word embeddings (WE) model and the last one is the proposed content based features model.

This system proposes the features set based on the following observations:

1. Messages in tweets written by users for same disaster type may have composed of same terms. It is usually the case that the same disasters have the same terms such as shake, strike, magnitude for disaster earthquake.
2. Different natural disaster related tweets may have composed of same terms such as need, pray, pray for, damage, death, destroy, survivor, etc. and may have same syntactical style such as POS tag.
3. Similar words have similar distributions of words to their immediate left and right [11].
4. If a tweet contain more than two hashtags in its content, it may not be information tweet.
5. In crisis related tweets, hashtag may be assumed as topic word or keyword of these tweets.
6. Informative tweets may contain numerical word and URL.

Based on these observations 1 and 2, we decided to create and use Disaster Lexicon and word Ngrams. According to observation 3, we use Brown Word cluster. We also use number of hashtags and hashtag term, URL and numeral features due to the observation 4, 5 and 6. Ngrams POS features are used according to the observation 2. The proposed features used in this system are shown in Table 1.

Table 1. Features used in the proposed system

Feature	Explanation
Brown Cluster Ngrams	Unigram and Bigram of 1000 Brown clusters in Twitter Word Clusters made available by CMU ARK group
Count of disaster related terms	Number of disaster related terms as informative words in a created lexicon for disaster tweets.
Total PMI Score of disaster related terms	Total PMI scores of unigrams and bigrams words that occurred in the tweet and listed as strongly correlated with natural disaster in a disaster lexicon for tweets
Count of non-informational terms	Lexicon creation function of this system also identifies a set of terms which appear only in Not-Informative tweets across all natural disasters datasets.
Total PMI Score of non-informational terms	Total PMI Score for each set of unigrams, bigrams that mostly occur in the Not informative tweet.
Count of numerals	Expected to be higher in situational tweets which contain information such as the number of casualties, emergency contact numbers.
POS tag	Unigram part of speech tags that occur in the Tweet generated by CMU ARK POS-Tagger
Word Ngrams	Unigram and bigram of terms from Disaster Lexicon

To extract neural word embeddings (WE) features for baseline model, this system used Word2vec model in Deep Learning4J [12]. Word2Vec is the representations of words with the help of vectors in such manner that semantic relationship between words preserved as basic linear algebra operations. The following parameters were used while training for Word2Vec: 100 dimensional space, 10 minimum words and 10 words in context. After transforming 100 dimension feature vector of each word in the corpus, this system used t-Distributed Stochastic Neighbor embedding (t-SNE) technique to reduce from 100 dimensions of each feature vector to 10 dimensions feature vector.

3.4. Extended Disaster Lexicon Creation

This system creates the disaster lexicon which contains specific natural disaster related terms with a point wise mutual information (PMI) based score and frequency distribution of these terms based on the set of annotated disaster datasets. This lexicon creation process follows the method of Olteanu et al. [13]. In this process, we exploit their natural disaster related datasets, the other available natural disaster related datasets [2] and newly annotated dataset such as Myanmar_Earthquake_2016 dataset which are collected by proposed system for lexicon expansion or keywords (disaster related terms) adaptation.

The disaster creation process consists of two main parts. At first, to obtain the most relevant disaster terms, we create various disaster lexicons based on different datasets of same disaster type. For example, this work uses all available earthquake datasets such as 2015_Neapl, 2014_Chile, 2014_Calfornia and 2016_Myanmar earthquakes for creation of Earthquake Lexicon. In this phase, we used equal number of Informative and Not informative tweets for each disaster datasets based on same disaster types.

At second, we combined the different disaster lexicons into one master disaster related lexicon with unique unigram and bigram terms. In this step, we calculate the mean PMI score for the terms which are contained in the two or more lexicons.

The score of a term could be calculated from the PMI value of a term t in an informative context PMI (t, informative) and the same term in a non-informative context PMI (t, non-informative) using the equation:

$$InfoScore = PMI\ (t, informative) - PMI\ (t, non\text{-}informative) \qquad (1)$$

Here PMI (t, informative) and PMI (t, non-informative) are calculated using:

$$PMI\ (t, orientation\ ) = log_2 \frac{freq(t, orientation).N}{freq(t).\ freq(orientation)} \qquad (2)$$

Where, freq (t) is the number of times term t appears in a tweet, while *N* is total number of terms in the tweet.

This automatically created lexicon is used in feature extraction process of the proposed system.

3.5. Feature Selection

Feature selection is an important problem for text classification. In feature selection, this work attempts to determine the features which are most relevant to the classification process. This is because some of the words are much more likely to be correlated to the class distribution than others. This system applied the information gain based feature selection method which is widely used for text classification.

Information gain (IG) measures the amount of information in bits about the class prediction, if the only information available is the presence of a feature and the corresponding class distribution. n this method, let P_i be the global probability of class *i*, and $P_i\ (w)$ be the probability of class *i*, given that the document contains the word w. Let *F(w)* be the fraction of the documents containing the word w. The information gain measure *I(w)* for a given word *w* is defined as follows:

$$I(w) = -\sum_{i=1}^{k} P_i\ .\log\ (P_i) +\ F(w).\sum_{i=1}^{k} P_i(\mathrm{w})\ .\log\ (P_i(w)) +$$

$$(1 - F(w)).\textstyle\sum_{i=1}^{k}(1 - P_i(\mathrm{w}))\ .\log\ (1 - (P_i(w)) \qquad (3)$$

The greater the value of the information gain *I(w)*, the greater the discriminatory power of the word *w*.

3.6. Annotation of Social Media Text

This system assess annotation of tweets by using supervised machine learning technique. This technique automatically classifies the information contained in tweets. To perform the annotation task, the proposed system trained a LIBLINEAR classifier operating on extracted features set. LIBLINEAR solves large-scale classification problems in many applications such as text classification. It is very efficient for training large scale. It takes only several seconds to train more than 600,000 examples while a Library for Support Vector Machines (LibSVM) takes several hours for same task [14].

Given a set of features and a learning corpus (i.e. the annotated dataset), the classifier trains a statistical model using the feature statistics extracted from the corpus and then annotates the tweets into Informative or Not Informative. This trained model is then employed in the classification of unknown tweets and, for each tweet, it assigns the probability of belonging to a class: Related and Informative as Informative, Not Related or Not applicable as Not Informative in first layer annotation. And then based on the Informative tweets, this system annotates again these informative into one of five types such as infrastructure damage, caution and advice, dead or injured people, needs and offer and Donations and volunteering as second layer annotation. The annotated datasets required by the system can be obtained from three sources such as AIDR, CrisisNLP which is the collection of tweets from 19 natural and man-made disasters and CrisisLexT26 which is the collection of tweets from 26 Crises [2, 3, 13]. This system uses datasets in English language only.

4. Architecture of the proposed System

The holy grail of text annotation is an automated system that accurately and reliably annotates very large numbers of cases using relatively small amounts of manually annotated training data. This work is intended to develop a two layer annotation system that automatically creates the different disaster datasets with annotated tweets. In this system, annotation is restricted to tweets in English language. Non-English tweets are not considered. Non-English tweets are not considered.

The system, illustrated in Figure, first collects the tweets from Twitter by using user desired query terms or target disaster related terms. After collecting the tweets, it removes the redundant tweets by using tweet_id and then it also eliminates the stop-words. In feature extraction, this system applies Linguistic features such as Brown cluster, Syntactic feature such as POS features, Lexical features using disaster lexicon and the other Twitter centric features such as Hashtags and URLs.

This system also analyzed which features are important in the data to annotation. It applied the annotated corpus to train a classifier that automatically annotates the tweets.

To improve model performance, the best set of 300 features were chosen by using Information gain theory based feature selection method.

In the ground truth annotation process, LIBLINEAR classifier uses these selected features subset for tweets categorization to create annotated corpus with Informative or Non-informative tweets and to provide informative tweets to the users.

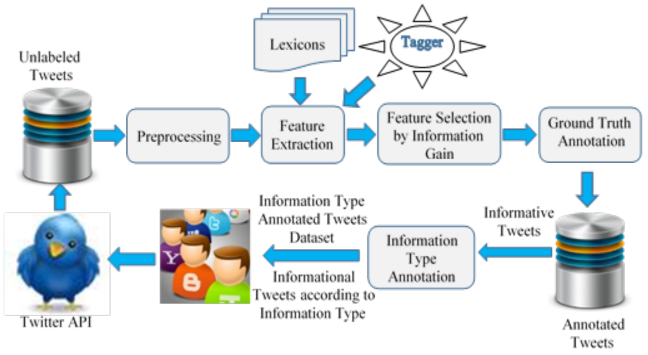

Unlabeled Tweets
Lexicons
Tagger
Preprocessing
Feature Extraction
Feature Selection by Information Gain
Ground Truth Annotation
Information Type Annotated Tweets Dataset
Informative Tweets
Information Type Annotation
Informational Tweets according to Information Type
Twitter API
Annotated Tweets

Figure 1: Architecture of the Proposed System

After annotating the collected tweets into one of the five information types, his system provides the informational tweets to users based on their desired type of information.

5. Experiments

This system performs a set of preliminary experiments to evaluate the effectiveness of feature extraction, feature selection model and classifier model on the performance of the proposed approach. For feature extraction, the proposed system applied three models such as neural word embedding, BOW with Unigram and Bigram model and the proposed model.

The final experiment is done under the best development settings in order to evaluate the classifier model with the best feature set. This section presents experiments and results for classification of four annotated datasets. The results along with the experimentation of different datasets are described based on accuracy, precision, recall, and F1 score of classifier model for feature extraction performance.

5.1. Datasets and Setting

In order to evaluate the effectiveness of the proposed social media text annotation strategies for identifying informative tweets during natural disasters events, the experiments of this system used people freely available 10 annotated natural disaster datasets . These datasets are already annotated with different information types.

To reduce the noise in training data, this system discarded all the following tweets.

1. The tweet where an information type clash is observed. An information type clash is a tweet that may happen two or more different type and may ambiguous in the dataset.

2. “Not labeled” tweets.

3. “Animal Management” are also eliminated.

The tweets with similar information types such as “Infrastructure damage” and “Infrastructure and utilities” are combined as “Infrastructure and utilities”. “Injured or dead people”, “missing or found people”, “displaced people and evacuation” and “personal updates” tweets are combined as “Affected individuals” and “donation needs or offer volunteering services” and “Money” are also combined as “Donations and volunteering”.

Before training the corpus for second layer annotation, the informative tweets with non-specific information type such as “Other Useful Information” are also discarded.

Detailed information of datasets is described in Table 2 and Table 3. In this table Type 1 refers to the information type “Affected individuals”, Type 2 refers to “Infrastructure and utilities”, Type 3 means “Donations and volunteering”, Type 4 is “Caution and advice”, and Type 5 refers to “Sympathy and emotional support”.

Table 2. Natural disaster datasets details including disaster type, name, number of informative tweets, number of Not Informative tweets and total tweets.

Type	Disaster Name	Info	Not-Info	Total
Floods	2013_Queensland_floods (QF)	728	281	1009
Bushfire	2013_Australia_bushfire (AB)	691	261	952
Typhoon	2013_Typhoon_Yolanda (TY)	765	175	940
Wildfire	2012_Colorado_wildfires (CW)	685	247	932
Earthquake	2014_Chile_earthquake (ChiE)	1834	179	2013
Floods	2013_Colorado_floods (CF)	589	190	779
Earthquake	2014_Costa_Rica_earthquake (CE)	842	170	912
Floods	2014_Manila_floods (MF)	628	293	921
Floods	2012_Phillipines_Floods (PF)	761	145	906
Floods	2013_Alberta Floods (AF)	684	297	981
Floods	2014_India_floods (IF)	940	396	1336

Table 3. Natural disaster datasets statistics for five information types

Dataset	Type 1	Type 2	Type 3	Type 4	Type 5
QF	207	113	55	114	17
AB	199	65	35	70	33
TY	77	106	383	20	63
CW	44	128	62	69	25
NE	6	165	239	1215	458
IF	30	792	42	51	25
ChiE	55	3	70	14	58

The proposed system performed 10 fold cross validation to test the efficiency of the feature extraction and the model. In the experiments of classification, the proposed system used the set of tweets from five natural disasters such as 2013_Queensland_floods denoted by QF, 2013_Australia Bushfire as AB, the set of tweets for 2013_Typoon_Yolanda as

TY, 2012_Colorado_wildfires as denoted by CW, and the set of tweets from 2012_Costa_Rica_Earthquake as CE respectively.

5.2. Effectiveness of Feature Extraction

To choose the best classification model, we tested the extracted feature set on four different classifiers such as Random Forest, Sequential Minimal Optimization (SMO) which is the fast training algorithm for Support Vector Machine (SVM), Naïve Bayes and our LIBLINEAR classifier that are well known in text classification process. Due to the experiments on the previous work, the performance of Random Forest, Naïve Bayes and SMO was sensitive to the large number of features. Therefore, this system used LIBLINEAR classifier with Information Gain based feature selection method to get better performance and to reduce inconsistent features. This wok uses a well-known WEKA machine learning tools for implementation of Random Forests, Naïve Bayes, SMO, LIBLINEAR and Information Gain based feature selection methods [15]. The results of our previous work are described in [10]. Due to these results, we selected the LIBLINEAR classifier as our classification model.

In this work, the proposed feature extraction method and two baseline methods are evaluated by experiments on ten datasets. To compare the performance of the different feature-models (using LIBLINEAR classifier) under three scenarios such as (i) in-domain classification, (ii) cross event classification and (iii) cross-domain classification, where the classifier is trained with tweets of one event, and tested on another event are considered in this system.

5.2.1. In Domain Classification

In this type of classification, the classifier is trained and tested with the tweets of the same event. To evaluate the in-domain performance of each model, the proposed system followed a 10-fold cross validation process: each dataset was randomly split in 10 different non overlapping training and test sets. The Accuracy, Precision, Recall and F-Measure were calculated as the weighted average of these values over all the 10 test sets.

Table 4. Classification results in terms of Precision, Recall, F-Measure and Accuracy across Informative and Not Informative classes.

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
QF	0.791	**0.813**	0.612	0.812	**0.822**	0.782	0.785	**0.816**	0.687
AB	**0.772**	0.754	0.567	**0.79**	0.765	0.753	0.768	0.758	0.647
TY	0.797	**0.802**	0.685	**0.825**	0.816	0.828	0.788	**0.807**	0.750
CW	0.813	**0.815**	0.676	0.818	**0.820**	0.714	0.711	**0.817**	0.633

Table 4 represents a summary of evaluation for in domain classification by weighted average Precision (P), Recall (R) and F-Measure (F-M) of the classification results on four datasets.

According to the results, BOW with hybrid unigram and bigram model would perform relatively well in in-domain classification, since the training event and test event share a common vocabulary. However, the performances of the proposed features model is as good as BOW method.

5.2.2. Cross Event Classification

In this type of classification, where the classifier is trained with tweets of one event, and tested on another event. The result is significant since it shows that good classification can be achieved even without considering the type of disasters.

Table 5, Table 6, Table 7 and Table 8 also show the cross event classification performance on AB, TY, CW and QF dataset as training and the remaining datasets as testing data using the features sets extracted by the baseline method, proposed method and neural word embeddings method. The results in these tables indicated that the proposed method yields a high accuracy by using the LIBLINEAR algorithm in predicting certain classes in cross event classification.

Table 5. Classification results in terms of Precision, Recall, F-Measure using (i) (BOW), (ii) Proposed (P) (iii) Word Embeddings (WE) for 2013_Australia_Bushfire as training set.

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
QF	0.77	**0.82**	0.76	0.73	**0.83**	0.50	0.74	**0.82**	0.52
TY	0.78	**0.81**	0.77	0.74	**0.83**	0.52	0.76	**0.81**	0.57
CW	0.79	**0.83**	0.75	0.80	**0.83**	0.66	0.80	**0.84**	0.66

Table 6. Classification results in terms of Precision, Recall, F-Measure across two classes using (i) (BOW), (ii) Proposed (P) (iii) Word Embeddings (WE) for 2013_Typhoon_Yolanda as training set.

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
QF	**0.77**	**0.77**	**0.77**	0.64	**0.793**	0.79	0.67	0.755	**0.77**
AB	0.72	**0.77**	0.76	0.554	**0.785**	0.78	0.579	0.748	**0.75**
CW	0.74	**0.796**	0.76	0.589	**0.804**	0.78	0.605	**0.788**	0.76

Table 7. Classification results in terms of Precision, Recall, F-Measure across two classes using (i) (BOW), (ii) Proposed (P) (iii) Word Embeddings (WE) for 2012_Colorado_Wildfire as training set.

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
QF	0.759	**0.795**	0.74	0.73	**0.808**	0.37	0.745	**0.80**	0.36
AB	0.769	**0.797**	0.72	0.756	**0.806**	0.38	0.762	**0.799**	0.35
TY	0.762	**0.810**	0.76	0.673	**0.809**	0.36	0.703	**0.809**	0.37

Table 8. Classification results across two classes using (i) (BOW), (ii) Proposed (P) (iii) Word Embeddings (WE) for 2013_Queensland_floods as training set.

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
AB	0.772	**0.817**	0.76	0.727	**0.831**	0.60	0.743	**0.818**	0.62
TY	0.771	**0.76**	0.76	0.641	**0.793**	0.65	0.672	**0.755**	0.68
CW	0.759	**0.795**	0.75	0.73	**0.808**	0.62	0.741	**0.799**	0.64

According to the experimental results, the performance of the BOW (hybrid unigram and bigram) and WE models is significantly inferior to the proposed model for cross-event classification. This is because the training and testing datasets (related to two different disaster events) have very different vocabularies. On the other hand, the classifier based on the proposed features significantly out-perform these two models in all cases. This implies that the selected features can separate between Informative and not-Informative tweets irrespective of the vocabulary and linguistic style related to specific events. Thus, classifiers can be trained over these features extracted from past disasters, and then deployed to classify tweets posted during future events.

5.2.3. *Cross Domain Classification*

In cross domain classification, to assign one of the two classes for first layer annotation and one of the five predefined categories (e.g. Affected individuals, Infrastructure and utilities, Donations and volunteering, Caution and advice, Sympathy and emotional support etc.) for second layer annotation to the tweet, the classifier requires sufficient training examples to learn about each pre-defined category. The proposed system used multiple past disasters of various types to train the classifier to robustly identify the different types of tweets for future natural disasters. The experiment for two classes classification (i.e. Informative and Not Informative), the proposed system used the set of tweets from Philippines (PF), Colorado (CF), and Queensland floods (QF) as the training set, denoted by PCQ, the set of tweets for Manila floods as the development set, denoted by MF, and the set of tweets from Alberta and Sardinia floods as two independent test sets, denoted by AF and SF, respectively. The results of this experiment are shown in Table 9.

Table 9. Experiments performed using the combined 2012_Philipinnes_flood, 2013_Colorado_floods and 2013_Queensland_floods as training set

Train /Test	Accuracy	Precision	Recall	F-Measure
PCQ/MF	80.34%	0.8609	0.803	0.808
PCQ/AF	76.47%	0.754	0.76476	0.7414
PCQ/SF	66.58%	0.6463	0.66579	0.6121

In the other experiments over five classes or multi-classes classification, this system combined the three or four datasets of different disaster types as training and the other one for testing data. For example, taking Colorado floods, Costa-Rica-earthquake, Philippine floods, Pablo-typhoon and Australia-Bushfire (CCPPA) as training dataset and the other datasets as individual test data. The classification results are shown in Table 10. Table 11 shows the classification results over the five classes of LIBLINEAR with 10-fold cross validation for six datasets by using three feature models.

Table 10. Classification results in terms of Precision, Recall, F-Measure across all five classes.

Train /Test	Accuracy	Precision	Recall	F-Measure
CCPPA/MF	72.3%	0.7308	0.723	0.719
CCPPA/TY	64.9%	0.72	0.649	0.663
CCPPA/CW	86.4%	86.8%	0.864	0.862
CCPPA/ChiE	79.5%	0.8214	0.795	0.7908

Table 11. Classification results across five classes using (i) (BOW), (ii) Proposed (P) (iii) Word Embeddings (WE)

Test Data	Feature Extraction Model								
	Weighted Avg. P			Weighted Avg. R			Weighted Avg. F-M		
	BOW	P	WE	BOW	P	WE	BOW	P	WE
QF	0.56	**0.581**	0.25	0.56	**0.586**	0.415	0.549	**0.576**	0.287
AB	0.537	**0.583**	0.22	0.569	**0.607**	0.465	0.532	**0.582**	0.295
TY	0.694	**0.701**	0.36	**0.724**	0.713	0.596	0.695	**0.705**	0.445
CW	0.653	**0.67**	0.18	0.64	**0.681**	0.427	0.631	**0.666**	0.255
NE	**0.749**	0.712	0.67	**0.761**	0.726	0.681	**0.738**	0.713	0.666
CE	0.757	0.78	0.58	0.783	**0.79**	0.76	0.754	**0.78**	0.658

According to the results in Table 10 and 11, the performance of LIBLINEAR classifier with proposed feature model outperforms the BOW and WE models in most cases and it can identify informational tweets at 67% accuracy on average. The performance of BOW and WE models sometime close to less than 20 %. This indicates that lexical features are critical to solve the ambiguous of information types.

5.3. *Effectiveness of Annotation*

In this section, the validation of this system on a real disaster study by classifying the data of Myanmar earthquake collected by Twitter API. The 6.8 magnitude earthquake that struck Myanmar on August 24th , 2016 is among the strongest in recent Myanmar history. The shaking was clearly perceived in all Central and Northern Myanmar and caused 4 deaths and several damage to the Pagodas of the area of Bagan. This dataset is crawled for a three days period from August 24th to 26th , 2016 by using the hashtags (#Myanmar, #Bagan, #earthquake, #Myanmarearthquake). And then it was randomly selected 1,800 tweets and was manually annotated based on the available news media in Myanmar such as Myanmar Times, The Global New Light of Myanmar and The Mirror.

Table 12. First Layer Annotation Results of Proposed features by LIBLINEAR

Dataset	Precision	Recall	F1	Accuracy
AB	0.897	0.895	0.895	89.45%
TY	0,912	0,92	0,913	92,02%
IF	0.908	0.908	0.908	90.82%
NE	0.748	0.751	0.749	75.05%

Table 13. Second Layer Annotation Results of Proposed features by LIBLINEAR

Dataset	Precision	Recall	F1	Accuracy
AB	0.682	0.693	0.685	69.31%
TY	0.768	0.786	0.774	78.60%
NE	0.815	0.843	0.828	84.26 %
IF	0.731	0.782	0.755	78.18 %

Base on cross domain classification over all five classes where we train the classifier on one dataset and test on another dataset, the experimental results using 2012-Costa-Rika- and 2014_Chile Earthquake as training data and Myanmar Earthquake as test data confirmed the expected classification of this work.

Myanmar_Earthquake_2016 was successfully annotated with predefined two labels at 75% accuracy on average and five labels also 74 % which is pretty high. The results of each annotation layer for four datasets are shown in Table 12 and Table 13.

5.4. Real Time Annotation

We showed a proof of real time model which takes a direct stream of new tweets as input test set and takes manually annotated tweets from previous disaster as training set, and then uses automated techniques to annotate the new tweets.

In the first layer, this system annotated the tweets into Informative if the tweets contain the information about the target disaster and Not informative for the remaining tweets. In the second layer, each informative tweet from previous layer is annotated into one of the five information types with respect to the information that contained in it.

To do this, we developed and deployed a real time system in the form of a Web application using Java 2 Platform Enterprise Edition (J2EE) and Twitter API binding library (Twitter4j). We analyzed the deployment and usage activity of our application from 24th August, 2016 to 28th August, 2016 which was the day of earthquake and days after an earthquake in Myanmar. For analysis and statistics, we collected the annotated datasets of our system for only three days period since August 24th to August 26th. In real time annotation process, we used the combined 2014_Chile_Earthquake and 2015_Nepal_Earthquake as training set. The number of collection and annotation times per day is 6. Each collection time is 4 minutes for 2000 tweets and annotation time is only 1 minute for each layer. Among the collected tweets, 62% of tweets are retweets and half of them are redundant tweets. Most of them are similar in text. Although the number of tweets collected in each time was 2000, the number of tweets annotated by our system is at most 750 because of the tweet cleaning step in preprocessing. The tweets in annotated datasets from each day are also redundant. The total unique automatically annotated tweets over three days, are nearly 2000.

After manually annotating the tweets, we compared the automatically annotated tweets from our system with manually annotated tweets to obtain the performance of our real time system. The manual annotation process is described in the previous section. According to the analysis results, the annotation accuracy of new tweets by our system is 80% in first layer and 74 % on second layer annotation.

5.5. Findings and Discussion

As mentioned above this system used 10 datasets for training and testing for evaluating classifier models and feature extraction models. Another new dataset for testing again for overall performance of the proposed system. According to the initial experimental results of BOW and word embeddings were very sensitive and depend on the vocabulary. The results of three feature extraction methods, the proposed method always outperforms the other two methods. Therefore, the proposed feature extraction model with LIBLINEAR classifier was chosen for second layer annotation process for categorizing the tweets into five specific frequently found information type.

6. Conclusion

Social media mining for disaster response and coordination has been receiving an increasing level of attention from the research community. It is still necessary to develop automated mechanisms to find critical and actionable information on Social Media in real-time. The proposed system combines effective feature extraction using NLP and machine learning approach to obtain the annotated datasets to improve disaster response efforts. Expanded disaster lexicon is also used to extract the relevant disaster related lexical features for annotation.

The proposed feature extraction method significantly outperforms the standard bag of words model and neural word embeddings model. By using LIBLINEAR classifier based on the proposed method, this system successfully annotated five information types to the Myanmar Earthquake data at 74% accuracy on average. In future, we will investigate the specific variation of terms over different disasters to perform annotation on all disaster types. We hope to formalize disaster lexicon in more detail to improve cross domain classification accuracy.

References

[1] S. S. M. Win, T.N. Aung, “Target Oriented Tweets Monitoring System during Natural Disasters”, In Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference, IEEE, 2017.

[2] M. Imran, C. Castillo, J. Lucas, P. Meier and S. Vieweg, “AIDR: Artificial intelligence for disaster response”, In Proc. of WWW (companion). IW3C2, 2014, pp. 159–162.

[3] M. Imran, P. Mitra, C. Castillo: "Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages." In Proceedings of the 10th Language Resources and Evaluation Conference (LREC), pp. 1638-1643. May 2016, Portorož, Slovenia.

[4] Z. Ashktorab, C. Brown, M. Nandi, A. Culotta, “Mining Twitter to Inform Disaster Response”, In Proceedings of the 11th International ISCRAM Conference – University Park, Pennsylvania, USA, May 2014.

[5] A. Gupta, P. Kumaraguru, “Credibility Ranking of Tweets during High Impact Events”, PSOSM’12, 2012.

[6] A. Gupta, P. Kumaraguru, C. Castillo, P. Meier, “TweetCred: Real-time credibility assessment of content on Twitter”, In Proc. Of SocInfo. Springer, 228–243, 2014.

[7] A. Stavrianou, C. Brun, T. Silander, C. Roux, “NLP-based Feature Extraction for Automated Tweet Classification”, Interactions between Data Mining and Natural Language Processing: 145.

[8] W. J. Corvey, S. Verma, S. Vieweg, M. Palmer and J. H. Martin, “Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events”, In Proc. International Conference on Language Resources and Evaluation (LREC'12), May 2012

[9] C. Castillo, M. Mendoza and B. Poblete, “Information Credibility on Twitter”, International World Wide Web Conference Committee (IW3C2), Hyderabad, India, 2011.

[10] C. C. Aggarwal and C.-X. Zhai, “Mining Text Data, Springer”, 2012.

[11] K. Gimpel et al., “Part-of-speech tagging for Twitter: Annotation, features, and experiments”, In Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume, Portland, June 2011.

[12] Deeplearning4j Development Team. Deeplearning4j: Open-source distributed deep learning for the JVM, Apache Software Foundation License 2.0. http://deeplearning4j.org.

[13] A. Olteanu, C. Castillo, F. Diaz and S. Vieweg, “CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises”, Association for the Advancement of Artificial Intelligence (www.aaai.org), 2014.

[14] R.E Fan, K.W Chang, C.J Hsieh, X.R Wang and C.J Lin, “LIBLINEAR: A Library for Large Linear Classification”, Journal of Machine Learning Research 9, 2008, pp. 1871-1874.

[15] E. Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data mining: Practical machine learning tools and techniques", Morgan Kaufmann, Fourth Edition, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 128-137 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Mission-Critical Systems Design Framework

Kyriakos Houliotis[*,1], Panagiotis Oikonomidis[1], Periklis Charchalakis[2], Elias Stipidis[3]

[1]*Research Fellow, Vetronics Research Centre, University of Brighton, BN2 4GJ, United Kingdom*

[2]*Principle Research Fellow, Vetronics Research Centre, University of Brighton, BN2 4GJ, United Kingdom*

[3]*Professor, Vetronics Research Centre, University of Brighton, BN2 4GJ, United Kingdom*

ARTICLE INFO

Article history:
Received: 16 December, 2017
Accepted: 01 February, 2018
Online: 12 March, 2018

Keywords:
Vetronics
Mission-Critical systems
Model driven architecture
NGVA
Threat estimation

ABSTRACT

Safety-critical systems are well documented and standardized (e.g. IEC 61508, RTCA DO-178B) within system design cycles. However in Defence and Security, systems that are critical to the success of a Mission are not defined within the literature nor are there any guidelines in defining criticality in their design or operational capabilities. When it comes to Vetronics (Vehicle Electronics), a mission-critical system, is a system with much complexity and mixed criticality levels that is a part of the overall platform (military vehicle) offering integrated system capabilities. In this paper, a framework is presented, providing guidelines in designing efficiently and effectively mission-critical systems considering principles of Interoperable Open Architectures (IOA), mission-critical integrity levels and following new standardization activities such as NATO Generic Vehicle Architecture (NGVA). A Defensive Aid Suite (DAS) system is used as a case study to illustrate how this framework can be exploited. The indention of this extension is to provide an approach to precisely estimate threats in order to de-risk missions in the very early stages.

1. Introduction

Modern military vehicles rely on mission-critical systems that enhance and guarantee successful mission capabilities. Currently, these mission-critical systems come as black boxes that are installed and maintained by the same manufacturer through the vehicle's life-cycle. These black boxes are built on proprietary technology that only the manufacturer has access to, thus limiting the choices of maintenance and upgrades. Furthermore, existing mission-critical systems are limited to communicating with other on-board systems resulting to a vehicle having multiple instances of the same equipment (e.g. GPS sensor). This presents a number of issues including having network complexity and reduced flexibility in vehicle systems configuration depending on operational requirements.

For this reason, there is a need for an innovative architecture approach that allows components from different manufacturers to be integrated, paying particular attention to the system's mission, safety, and security. When building a mission-critical system, the system designer should have the freedom to choose components that fit appropriately to the intended use as well as enable integration to any legacy mission-critical system or sensors/actuators that exist on-board the vehicle.

In this study, a thorough investigation is conducted offering a new open modular architectural approach on mission-critical systems including a case study on Defensive Aid Suites (DAS), aided to extract the necessary technical and functional requirements directly related to the system. This presented an opportunity to research a conceptual approach to the bespoke system whereby a constructive framework could be established with firm recommendations on a high level (abstracted) design so that a target platform can be equipped with a tailored mission-critical system to meet its specific requirements.

The novel mission-critical system architecture for military platforms adopts an open and modular design approach offering flexibility in configuration, upgradability, and integration. This enables a better operational and functional understanding of the mission-critical system, increasing integrated survivability capabilities [1].

Following the presented framework of this work, qualitative and quantitative results are extracted in order to provide mission

[*] Kyriakos Houliotis, Vetronics Research Centre, University of Brighton, BN2 4GJ, United Kingdom, +44 (0) 1273 642251, k.houliotis@vetronics.org

https://dx.doi.org/10.25046/aj030215

functional concepts based on mission-critical systems and threats. Additionally from the results, an early de-risking estimation can be observed, that could be beneficial for stakeholders, systems engineers and architects to decide the appropriate elements for designing mission-critical systems in the very early stages of the overall system's life-cycle. The rest of the paper is structured as shown below:

- Section 2 presents some representative key questions that offer the direction of designing mission-critical systems.
- Sections 3 and 4 provides a definition for mission-critical systems in line with standardisation activities around military vehicle architecture design approach.
- Section 5 offers the Generic Architecture Framework.
- Section 6 identifies some key considerations on safety and security in aligning and defining mission criticality levels.
- Section 7 provides an approach of calculating threats using numerical values and mathematical equations.
- Sections 8, 9, 10 and 11 present a case study on how the framework can be used to define a DAS Architecture including qualitative and quantitative results.
- Section 12 concludes the paper with some indication of next steps to this research.

This paper is an extension of work originally presented in International Conference on Military Technologies (ICMT) 2017 [2].

2. Key Questions

2.1. Modularity and Openness

What approaches are needed for a mission-critical system to be feasible and extract advantages such as modularity and openness?

2.2. Construction, Maintenance and Safety Certification

What are the benefits of using modules sharing specific functionalities on a mission-critical system and how efficient and effective can become?

2.3. Alongside benefits from the framework

How the framework's modularity extracts through the mission-critical system development benefits such as safety cases and certifications?

2.4. Low-cost tools

What are the necessary tools needed to accomplish a rapid prototyping testing and permitting software functionality and operation of a mission-critical system using low-cost components?

2.5. Migration of low cost to a safety-critical performance verification testbed

How to achieve a transition from a low-cost functional testbed to a more elaborated safety-critical performance verification testbed?

3. Mission-Critical Systems

In general, the mission is the formal summary of the aims and values of an activity. The activity can be achieved with specific mission-critical elements. Those mission-critical elements are defined as vital to the functioning of an activity. Meaning that, a successful mission can be achieved when only the right mission-critical elements are applied. There are two attributes that make the specific mission-critical elements to be applied and to be right.

First, is usually when there is maturity in the applied mission-critical elements. The maturity must reach into a level that is satisfactory in each of the involved disciplines. When this level is reached, the expected outcome is sufficient and hence, the mission can be considered successful.

The second attribute is when enough knowledge is accumulated to allow for the prediction of a mission outcome to be more accurate. To achieve this, consideration needs to be given to all possible factors involved on the specific mission. Those factors are usually known or unknown and could be anything related to the mission. The knowledge can be gained when those factors are asked and answered using three main engineering questions; "What", "Why" and "How". Once these answers are mature enough and understandable the mission-critical elements can be referred to as vital and therefore, provide success to the mission.

Today, the technology has been developed in such a way that many missions could be successfully completed with the aid of systems. Those systems are referred to as mission-critical systems. A general definition of a mission-critical system is [3]:

"*A system that is essential to the survival of a service, and whose failure or interruption significantly impacts the mission*".

A mission-critical system for a typical land military vehicle is:

"*A system that is essential to complete the mission successfully*".

A mission-critical system in land military platform is composed of many discrete Vetronics (vehicle electronics) sub-systems and components including sensors, actuators, effectors, radars and processing resources. Each of these sub-systems may contain further sub-systems and components including mechanical parts.

In Vetronics the mission can be designed, described and/or accomplished either in simple or complex terms. This differentiation resulted from the characteristics that a Vetronics system has. A simple mission for a Vetronics system is when not many factors are involved. It is also simple, when a clear and an easy step-by-step procedure is provided. For instance, a mission-critical system has to transmit data from node A to node B. That can be described as a simple mission since there is only one task to be completed and if the right mission-critical elements are used.

However, in Vetronics, for a data to be transmitted from one node to another in reality it is more complicated than the previous example. What makes the mission more complicated in Vetronics mission-critical systems is when a number of multiple disciplines, such as safety, security and survivability, are involved to achieve the mission. A more desirable, refined and detailed mission procedure is required.

Assume each of the aforementioned disciplines require to complete a specific goal on the same mission. The safety prioritises the safety of people and environment; the security prioritises the protection of data from various threats; and survivability prioritise the whole mission envelop. This makes the mission more difficult to accomplish if there is neither enough maturity nor confidence on the applied mission-critical elements.

In conclusion, "mission in military applications cannot be specified or narrowed down into a single element that easily". Therefore, an innovative unified framework is required to define and guide Mission-Critical Systems development so as to enhance mission success.

4. IOA International Activities

Today within a modern military platform, land, naval and air force, have adapted the principles of the Interoperable Open Architecture (IOA) in their system design to speed up acquisition and upgrading alongside with reducing life-cycle costs through data modelling. Below a selection of significant activities in the area of architectures and standardisation with IOA is presented.

4.1. Generic Vehicle Architecture (GVA)

The Generic Vehicle Architecture is an approach taken by the UK Ministry of Defence (MOD) to the design of electronic and power architectures for military vehicles. The approach is based on establishing system engineering principles to define a generic architecture that requires open implementation standards, Def-Stan 23-009, to support cost-effective integration of sub-systems on land platforms, electronically, electrically and physically. Any equipment shall be integrated with the GVA military land platforms must be designed in the Land Data Model (LDM) which is a (sub)-system standardisation process [4].

4.2. NATO Generic Vehicle Architecture (NGVA)

The NGVA is an approach to ensure interoperability among military land vehicles equipment. The NGVA follows a similar line to the GVA, incorporating a new method of verification and validation and by maturing the NGVA Data Model concepts, focus and implementation can be achieved [5], [6].

4.3. Future Airborne Capability Environment (FACE)

The FACE approach is an aviation US government-industry software standard and business strategy for acquisition of affordable software systems that promotes innovation and rapid integration of portable capabilities across global defence programs. The main objective of this approach is to make military operations more robust, interoperable and secure using open standards [7].

4.4. Vehicle Integration for C4ISR/EW Interoperability (VICTORY)

VICTORY is a US army vehicle's open standard for physical and logical interfaces between systems and C4ISR/EW components. The VICTORY architecture targets to provide a clear picture between to the users and the developers. Throughout the usage of an open architecture, the platforms can accept upgrades without a significant impact on the design.

5. Generic Architecture Approach

Model Driven Architecture (MDA) is an approach that is used in the system engineering domain to improve product development and delivery. The approach was initially launched in 2001 by the Object Management Group (OMG) to support the model-driven engineering of software systems. The main objective of the MDA is to provide a set of specifications of system's functionality and behaviour expressed in models. Instead of writing the code manually, the MDA approach with the help of a data modelling tool it is possible to regenerate automatically an application code. Additionally, this approach reduces implementation and integration risks when an activity is designed.

In Figure 1 an illustration of the MDA process is presented. Initially, the system requirements can be defined and specified into a Platform Independent Model (PIM) model. Using a specific standard and specification, the model can be constructed in a formal way such as the Unified Modelling Language (UML). The objective of the PIM model is to specify data, operations, functions and modes of a system, independent of the platform in which it may be integrated. In order to organise and standardise the data as well as facilitate a long-term improvement in interoperability and upgradability within the model, a data model is essential. A Platform Specific Model (PSM) model contains elements of a specific software platform. It can be generated from the PIM model either manually or automatically if appropriate tools are used. The PSM embeds the chosen software architecture strategies which refine the PIM model based on specifications. The Platform Specific Implementation (PSI) embeds the chosen middleware technology and explains the usage of a specific platform.

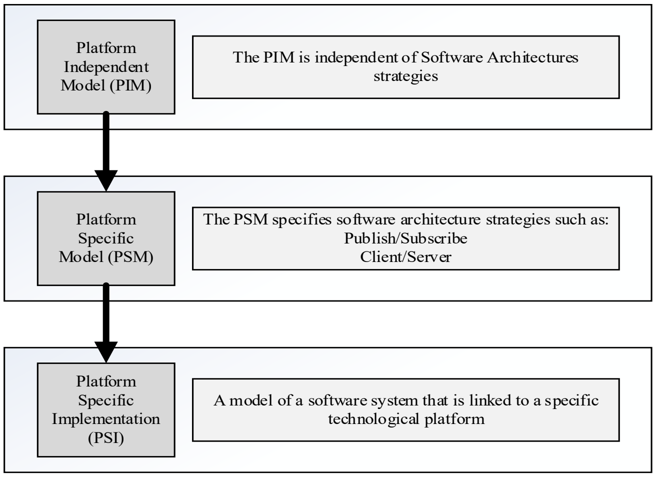

Platform Independent Model (PIM)
The PIM is independent of Software Architectures strategies
Platform Specific Model (PSM)
The PSM specifies software architecture strategies such as: Publish/Subscribe Client/Server
Platform Specific Implementation (PSI)
A model of a software system that is linked to a specific technological platform

Figure 1. Model Driven Architecture Approach.

Each electronic device is developed to satisfy one or more specific task(s) thus, the devices can only generate or receive a set of specific data for their operation. Therefore, when these devices are integrated into an IOA architecture, the broadcasted data cannot be ensured if is critical or not. For instance, the real-time level cannot be defined just by the device itself but is needed to be declared from the designers. Below there is a brief explanation of the real-time levels.

Real-Time Level – In any electronic architecture there is a set of data or information it might be critical or noncritical. Real-time can enable non, soft and hard responsiveness depending on what level, prioritisation, and importance the data is designed for. In Vetronics different real-time levels are applied for satisfying different level processes or events. The definitions of real-time levels are:

- Non-Real-Time (**NRT**) - Best Effort Service with no time constraints.

- Soft Real-Time (**SRT**) - Relaxed time (latency) requirements.
- Hard Real-Time (**HRT**) - Fixed time requirements. [8]

Due to the different critical level of data in Vetronics, this paper proposes a novel framework to support the developers to design and decide whether the data is critical or noncritical. With the aid of a flowchart, Figure 2 demonstrates a modular framework that is aimed to accurately design a data for a mission-critical system. The designer must firstly define the data attribute and if is already existing on the model. When the data attribute is declared, the designer must decide in what level of real-time responsiveness the data is belonging to. The different levels the data can take are the Hard, Soft, and Non-Real-Time. This can be useful for the developers to choose the appropriate network communication technology. Next, the designer must declare the criticality level of the data by choosing between Mission, Safety, and Security Critical levels. The critical levels shown in the diagram are the most commonly used levels used in the Vetronics systems. The user has the freedom to add/remove other critical levels such as survivability or business critical.

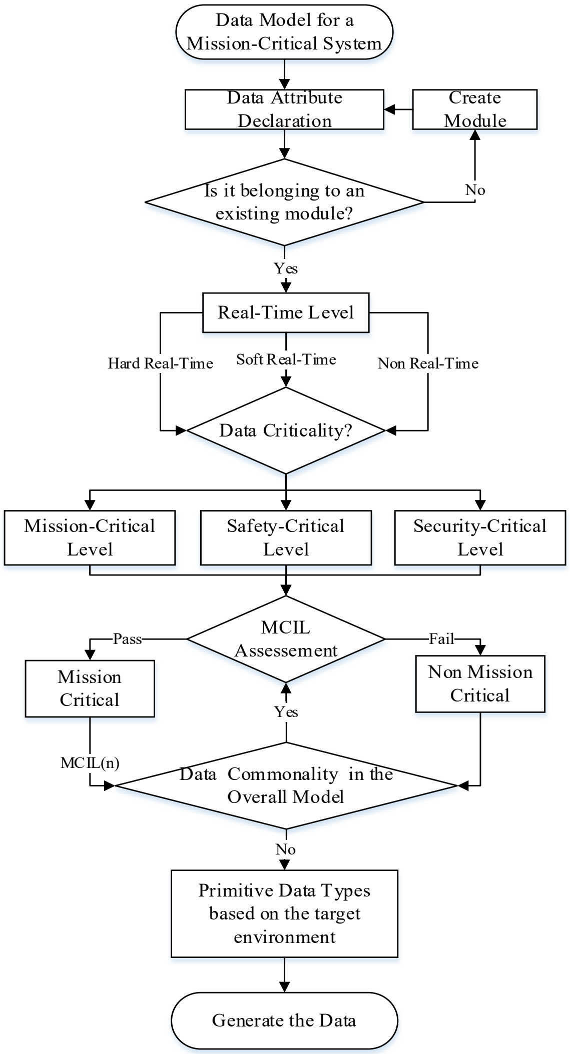


Figure 2. Mission-Critical Modelling Framework.

The next step is an assessment that is applied for the mission planning and goals based on risk assessments, a detailed explanation will be discussed in Sections 10 and 11. If the data fails the assessment is considered as non-mission-critical and if the data meets the requirements then is considered as mission-critical. In the mission-critical block, the data will be labelled with the Mission-Critical Integrity Level (MCIL). To ensure that the model is not polluted with data having similar attributes, a data commonality assessment is provided for addressing that issue. Finally, the data must have its own data type and must be checked whether the data type can be supported or not by the targeted programming language.

In a communication network, there are multiple connection points that are able to receive, store and send data across. These connection points are known as network nodes. Figure 3 depicts a generic network node which is divided into three elements, communication, processing, and application. The application element, is a computer program designed to perform a group of coordinated functions, tasks or activities for the benefit of the end user. The processing in a networked node, is a combination of machines, people, and processes that for a set of inputs produces a defined set of outputs. And the communication is the communication endpoint. The transmission of data from one computer to another is achieved by the communication device using various communication technologies [9].

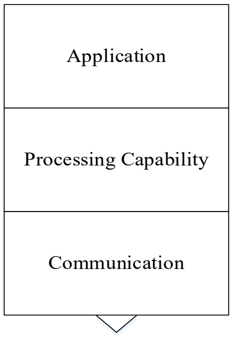


Figure 3. Network Node.

In the military vehicles, multiple (sub)-systems are integrated on a single platform. In a mission-critical system the most commonly used network nodes are sensors, actuators, effectors, processing nodes and so on. All the network nodes are interacting together through a virtual networking, as shown in Figure 4. A virtual network is capable of controlling one or more nodes over a logical or virtual networks that are decoupled from the underlying network hardware. This is used to ensure that the network nodes can efficiently integrate and perform on a single network. Common data sharing technologies used for military applications are the Data Distribution Service (DDS) and the Message Queuing Telemetry Transport (MQTT). The gateway block is integrated to support system legacy and allow different existing on-board systems to provide and receive services.

6. The need of Safety and Security

In land military platforms, there are other existing critical systems. These critical systems are for the safety and for the security of the platform and the crew. Safety critical systems are the systems whose failure may endanger human life, economics or the environment. Examples of safety critical systems in military vehicles are the vehicle's steering and fire control. Security critical systems deal with the integrity and loss of sensitive data through theft or accidental loss [10].

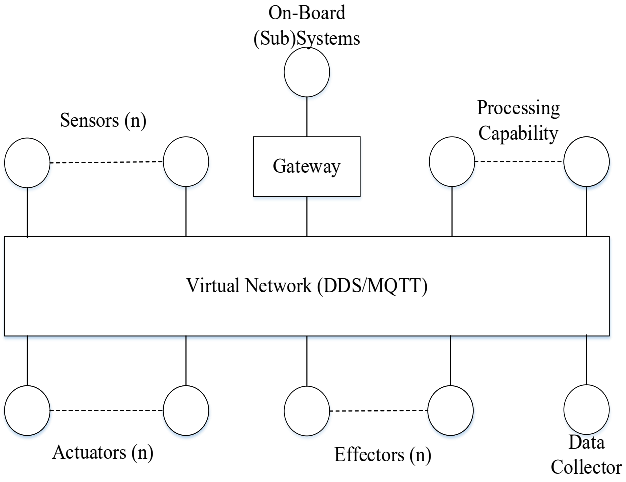


Figure 4. Virtual Network.

Safety and security critical systems can be used to extract logical capabilities, in which they can be mapped and used as the essentials of a successful mission-critical system. Considering safety and security capabilities on the development of a mission-critical system, it is possible to achieve a complete integrated survivability system [11]. It is vital to increase the dimensions and properties of safety and security in a mission-critical system to address issues such as intrusion detection, component fault/failure detection system behaviour, and restoring essential services in case of a security attack [12].

The developer must specify and approve the criticality level of each data. This can be varied depending on the application and technical specification related to user and system requirements. Although it is possible to declare and assess the data's mission-critical level using the Tables 1-4. The tables are essential to achieving a successful mission-critical system in which the mission goals are based on risk assessments. Risk assessments should be set and then that, the rigour of management and processes should be appropriate to meeting them.

Assume a civilian vehicle used for a military application; when the vehicle's passenger opens the vehicle's door some interior lights are switched on. The interior lights are designed to provide luminosity in the vehicle at night or in dark environments. During the day or at bright environments the light does not significantly impact any process thus the related data for the light can be categorised as negligible or MCIL4. During the night the light may provide luminosity to the passengers but it also indicates the vehicle's location. If the vehicle is used for a mission during night, it is likely the mission will fail, thus the corresponding data, for the light, can be categorised to as catastrophic or MCIL1. However, any data must be filtered and tested through the risk assessments before is integrated into the platform. Furthermore, the framework requires, that hazard and risk assessments be executed for the analysis of the likelihood of occurrence, consequences and detection levels provided by the tables below.

Table 1 Categories of likelihood of failure occurrence

Category	**Definition**	**Range** **(Mission Failure)**
Frequent	Many times in missions	> 10^{-3}
Probable	Several time in missions	10^{-3} to 10^{-4}
Occasional	Once in mission	10^{-4} to 10^{-5}
Remote	Unlikely in missions	10^{-5} to 10^{-6}
Improbable	Very Unlikely	10^{-6} to 10^{-7}
Incredible	Cannot believe that it could occur	< 10^{-7}

Table 2 Consequence categories

Category	**Definition**
Catastrophic	Complete mission failure
Critical	Impacts mission but not complete failure
Marginal	Major mission issues
Negligible	Minor mission issues

Table 3 Risk class matrix

Likelihood	**Consequence**			
	Catastrophic	***Critical***	***Marginal***	***Negligible***
Frequent	Class 1	1	1	2
Probable	1	1	2	3
Occasional	1	2	3	3
Remote	2	3	3	4
Improbable	3	3	4	4
Incredible	4	4	4	Class 4

The classification of the consequences are as follow:

- **Class 1**: Unacceptable in any circumstance
- **Class 2**: Undesirable: tolerable only if risk reduction is impracticable or if the costs are grossly disproportionate to the improvement gained.
- **Class 3**: Tolerable if the cost of risk reduction would exceed the improvement.
- **Class 4**: Acceptable as it stands, though it may need to be monitored.

Once the hazard and risk assessments are identified, each of the threats should also be assigned with a detection level, as given in Table 4, in order to provide a definition in which degree a threat can be detected.

Table 4 Detection Levels

Difficulty Level	Definition
No effort	Very likely to be detected
Very Easy	With almost no effort
Easy	Without great effort
Normal	Conforming to a standard
Hard	With a great deal or effort
Very Hard	Not likely to be detected

Mission-critical levels, see Table 5, offers the ability to attain in regards to mission-critical system development and related to the Classification Matrix given in Table 3. Throughout risk assessments the target MCIL can be identified and thus, it can be converted as a requirement for the mission-critical system. The derived requirements can provide an efficient data model development that can be used and ensure that the mission-critical system can succeed a mission.

Table 5 Mission-Critical Levels

Mission-Critical Integrity Level	Mission Failure Factor	Risk Classification
MCIL 4	100,000 to 10,000	Class 4
MCIL 3	10,000 to 1,000	Class 3
MCIL 2	1,000 to 100	Class 2
MCIL 1	100 to 10	Class 1

Safety (mission) critical applications, such as in [13] and [14], are using safety standards, such as the IEC 61508 and the RTCA DO-178B, in which are oriented for people's safety when a system is designed. The standards are intended to be a basic functional safety standard applicable to cover the safety management of electrical, electronic and programmable electronic systems throughout their lives. If the development of a mission-critical system involves human factors then the system should be considered as a safety critical system.

7. Threat Estimation

First, it is important to note that Tables 1, 2 and 4 are abbreviated as, Occurrence – O[n], Severity – SE[n], Detection – D[n] and threat estimation as T[n], with "[n]" representing a natural number [1] of each element or requirement. For example, if one threat is identified within the framework, then the threat will be assigned to as T[1]. If threat T[1] has sub-requirements, then it will be assigned to as T[1][n].

In order to evaluate or estimate the criticality of the identified threat T[n], the aforementioned tables and their elements must be assigned with numerical values, as depicted in Table 6. The idea behind the values is indicative (assumption), therefore, the maximum value of the threat T[n] can be roughly 99.9% and the lowest 0%. These values will indicate the probability of the threat affecting the mission.

Therefore, the threat level of the identified threat T[n], TL_T[n] can be calculated using the following expression,

$$TL\_T[n] = O[n] + SE[n] + D[n] \qquad (1)$$

Where,

TL_T[n]: The threat level of the identified threat.

O[n]: The occurrence value of the O[n].

SE[n]: The severity value of the SE[n].

D[n]: The detection value of the D[n].

In the event of having multiple threats or sub-threats, an estimation of an overall threat must be calculated in order to predict the probability of the mission success. A representation of this is shown in Table 7.

Where,

Req_core: The core requirement.

1st Sub: The first sub_core_requirement.

n Sub: Indicated the last sub-requirement.

n: Represents the real number.

i: Represents the sequential number of sub-requirements.

An approach on how to calculate the average value of two or more requirements of the same degree is as follows, (2),

Table 6 The Assigned Values for Calculating Threat T[n]

Occurrence – O[n]	Occurrence (%)	Severity – SE[n]	Severity (%)	Detection – D[n]	Detection (%)
Frequent	25	Catastrophic	≤50	Very Hard	25
Probable	20	Critical	~33.4[3.s.f]	Hard	20
Occasional	15	Marginal	~16.7[3.s.f]	Normal	15
Remote	10	Negligible	0	Easy	10
Improbable	5			Very Easy	5
Incredible	0			No Effort	0

Table 7 Requirement Sequence

Req	**1st Sub**	**n Sub**
	Req[1][1]	Req[1]…i…[1]
Req[1]	…	…
	Req[1][n]	Req[1]…i…[n]
	…	…
	Req[n][1]	Req[n]…i…[1]
Req[n]	…	…
	Req[n][n]	Req[n]…i…[n]

$$Req\_core = \sum_{i=1}^{n} \frac{Req[i]}{n} \qquad (2)$$

Where:

Req_core: The overall average value of the core requirement.

i: Lower limit number of requirement.

n: Upper limit number of requirement.

8. Case Study: Defensive Aid Suite (DAS) System

A most commonly used mission-critical system in the military platforms, is the Defensive Aid Suite (DAS) system. Is a survivability system, and can be used as a potential case study that addresses similar complexity issues to the aforementioned mission-critical systems and integration levels. A DAS is composed of sensors, effectors, algorithms and Human Machine Interfaces (HMI) that enhances the integrated survivability of a military vehicle. DAS also consists of different decoupling physical and logical capability networks applied for specific tasks or applications.

The DAS system can be a semi-autonomous or autonomous system that is capable of detecting, recognising and addressing threats. DAS elements are classified into two major categories Soft-Kill and Hard-Kill; other action categories are also possible where effective countermeasures can be deployed by other assets.

A **Hard-Kill** system engages and destroys threats. It creates an active fire zone of protection at a safe distance near the vehicle.

A **Soft-Kill** system is designed to avoid threats by confusing or re-directing the threats using jammers, decoys, and signature reduction measures.

An existing soft-kill DAS system from [14] has been selected for this case study, using existing DAS system components. In this DAS system various electronic components can be addressed, which are used in a Light Armed Vehicle (LAV) vehicle. A component such as the Long Range Passive Sensing (LRPS) sensor. The threat can be detected using optical systems with either Wide Field Of View (WFOV) or Narrowed Field Of View (NFOV) mounted on the platform. Each of the detected threats produces a signature identifying a potential threat that may be weapon systems such as guns and anti-tank rocket-propelled grenade launchers (i.e. M-712 and RPG-7). The detection range of the WFOV and NFOV optics are represented in Table 8. The table can be used as a message specification for a generic DAS system.

Table 8 Sensor Camera and Threats Attributes [15]

Anti-Armour Threats	***Threat, Calibre***	M-712, LSAH, 155mm	RPG-7, 80mm	Gun, 20mm, APDS
IR WFOV	***Distance, [m]***	400	470	5480
IR NFOV	***Distance, [m]***	3600	4200	340
LI/RG Camera	***Threat, [Pixels]***	1.3	42 x 42	0.8
	Target, [Pixels]	25x20	234 x 187	118 x 60
Threat Variables	***Dimensions, [m]***	0.155 dia.	0.18 dia.	2.1 dia.
	Range, [m]	14000	500	2000
	Velocity, [m/s]	255	255	815

9. Proposed DAS Architecture and Modules

This section presents a novel DAS system architecture that could be potentially used to satisfy the questions in Section II as presented earlier in this paper. For the system to facilitate a modular framework the Model Driven Architecture approach is used to construct modules that enable upgradability, maintainability and system legacy with technologies, devices and operational or functional capabilities. This proposed DAS architecture aims to fuse systems and software modelling and simulation capabilities, modular open system architectures and device integration techniques into a single package to enable rapid design, development, verification, certification and deployment of interoperable, platform portable and manoeuvre embedded mission criticality. The following sub-sections are classified as module models used for a DAS system and can also be applied for any mission-critical system.

9.1. Threats

Threats are all the known causes that can damage the platforms. Threats can be either external or internal; internal threats could be cyber-attacks, malfunction etc. External threats could be missiles or mines.

9.2. Sensors

Sensors module represents all the candidate sensors used for the DAS system. This module is for detecting and responding to any incoming threat from the physical environment. The specific input may be motion, heat, light or any of other environmental phenomena. The output is a signal that is converted into a format that is human readable or machine readable electronically transmitted over a network for exploitation and further processing.

The Sensors module can be directly connected to the DAS Sensor Processing Module to apply the safety, security, and performance on the indented design.

9.3. DAS Sensor Processing Module (DASSPM)

The DASSPM is the module that is capable of receiving the data/information from the DAS sensor. The DASSPM converts the data into a usable format that can classify, position and eventually eliminate or avoid the detected threat. The module will be able to detect also the attributes of the candidate DAS sensor in terms of type and capability if the appropriate message specification is constructed.

The DASSPM, therefore, receives the data/information from the candidate sensor and extra qualities can be added. Developers can customise the data and modify it into the corresponding real-time environment, criticality level, and mission-critical integrity level. The DASSPM will also communicate with the DASECM.

9.4. DAS Computer Module (DASCM)

The DASCM collects and processes the information from the DASSPM, to decide the best action for each available threat. This action may require user interaction or it can be fully automated. The modular design allows the system designer to use a single or multiple DASC that may deal with the same or different types of threats at the same time, providing fault tolerant and distributed design. Then the DASC communicated with the appropriate DAS Effector Control Module (DASECM), depending on the type of action decided. When there is no DASC available, the DASECM takes responsibility for the appropriate countermeasure action.

9.5. DAS Effector Control Module (DASECM)

The DASECM is responsible for controlling the DAS effectors according to the commands received from the DASCM. This module has many commonalities in data and functionality with the DASSPM. In the case of DASC failure, the DASECM can be actioned to directly deal with the threat. Additionally, it offers configuration options for the effector to achieve modularity and dynamic configuration.

9.6. Effectors

The Effectors Module is the representation of the all the candidate effectors. This module is constructed upon the available effectors needed to be installed on the platform. Each of the candidate effectors carries specific attribute thus, the module is using message specifications common to the overall architecture.

10. Top Level of a DAS Data Model

Due to the increased number of electronic components designed from different manufacturers who use different technologies, approaches, standards and architectures, the systems integration process becomes rather complex. A suggested approach to improve the vehicle's performance and survivability is to use, information management techniques. In this paper, the potential approach to such component's capability acquisition is proposed by structuring a set of message specifications and develop the DAS Data Model, represented in Figure 5.

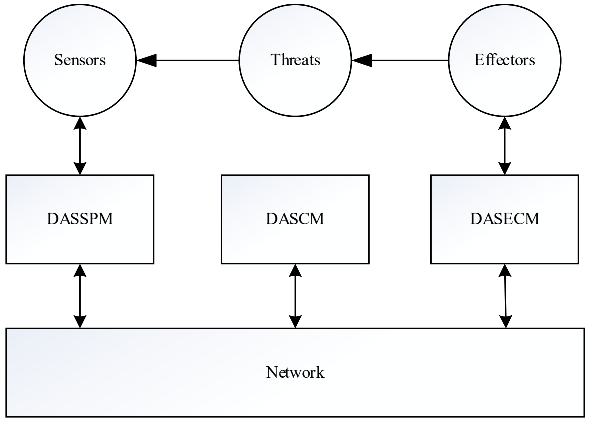


Figure 5. Top Level of DAS Architecture.

Following the proposed framework and the proposed DAS architecture and modules, the system will be able to use different battlefield scenarios, operational modes and different electronic devices from different manufacturers. With this development and the combination of the proposed approaches, the DAS Data Model will provide the different criticality levels and the different real-time environments that will be the essentials for a successful mission.

10.1. DAS PIM

Using the IBM Rational Rhapsody Developer for C++ tool, the message specification for the soft-kill DAS system sensors is represented in a UML model as shown in Figure 6. The tool is selected because it is the most commonly used for land vehicle electronic architectures, specifically in the GVA approach implementation [4]. This a PIM in which it has the ability to be redesigned when additional modifications are required during the development, or additional sensors are included in the DAS system. This PIM model is constructed using the attributes in Table 8 and with the usage of the proposed MCIL framework and DAS architecture, the mission-critical message specifications of a soft kill DAS can be created.

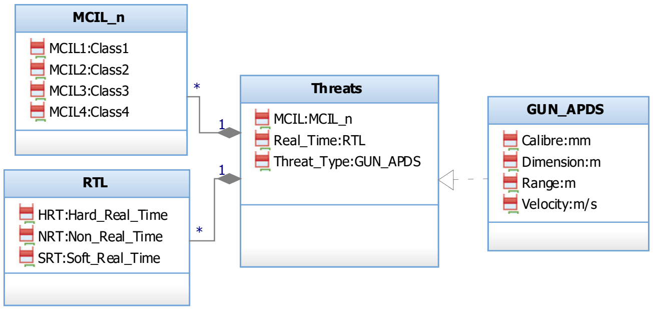


Figure 6. Threat PIM.

The message specification in Figure 6 has its own Mission-Critical Integrity Level (MCIL_n) and Real-Time (RTL) levels.

The developer will have the freedom to choose between the aforementioned levels when applying the paper's approaches. For instance, if the vehicle uses passive armour the specific threat is not lethal, therefore, the Mission-Critical Integrity level of the message can be classified as Class 4. If is a light armoured vehicle then the Mission Criticality Integrity level can be classified as Class 2. The same can be applied for the Real-Time responsiveness.

10.2. DAS PSM

Once the PIM model is designed and specified, the model can be translated into PSM for integration into a specific architecture. The middleware technologies that can be used are the Data Distribution Service (DDS) or Message Queue Telemetry Transport (MQTT). DDS is an OMG machine-to-machine middleware standard that offers scalability, real-time, Quality of Service (QoS), high performance and interoperability data exchange between data publishers and subscribers. Publish/Subscribe message patterns are used for sharing the DAS system data in order to minimise the impact of adding new sub-systems [16]. MQTT is a lightweight messaging protocol that offers bi-directional communication to nodes. Its design has been created to minimise network bandwidth that uses the messages in a reliable degree of delivery.

When each attribute in the PSM model has its own data type, it must be specified and validated from the supported primitive data types of each target environment of DDS. The message specification file is used to describe the software component's application in order to enable communication between software components from different programming languages. However, the PSM transformation can be translated into a PSI model for the simulations and apply case studies of any mission-critical environments.

11. Qualitative and Quantitative Results

Using the example in Figure 6 (Section 10.1) and applying the threat estimation procedure discussed in Section 7, the threat can be calculated and then estimate the effect level of the mission. The example stated the following, "if the vehicle uses passive armour the specific threat (see Figure 6) is not lethal, therefore, the Mission-Critical Integrity level of the message can be classified as Class 4. For a light armoured vehicle then the Mission Criticality Integrity level can be classified as Class 2". In this section an estimation of the effect level of the stated threat is presented using anticipated values for each element.

11.1. Case 1: Passive Armour Vehicle

Case 1: If the vehicle uses passive armour then threat is as depicted in Figure 7.

The occurrence O[1] of the potential cause can be occasional. Therefore, the occurrence selected from Tables 1 and 6 is,

Occurrence: **O[1]** – Occasional (15%)

Using Table 6, the O[1] is 15%.

Considering that the threat does not significantly impact the mission or the vehicle, the severity of the specified threat can be negligible. Using Tables 2 and 6,

Severity: **SE[1]** – Negligible (0%)

Assuming that the vehicle uses passive armour instead of active, the threat T[1] will be detected from the crew. Therefore, the detection might be considered to as hard to detect, and using Tables 4 and 6,

Detection: **D[1]** – Hard (20%)

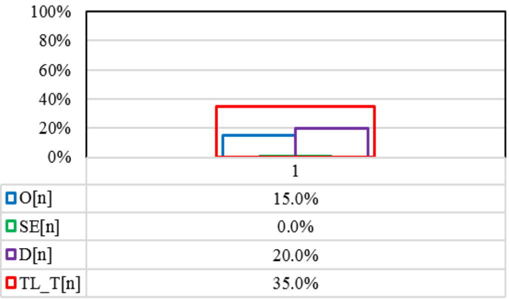


Figure 7 Case 1: T[n] Level

TL_T[1]: [(O[1]:15%)+(SE[1]:0%)+(D[1]:20%)= 35%

After the threat analysis of the specific threat for this case study, it has been identified to be 35% hazardous against the mission.

11.2. Case 2: Light Armoured Vehicle

Case 2: If the vehicle is a light armoured vehicle then the threat will be as in Figure 8. The occurrence O[1] of the potential cause can be occasional similar to the Case 1. Therefore, the occurrence selected from Tables 1 and 6 is,

Occurrence: **O[1]** – Occasional (15%)

Using Table 6, the O[1] is 15%.

Considering that the threat significantly impacts the mission or the vehicle then the severity of the specified threat can be catastrophic. Using the table 2 and 6,

Severity: **SE[1]** – Negligible (50%)

As aforementioned, the vehicle is a light armoured vehicle and assume that it has an integrated DAS system. Therefore, the threat T[1] to be detected can be categorised as easy. Using the Tables 4 and 6, detection is,

Detection: **D[1]** – Easy (10%)

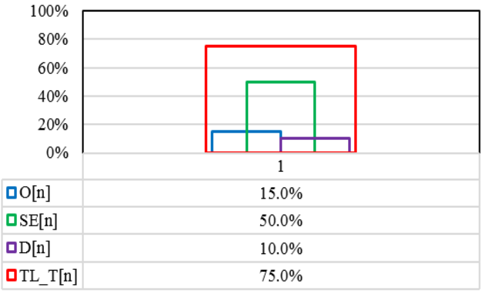


Figure 8 Case 2: T[n] Level

TL_T[1]: [(O[1]:15%)+(SE[1]:50%)+(D[1]:10%)= 75%.

The same threat as identified in Case 1 has being calculated to be 75% within this case study (2); whilst the same threat for a different mission scenario has increased in terms of threatening level. This in turn, elements that can be identified for a mission shall be re-used and analysed accordingly using this preliminary mission analysis at this early stage of a mission-critical system.

12. Conclusion and Future Work

At present there are no current development activities employed for mission-critical systems. In this paper, the importance of those activities is discussed. The objective of the proposed framework, is to enable system engineers achieve the mission development of any critical system's life-cycle, efficiently and effectively. This is achieved through a single package, using the three following main capabilities.

Firstly, a basic functional "mission standard" that enables the mission life-cycle development in a critical system; Secondly, the ability of mission interoperability for services and functionalities between systems and sub-systems built and procured in different times; and finally, the process that can support to define and analyse the mission data requirements of various critical systems.

The paper covered only the primary mission-critical attributes of a critical system, extracting the basic functional requirements. The scope of those requirements is paying particular attention on the mission, safety and security-critical attributes.

Furthermore, the extended part of this work, proposed an approach to estimate threats and their impact on a mission so as to early de-risk missions and systems (mission-critical) whilst in their design stages.

The next step is to further develop and refine the proposed using detail low level design case studies along with measures of performance.

References

[1] C. Ponsard, P. Massonet, A. Rifaut and J. F. Molderez, "Early Verification and Validation of Mission Critical Systems," Electronic Notes in Theoretical Computer Science 133, pp. 237–254, 2005

[2] K. Houliotis, P. Oikonomidis, P.Charchalakis, E. Stipidis, "An Efficient Approach to Designing Mission-Critical Systems, Case Study: Defensive Aid Suite (DAS) Systems", 2017 International Conference on Military Technologies (ICMT), Brno, Czech Republic May 31 – June 2, pp 402-409, 2017

[3] C. Ponsard, P. Massonet, A. Rifaut and J. F. Molderez, "Early Verification and Validation of Mission Critical Systems," Electronic Notes in Theoretical Computer Science 133, pp. 237–254, 2005

[4] F. Ciccozzi, I. Crnkovic, D. Di Ruscio, I. Malavolta, P. Pelliccione, R. Spalazzese, "Model-Driven Engineering for Mission-Critical IoT Systems." in IEEE Software, vol. 34, no. 1, pp, 46-53, Jan.-Feb.2017

[5] UK Ministry of Defence (MOD), "Generic Vehicle Architecture (GVA)," 2010

[6] NATO, STANAG 4754, "NATO Generic Systems Architecture (NGVA) for Land Systems," Edition 1, Ratification draft, August 2015

[7] M. Pradhan and D. Ota, "An adaptable multimodal crew assistance system for NATO generic vehicle architecture," 2016 International Conference on Military Communications and Information Systems (ICMCIS), Brussels, pp. 1-8, 2016

[8] M. Williamson, "Future Airborne Capability Environment (FACE),"2010, from"http://www.defensedaily.com/Assets/Williamson%20Panel%204%20.pptx

[9] S. A. Brandt, S. Banachowski, C. Lin, T. Bisson, "Dynamic Integrated Scheduling of Hard Real-Time, Soft Real-Time and Non-Real-Time Processes", Proceedings of the 24th IEEE International Real-Time Systems Symposium (RTSS'03), 2003

[10] R. M. Connor, "VSI Vetronics Standards and Guidelines", QINETIQ/EMEA/TS/CR0702540 Issue 3, June 2009

[11] J. P. Lobo, P. Charchalakis, and E. Stipidis, "Safety and security aware framework for the development of feedback control systems," 10th IET System Safety and Cyber-Security Conference 2015, 2015

[12] A. Deshpande, O. Obi, E. Stipidis, and P. Charchalakis, "Integrated vetronics survivability : Requirements for vetronics survivability strategies," 6th IET International Conference on System Safety, Birmingham, 2011

[13] O. Obi, a Deshpande, E. Stipidis, and P. Charchalakis, "Intrusion Tolerant System for Integrated Vetronics Survivability Strategy," 8th IET International Safety Conference incorporating the Cyber Security Conference, Cardiff, 2013

[14] A. Larrucea, J. Perez, and R. Obermaisser, "A Modular Safety Case for an IEC-61508 Compliant Generic COTS Processor," IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, pp. 1788–1795, 2015

[15] J. Kong and H. Yan, "Comparisons and analyses between RTCA DO-178B and GJB5000A, and integration of software process control," 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, 2010

[16] [15] J. L. Rapanotti, "Developing soft-kill capability for light armoured vehicles through battlefield simulations," Defence R&D Canada – Valcartier Technical Memorandum DRDC Valcartier TM 2003-276, 2007

[17] USA Department of Defence (DOD), "The Data Distribution Service Reducing Cost through Agile Integration", 2011, from "http://www.twinoakscomputing.com/wp/DDS_Exec_Brief_v20l-public.pdf

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 2, No. 3, 138-145 (2018)
www.astesj.com
Special Issue on Advancement in Engineering Technology

ASTES Journal
ISSN: 2415-6698

Efficient Limited Feedback Technique for FDD MIMO Systems

Papis Ndiaye, Moussa Diallo, Idy Diop

Department of Computer Science, Polytechnic Institute (ESP), Université Cheikh Anta Diop de Dakar, Senegal

ARTICLEINFO

Article history:
Received: 15 November, 2017
Accepted: 05 February, 2018
Online: 17 March, 2018

Keywords:
MIMO OFDM
Beamforming
Feedback overhead

ABSTRACT

In this paper an efficient feedback quantization technic for beamforming in MIMO systems is presented. The proposed technic named time domain quantization TD-Q is based on the feedback of time domain parameters necessary for the reproduction of the beamforming matrix at the transmitter. This TD-Q presents the same performance than the conventional Givens rotation quantization GR-Q approach which is adopted in IEEE 802.11ac standardand. The performance and amount of feedback of the proposed TD-Q are studied and compared with the GR-Q in IEEE 802.11ac context.

1 Introduction

The combination of Multiple-Input Multiple-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) technologies (MIMO-OFDM) is now adopted in several communication standards, including the 5^{th} generation of mobile communication network (5G) [1], the IEEE WLAN 802.11ac [2] and the IEEE 802.16 standards (WiMax)[3].

On the one hand, OFDM is a worthwhile tradeoff between bit-error rate performance and spectral efficiency. OFDM consumes part of the channel bandwidth, but it is robust to frequency selective fading environment. In addition, it enables the use of several advanced technics to further enhance the system throughput, as for instance the bit loading technic [4] and the subcarriers allocation in orthogonal frequency division multiple access (OFDMA) [5].

On the other hand, the MIMO system has the potential to improve the system capacity. In IEEE 802.11ac wireless local area network (WLAN) standard, there are five transmit and receive MIMO technics: Cyclic Shift Diversity (CSD), Space Time Block Coding (STBC), Spatial Division Multiplexing (SDM), Maximal Rotation Combining (MRC) and Transmit Beamforming (TxBF) [2]. Among these five technics, this is the Transmit Beamforming which can maximize the system capacity. Indeed, the Beamforming with precoding and postcoding eliminates the co-channel interferences (CCI) which are the fundamental problem faced by the practical MIMO system [6-8].

However, in beamforming technic the channel state information (CSI) must be available at the transmitter and the receiver. This CSI is estimated at the receiver and fed back to the transmitter. The CSI, which indicates amplitude and phase for each transmit antenna, receive antenna, and each OFDM subcarrier in the RF channel may reduce the overall throughput. To reduce the feedback amount, the receiver computes the beamforming matrix (precoding matrix) which is an unitary matrix and compresses it before sending back to the transmitter.

There are several proposals in the literature for the beamforming matrix compression. Many authors propose codebook based approach. In codebooks based technics, the channel distribution is taken into account during the codebook design. Instead of sending the precoding matrix, only the index of the selected precoding matrix (after channel estimation and SVD decomposition) is sent. Intrinsically, the codebook is restricted to have fixed cardinality. Thereby, the selected pre-coding matrix, which is the most similar is not necessarily the most optimal. Unfortunately, the codebook size has an impact on the system performance and requires high storage. For traditional MIMO systems, several codebooks have been proposed, such as Kerdock codebook [9], codebooks based on vector quantization [10], Grassmannian packing [11], discrete Fourier transform (DFT) [12] and quadrature amplitude modulation [13]. The coodbook principle is adopted in LTE [14]. In [15], the codebook was designed and optimized by selecting matrices having 8 phase-shift keying (PSK) entries and coping with a large range of propagation conditions. However the cost of this scheme is that the performance gain decreases dramatically due to the channel quantized er-

https://dx.doi.org/10.25046/aj030216

ror caused by the codebook feedback[16].

Another alternative is the Givens rotation (GR) approach. Here, Givens rotation is used to decompose the unitary beamforming matrix. After the decomposition, the receiver only feeds back the GR parameters necessary for the reproduction of the beamforming matrix by the transmitter [17]. The GR approach is adopted in IEEE 802.11ac standard for TxBF mode[2]. The compression ratio of the GR quantization is 75% for a MIMO 2×2 configuration [18]. However, this ratio decreases depending on the number of antennas. It then becomes necessary to find an alternative to the GR qunatization since the $802.11ac$ standard has adopted an 8×8 MIMO configuration which may reach $10Gbps$ and this number of antennas will undoubtedly increase due to the growing demand for bitrate.

This paper proposes a quantization approach named time domain quantization (TD-Q). Instead of the quantized precoder matrix as in GR, we propose the feedback of the quantized time domain channel coefficients. In addition, an optimization based on a metric updated during the channel estimation process and the MIMO channels impulse response is also proposed in order to further reduce the feedback overhead.

The rest of the paper is organized as follows. The section 2 presents the beamforming in MIMO OFDM systems. Section 3 is dedicated to the GR-Q which is adopted in the IEEE 802.11ac standard. Following that, the proposed TD-Q is presented in the section 4. Next, simulation results and comparison between GR-Q and TD-Q are shown in the section 5. Finally, the conclusion on the works presented in this paper is done in the section 6.

2 System Model

Consider a MIMO channel with N_t transmit antennas and N_r receiver antennas. In classical MIMO-OFDM context, for each subcarrier, the $N_r \times 1$ received signal y can be expressed as:

$$y = Hx + n \tag{1}$$

where H is the $N_r \times N_t$ frequency channel coefficients, x the $N_t \times 1$ transmitted signal and n the $N_r \times 1$ zero mean Gaussian noise vector.

By using singular value decomposition (SVD), the matrix H can be decomposed into:

$$H = UDV^H \tag{2}$$

where $U \in \mathbb{C}^{N_r \times R}$ and $V \in \mathbb{C}^{N_t \times R}$ are unitary matrix, $D \in \mathbb{C}^{R \times R}$ is diagonal, $(.)^H$ denotes the hermitian of (.) and R the rank of the matrix H.

In MIMO-OFDM with beamforming context as in TxBF mode of IEEE 802.11ac, the received signal y is expressed this time by:

$$y = U^H HVx + U^H n = U^H UDV^H Vx + U^H n = Dx + U^H n \tag{3}$$

where V and U^H are used as a precoder matrix and a postcoder matrix respectively. The postcoder matrix U^H is an unitary matrix and thus the noise is neither colored nor enhanced.

Since D is exactly diagonal, there is no CCI and the MIMO channel matrix is reduced on R separate and independent SISO channels.

However, the precoder matrix V is needed at the transmitter. Two solutions are available in order to provide CSI at the transmitter. The first solution consists of exploiting the channel reciprocity in case of time division duplex (TDD) transmission modes. But, it is not straightforward due to the mismatch in the radio front end. The alternative, which is the most appropriate and that can be applied for both TDD and frequency division duplex (FDD) transmission modes, consists in using the reverse link to feed back the matrix V to the transmitter. Note that, for each subcarrier the unitary matrix V has to be quantized and fed back to the transmitter.

3 Givens Rotation Quantization and feedback compression ratio analysis

3.1 Givens Rotation Quantization principle

Givens Rotation Quantization GR-Q is proposed in [17]. The authors propose a reduction of the feedback overhead by exploiting the unitary property of the precoder matrix V. The idea is to represent the matrix V as a special form of Givens rotation with a complex diagonal matrix. Thereby, the $N_t \times R$ unitary matrix V can be factorized as follows.

$$V = \prod_{i=1}^{min(R,N_t-1)} [D_i \prod_{l=i+1}^{N_t} G_{li}^T(\psi_{li})] \times I_{N_t \times R} \tag{4}$$

where

$$D_i = \begin{bmatrix} I_{i-1} & 0 & 0 & \dots & 0 \\ 0 & e^{j\phi_{i,i}} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & e^{j\phi_{N_t-1,i}} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$G_{li} = \begin{bmatrix} I_{i-1} & 0 & 0 & \dots & 0 \\ 0 & cos(\psi_{li}) & 0 & sin(\psi_{li}) & 0 \\ 0 & 0 & I_{l-i-1} & 0 & 0 \\ 0 & -sin(\psi_{li}) & 0 & cos(\psi_{li}) & 0 \\ 0 & 0 & 0 & 0 & I_{N_t-l} \end{bmatrix} \tag{6}$$

The matrix D_i is an $N_t \times N_t$ diagonal matrix, $I_{N_t \times R}$ is an $N_t \times R$ identity matrix and G_{li} is an $N_t \times N_t$ Givens rotation matrix.

According to the equations (4-6), the parameters to be determined to identify the precoder matrix V are:

$$\begin{array}{llll} \psi_{li} & for & i = 1, 2, \dots, m & and \quad i < l \leq N_t \\ \phi_{j,i} & for & i = 1, 2, \dots, m & and \quad i < j \leq N_t - 1 \end{array} \tag{7}$$

where $m = min(R, N_t - 1)$.

Take a 3×3 unitary matrix V as an example, it can be expressed as:

$$V = D_1 \times G_{21}^T(\psi_{21}) \times G_{31}^T(\psi_{31}) \times D_2 \times G_{32}^T(\psi_{32}) \times I_{3\times 3} \quad (8)$$

The precoder matrix V can then be reconstructed through the six parameters $\psi_{21}, \psi_{31}, \psi_{32}, \phi_{1,1}, \phi_{2,1}$ and $\phi_{2,2}$.

Therefore, instead of all the elements of the matrix V, it is sufficient to considere the parameters ψ_{li} and $\phi_{j,i}$. These parameters can vary from 0 to 2π for ϕ and from 0 to $\pi/2$ for ψ.

Now, ψ and ϕ can be quantized according to equations (9) and (10) where $\widehat{\psi}$ and $\widehat{\phi}$ represent the quantized angles.

$$\widehat{\psi} = \frac{k\pi}{2^{nb_\psi+1}} + \frac{\pi}{2^{nb_\psi+2}} \quad (9)$$

where $k = 1, 2, \ldots, 2^{nb_\psi} - 1$ and nb_ψ the number of bits per angle ψ

$$\widehat{\phi} = \frac{k\pi}{2^{nb_\phi-1}} + \frac{\pi}{2^{nb_\phi}} \quad (10)$$

where $k = 1, 2, \ldots, 2^{nb_\phi} - 1$ and nb_ϕ the number of bits per angle ϕ

Finally, the receiver feeds back the quantized parameters $\widehat{\psi}_{li}$ and $\widehat{\phi}_{j,i}$ to the transmitter which can recover the quantized precoder matrix $\widehat{V}$ by using:

$$\widehat{V} = \prod_{i=1}^{min(R,N_t-1)} [\widehat{D}_i \prod_{l=i+1}^{N_t} \widehat{G}_{li}^T(\widehat{\psi}_{li})] \times I_{N_t \times R} \quad (11)$$

where

$$\widehat{D}_i = \begin{bmatrix} I_{i-1} & 0 & 0 & \ldots & 0 \\ 0 & e^{j\widehat{\phi}_{i,i}} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & e^{j\widehat{\phi}_{N_t-1,i}} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

It is important to note that:

- The receiver has to feed back a precoder matrix $\widehat{V}$ for each subcarrier.
- The quantization error Σ with $\widehat{V} = V + \Sigma$ exists on each subcarrier.

Consider the equation (3) with the quantized error Σ and $\widehat{V}$ as a precoder matrix.

$$\begin{aligned} y &= U^H UDV^H \widehat{V} x + U^H n \\ &= U^H UDV^H (V + \Sigma) x + U^H n \\ &= Dx + \overline{D} V^H \Sigma x + U^H n \\ &= Dx + M_{CCI} x + U^H n \end{aligned} \quad (13)$$

The term $M_{CCI}x = (DV^H\widehat{V} - Dx$ is the corresponding CCI caused by the quantization error. However, in telecommunication standrd, one always chooses a number of quantization bits (nb_ψ and nb_ϕ) that allows the minimization of resulting CCI. This is the case of IEEE 802.11ac where $nb_\psi = nb_\phi = 10$.

Finally, the feedback overhead length (in bit) noted by N_{FGR} can be calculated using the formula $N_{FGR} = N_u * nb * N_t(N_t - 1)$ assuming that $N_r = N_t$, $nb_\psi = nb_\phi = nb$ and N_u the number of used subcarriers.

3.2 Feedback overhead length analysis in IEEE 802.11ac context

The bandwidth in IEEE802.11ac has increased from a maximum of 40 MHz with the old standard up to 80 or even 160 MHz. The number of subcarriers and used one according to the bandwidth are grouped in the Table 1.

Table 1: Number of used subcarriers according to the de bandwidth in IEEE 802.11ac

Bandwidths	N	N_u
20 MHz	64	52
40 MHz	128	108
80 MHz	256	234
160 MHz	512	486

Table 2 contains for any MIMO configurations and bandwidth:

- the feedback overhead length when no compression is performed,
- the feedback overhead length when GR quantization is performed,
- the compression ratio achieved by GR quantization.

As we can see, the GR quantization can reach a compression ratio up to 75% for a MIMO 2×2 configuration. This compression ratio decreases when the number of antennas becomes large. Because of the high demand in terms of transmission throughput, the number of antennas will increase in the evolution of IEEE 802.11 standard. The next WLAN generation will undoubtedly consider the massive MIMO concept which is more and more studied for WLAN [19-20]. In this context, it is necessary to find an alternative to the GR quantization because its compression ratio, as demonstrated by the equation (14), will be around 50%:

$$\lim_{N_t=N_r\to+\infty} \frac{N_t(N_t-1)}{N_t * N_t * 2} = \frac{1}{2} \quad (14)$$

4 Time domain quantization

4.1 Time domain quantization principle

Consider a MIMO channel with N_t transmit antennas and N_r receive antennas. The discrete time domain channel response between the transmitting antenna i

Table 2: Feedback overhead (in bit) with and without GR quantization for any configurations ($N_t \times N_r$ and bandwidth).

MIMO configuration	Number of GR parameters	feedback overhead N_{FGR} for GR				feedback overhead without quantization				Ratio
		20 MHz	40 MHz	80 MHz	160 MHz	20 MHz	40 MHz	80 MHz	160 MHz	
2x2	2	1040	2160	4680	9720	4160	8640	18720	38880	75%
3x3	6	3120	6480	14042	29160	9360	19365	42120	87480	66.67%
4x4	12	6240	12960	28080	58320	16640	34560	74880	155520	62.5%
6x6	30	15600	32400	70200	145800	37440	77760	169730	349920	58.33%
8x8	56	29120	60480	131040	271410	53248	110592	239616	622080	56.25%

and the receiving antenna j under the multipath fading environments can be expressed as:

$$h_{ij}(n) = \sum_{l=0}^{L-1} h_{ij,l}\delta(n - \tau_{ij,l}) \tag{15}$$

where L is the number of paths, $h_{ij,l}$ and $\tau_{ij,l}$ the complex time varying channel coefficient and delay of the l^{th} path.

Note that in practical OFDM system, in order to eliminate the intersymbol interference (ISI) between consecutive OFDM symbols, the maximum multipath delay is within the cyclic prefix, of length (CP) *ie* $L \leq CP$. Thereby, we propose, instead of the quantized precoder matrix V for each subcarrier, to feed back the quantized time domain channel coefficients for the entire system. The main goal of this proposition is to considerably reduce the feedback overhead because in this case, the feedback is composed at most by $N_t * N_r * CP$ complex coefficients. As can be seen, the number of OFDM subcarriers has no impact on the feedback overhead length. In addition, this proposed time domain quatization method can be performed jointly with the channel estimate.

4.2 Joint time domaine channel estimation and feedback quantization

In closed loop MIMO-OFDM systems, a preamble is inserted at the beginning of frames to facilitate the channel estimation. Comb-type pilot method can also be used on the rest of the frame (as in IEEE 802.11ac [2] or LTE [1]) to improve the accuracy of the channel estimation during the transmission. In this paper, we focus on time domain channel estimation (TD-CE). This technique is known to provide very good results by significantly reducing the noise on the estimated channel coefficients [21][22].

Firstly, without using any knowledge of the statistics on the channels, least square (LS) channel estimation can be performed at the receiver side using the demodulated pilots signal and the known pilots symbol in the frequency domain for each subcarrier.

Next assuming that $H_{ij,k}$ is the discrete LS estimated channel response on subcarriers k between the ith transmit antenna and the jth receive antenna, the CSI can be converted into the time domain by the inverse discrete fourier transform (IDFT) algorithm as:

$$\begin{aligned} h_{ij}(n) &= IDFT(H_{ij}) \\ &= \sqrt{\tfrac{1}{N_s}} \textstyle\sum_{k=0}^{N_s-1} H_{ij,k} e^{\frac{j2nk\pi}{N_s}} \end{aligned} \tag{16}$$

where N_s is the number of subcarriers and $n = 1, \ldots, N_s$.

Note that the LS method is widely used due to its simplicity and minimum requirements for the knowledge of channel statistics . However, the LS estimator does not consider the noise effect. That is why its performance is often degraded by the noise. By taking into account the equation (15) and the fact that $L \leq CP$, the time domain samples $h_{ij}(n)$ from ((16)) can be divided into two parts:

- the first CP samples in which are the paths of the channel,
- the other samples which are only composed by noise.

A smothing process which keeps only the first CP samples by eliminating the other one can considerably improve the accurate of the estimated channel [21]. However, it is possible to estimate a mean of the noise noted by $\overline{m}$ from the noise samples before its deletion.

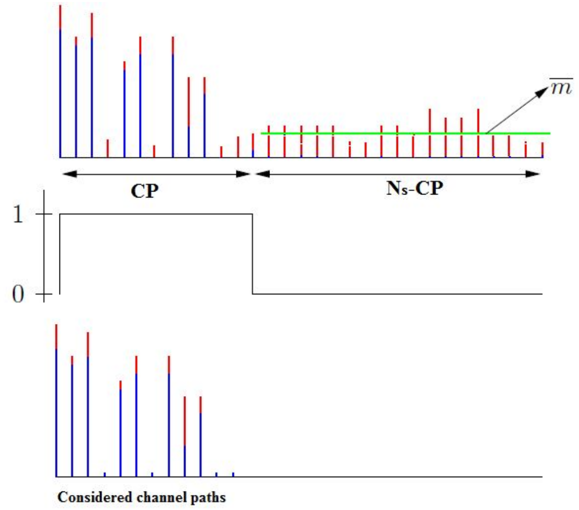


Figure 1: Mean Noise

This proposed process is illustrated in the figure 1. the metric $\overline{m}$ estimated from the noise samples allows to keep only the channel paths which are among the first CP samples. As can be seen, all samples below the $\overline{m}$ were eliminated.

Now, the receiver have to do the two following actions:

- Perform a DFT of size N_s by considering only the remaining samples $h_{ij}(n)$. The resulting estimated channel frequency response is then more accurate than the LS one.

- Perform the feedback of the remaining samples $h_{ij}(n)$. To achieve this goal, each of the remaining samples $h_{ij}(n)$ whose number is lower than CP is expressed as:

  $$h_{ij}(n) = \rho_{ij} e^{\phi_{ij,n}} = tan(\psi_{ij,n}) e^{\phi_{ij,n}} \quad (17)$$

  where $n \leq CP$, $\psi_{ij,n}$ and $\phi_{ij,n}$ angles parameters can vary from 0 to $\pi/2$ and 0 to 2π respectively.

  After that, $\psi_{ij,n}$ and $\phi_{ij,n}$ is quantized and fed back to the transmitter according to equations (9) and (10) where $\widehat{\psi}_{ij,n}$ and $\widehat{\phi}_{ij,n}$ represent the quantized angles parameters.

Thus, the feedback overhead length (in bit) noted by N_{FTD} can be calculated using the formula $N_{FTD} = N_{rs} * 2 * nb * N_t^2$ assuming that $N_r = N_t$, $nb_\psi = nb_\phi = nb$ and N_{rs} the number of remaining samples.

Thanks to this feedback, the transmitter can reconstitute the accurate estimated channel. Thereby, the MIMO channel response is available both at the reciever and the transmitter side. Finally, transmitter and reciever perform SVD decomposition which allows it to respectively have the pre-coder and the post-coder.

5 Performance comparison in IEEE 802.11ac context

In this section, we compare the Givens rotation quantization GR-Q and the proposed time domain quantization TD-Q one in terms of feedback overhead and BER performance in IEEE 802.11ac context [2]. Note that the bandwidth in IEEE802.11ac has increased from a maximum of 40 MHz with the old standard up to 80 or even 160 MHz. For each bandwidth, the cyclic prefix length (CP) is equal to a quarter of the number of subcarriers (in Table 1). We consider in the similations MIMO 2×2, 3×3, 4×4, 6×6 and 8×8 with $4 - QAM$ and $64 - QAM$ modulations.

In the sequel, all simulations are performed using the ITU Channel Model for Outdoor to Indoor and Pedestrian Test Environment. Since the delay spread can vary significantly, the ITU recommendation specifies two different delay spreads for each test environment: low delay spread (channel model A), and medium delay spread (channel model B). The parameters of these channel's model are grouped in Table 3. In this paper we focus on the channel model B.

5.1 feedback overhead comparison

Figures 2 to 6 show the evolution of the total number of feedback complex coefficients of the proposed time domain quantization according to the signal to noise ratio (SNR) for MIMO 2×2, 3×3, 4×4, 6×6 and 8×8 configurations. For each MIMO configuration, simulations are performed for the 4 existing bandwidth of the IEEE 802.11 ac standard (20 MHz,40 MHz,80 MHz and 160 MHz). Note that the total number of feedback complex coefficients represents the sum of all the remaining samples and it can be calculated by $N_{rs} * N_t^2$ assuming that $N_r = N_t$.

As can be seen from these figures:

- whatever the MIMO configuration, the total number of feedback complex coefficients increases according to the bandwidth and the number of antennas.

- whatever the SNR, the plotted cumulative distribution function (CDF) shows that it is statistically possible for each MIMO configuration and bandwidth to know the maximum value of the total number of feedback complex coefficients. Following this study the total number of feedback complex coefficients for any configurations ($N_t \times N_r$ and bandwidth) in IEEE 802.11ac ontext are grouped in the table 4.

Therefore, according to equation (17) the feedback overhead length of the proposed time domaine quatization (in bit) noted by N_{FTD} can be calculated using $N_{FTD} = N_{rs} * 2 * nb * N_t^2$ and the number of remaining samples N_{rs} in the Table 4.

Except for MIMO 2×2 configuration and bandwidth $20MHz$, the feedback overhead of the proposed time domain quantization is much shorter than that of the Givens Rotation quantization. The compression ratio of the proposed time domain quantization compared to Givens Rotation quantization can be up to almost 70%.

The most important to note here is the fact that unlike the Givens Rotation quantization for which the compression ratio decreases when the number of antennas becomes large (see section 3.2), the proposed time domain compression becomes more efficient. This last remark makes the proposed time domain compression very interesting for the next WLAN generation which will undoubtedly consider the massive MIMO concept which is more and more studied for WLAN [19-20].

Table 3: ITU Channel Model for Outdoor to Indoor and Pedestrian Test Environment.

Tap	Channel A		Channel B	
	Relative delay(ns)	Average power (dB)	Relative delay(ns)	Average power (dB)
1	0	0	0	0
2	110	−9.7	220	−0.9
3	190	−19.2	800	−4.9
4	410	−22.8	1200	−8
5			2300	−7.8
6			3700	−23.9

Table 4: Total number of feedback complex coefficients of the proposed time domain quantization for any configurations ($N_t \times N_r$ and bandwidth) in IEEE 802.11ac ontext.

MIMO configuration	feedback overhead N_{FTR}			
	20 MHz	40 MHz	80 MHz	160 MHz
2x2	55	90	161	293
3x3	111	192	334	613
4x4	191	323	582	1069
6x6	415	703	1254	2337
8x8	718	1215	2205	4106

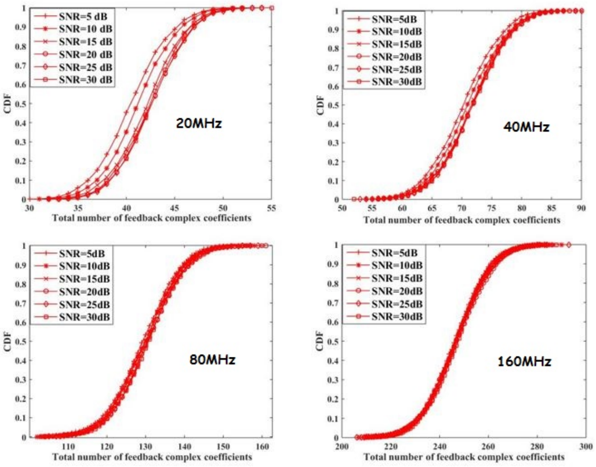


Figure 2: The Cumulative Distribution Function of the total number of feedback complex coefficients according to the Signal to Noise ratio and Bandwidth for MIMO 2 × 2 configutation

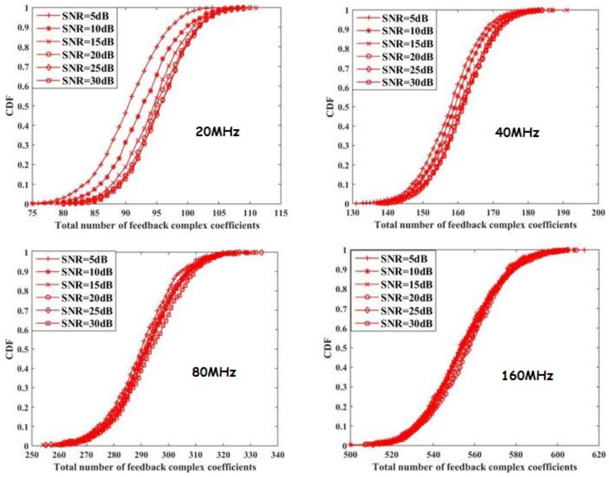


Figure 3: The Cumulative Distribution Function of the total number of feedback complex coefficients according to the Signal to Noise ratio and Bandwidth for MIMO 3 × 3 configutation

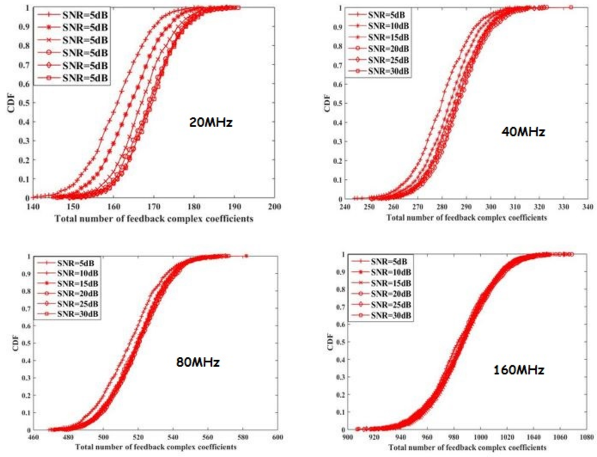


Figure 4: The Cumulative Distribution Function of the total number of feedback complex coefficients according to the Signal to Noise ratio and Bandwidth for MIMO 4 × 4 configutation

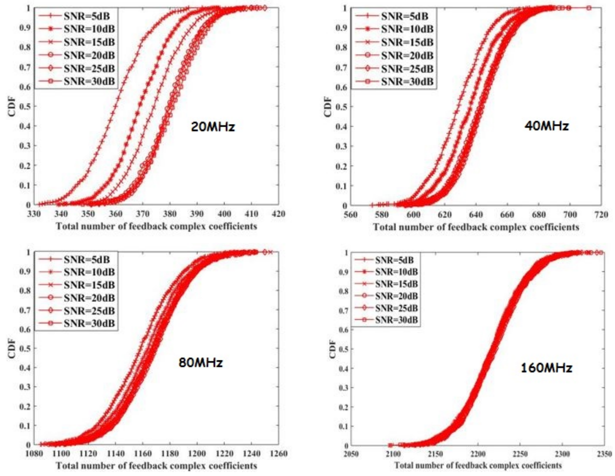


Figure 5: The Cumulative Distribution Function of the total number of feedback complex coefficients according to the Signal to Noise ratio and Bandwidth for MIMO 6 × 6 configutation

5.2 BER performance

In this subsection, we compare the beamforming with Givens rotation quantization (GR-Q) and that with the time domain quantization in terms of bit error rate (BER) performance. The considered simulation parameters are listed in Table 6.

Table 5: Feedback overhead (in bit) comparison between Givens Rotation quantization and the proposed time domain quatization for any configurations ($N_t \times N_r$ and bandwidth).

MIMO configuration	feedback overhead N_{FGR} for Givens Rotation quantization				feedback overhead N_{FTR} for the proposed time domain quatization				Ratio	
	20 MHz	40 MHz	80 MHz	160 MHz	20 MHz	40 MHz	80 MHz	160 MHz	min	max
2x2	1040	2160	4680	9720	1100	1800	3220	5860	−0.5%	39.7%
3x3	3120	6480	14042	29160	2220	3840	6680	12260	28.8%	58%
4x4	6240	12960	28080	58320	3820	6460	11640	21380	38.7%	63.3%
6x6	15600	32400	70200	145800	8300	14060	25080	46740	46.8%	68%
8x8	29120	60480	131040	271410	14360	24300	44100	82120	50.6%	69.7%

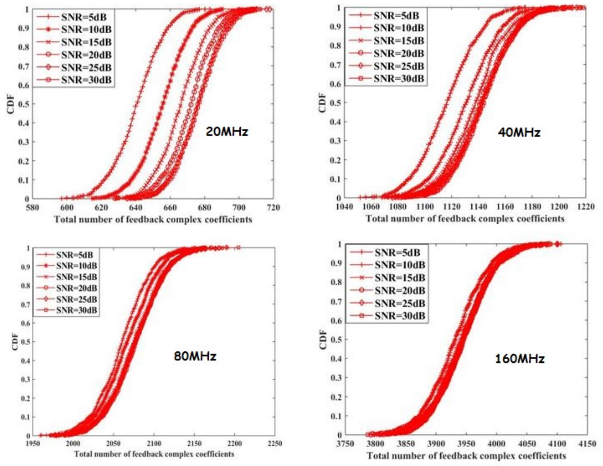


Figure 6: The Cumulative Distribution Function of the total number of feedback complex coefficients according to the Signal to Noise ratio and Bandwidth for MIMO 8 × 8 configutation

Table 6: Simulation Parameters

Channel Model	ITU Channel Model
Bandwidth	40MHz
Number of FFT points (N)	128
Number used subcarriers	108
Number of bits per angle	10
cyclic prefix	*Long ie* 32 samples
Number of $N_t \times N_r$ antennas	2×2 and 4×4
Modulation	QPSK and 64-QAM
FEC	Convolutional
Coding Rate	1/2

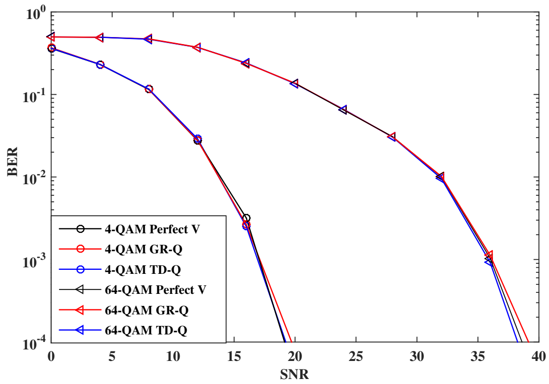


Figure 7: BER performance for perfect beamforming (Perfect V), beamforming with GR-Q and beamforming with the prposed time domain quantization (TD-Q) in MIMO 2 × 2 context. The ITU Channel Model is considered for simulations.

The BER performance is shown in Figures 7 and 8 for 4–QAM and 64–QAM in MIMO 2 × 2 and MIMO 4 × 4 context respectively. The number of bits per angle feedback is taken to 10. It can be seen that the proposed TD-Q presents the same performance, for all considered configurations in these simulations, than the GR-Q which is adopted in IEEE 802.11ac standard. As considered in the 802.11ac standard, the simulations confirm that 10 bits per angle feedback minimizes the quantization error and allows the quantized beamforming (GR-Q and TD-Q) to have the same performance than the perfect one.

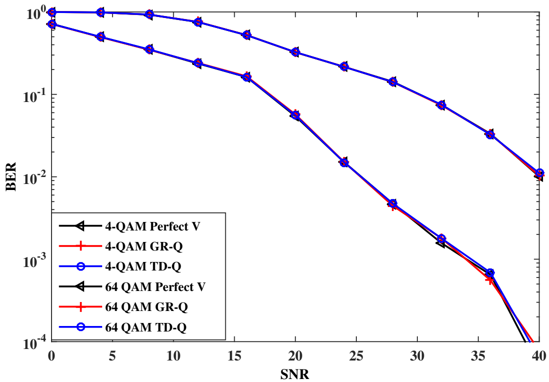


Figure 8: BER performance for perfect beamforming (Perfect V), beamforming with GR-Q and beamforming with the prposed time domain quantization (TD-Q) in MIMO 4 × 4 context. The ITU Channel Model is considered for simulations.

6 Conclusion

The proposed TD-Q requires less amount of feedback than the GR-Q one in IEEE 802.11ac TxBF mode. In addition, simulations in 802.11ac context with standard TGn channel model show that the TD-Q beamforming allows to have the same performance than the adopted GR-Q. The highlight of this proposal is that the proposed TD-Q can be considered as a credible alternative to the GR-Q for the IEEE 802.11ac evolution. The proposed quantization technic is not only for WLAN systems. It can be adapted for all SISO or MIMO system using closed loop beamforming.

References

1. Q.Cui, H. Wang, P. Hu, X. Tao, P.Zhang, J. Hamalainen and L. Xia, "Evolution of limited feedback CoMP systems from 4G

to 5G " IEEE vehicular technologie magazine, **9**(3), 94–103, 2014. DOI: 10.1109/MVT.2014.2334451

2. "IEEE P802.11ac/D0.2" IEEE standard Draft, 2011.
3. "Mobile WiMAX Part II: A Comparative Analysis" IWiMAX Forum, 2006
4. D. Wang, Y. Cao and L. Zheng, "Efficient Two-Stage Discrete Bit-Loading Algorithms for OFDM Systems" IEEE Trans. on Vehicular Technology, **59**(7), 3407–3416, 2010. DOI: 10.1109/TVT.2010.2052937
5. T. Ramji, B. Ramkumar and M. S. Manikandan "Resource and subcarriers allocation for OFDMA based wireless distributed computing system" IEEE Intern Adv Comp. Conf. 2014. DOI: 10.1109/IAdCC.2014.6779345
6. A.Goldsmith, S. A. Jafar, A. Jindal and S. Vishwanath "ACapacity limits of MIMO Channels" IEEE J. Select. Areas Commun **21**, 684-702, 2003.
7. H. Kim, Y. Jung, J. Park and J. Kim "Adaptative CSI Feedback Scheme to maximize the throughput in IEEE 802.11ac system," IEEE ISCE 1-2, Jeju, Korea, June 2014.
8. M. Mbaye, M. diallo and M. Mboup"LU based Beamforming schemes for MIMO systems" IEEE Trans. on Vehicular Technology, **66**(3), 1-9, 2016 DOI: 10.1109/TVT.2016.2573046.
9. T. Inoue and R. W. Heath Jr"Kerdock codes for limited feedback precoded MIMO systems" IEEE Transactions on Signal Processing, **57**(9), 37113716, 2009.
10. J. C. Roh and B. D. Rao"Design and analysis of MIMO spatial multiplexing systems with quantized feedback" IEEE Transactions on Signal Processing, **54**(8), 28742886, 2006.
11. D. J. Love and R. W. Heath Jr"Limited feedback unitary precoding for spatial multiplexing systems" IEEE Transactions on Information Theory, **51**(8), 29672976, 2005.
12. K. Schober, R. Wichman, and T. Koivisto"MIMO adaptive codebook for closely spaced antenna arrays" European Signal Processing Conference, 1-5, 2011.
13. D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig"QAM codebooks for low-complexity limited feedback MIMO beamforming" IEEE International Conference on Communications, 41624167, 2007.
14. LTE Physical layer procedures "Evolved Universal Terrestrial Radio Access (E-UTRA)" 3GPP TS 36.213 version 13.4.0 Release 13, 2017.
15. G. Berardinelli and T. B. Sorensen and P. Mogensen and K. Pajukoski "SVD-Based vs. Release 8 Codebooks for Single User MIMO LTE-A Uplink" IEEE 71st Vehicular Technology Conference, 1-5, 2010, DOI: 10.1109/VETECS.2010.5493709.
16. Zheng Jiang; Bin Han; Liang Lin; Peng Chen; Fengyi Yang; Qi Bi "On compressive CSI feedback beamforming scheme for FDD massive MIMO" IEEE International Conference on Communications in China, 1-5, 2016.
17. C. Yuen, S. Sun, M. Ho and Z. Zhang "Beamforming matrix quantization with variable feedback rate" EURASIP journal on wireless communications and networking, 2012.
18. J. Kim and C. Aldana"Efficient feedback of the channel information for closedloop beamforming in WLAN" IEEE VTC-spring, pp. 2227-2230, Melbourne, 2006.
19. Y. Morino, T. Hiraguri, H. Yoshino and K. Nishimori"Proposal of overhead-less access control scheme for multi-beam massive MIMO transmission in WLAN systems" Med-Hoc-Net, pp. 1-5, Budva, 2017.
20. J. Lee, K. J. Choi and K. S. Kim"Massive MIMO full-duplex for high-efficiency next generation WLAN systems" International Conference on Information and Communication Technology Convergence, pp. 1152-1154, Jeju, 2016.
21. Moussa Diallo; Laurent Boher; Rodrigue Rabineau; Laurent Cariou; Maryline Helard "Transform domain channel estimation with null subcarriers for MIMO-OFDM systems" IEEE International Symposium on Wireless Communication Systemse, pp. 209-213, Reykjavik, 2008.
22. Zhi Zheng, Caiyong Hao and Xuemin Yang"Least squares channel estimation with noise suppression for OFDM systems" IEEE ELECTRONICS LETTERS, Vol. 52 No. 1, pp. 3739, 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 146-156 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Limitations of HVAC Offshore Cables in Large Scale Offshore Wind Farm Applications

Tiago Antunes[*,1], Tiago Alexandre dos Reis Antunes[1], Paulo Jorge da Costa Santos[2], Armando José Pinheiro Marques Pires[3]

[1]Siemens, S.A. & MIT Portugal Program, Sustainable Energy Systems, Universidade Técnica de Lisboa, 2744-016, Portugal

[2]EST Setúbal Instituto Politécnico de Setúbal / INESC Coimbra, Electrical Engineering, 2914-76, Portugal

[3]EST Setúbal Instituto Politécnico de Setúbal / UNINOVA, Electrical Engineering, 2914-76, Portugal

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 24 January, 2018
Online: 17 March, 2018

Keywords:
Alternate Current
Transmission Line Simulation
Offshore Wind Farms

ABSTRACT

The energy marathon is becoming increasingly based on renewable sources, whereas the continuous decrease on the cost of energy production has supported in last decade, for instance, the development of large-scale offshore wind applications. The consistency and availability of the AC-working equipment portfolio is only limited by physical application boundaries, which are quite evident in this sort of accomplishments. The focus of this article is to present a tool developed under the MATLAB environment which allows for a quick and real-time analysis of HVAC links, assessing the impact of voltage, current, power factor or distance conditions. The conclusions are drawn directly by means of a stress-point and operational acceptance range.

1. Introduction

This paper is an extension of work originally presented in 2017, 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG) [1]. Offshore wind applications are clearly under the spotlight. The wind conditions, the ability to increase the rated power of the turbines and thus the size of the plant, in addition to the operational and technical advantages of having the units dispersed over the sea, provide the arguments to position it as one of the viable alternatives for the energy demands of the future [2]. However, there are known limitations for the Alternate Current (AC) Transmission Links.

As wind turbines now compete with gas turbines – in rated power – and the cost of production is consistently approaching the one of traditional carbon-based plants [3], innovation pushes the offshore power flow to the MW-unit level. That, added to the long distance the submarine cables should meet, reduces, in a sense, the reliability of traditional AC solutions [4]. This is the starting-point of this investigation. Would that information provide the basics, the goal is clear – the model aims to provide the operational range and the restrictions for any given AC link (meaning a set of initial conditions). Moreover, that was proven based on the results obtained shown later.

AC-based plant-to-shore links do not provide a one-size fits-all option and a careful analysis still is mandatory to ensure the proper conversion of the bulk wind onto usable power preferably with limited losses.

Although it is common knowledge that current HVAC Overhead Lines (OHL) can easily be deployed in distances in upwards of 100km, the same is not true for insulated-cable based lines, especially in offshore applications, which are based in submarine cables. For both of these cable types, their fabrication characteristics – hence the electromagnetic characteristics seriously restrict the application of AC on the offshore [5]. Once more, the goal of this paper is to present a model that allows for a reliable estimation of the maximum operational distance in such regard without requiring a time-consuming case-by-case investigation and still providing detailed results.

This paper divides in seven sections. The first part provides a description of the wind farm (WF) model under appreciation, which constitutes the baseline for the operation of the model. The intention is to provide a set of results, for equivalent air and cable-based lines, and ultimately, also to allow for an extrapolation for other configurations. The transmission-line equivalent and the GUI-based analysis models are presented in sections 2 and 3, respectively.

Sections 4 to 6 present the experimental results of the analysis performed, for standard OHL submarine cables, as well as, for

[*] Tiago Antunes, Email: tiago.antunes@siemens.com

https://dx.doi.org/10.25046/aj030217

offshore compensated-cable solutions. The last chapter includes the conclusions.

2. Simplified Model of the Offshore Wind Farm

2.1. Characteristics of the Wind Farm

This section consists of two main parts; The first one aims to present the simplified model of the wind farm under consideration, whereas the second one introduces the line model, later on used on the MATLAB-based software (section 3). Based on actual available portfolio [5] and existing projects [6], a decision was made to assess the link of an 80MW offshore wind cluster, composed by ten 8MW turbines. The schematic of the model is shown on Figure 1.

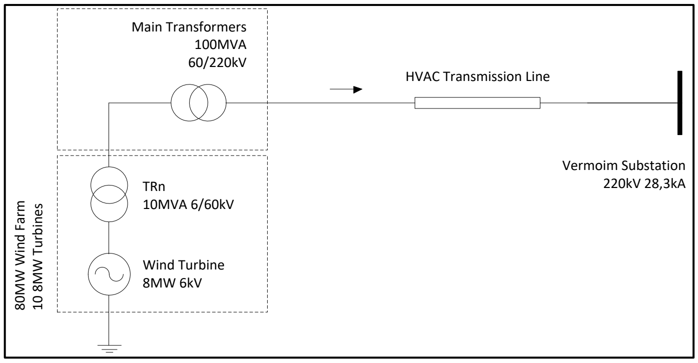


Figure 1. Layout of the Wind Farm

As shown on the figure above, the authors chose to use a 220kV link to the shore, given the rated power of the wind farm and the existing transmission infrastructure on that region. REN' Vermoim Substation located approximately 40km from the district capital Oporto was the onshore Point of Connection (PoC). The tables 1 and 2 present, respectively, the characteristics of the wind farm and the ratings at the given substation (at the 220kV busbar).

Table 1. Characteristics of the Wind Farm

Nr. Units	$S_{N\text{-Turb}}$ [MW]	$S_{N\text{-Farm}}$ [MW]	L_{Link} [km]
10	8	80	200

Table 2. Short-circuit ratings at Vermoim 220kV Busbar

$S_{CC\ MAX}$ [MW]	$S_{CC\ MIN}$ [MW]	$I_{CC\ MAX}$ [kA]	$I_{CC\ MIN}$ [kA]
10.7	8.3	28.3	21.7

2.2. Characteristics of the HVAC Link

The usage of insulated cables as the means to transmit power in HVAC is understandingly common. Nonetheless, offshore (submarine) applications have a different behavior that ultimately and most importantly, significantly reduce the acceptable distance of application. Limitations, such as high charging current and higher *Ferranti* effect [8], are widely known, the model must allow for the analysis of several scenarios, with real-time adjustment and performance assessment, still ensuring a less time-consuming computation of each one.

There are two main cable-sizing criteria, which are the rated voltage and admissible temperature [8] and the model is based on this simplification.

The nominal current is given directly by the rating of the wind farm whereas the fault current value is given at the HV OHL-bay at the onshore substation as it is the most significant contributor in case of short-circuit at the PoC. Both allow for an estimation of the minimum cable cross-section [7]:

$$I_N = \frac{P_N}{\sqrt{3}\,V_N \cos\varphi} = 262{,}43 \text{ A} \tag{1}$$

$$s_{min} = \frac{I_{CC}\sqrt{t}}{K_c} = 521{,}3 \text{ mm}^2 \tag{2}$$

Where *t* represents the fault-clearance time.

Having this as a reference, and given the cable standardization, a core section of 630mm^2 shall be selected. The OHL and cable will be selected having such as a reference. The upstream grid is not taken into consideration on the analysis (meaning the onshore PoC is assumed a PV-type node).

3. Software Implementation

3.1. Electrical Line-equivalent Model

When in comparison between lumped-sum or parabolic line-equivalent models, the differences are significant. In the case of OHL, the parabolic components of the equivalent are almost unitary, which allows for simplification and a lump-sum provides acceptable results [8].

In cable applications in general, the wave impedance is higher and as so the propagation constant decreases, resulting in non-unitary trigonometric components. These models compute the opposite-end results based on a linear and concentrated impedance section, the propagation speed is not taken into account (which drives from the simplification itself).

In order to proper translate the effects of submarine cable applications, the model has to take into account the influence of the propagation speed and the geometrical distance. An accurate model has to take effectively into account the distribution and variation of the parameters along the line, being therefore distance and transmission velocity- dependent (which lastly decreases for a higher cable length).

This knowledge is widely applied in very-fast transient analysis for atmospheric discharges or other high-velocity voltage and/or current variations on transmission lines. An *x* section of the line is presented in Figure 2.

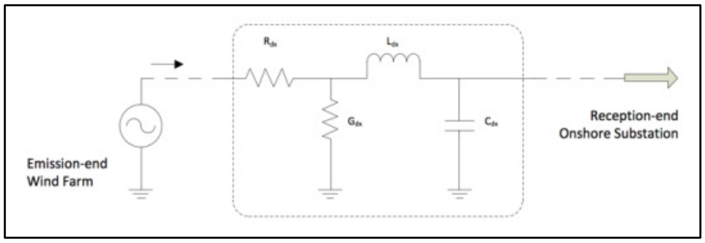


Figure 2. Infinitesimal section of the cable ([9], adapted)

As the results of that equivalent scheme are much more accurate, the respective equations (3) to (6) are used on the MATLAB model. These, on the first approach, provide a distance-based analysis of a given transmission link in regards to estimated voltage and current:

$$Z_L = (R + jX)\, l \text{ [ad]} \tag{3}$$

$$Y_T = (G + jB)\, l \text{ [ad]} \tag{4}$$

The voltage and current for *x* distance are given by:

$$V_x = V_1 \cosh(\gamma x) - Z_0\, I_1 \sinh(\gamma x) \text{ [kV]} \tag{5}$$

$$I_x = -V_1 \frac{1}{Z_0} \sinh(\gamma x) + I_1 \cosh(\gamma x) \text{ [A]} \tag{6}$$

All data shown on the GUI (hence plot options) are also based on a set of calculations detailed on the script. The calculations are performed for any given initial conditions, non-related to any previous tests – although a multi-set analysis can be provided, as shown on sections 4 and 5.

3.2. Model Design

The necessity to address a multitude of scenarios to draft earlier conclusions has led itself to the conclusion that a software, rather than casual calculation, had to be taken into practice. The development of a fast and custom-made tool to the analysis the design of HV transmission lines led the authors almost directly to the implementation of a GUI-based MATLAB model.

That allowed for the creation of a versatile and digital tool, which allows for a fast and straightforward verification, given a set of any initial conditions, if power transmission via a selected line (or power cable) is acceptable in technical terms. And, the model spans virtually to any range of, not only cables, but voltage, current, power factor or distance.

The possibility of checking a certain scenario in real-time and in on-hand is another plus, possible via the multitude of adjustments possible on the model itself, with instantaneous calculation and results. As the GUIDE is, basically, the user-friendly window of MATLAB C-based programming, also quite an user-friendly design was achieved.

In practical terms, the model operates by taking into account a set of transmission line conditions – stored on pre-loaded, yet adjustable .XLS sheet, of OHL and cables, then applies the equivalent model earlier addressed (section 3), and computes all the results that illustrate the steady-state operation of a given transmission "line". That includes voltage, current, no-load voltage & current, voltage drop, efficiency (losses evaluation), power factor, stability, surge impedance level (SIL) and model parameters (Z_L & Y_T).

3.3. Model Key Advantages

The segmentation and expandability of the model can be defined as two of the most important characteristics. All the information used and key features, aside from cables, are actually .XLS-based. The model starts and loads-up the majority of its functionalities with reference to the options and fields set on the configuration Excel sheets.

Features such as the stress parameters used for the calculation – further on shown – or the actual regulations, such as voltage or reactive power boundaries, can be adjusted with ease. The model also loads using acceptable and recommended voltage, current, power factor and distance values, further facilitating the usage. The plot section, on the other hand, allows for the simultaneous visualization of three different calculation windows, thus allowing the user do address a scenario in a much more complete manner.

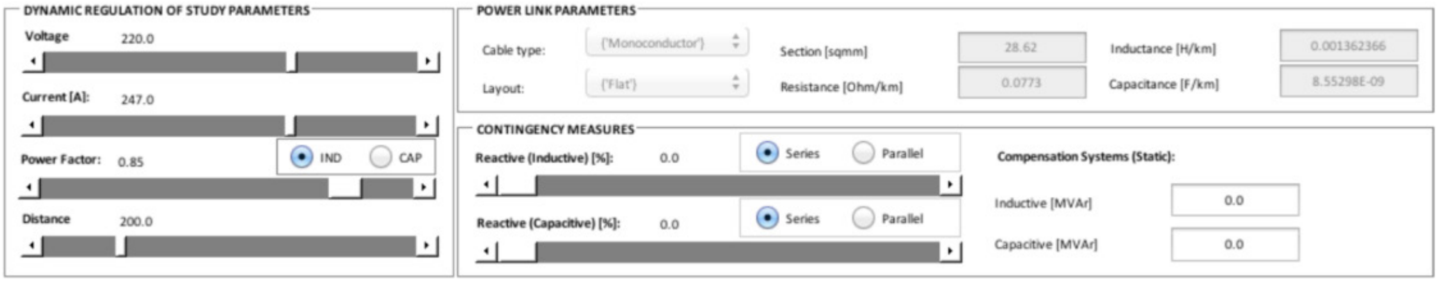
Finally, the information provided has a key focus. As said, the ability to implement the line-equivalent is wide-spread in literature. The difference is to provide such information and the data outcomes in simpler yet much more complete manner. That information is provided to the user using overlay layers on the results plotted, and may include operational boundaries, compensation expectations, stress-range distance pin-pointing, or even, simultaneous calculation results – for example, for difference one-end power factors. The summary of the available results is shown on Table 3.

Table 3. Available Results of the Model

Ref	Description
1	Powerline Voltage
2	Powerline Current
3	No-load Current
4	No-load Voltage
5	Voltage Angle
6	Current Angle
7	Voltage Phase Displacement
8	Current Phase Displacement
9	Active Power
10	Reactive Power
11	Apparent Power
12	Transmission Losses
13	Transmission Efficiency
14	Voltage Drop
15	Power Factor
16	Tan-Phi
17	Stability Limits (Pe-Phi)
18	Voltage Drop (DV-PF)
19	SIL (impedance)
20	SIL (apparent power)
21	Longitudinal Impedance
22	Transverse Admittance

The critical factors that guide the operational ranges are shown in Table 5. Generically (and, in fact changeable), the critical factors are presented based on two conditions: "C" (meaning Critical, for 80% of target) and "F" (for Failure, or 100% of target value). Other conditions may apply as mention on Table 5. The numbers refer to the references and 80/100 thresholds can be adjusted as the intention of the user with minor core programming adaptation.

Once again, model design is as simple yet flexible and expandable as possible. The majority of the reference and configuration-data is XLS-based and new features are easy to implement. The same can be done for the (line) data, as the routine reads any given organized sheet and performs the calculation. After the automated analysis is complete, the user can make a search, distance or variable-dependent, for a value-at-point or given location of the "line".



DYNAMIC REGULATION OF STUDY PARAMETERS

Voltage 220.0

Current [A]: 247.0

Power Factor: 0.85 ● IND ○ CAP

Distance 200.0

POWER LINK PARAMETERS

Cable type: {'Monoconductor'}

Layout: {'Flat'}

Section [sqmm] 28.62

Resistance [Ohm/km] 0.0773

Inductance [H/km] 0.001362366

Capacitance [F/km] 8.55298E-09

CONTINGENCY MEASURES

Reactive (Inductive) [%]: 0.0 ● Series ○ Parallel

Reactive (Capacitive) [%]: 0.0 ● Series ○ Parallel

Compensation Systems (Static):

Inductive [MVAr] 0.0

Capacitive [MVAr] 0.0

As observed in Figure 3, the real-time interface is possible via the adjustable sliders – which adapt the basic parameters. Re-calculation is done fast and automatically after any change on those. The advantages of a tailor-made GUI model span for other areas. The plotting windows included have been adapted to allow for the *online* setting of the units to be used: International system (SI) units, per-unit (P.U.) or %-based results.

The adjustment of the ranges and precision of the axis is adjusted directly on each respective axis line. Exporting is also made directly from the working plotting window, allowing for quick integration on written reports. This is, once more, allowed by a strong background interface with .XLS files, and further on, with facilities such as additional interfaces, adaptation, parameter-variation mode and export facilities, are all intricate on the model.

Table 4. Critical Factor References for Model

Ref	Variable Checked	Slope	Critical Value	Failure Value
1-4	Line Voltage	>=	REN regulations	
5-8	Line Voltage	<=	REN regulations	
9	Line Current	>=	Thermal limit of cable	
10-13	No-load Voltage	>=	REN regulations	
14-17	No-load Voltage	<=	REN regulations	
18	No-load Current	>=	50%	100%
19	Reactive Power	>=	40%	50%
20	Losses	>=	40%	50%
21-22	Voltage Angle	>=	±30º	±40º
1-4	Line Voltage	>=	REN regulations	
5-8	Line Voltage	<=	REN regulations	
9	Line Current	>=	Thermal limit of cable	
10-13	No-load Voltage	>=	REN regulations	
14-17	No-load Voltage	<=	REN regulations	

What is also worth mentioning is the ability to automatically see the stress ranges on a color-coded plot-overlay for any given calculation. These ranges are recognized and the maximum operating distances are identified. Ranges are loaded based on the Portuguese regulations (e.g. REN) and those can also be adjusted.

For a comparison to be made on different transmission solutions a simulation is made on a 220kV air-insulated single-circuit single-pole transmission line. The rated voltage is selected taking once again into account the grid topology at the PoC.

4. Evaluation of Air-Insulated Connection

4.1. Parameter Calculation

The overall manufacturer-issued characteristics of such cable, later on introduced on the MATLAB model, are shown on Table 5. The maximum active power output is stationed at 80MW (having a cos φ of 0,85 lag). The matrix therefore can be calculated as presented in equations (7) and (8).

Table 5. Characteristics of the simulated Overhead Line

U_N [MW]	r [mΩ/km]	L [mH/km]	C [nF/km]
220.0	77.3	1.362	8.553

$$Z_L = (R + jX)\,l = 10.381 + \mathrm{j}26.3894\ [\mathrm{ad}] \tag{7}$$

$$Y_T = (G + jB)\,l = \mathrm{j}10.1 \times 10^{-3}\ [\mathrm{ad}] \tag{8}$$

These values are automatically calculated on the model.

4.2. Experimental Results

Computation is done assuming the reference data at the reception-end (onshore PoC). The simulation of the voltage and current at the emission-end of the line are shown on Figures 4 and 5. Ultimately, the results confirm the reliability of the model with existing accurate OHL-analysis data [9]. The calculation method is than transposed to an offshore (submarine) framework, as presented on section 5.

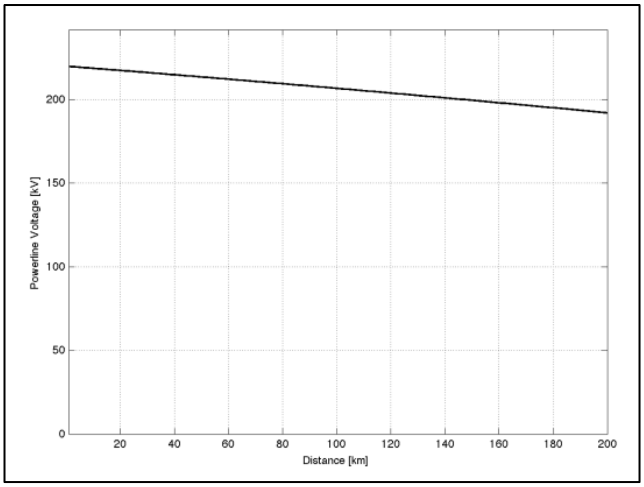


Figure 4. Powerline voltage simulation (220kV OHL)

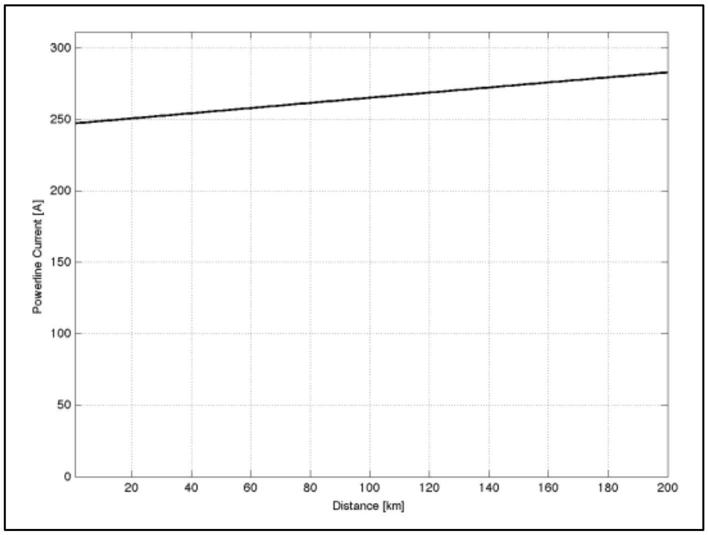


Figure 5. Powerline current simulation (220kV OHL)

The analysis of the submarine cable, being the focus, shall include much more significant data, however all being supported on the positive result of the voltage & current simulation shown on these two test plots.

5. Evaluation of Submarine Cable Connection

5.1. Parameter Calculation

The analysis of the submarine cable, hence parameter calculation, is done on the same basis as before. The cable data obtained from the manufacturer (ABB) is presented in Table 6. Cables are laid in a single pole linear and evenly distributed along the seabed.

Table 6. Characteristics of the simulated Submarine Cable

V_N [kV]	I_N [A]	Size [mm2]	R [Ω/km]	L [mH/km]	C [μF/km]
220	600	630	0.04935	1.4000	0.1600

$$Z_L = (R + jX)\mathrm{l} = 9.87 + \mathrm{j}87.9646\ [\mathrm{ad}] \tag{9}$$

$$Y_T = (G + jB) = \mathrm{j}10.1 \times 10^3[\mathrm{ad}] \tag{10}$$

Once again, these are automatically calculated on the model. Equations (5) and (6) apply for the calculation of all data plotted, including the calculation of the ABCD equivalent model parameters.

5.2. Voltage & Current Results

The operation, critical failure ranges can be seen on the figures shown. The reference for criticality/failure leads back to the REN regulations on the voltage boundaries. The same is shown on the current plot. A distance shorter than 100km is also sufficient to show the relevant information yet that range facilitates the condensation of data on the plot snapshot.

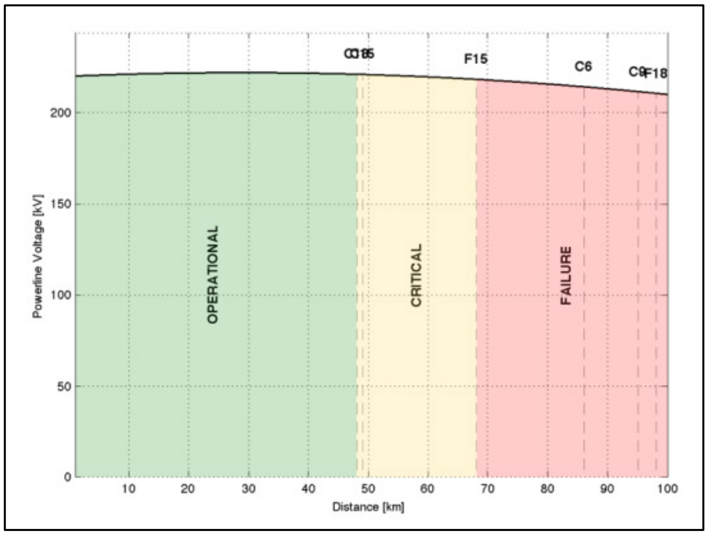


Figure 6. Powerline voltage simulation (220kV Submarine Cable)

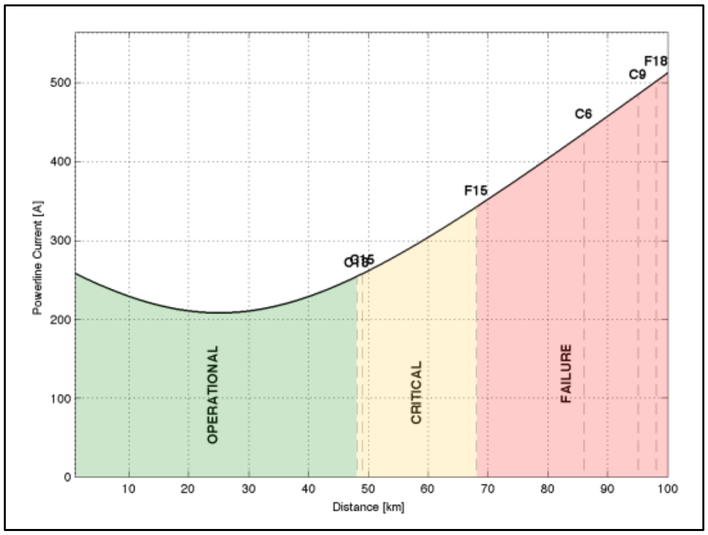


Figure 7. Powerline current simulation (220kV Submarine Cable)

The emission-end (WF) values are 210kV (Fig. 6) and 513A (Fig. 7). As it can be verified, the voltage drop reaches close to 5%. As for the current, it's final value is around two times the initial current value at the reception. This allows for an insight of one of the most important downsides of cables – the charging effect.

As far as the current values are concerned, the high capacitive characteristics of the submarine cable, lead to a significant increase of the line' current and emphasize the reactive power flow. This not only increases the power losses along the line but also confirms that cable solutions are not feasible after 30-50km [10] alongside theoretical information available.

5.3. Voltage Drop on Underground & Submarine Cables

As a basis of comparison, Figures 8 and 9 include the voltage drop along transmission cables using underground (8) and submarine (9) solutions. Evidently, as expected, the reduction of the voltage is much more significant on the second option.

The authors emphasize that, on both land and sea cables, the voltage drop encountered is actually a voltage rise along the line, due to the capacitive characteristics of the cables themselves. This has to be later on compensated on the WF-side, forcing the generators to operate under very low voltages.

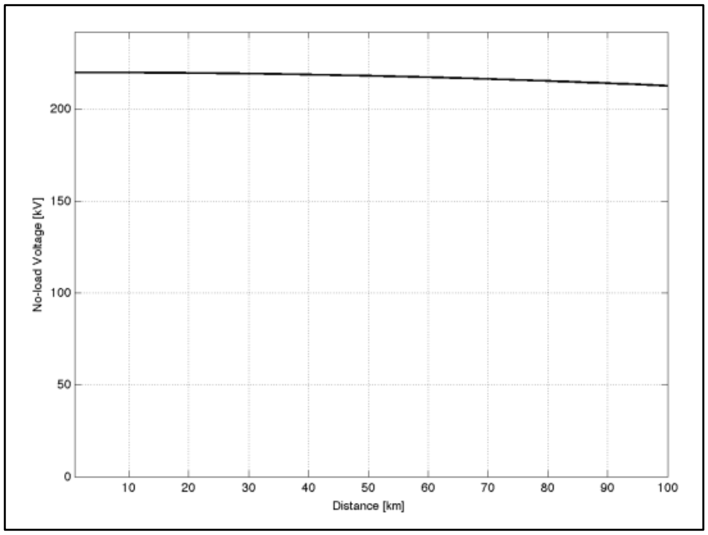


Figure 8. No-load voltage simulation (220kV Underground Cable)

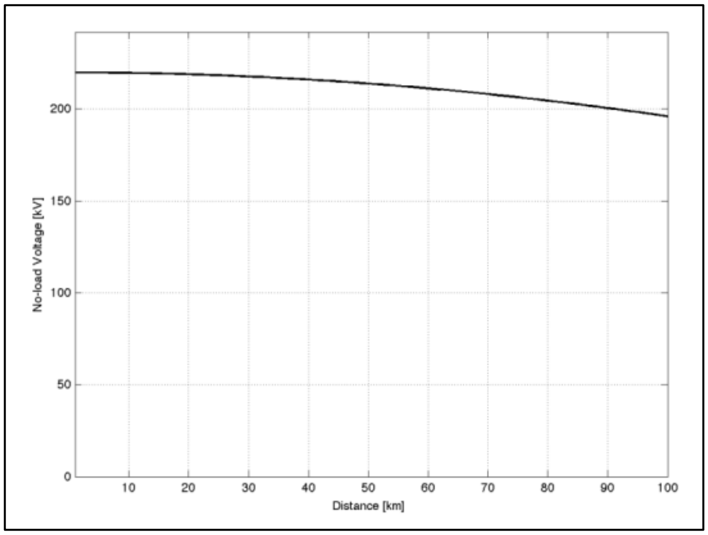


Figure 9. No-load voltage simulation (220kV Submarine Cable)

The inductance value has a tremendous influence on the voltage and current behaviour of the transmission line, but with more focus on the first one mentioned. Aside from the voltage decline – and as addressed earlier – with a higher impact, the charging and Ferranti effects are evaluated for this particular cable.

5.4. Charging Current & Ferranti Effect

The charging and Ferranti effects are the two most distinguishable plots which can be made from the behavior of insulated cables (both underground and submarine. The high capacitance of these cables leeds to a early increase on the current flow and therefore the transmittable power is severely restricted. No-load current, based on the F18 marker on the plot, is actually the failure cause of the cable under investigation.

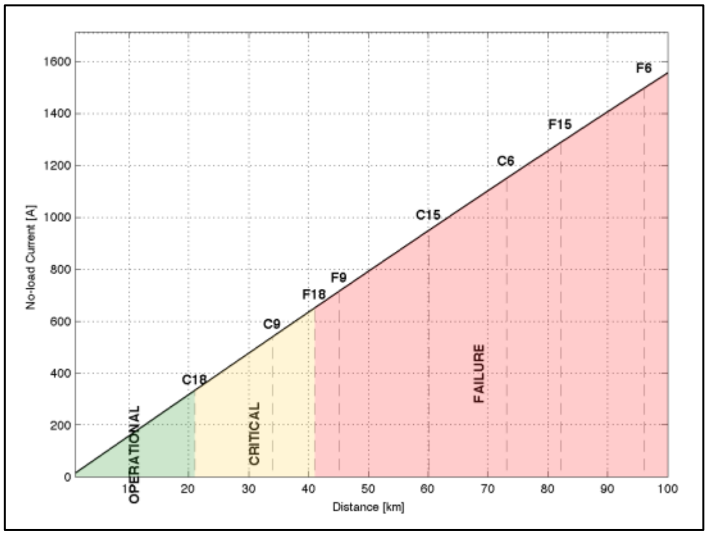


Figure 10. No-load current simulation (220kV Submarine Cable)

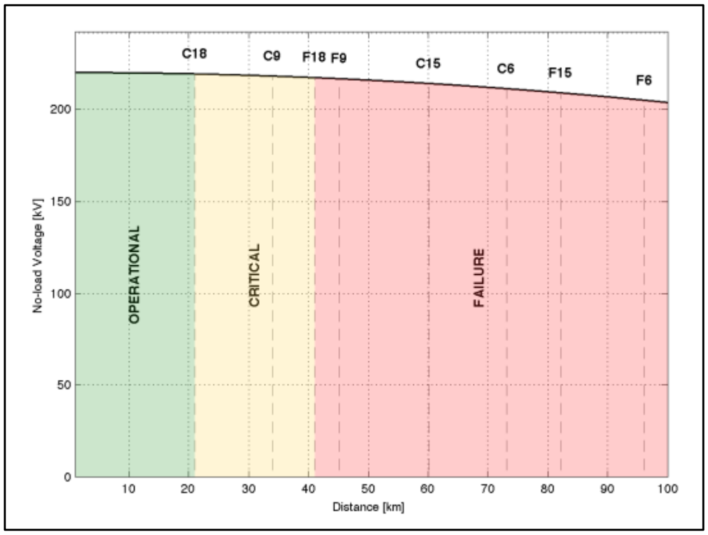


Figure 11. No-load voltage simulation (220kV Submarine Cable)

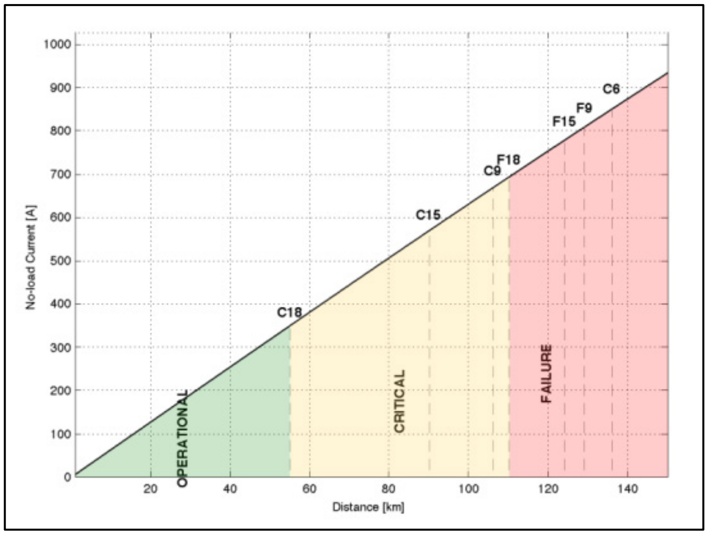


Figure 12. No-load voltage simulation (220kV Submarine Cable)

When compared to the equivalent underground cable included on the examples of the model, the differences are tremendous. The effect on the submarine option is almost 100% more. The operation of such "large capacitor" under no-load conditions also translates onto a voltage increase along the same cable – as shown previously on Figure 9.

5.5. Power Factor & Reactive Power Results

The behaviour of the power factor for a given section of the transmission cable is highly dependent on the initial conditions, especially the reception-end (PoC) power factor. For that reason, this section also includes a multi-scenario analysis, thus allowing for a better understanding of such impact.

A very high drop on the power factor is observed (close to 50% of the initial value), to a minimum of 0,44. Due to the increasing current, the active power still increases. However, it is the reactive power which increases the most, leading again to the failure of this setup for more than 70 kilometres (F15).

What is also verified is that, up to a certain distance, the capacitive characteristics of the line help to compensate for the inductive grade of the load (0,8 lag). Once the characteristics of the transmission line match the PoC, there is an inversion on the slope.

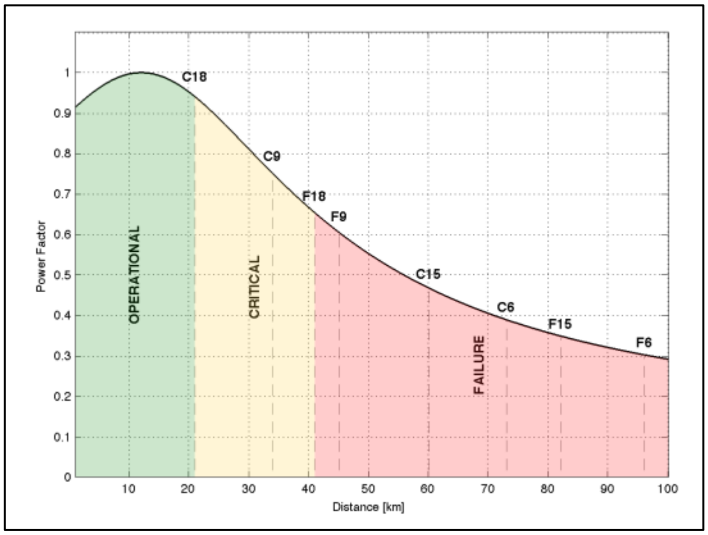


Figure 13. Power factor simulation (220kV Submarine Cable)

What is also verified is that, up to a certain distance, the capacitive characteristics of the line help to compensate for the inductive grade of the load (0,8 lag). Once the characteristics of the transmission line match the PoC, there is an inversion on the slope. After a certain distance (around 25km), the cable is actually acting as a reactive power provider. This fault is not triggered because, in percentage, it is still below the 50% threshold shown on Table 4.

What can also be understood from Fig. 8 is that the unitary power factor is obtained for a distance just under 15km. Despite failure only occurs after 40km (once more confirmed by literature), because the line has a significantly higher capacitance than the OHL, it acts much like a very large capacitor, consuming a highly capacitive current during operation. After 100km, the current required is about the same as the rating of the load (Figure. 7).

These results are quite conclusive on the balance that must be met. They also provide a further understanding on how much the compensation solutions must be tailor-made and real-time adjustible to proper ensure the safety of the link.

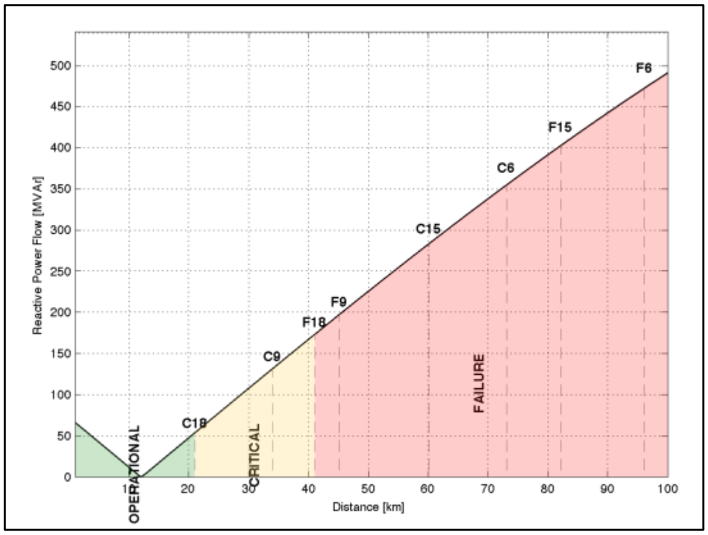


Figure 14. Reactive Power simulation (220kV Submarine Cable)

5.6. Losses & Efficiency

Losses are a significant part of the link evaluation. The model includes several paths to illustrate the losses. Two of the options are provided here: transmission efficiency and transmission (Joule) losses, plotted in Figures 15 and 16.

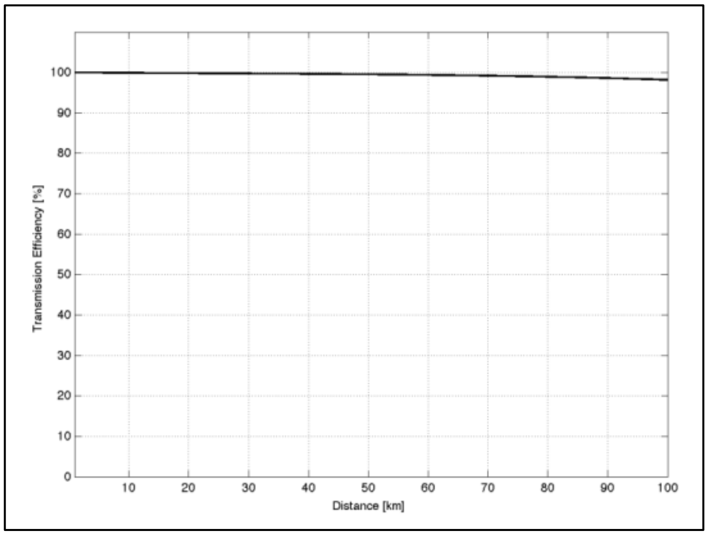


Figure 15. Transmission efficiency simulation (220kV Submarine Cable)

The efficiency remains under an acceptable range (above 90%). This easily confirms that the insulated cables have equivalent losses to the OHL, whereas the capacitive restrictions are the actual restrictions to the operation of these AC offshore links.

In a way, due to the higher inductance value, the capacitive current is not so high and Joule losses are slightly reduced. Nevertheless, the values commented are still very high and above acceptable boundaries. Losses are also similar to the equivalent land cable.

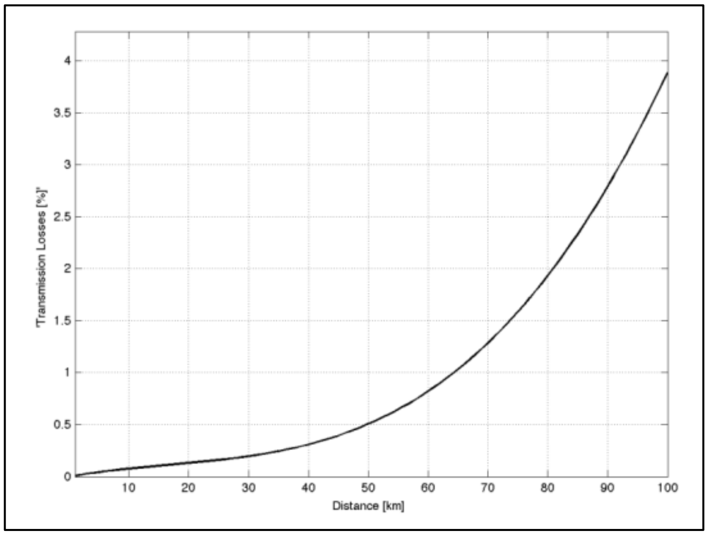


Figure 16. Transmission losses simulation (220kV Submarine Cable)

5.7. Voltage, Current & Displacement Angles

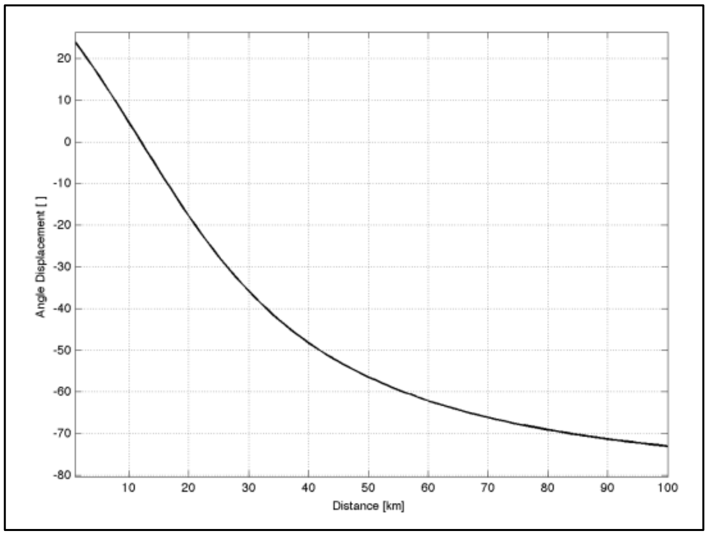


Figure 17. Angle displacement simulation (220kV Submarine Cable)

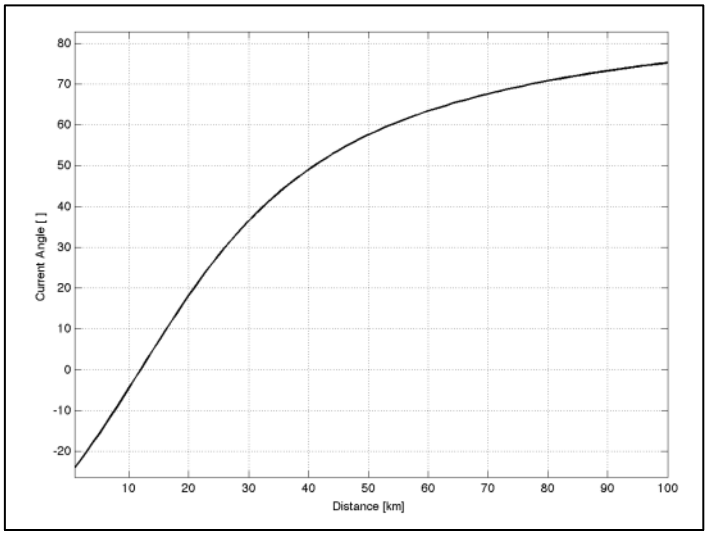


Figure 18. Current angle displacement simulation (220kV Submarine Cable)

The plot of the voltage and current angle displacements between the two ends of the cable is also a clear sign that, above the mentioned 40km, the cable is working under unacceptable operational conditions. The voltage is assuming a strongly capacitive characteristic. A comparison on the voltage angle variation for equivalent land and submarine cables is on Fig. 19.

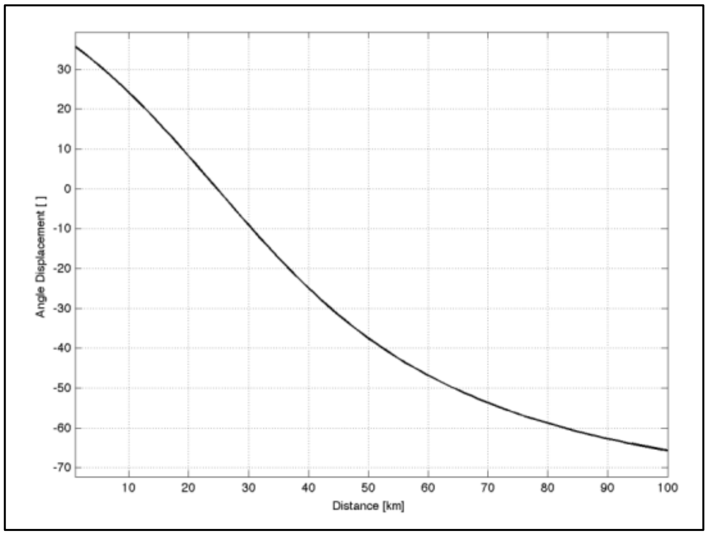


Figure 19. Angle displacement simulation (220kV Submarine Cable)

The angle and power factor values provide once more a good understanding path of the range of operation of the submarine link. The tendency of both curves is much alike the equivalent land cable. The power factor, here stated at 0,33 continues to be very low, inducing that the cable still behaves like a large capacitor. The analysis performed for the land cable, in order to assess the natural load and connectable distance is now demonstrated.

5.8. Variation of Cable Loading

With an inherit highlight of one of the additional major features of the model, this part presents the effects on the variation of the cable loading – current at the reception-end, once more, in regards to power factor and line current. The results are shown next. Hatches were disable to facilitate the visualization.

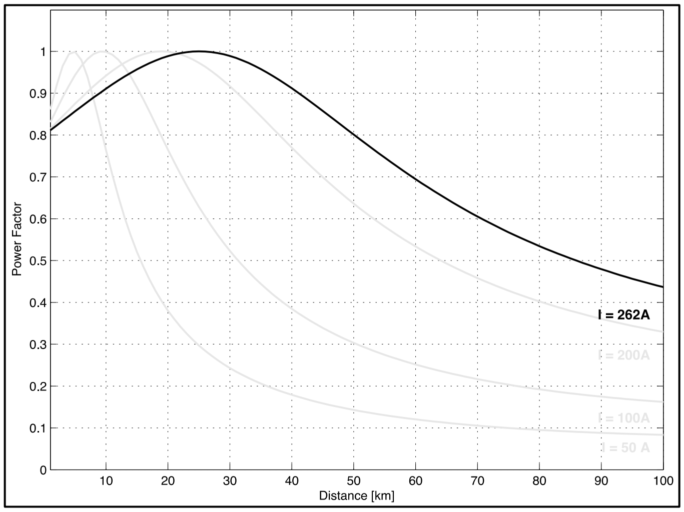


Figure 20. Power factor (load variation) simulation (220kV Submarine Cable)

If the impedance of the load is below the line's impedance, it means that the active power delivered is higher than the natural power for such voltage / cable. Land cables, which have a lower inductance and higher capacitance, are usually operated below their natural power [9], which is calculated, for this particular case, as follows:

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{87.9646}{10.1 \times 10^{-3}}} = 93.32\ \Omega \tag{11}$$

$$P_n = \frac{V^2}{Z_0} = \frac{200 \times 10^3}{93.32} = 518.65\ MW \tag{12}$$

However, subsea cables are characterized to a much higher inductance value, thus changing that approach. With a higher rating yet under the thermal capabilities of the cable, such cable can operate and obtain better operational values.

Addressing the effect on the cable loading, both in terms of rated power, but also, in terms of power factor is crucial. The higher the current provided by the wind farm, a negative shift of the peak power factor occurs, thus leading to conclude that the cable has a wider operation range if the loading (current) is close to its rated value.

On the other hand, for an equivalent unitary capacitance, the sea cables have a much higher inductance, which for certain ranges and loads, it can help to compensate the power factor variation on the line.

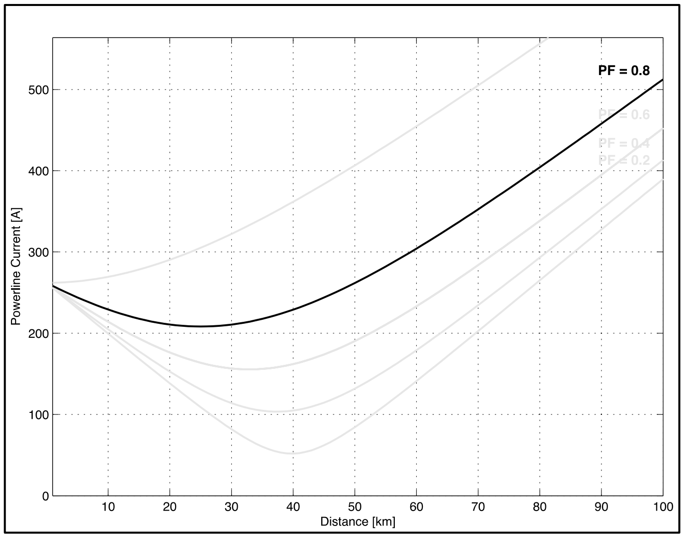


Figure 21. Current (load variation) simulation (220kV Submarine Cable)

As shown on Fig. 14, the lower the power factor of at the WF-side, the lower the current fed from the WF itself. On the other hand, the impact on the reference voltage is important. The voltage and current angles are also exposed to significant changes due to the variation of the initial power factor. The results are included on Figures 16 and 17.

This means that the lower the power factor (lag) is assumed, the voltage and current angles have smother slopes. This translates the mutual "compensation" between the power factor and the PoC and the capacitive characteristics of the long submarine cable.

For the current angle, the values at the end of scale (full distance) do not pose significant differences. But, if the center range of the angle curve is observed, it is clear that the acceptable operational boundary is improved with such power factor decrease.

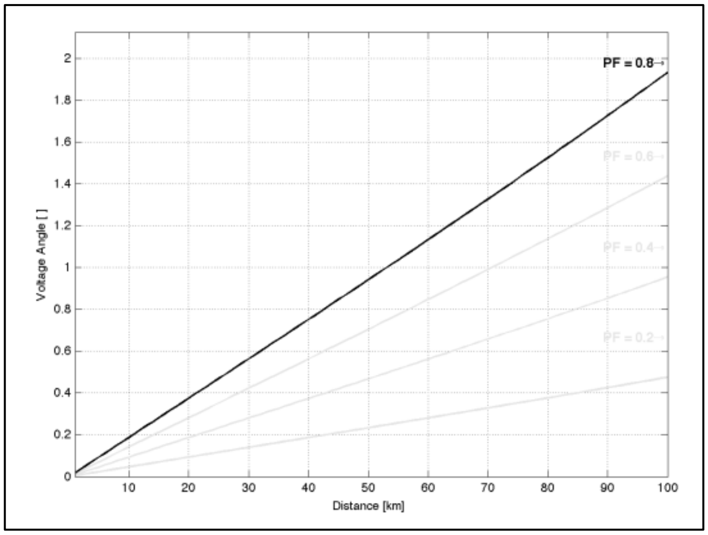


Figure 22. Voltage angle (load variation) simulation (220kV Submarine Cable)

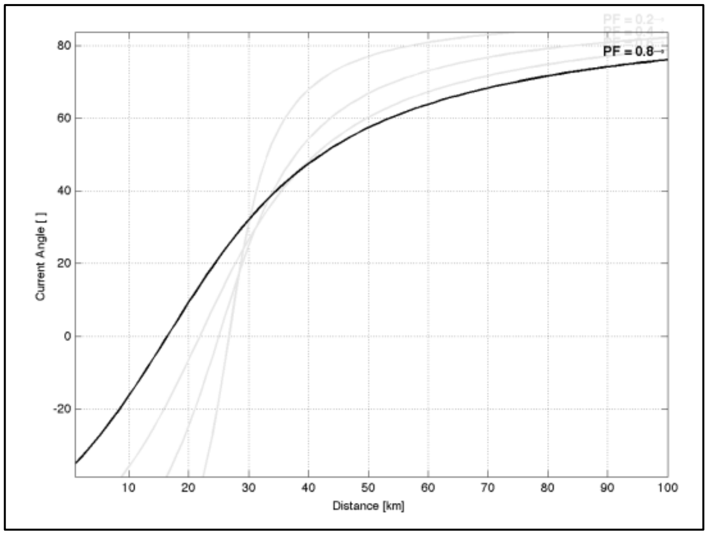


Figure 23. Current angle (load variation) simulation (220kV Submarine Cable)

Bearing in mind the purely academic approach for this – as cables used differ for distinct nominal voltages, what the model quickly provides is a framework on how an higher voltage can facilitate higher power flow. This is addressed via the SIL plot shown in Fig. 18.

The results on Fig. 16. further prove that model designed, as the conclusion, voltage increase allows for a higher SIL rating of the cable, which is aligned with the bibliography. The main result however is the easiness how this is selected on the GUI, by using the "Label" and "Freeze" capabilities. This allows the user to shown several identified scenarios {voltage, current, power factor, distance and compensation}.

Regardless, the what the authors conclude is that is now much easier, both in an academic or (risks measured) practical manner, to quickly assess the best conditions for which an AC offshore link should be designed. One route for the improvement of these results (links) is compensation.

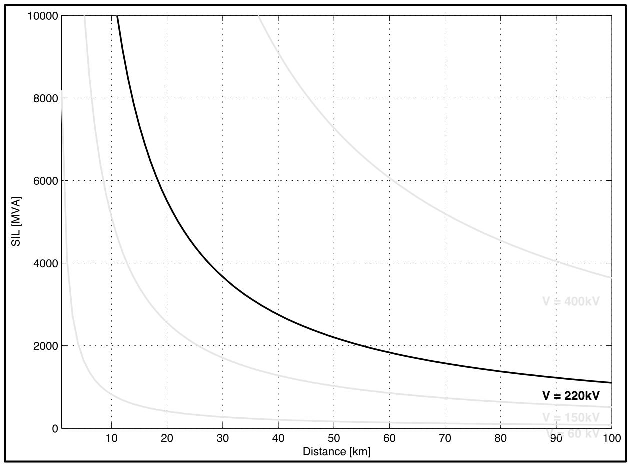


Figure 24. Current (load variation) simulation (220kV Submarine Cable)

6. MATLAB Compensated Results

The analysis of the compensated transmission cables is done still using the multi-scenario approach. As provided earlier, as the parameter changed in each scenario is automacally overlayd on the plot itself, concludes are easily drafted.

6.1. Estimation of Compensation

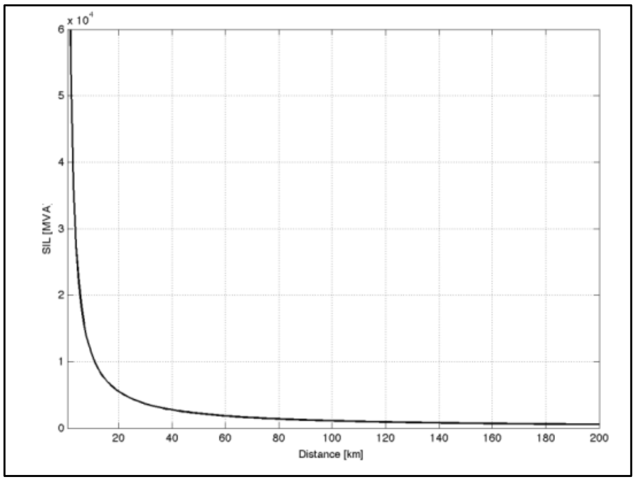


Figure 25. SIL simulation (220kV Submarine Cable)

Based on the results of the reactive power and considering a unitary power factor load, it is possible to estimate the rating of a shunt reactive compensation system – previously decided by the authors after execution a several scenario investigations using the same model. The reactive power is included in Figure 14 whereas the SIL (loading) evaluation is provided in Figure 25.

Given such and based on the experimental values shown on Fig. 16, 265MVAr was put into experiment. The effects on the voltage and power factor, two of the main images of the cable, are shown on the next section.

6.2. Voltage after Compensation

What is immediately observed is a shift of the stress (failure) point earlier identified, thus a clear improvement on the operational range. The reactive power at the emission, for nod distance, represents the full rating of the compensation.

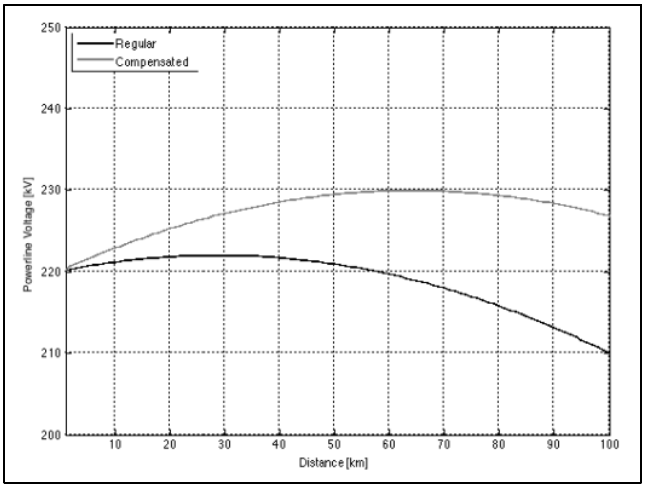


Figure 26. Voltage (after compensation) simulation (220kV Submarine Cable)

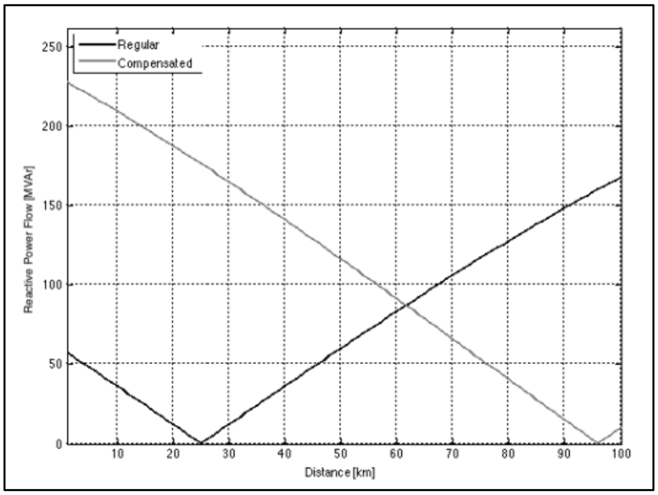


Figure 27. Current (after compensation) simulation (220kV Submarine Cable)

The GUI interface easily presents several scenarios, without going through a labour-intensive calculation routine. Taking once more the approach of adapting the cable loading (current at the reception-end), additional conclusions can be gathered (as provided on Figures 27 and 28).

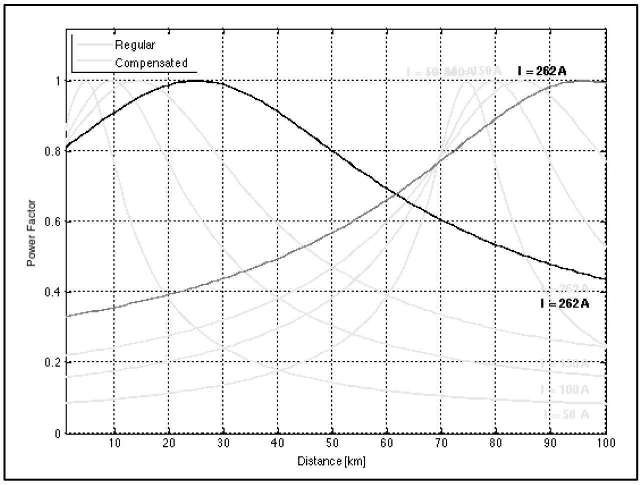


Figure 28. Power factor (load variation after compensation) (220kV Cable)

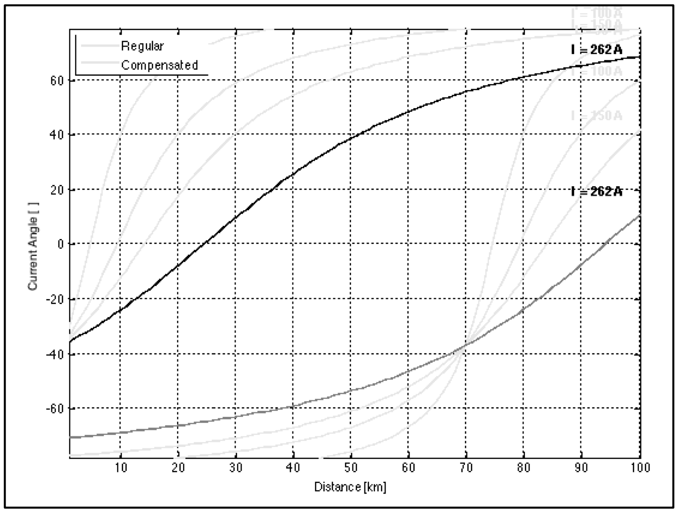


Figure 29. Current (load variation after compensation) (220kV Submarine Cable)

Voltage, reactive power and the power factor are now within perfectly acceptable boundaries, for the full 100km length. The power factor must be analysed depending on the point of connection of the compensation system, eventually using a split-off between the two ends.

Especially after the introduction of compensation, several loading scenarios have to be taken into account, once more, taking advantage of the GUI-model. For lower loading scenarios, the compensation system output has to be reduced or, eventually, taken out of service. Otherwise, what is identified in Figs. 19 and 20 is a clear overcompensation for a load drop (current). Nowadays, online FACTS systems effectively support the reactive power flow demanded by large-scale offshore wind farms. Although not addressed on this paper, once again, the model points to a quick rough estimation of a possible compensation solution.

7. Final Remarks & Conclusion

The authors firmly believe that this computational simulation tool is an important guide on the assessment of HVAC transmission links, especially for offshore applications such as wind farms.

As expected, the submarine cables presented the authors with the worst operational conditions from all the three options addressed on this document. The high inductance of such cables, mostly due to their type of insulation, introduces serious limitations, which translate in very short feasible distances. If this was the baseline for the investigation, the model as shown that the conclusions are obtained experimentally with ease.

The authors believe that the results shown before allow for a sufficient and accurate perception of the problematic associated with large HVAC transmission lines. The results were conclusive, with minor errors taking into consideration the objective proposed. So much more can be done in a further stage.

There is however still room for further improvement on the existing model, addressing topics such as different types of compensation of lines or introducing further characteristics of the line, of course, that can affect the line's behavior in some way.

The compensation system here studied also has some limitations, namely in terms of type (only fixed-step compensation is addressed) and in terms of distribution, eventually on several stages of the cable or transmission line. Future developments of the model should address this topic and allow for a more broad study of compensation effects.

The work can also extended and applied across other industries, such as Oil&Gas – for offshore energy supply to platforms and interplatform energy safety-grids. The goal of providing a safe and reliable tool, able to provide link estimations in a window of seconds to minutes was achieved.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors thank EDP Distribution (the Distribution System Operator in Portugal). This work also has been once more partially supported by the Portuguese Foundation for Science and Technology under projects grants PEst-OE/ EEI/UI308/2014 and UID/EEA/00066/2013.

References

[1] T. A. Antunes, P. J. Santos and A. J. Pires, "HVAC transmission restrictions in large scale offshore wind farm applications," in *2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, Cadiz, Spain, 2017.

[2] H. Holttinen, J. Kiviluoma, J. McCann, M. Clancy, M. Millgan, I. Pineda, P. B. Eriksen, A. Orths and O. Wolfgang, "Reduction of CO2 emissions due to wind energy - methods and issues in estimating operational emission reductions," in *Power & Energy Society General Meeting, 2015 IEEE*, Denver, USA, 2015.

[3] International Renewagle Energy Association, "The Power to Change: Solar and Wind Cost Reduction Potential to 2025," IRENA, 2016.

[4] B. Gustavsen and O. Mo, "Variable Transmission Voltage for Loss Minimization in Long Offshore Wind Farm AC Export Cables," in *IEEE Transactions on Power Delivery* , 2016.

[5] Siemens AG, *Wind Turbine SWT-8.0-154: Technical Specifications,* Hamburg: Wind Power and Renewables Division, 2016.

[6] Siemens AG, "Powered by partnership: Sustainable solutions for your offshore wind power project," 2016.

[7] ABB, "XLPE Cables up to 420kV," 2013.

[8] D. Moura, Técnicas de Alta Tensão - Curso Introdutório, Lisboa: Técnica - Revista de Engenharia, Lda., 1980.

[9] J. P. Sucena Paiva, Redes de Energia Eléctrica, Uma Análise Sistémica, 2nd Edition ed., vol. 1, I. Press, Ed., Lisbon: IST Press, 2007.

[10] Siemens, "High Voltage Direct Current Transmission: Proven Technology for Power Exchange," Siemens AG, Alemanha, 2011.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 157-163 (2018)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Modeling of the wave functions and of the energy states of hydrogen stored in a spherical cavity

Kamel Idris-Bey*

Matter Science Department, Faculty of Sciences, Laboratory of Experimental Physics and Applications University Yahia Fares- Medea, 26000, Algeria

ARTICLE INFO

Article history:
Received: 03 April, 2017
Accepted: 20 May, 2017
Online: 17 March, 2018

Keywords :
Hydrogen storage
Wave functions, energies
Transcendental equation

ABSTRACT

This article examines the hydrogen storage phenomenon in a spherical cavity. The hydrogen gas or liquid is subjected to high pressures, leading to significant loss of mass of hydrogen, and requires materials that can withstand these high pressures also minimize losses. For all these reasons, the problem is considered at the quantum scale. So in quantum mechanics it studies the theory of wave functions corresponding to the hydrogen with the correct expressions development of the radial functions and the spherical harmonics, and also the energy stored, and then the graphic applications that gives a spatial representation of each function with a program informatics.

1. Introduction

Hydrogen is the lightest of gases and possesses the lowest density. However at ambient temperature and pressure it occupies a large volume, a car, whose weekly need is on average 5 Kg of hydrogen, must have a spherical tank of radius r # 2.44 m. This necessitates compressing it at high pressures up to P # 800 Bars [1]. The immense interest generated by hydrogen comes from the fact that it has the best energy per weight ratio of all fuels and the ecological nature of the combustion product (water vapor). But much of the energy is lost during storage. From the pedagogical point of view, it is also the most taught and involved in research. In particular in quantum mechanics, this deals with the state of particles such as the electron. This aspect is treated in this article in order to solve the problem of storage of the hydrogen by minimizing losses as much as possible. The solutions envisaged are, first, the improvement of the theory to understand the physical phenomena that occur in the physical system, especially the resolution of the transcendental equation, and then the means of perfecting the materials constituting the cavity (tank).

The confinement of atoms or molecules in a finite region of space is an important subject, for example, in the context of quantum dots, encapsulation in fullerenes, or other aspects of nanotechnology. Confinement in a cavity has also been used to simulate the effects of high pressure. A hydrogen atom at the center of a spherical cavity was first studied by Michels & al. in 1937 [2] in order to model hydrogen at high pressure, as well as by Sommerfeld and Welker in 1938 [3], and also in an extended body of subsequent work [4, 5, 6, 7].

Most of these works use the standard Dirichlet boundary condition with a vanishing wave function at the cavity wall, while some use Neumann boundary conditions, i.e. a vanishing gradient of the wave function perpendicular to the wall. A notable exception is where the most general so called Robin boundary condition has been considered [8].

Al-Hashimi in a recent article 2012 [9], has well studied the problem of hydrogen which is confined in a spherical cavity using the conditions at the limits of Robin, he also introduced a transcendental equation for the calculation of the energies.

However to solve theoretically, hydrogen gas transmission phenomenon that is enclosed in a spherical cavity or into a conical cavity and under high pressure, two equations are necessary: the Schrödinger equation, equation (1), which is well known and, the Victor Gustave Robin boundary condition equation (2)

$$H.\psi(\vec{x}) = \mathrm{E}(\vec{\mathrm{k}}).\psi(\vec{x}) \quad (1)$$

*Kamel Idris-Bey, Medea-Algeria-, 0668614740, idrisbeykamel@gmail.com

https://dx.doi.org/10.25046/aj030218

$$\gamma(\vec{x})\psi(\vec{x}) + \vec{n}(\vec{x}).\vec{\nabla}\psi(\vec{x}) = 0, \quad \vec{x} \in \partial\Omega \tag{2}$$

The self-adjoint extension parameter $\gamma(\vec{x})$ takes into account the constituent material of the cavity, $\psi(\vec{x})$ is the wave function, $\partial\Omega$ is the limit of a spatial region Ω and $\vec{n}(\vec{x})$ is the unit vector perpendicular to the surface. As usual, the wave function can be factored as the product of a radial function $\psi_{v\ell}(r)$ with a spherical harmonic function $Y_{lm}(\theta,\varphi)$ according to the following expression

$$\psi(\vec{x}) = \psi_{v\ell}(r).Y_{\ell m}(\theta,\varphi) \tag{3}$$

This equation is difficult to solve because of the size, effectively the radial function is graphically traceable in 2D graph, and the spherical harmonic function in 3D dimension. And thus the product, ie the wave function, is of dimension greater than or equal to four.

2. Theory of spherical harmonics

The spherical harmonic functions are defined either as the eigenfunctions of angular momentum in quantum physics, or as the solutions of the Laplace equation $\nabla^2 f = 0$ [10, 11, 12, 13]. Solving this equation in spherical coordinates leads to the following expressions [14]:

$$Y_\ell^m(\theta,\varphi) = (-1)^m k(\ell,m) P_\ell^m(cos\theta) e^{im\varphi} \tag{4}$$

where P_ℓ^m is the Legendre associated polynomial of degree ℓ and order m

$$P_\ell^m(cos\theta) = \frac{(-1)^m}{2^\ell \ell!}(1-(cos\theta)^2)^{\frac{m}{2}} \frac{d^{\ell+m}}{dx^{\ell+m}}((cos\theta)^2-1)^\ell \tag{5}$$

and $k(\ell,m)$ is the normalization function:

$$k(\ell,m) = (-1)^m \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} \tag{6}$$

2.1. Expressions of the spherical harmonics

The terms of spherical harmonics which have azimuthally quantum numbers ℓ = 1, ℓ = 2 and ℓ = 3 are defined, in quantum mechanics, as follows:

$\boldsymbol{\ell = 0}$ **:** Necessarily implies that the magnetic quantum number is zero (m=0).

$$Y_0^0(\theta,\varphi) = \frac{1}{\sqrt{4\pi}} \tag{7}$$

$\boldsymbol{\ell = 1}$: There are three possible values for m: m = 1, m = 0 and m = -1.

$$Y_1^1(\theta,\varphi) = -\sqrt{\frac{3}{8\pi}} \sin\theta \, e^{i\varphi} \tag{8}$$

$$Y_1^0(\theta, \ \varphi) = \sqrt{\frac{3}{4\pi}} \cos\theta \tag{9}$$

$$Y_1^{-1}(\theta,\varphi) = \sqrt{\frac{3}{8\pi}} \sin\theta \, e^{-i\varphi} \tag{10}$$

$\boldsymbol{\ell = 2}$: There are five possible values for m: m = 2, m = 1, m = 0, m = -1 and m = -2.

$$Y_2^2(\theta,\varphi) = \sqrt{\frac{15}{32\pi}} sin^2\theta e^{i2\varphi} \tag{11}$$

$$Y_2^1(\theta,\varphi) = -\sqrt{\frac{15}{8\pi}} \sin\theta\cos\theta \, e^{i\varphi} \tag{12}$$

$$Y_2^0(\theta,\varphi) = \sqrt{\frac{5}{16\pi}}(3cos^2\theta - 1) \tag{13}$$

$$Y_2^{-1}(\theta,\varphi) = \sqrt{\frac{15}{8\pi}} \sin\theta\cos\theta \, e^{-i\varphi} \tag{14}$$

$$Y_2^{-2}(\theta,\varphi) = \sqrt{\frac{15}{32\pi}} sin^2\theta e^{-i2\varphi} \tag{15}$$

$\boldsymbol{\ell = 3}$: There are seven possible values for the magnetic quantum number m: m = 3, m = 2, m = 1, m = 0, m = -1, m = -2 and m = -3.

$$Y_3^3(\theta,\varphi) = -\sqrt{\frac{35}{64\pi}} \sin^3\theta \, e^{i3\varphi} \tag{16}$$

$$Y_3^2(\theta,\varphi) = \sqrt{\frac{105}{32\pi}} \cos\theta \sin^2\theta e^{i2\varphi} \tag{17}$$

$$Y_3^1(\theta,\varphi) = -\sqrt{\frac{21}{64\pi}} \sin\theta \, (5cos^2\theta - 1) e^{i\varphi} \tag{18}$$

$$Y_3^0(\theta,\varphi) = \sqrt{\frac{7}{16\pi}}(5cos^3\theta - 3\cos\theta) \tag{19}$$

$$Y_3^{-1}(\theta,\varphi) = \sqrt{\frac{21}{64\pi}} \sin\theta \, (5cos^2\theta - 1) e^{-i\varphi} \tag{20}$$

$$Y_3^{-2}(\theta,\varphi) = \sqrt{\frac{105}{32\pi}} \cos\theta \sin^2\theta \, e^{-i2\varphi} \tag{21}$$

$$Y_3^{-3}(\theta,\varphi) = \sqrt{\frac{35}{64\pi}} \sin^3\theta \, e^{-i3\varphi} \tag{22}$$

3. Theory of the radial functions

As a preparation for the hydrogen problem, in this section we consider a ‘‘free’’ particle in a spherical cavity with general reflecting boundary conditions specified by the self-adjoint extension parameter γ. And after that, we study the problem of the hydrogen atom in a spherical cavity with general reflecting boundary conditions, again specified by the self-adjoint extension parameter γ ∈ ℝ (real numbers).

3.1. Particle in a spherical cavity with general reflecting boundaries

Let us consider the Hamiltonian of a free particle of mass M in spherical coordinates:

$$H = \frac{\vec{P}^2}{2M} = -\frac{\hbar^2}{2M}\Delta = -\frac{\hbar^2}{2\mathrm{M}}\left(\partial_\mathrm{r}^2 + \frac{2}{\mathrm{r}}\partial_\mathrm{r} - \frac{\overrightarrow{\mathrm{L}^2}}{\mathrm{r}^2}\right) \tag{23}$$

with angular momentum $\vec{L}$ in a spherical cavity of radius R. As usual, the wave function can be factorized as:

$$\psi(\vec{x}) = \psi_{k\ell}(r).Y_{\ell m}(\theta,\varphi) \qquad (24)$$

where the angular dependence is described by the spherical harmonics $Y_{lm}(\theta,\varphi)$. For $\hbar = 1$, the radial wave function obeys to:

$$-\frac{1}{2\mathrm{M}}\left(\partial_{\mathrm{r}}^2+\frac{2}{\mathrm{r}}\partial_{\mathrm{r}}-\frac{\ell(\ell+1)}{\mathrm{r}^2}\right)\psi_{k\ell}(r) = E\psi_{k\ell}(r) \qquad (25)$$

$$with\ E = \frac{k^2}{2M}$$

For a spherical cavity, the most general perfectly reflecting boundary condition of equation (2) takes the form:

$$\gamma\psi_{k\ell}(R) + \partial_{\mathrm{r}}\psi_{k\ell}(R) = 0 \qquad (26)$$

For positive energy the normalizable wave function is given by the spherical Bessel functions:

$$\psi_{k\ell}(r) = A\,J_\ell(kr) \qquad (27)$$

For general ℓ at $\gamma = -\ell/\mathrm{R}$, the ground state has zero energy with the radial wave function given by:

$$\psi(r) = \sqrt{\frac{2\ell+3}{R^3}}\left(\frac{r}{R}\right)^{\ell} \qquad (28)$$

Consider the following cases:

$$\begin{cases} \ell = 0:\ \psi(r) = \sqrt{\frac{3}{R^3}} \\ \ell = 1:\ \ \psi(r) = \sqrt{\frac{5}{R^3}}\left(\frac{r}{R}\right) \\ \ell = 2:\ \psi(r) = \sqrt{\frac{7}{R^3}}\left(\frac{r}{R}\right)^2 \\ \ell = 3:\ \psi(r) = \sqrt{\frac{9}{R^3}}\left(\frac{r}{R}\right)^3 \end{cases} \qquad (29)$$

3.2. Hydrogen atom in a spherical cavity with general reflecting boundaries

In this section consider an electron bound to a proton that is localized at the center of a spherical cavity with general reflecting boundary conditions, again specified by the self-adjoint extension parameter $\gamma \in \mathbb{R}$. The Hamiltonian radial equation of the hydrogen atom, in spherical coordinates, takes the expression:

$$-\frac{1}{2\mathrm{M}}\left(\partial_{\mathrm{r}}^2+\frac{2}{\mathrm{r}}\partial_{\mathrm{r}}-\frac{\ell(\ell+1)}{\mathrm{r}^2}-\frac{\mathrm{e}^2}{\mathrm{r}}\right)\psi_{\nu\ell}(r) = E\psi_{\nu\ell}(r) \qquad (30)$$

and the normalizable wave function is given by:

$$\psi_{\nu\ell}(r) = A\left(\frac{2r}{\nu a}\right)^{\ell} L_{\nu-\ell-1}^{2\ell+1}\left(\frac{2r}{\nu a}\right) exp\left(\frac{-r}{\nu a}\right) \qquad (31)$$

where $L_{\nu-\ell-1}^{2\ell+1}\left(\frac{2r}{\nu a}\right)$ is an associated Laguerre function [15] and [16], and a is the Bohr radius and A is a constant.

Consider the following cases:

$$\begin{cases} \ell = 0:\ \psi_{\nu 0}(r) = AL_{\nu-1}^{1}\left(\frac{2r}{\nu a}\right) exp\left(\frac{-r}{\nu a}\right) \\ \ell = 1:\ \psi_{\nu 1}(r) = A\left(\frac{2r}{\nu a}\right) L_{\nu-2}^{3}\left(\frac{2r}{\nu a}\right) exp\left(\frac{-r}{\nu a}\right) \\ \ell = 2:\ \psi_{\nu 2}(r) = A\left(\frac{2r}{\nu a}\right)^2 L_{\nu-3}^{5}\left(\frac{2r}{\nu a}\right) exp\left(\frac{-r}{\nu a}\right) \\ \ell = 3:\ \psi_{\nu 3}(r) = A\left(\frac{2r}{\nu a}\right)^3 L_{\nu-4}^{7}\left(\frac{2r}{\nu a}\right) exp\left(\frac{-r}{\nu a}\right) \end{cases} \qquad (32)$$

And for $\nu = 4$, expressions (32) take the form :

$$\begin{cases} \ell = 0:\ \psi_{40}(r) = AL_{3}^{1}\left(\frac{r}{2a}\right) exp\left(\frac{-r}{4a}\right) \\ \ell = 1:\ \psi_{41}(r) = A\left(\frac{r}{2a}\right) L_{2}^{3}\left(\frac{r}{2a}\right) exp\left(\frac{-r}{4a}\right) \\ \ell = 2:\ \psi_{42}(r) = A\left(\frac{r}{2a}\right)^2 L_{1}^{5}\left(\frac{r}{2a}\right) exp\left(\frac{-r}{4a}\right) \\ \ell = 3:\ \psi_{43}(r) = A\left(\frac{r}{2a}\right)^3 L_{0}^{7}\left(\frac{r}{2a}\right) exp\left(\frac{-r}{4a}\right) \end{cases} \qquad (33)$$

4. Implementation and graphs of the spherical harmonics functions [17]

The graphs of spherical harmonic functions, calculated above, were modeled and mapped with “SPharm” program Matlab environment 7.8.0 (R2009a). The essential functions are plotted on figure (1) and figure (2) below:

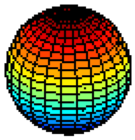
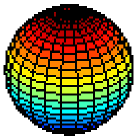
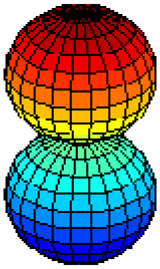
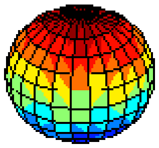
$Y_l^m(\theta,\varphi)$	$\lvert Y_l^m(\theta,\varphi\rvert$	$Real\ Y_l^m(\theta,\varphi)$
$\boldsymbol{Y_0^0}$		
$\boldsymbol{Y_1^0}$		

Figure (1): The module and the real part of the spherical harmonic functions: Y_0^0, Y_1^0.

4.1. Graphs of the radial functions and the density functions for the hydrogen atom

Knowing the associated Laguerre functions expressions:

$$\begin{cases} L_3^1\left(\frac{r}{2a}\right) = -\frac{1}{6}\left(\frac{r}{2a}\right)^3 + 2\left(\frac{r}{2a}\right)^2 - 6\left(\frac{r}{2a}\right) + 4 \\ L_2^3\left(\frac{r}{2a}\right) = \frac{1}{2}\left(\frac{r}{2a}\right)^2 - 5\left(\frac{r}{2a}\right) + 10 \\ L_1^5\left(\frac{r}{2a}\right) = -\left(\frac{r}{2a}\right) + 6 \\ L_0^7 = 1 \end{cases} \quad (34)$$

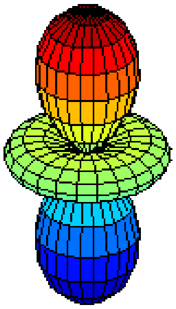
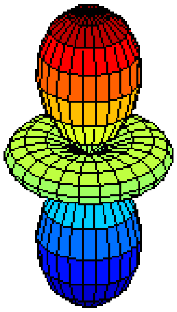
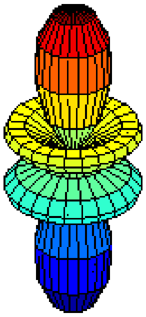
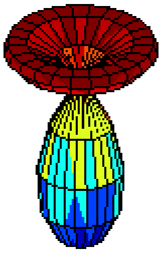
$Y_l^m(\theta,$	$\lvert Y_l^m(\theta,\varphi\rvert$	$Real\ Y_l^m(\theta,\varphi)$
$\boldsymbol{Y_2^0}$		
$\boldsymbol{Y_3^0}$		

Figure (2): The module and the real part of the spherical harmonic functions: Y_2^0, Y_3^0.

The radial functions take the expressions:

$$\ell = 0: \ \psi_{40}(r) = A_0\left(-\frac{1}{6}\left(\frac{r}{2a}\right)^4 + 2\left(\frac{r}{2a}\right)^3 - 6\left(\frac{r}{2a}\right)^2 + 4\left(\frac{r}{2a}\right)\right).exp\left(\frac{-r}{4a}\right) \quad (35)$$

$$\ell = 1: \ \psi_{41}(r) = A_1\left(\frac{1}{2}\left(\frac{r}{2a}\right)^4 - 5\left(\frac{r}{2a}\right)^3 + 10\left(\frac{r}{2a}\right)^2\right).exp\left(\frac{-r}{4a}\right) \quad (36)$$

$$\ell = 2: \ \psi_{42}(r) = A_2\left(-\left(\frac{r}{2a}\right)^4 + 6\left(\frac{r}{2a}\right)^3\right).exp\left(\frac{-r}{4a}\right) \quad (37)$$

$$\ell = 3: \ \psi_{43}(r) = A_3\left(\frac{r}{2a}\right)^4.exp\left(\frac{-r}{4a}\right) \quad (38)$$

A_0, A_1, A_2, A_3 are constants which can be calculated with the normalization condition:

$$\int_0^{+\infty} r^2\, |\psi_{vl}(r)|^2 dr = 1 \quad (39)$$

The calculations were made with $A_0 = A_1 = A_2 = A_3 = 2$, the graphs are shown in figure (3) below:

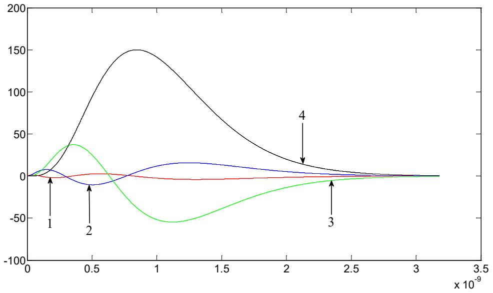


Figure (3): The radial functions: $\psi_{40}(r)$, 1-red; $\psi_{41}(r)$, 2-bleu; $\psi_{42}(r)$, 3-green; $\psi_{43}(r)$, 4-black (the x-axis in 10^{-9} meter).

The probability density function gives the maximum probability of finding the electron at a position $r = r_0$ on the radius of the sphere; this is reflected by a maximum of this function at the point r_0. The graphs of probability density functions $u_{\nu\ell}^2(r) = r^2\, {\psi_{\nu\ell}}^2(r)$, for the hydrogen atom, are shown in figures (4, 5, 6 and 7):

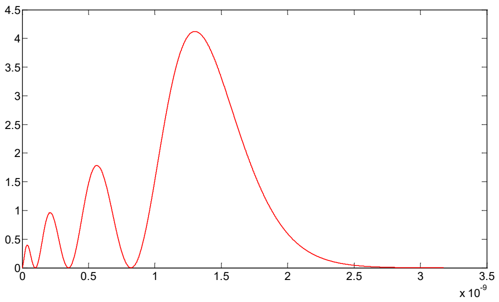


Figure (4): density $u_{40}^2(r)$which has four maximums.

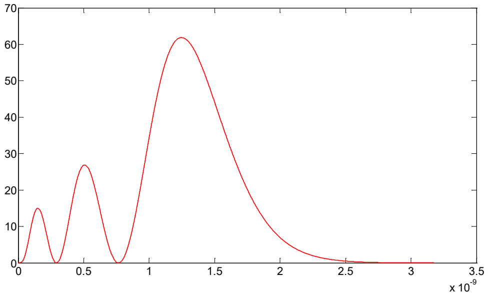


Figure (5): density $u_{41}^2(r)$ which has three maximums.

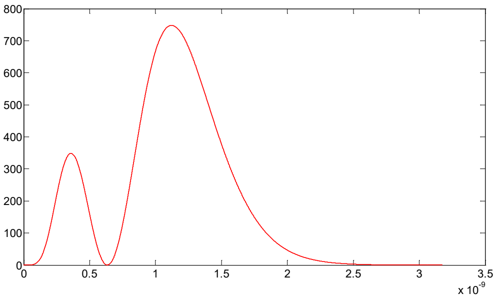


Figure (6): density $u_{42}^2(r)$ which has two maximums.

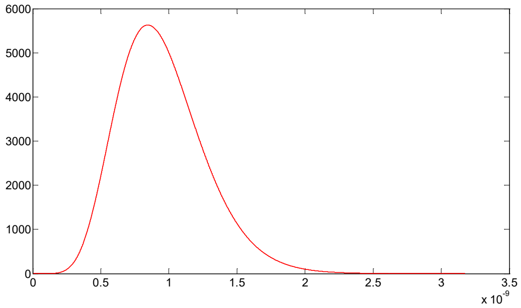


Figure (7): density $u_{43}^2(r)$ which has one maximum.

5. The transcendental equation of the energy spectrum

Knowing the most general perfectly reflecting boundary condition by the equation (26), the energy spectrum is thus determined from the transcendental equation

$$\gamma J_\ell(kr) + \partial_{\mathrm{r}} J_\ell(kr) = \left(\gamma + \frac{\ell}{r}\right) J_\ell(kr) - kJ_{\ell+1}(kr) \ = 0 \quad (40)$$

This transcendental equation, which expresses the energy of the physical system, arises from the resolution of the Schrödinger equation and takes into account parameters relating to the quantification of the energy considered to be the real number k, and also of a second parameter ℓ, which is the azimuthally quantum number which generates the degeneration of the stationary states of energy. A third parameter, the self-adjoint extension parameter γ, is also considered and concerns surface boundary conditions, taking into account the state of the surface as well as the material constituting the inner envelope of the cavity. In fact, this envelope must prevent the loss of energy at the atomic scale; the inside of the envelope is therefore made reflective so that all the particles arriving at the wall are deflected towards the inside of the cavity.

In the following, the expressions of the Bessel functions for ℓ and (ℓ +1) orders, and also the derivative:

$$J_\ell(kr) = \left(\frac{kr}{2}\right)^\ell \sum_{n=0}^{+\infty} \frac{(-1)^n}{n!(n+\ell)!} \left(\frac{kr}{2}\right)^{2n} \quad (41)$$

$$J_{\ell+1}(kr) = \left(\frac{kr}{2}\right)^{\ell+1} \sum_{n=0}^{+\infty} \frac{(-1)^n}{n!(n+\ell+1)!} \left(\frac{kr}{2}\right)^{2n} \quad (42)$$

$$\frac{dJ_\ell(kr)}{dr} = \frac{dJ_\ell(kr)}{d(kr)} \frac{d(kr)}{dr} = \frac{d(kr)}{dr} = k \frac{dJ_\ell(kr)}{d(kr)}$$

$$= \frac{\ell}{r} J_\ell(kr) + \frac{2}{r} \left(\frac{kr}{2}\right)^\ell \sum_{n=0}^{+\infty} \frac{(-1)^n}{(n-1)!(n+\ell)!} \left(\frac{kr}{2}\right)^{2n} \quad (43)$$

Substituting these expressions, equations (41, 42 and 43) in the equation (40), the transcendental equation become:

$$\gamma J_\ell(kr) + \frac{\ell}{r} J_\ell(kr) + \frac{2}{r} \left(\frac{kr}{2}\right)^\ell \sum_{n=0}^{+\infty} \frac{(-1)^n}{(n-1)!\,(n+\ell)!} \left(\frac{kr}{2}\right)^{2n} =$$

$$\left(\gamma + \frac{\ell}{r}\right) J_\ell(kr) - k \left(\frac{kr}{2}\right)^{\ell+1} \sum_{n=0}^{+\infty} \frac{(-1)^n}{n!(n+\ell+1)!} \left(\frac{kr}{2}\right)^{2n} = 0 \quad (44)$$

5.1. Calcul of the energies states

Consider now the equation giving the boundary condition for an indifferent position r:

$$\gamma \psi_{k\ell}(r) + \partial_{\mathrm{r}} \psi_{k\ell}(r) = 0 \quad (45)$$

The solution of this equation (45) was find as:

$$\psi_{k\ell}(r) = \psi_0 \exp\left(-\gamma . r\right) \quad (46)$$

With applying the conditions at the wall of the cavity for r = R and at the center of the cavity for r = 0:

$$\psi_{k\ell}(R) = 0 \ \ and \ \ \psi_{k\ell}(0) = \psi_0 \quad (47)$$

The exact wave function, which takes into account the self-adjoint extension parameter γ , is then found like:

$$\psi_{k\ell}(R) = \psi_0 \exp\left(-\gamma . R\right) \quad (48)$$

For $\gamma \to +\infty$, the boundary condition reduces to $\psi_{k\ell}(R) = 0$, as well as the textbook case.

It is well know that, the wave function which is depending on the wave vector $\vec{k}$ and on the space vector $\vec{r}$, is written as in the following form:

$$\psi_{k\ell}(\vec{r}) = A(\vec{r}).\exp\left(-i.\vec{k}.\vec{r}\right) \quad (49)$$

And for the radial component for r = R:

$$\psi_{k\ell}(R) = A(R).\exp(-i.k.R) = 0 \quad (50)$$

$$A(R) \neq 0, then \exp(-i.k.R) = 0 \implies \begin{cases} \cos(kR) = 0 \\ \sin(kR) = 0 \end{cases} \quad (51)$$

Thus:

$$\begin{cases} kR = (2n+1)\frac{\pi}{2} \Longrightarrow 2kR = (2n+1)\pi \\ \qquad\qquad kR = n\pi \end{cases} \quad (52)$$

Therefore:

$$2kR - kR = (2n+1)\pi - n\pi = (n+1)\pi \quad (53)$$

At the end, the wave vector k takes the following form:

$$k = \frac{(n+1)\pi}{R} \quad (54)$$

The corresponding energies are then given by:

$$E_{n0}(k) = \frac{k^2}{2M} = \frac{(n+1)^2\pi^2}{2MR^2} \quad with \ \hbar^2 = 1 \quad (55)$$

This equation (55) is also valid for $\ell = 1$ and $\gamma = \frac{2}{R}$, so:

$$E_{n0}(k) = E_{n1}(k) = \frac{(n+1)^2\pi^2}{2MR^2} \quad with\, \gamma = \frac{2}{R} \qquad (56)$$

For $\ell = 0$, the equation (44) become with using the equation (40):

$$\gamma\, J_0(kr) - k\, J_1(kr) = 0 \qquad (57)$$

By substituting the expressions of $J_0(kr)\; and\, J_1(kr)$ one can arrive at the following relation:

$$\gamma = \frac{k^2 r}{2(n+1)} \qquad (58)$$

Then at the wall of the cavity for, r = R and taking into account the equation (54), the equation (57) becomes:

$$\gamma = \frac{(n+1)^2\pi^2}{2(n+1)R} = \frac{(n+1)\pi^2}{2R} \qquad (59)$$

In order to know the meaning of the self-extension parameter, it is important to express the natural number n as a function of γ according to the following relation:

$$n = \left(\frac{2\gamma R}{\pi^2} - 1\right) \qquad (60)$$

and consequently, the energy as a function of γ arises from equation (55) as:

$$E_{\gamma 0} = \frac{\left(\frac{2\gamma R}{\pi^2}\right)^2 \pi^2}{2MR^2} = \frac{2\gamma^2}{\pi^2 M} \qquad (61)$$

It is then easy to obtain the two conditions:

$$\begin{cases} if\;\; \gamma \to +\infty \;\; then\; E_{\gamma 0}(k) \;\to +\infty \\ \qquad\qquad and \\ \quad if\;\; \gamma \to 0 \;\; then\; E_{\gamma 0}(k) \;\to 0 \end{cases} \qquad (62)$$

These conditions, equation (61), mean that when the self-adjoint extension parameter γ is high, tends to the infinite, the storage of the hydrogen energy is well good and there are little or no losses. On the other hand, when the self-adjoint extension parameter γ is small, tends to zero, the hydrogen energy storage make many losses, and it is not good.

5.2. Graphs of the energies states

The energy graph as a function of the radius R of the spherical cavity, $E_{n0} = f(R)$, for the following values of n ($n = 0, 1, 2, 3, 4\; and\; 5$) is shown in the figure (8).

For radii$R = 1, 2, 3, 4\; and\; 5$, the energies of the lowest level $n = 0$ are respectively (in Joules):

$$5{,}42.\,10^{30}; 1{,}36.\,10^{30}; 0{,}60.\,10^{30}; 0{,}34.\,10^{30}; 0{,}22.\,10^{30} \qquad (63)$$

The graph of energy as a function of the self-adjoint parameter, $E\gamma 0 = \mathrm{f}\,(\gamma)$, is shown in figure (9).

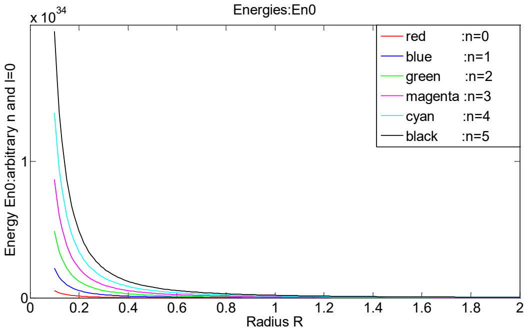


Figure (8): Energy E_{n0} with function of R and n (n=0, 1, 2, 3, 4 and 5).

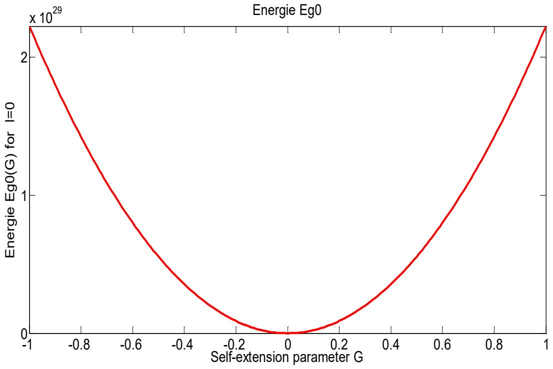


Figure (9): Energy $E_{\gamma 0}$ with function of γ.

The case, gamma tends towards infinity is treated previously and the energy depends on the number n. There is no contradiction since energy tends to infinity in both cases: when n and γ tend towards infinity.

6. Conclusion

This article studies both a free particle and an electron bound in a hydrogen atom confined to a spherical cavity with general perfectly reflecting boundary conditions characterized by a self-adjoint extension parameter. It is well known that hydrogen gas need high pressure, about 700 to 800 bars, to minimize the volume because the hydrogen density is very small. And also the material of the cavity must be lightweight, like fullerenes or other aspects of nanotechnology, with a perfectly reflecting wall inside.

The subject requires much knowledge to take into account the thermodynamics effect like the pressure, so this modeling is based on the calculation of the wave functions especially the radial function. And also a calculation of the clean energies that are stored in the cavity was approached.

The solutions depend on the self-adjoint extension parameter which takes into account the physical properties of the cavity wall. For some values of this parameter, the radial wave function is expressed with function of the Laguerre polynomials. These radial functions tend quickly to zero (equal to zero at about 3.10^{-9}

meter), this proves that the cavity must be made with a nanotechnology material.

The calculation of the clean energy inside the cavity indicates that the maximum storage is reached when the self-adjoint extension parameter tends to infinity this corresponds to R tending to zero. This is done experimentally by making the inside surface of the cavity perfectly reflective, and also impervious to the passage of electrons to the outside of the cavity. On the other hand, if this parameter tends towards (or equal) zero, particle leakage is important, which causes significant losses of energies (energy stored equal zero for$\gamma = 0$).

7. References

[1] J.-M. Joubert, F.Cuevas, M. Latroche et A. Percheron-Guégan, Stockage de l'hydrogène et risques, Journée du CUEPE 2005 « L'hydrogène, futur vecteur énergétique ? » Genève le 13 mai 2005.

[2] A. Michels, J. de Boer, A. Bijl, Remarks concerning molecular interaction and their influence on the polarisability, Physica 4, 1937, 981.

[3] A.Sommerfeld, H. Welker, Advances in Quantum Chemistry: Theory of Confined Quantum Systems - Part One, Ann. Phys. 424, 1938, 56.

[4] S. R. de Groot, C.A. Ten Seldam, On the energy levels of a model of the compressed hydrogen atom, Physica 12, 1946, 669.

[5] E.P. Wigner, Application of the Rayleigh-Schrödinger Perturbation Theory to the Hydrogen Atom, Phys. Rev. 94, 1954, 77.

[6] P. W. Fowler, Energy, polarizability and size of confined one-electron systems, Mol. Phys. 53, 1984, 865.

[7] P.O. Frömann, S. Yngve, N. Frömann, The energy levels and the corresponding normalized wave functions for a model of a compressed atom, J. Math. Phys. 28, 1987, 1813.

[8] A.V. Scherbinin, VI. Pupyshev, Electronic Structure of Quantum Confined Atoms and Molecules, Russ. J. Phys. Chem. 74, 2000, 292.

[9] M.H. Al-Hashimi, U.-J. Wiese, Self-adjoint extensions for confined electrons: From a particle in a spherical cavity to the hydrogen atom in a sphere and on a cone, Annals of Physics 327, 2012, 2742-2759.

[10] M.H. Mousa. Calcul efficace et direct des représentations de maillage 3D utilisant les harmoniques sphériques. Thèse de doctorat en informatique, Université Claude Bernard Lyon 1, septembre 2007.

[11] Claude Cohen-Tannoudji, Bernard Diu & Franck Laloë ; Mécanique Quantique, (1973).

[12] Albert Messiah ; Mécanique Quantique, 2 volumes, Dunod (1959). Réédité en 1995.

[13] W. E. Byerly. Spherical Harmonics, chapter 6, pages 195–218. New York : Dover, 1959. An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.

[14] E. W. Hobson. The Theory of Spherical and Ellipsoidal Harmonics. New York : Chelsea, 1955.

[15] L. Landau and E. Lifchitz, Mécanique quantique. Edition MIR, Moscou (1966).

[16] L. Landau and E. Lifchitz, Théorie quantique relativiste. Edition MIR, Moscou 1972.

[17] Kamel. Idris-Bey, Quantum study of hydrogen stored under high pressure in a spherical cavity, 8th International Conference on Modelling, Identification and Control (ICMIC-2016) Algiers, Algeria- November 15-17, 978-0-9567157-6-0 © IEEE 2016.

[18] Ravindra Shinde and Meenakshi Tayade; Remarkable Hydrogen Storage on Beryllium Oxide Clusters: First Principles Calculations; Department of Physics, Indian Institute of Technology Bombay, Mumbai, Maharashtra 400076, INDIA., and Department of Chemistry, Institute of Chemical Technology, Mumbai, Maharashtra 400019, INDIA, 2016.

[19] Lueking, A.; Yang, R. T. Hydrogen Spillover from a Metal Oxide Catalyst onto CarbonNanotubes—Implications for Hydrogen Storage. J. Catal. 2002, 206, 165–168.

[20] Li, Y.; Yang, R. T. Hydrogen Storage in Metal–Organic Frameworks by Bridged Hydrogen Spillover. J. Am. Chem. Soc. 2006, 128, 8136–8137.

[21] Zhou, J.; Wang, Q.; Sun, Q.; Jena, P. Enhanced Hydrogen Storage on Li Functionalized BC3 Nanotube. J. Phys. Chem. C 2011, 115, 6136–6140.

[22] Dodziuk, H.; Dolgonos, G. Molecular modeling study of hydrogen storage in carbon nanotubes. Chem. Phys. Lett. 2002, 356, 79 – 83.

[23] Li, C.; Li, J.; Wu, F.; Li, S.-S.; Xia, J.-B.; Wang, L.-W. High Capacity Hydrogen Storage in Ca Decorated Graphyne: A First-Principles Study. J. Phys. Chem. C 2011, 115, 23221– 23225.

[24] Liu, C.; Chen, Y.; Wu, C.-Z.; Xu, S.-T.; Cheng, H.-M. Hydrogen storage in carbon nanotubes revisited. Carbon 2010, 48, 452 – 455.

[25] Wu, H.-Y.; Fan, X.; Kuo, J.-L.; Deng, W.-Q. DFT Study of Hydrogen storage by Spillover on Graphene with Boron Substitution. J. Phys. Chem. C 2011, 115, 9241–9249.

[26] J.R. Sabin, E. Brändas, S.A. Cruz (Eds.), Theory of Confined Quantum Systems, in: Advances in Quantum Chemistry, vol. 57, Academic Press, Elsevier, 2009.

[27] G. ZEPP; Exercices de Mécanique Quantique, Vuibert (1975).

[28] Eric W. Weisstein. Legendre Polynomial. From MathWorld : A Wolfram Web resource.

[29] J. Hladik, M.Chrysos, P. E. Hladik, L. U. Ancarani, Mécanique quantique : Atomes et noyaux, applications technologiques, 3^e Edition. DUNOD (2009).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 164-173 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Improving Patient Outcomes Through Untethered Patient-Centered Health Records

Mohammed Abdulkareem Alyami[*,1], Majed Almotairi[1], Alberto R. Yataco[2], Yeong-Tae Song[1]

[1]*Dept. of Computer & Information Sciences, Towson University, 21252, USA*

[2]*Medical Director Get Well Immediate Care Towson, 21204, USA*

ARTICLE INFO

Article history:
Received: 16 November, 2017
Accepted: 11 March, 2018
Online: 27 March, 2018

Keywords:
Personal Health Records (PHRs)
Personal Health Records Systems (PHRS)
Dublin Core Metadata
Cloud storage
Standard Medical Codes

ABSTRACT

Patient generated data, or personal clinical data, is considered an important aspect in improving patient outcomes. However, personal clinical data is difficult to collect and manage due to its distributed nature. For example, they can be located in multiple places such as doctors' offices, radiology centers, hospitals, or some clinics. Another factor that can make personal clinical data difficult to manage is that it can be heterogeneous data types such as text, images, charts, or paper-based documents. In case of emergencies, this situation makes personal clinical data retrieval very difficult. In addition, since the amount and types of personal clinical data continue to grow, finding relevant clinical data when needed is getting more difficult if no action is taken. In response to such scenarios, we propose an untethered patient health record system that manages personal health data by utilizing meta-data that enables easy retrieval of clinical data. We incorporate cloud-based storage for easy access and sharing with caregivers to implement continuity of care and evidence-based treatment. In emergency cases, we make critical medical information such as current medications and allergies available to relevant caregivers with valid license numbers only. Clinical data needs to be stored or made accessible from one place for easy sharing and retrieval. Well-managed personal cloud space could outlive the lifetime of personal health records system (PHRS) since the discontinuity of the service does not affect the data stored in the cloud space. In our approach, we separate the clinical data from applications in order to make the data independent from the application. Also, the users can have alternative applications for their clinical data. Such independence motivates users to use PHRS with flexibility.

1. Introduction

For most people, healthcare is considered important as there has been significant increase in chronic diseases such as heart disease, cancer, diabetes and asthma. This requires continuous treatment, reduces quality of life, and increases overall medical expenses (The Growing Crisis of Chronic Disease in the United States) [1]. According to the Centers for Disease Control and Prevention (CDC), in the U.S. about 610,000 people die of heart disease every year. In addition, 26 million people suffer from Type I or Type II Diabetes, around 14 million have severe chronic respiratory problems such as Chronic Obstructive Pulmonary Disease (COPD), and 68 million have been diagnosed with hypertension [2]. However, many of these diseases can be prevented and managed through early detection, physical activities, a balanced diet and treatment therapy. Adopting PHRS could help improve patient outcomes. Recently, there has been more focus on preventive care and proactive measures such as monitoring and controlling patients' symptoms. Nowadays, there are many mobile health applications and sensors such as blood pressure sensors, electrocardiogram sensors, blood glucose measuring devices, and others that are used for monitoring and controlling personal health. These apps and sensors produce personal health data that can be used for treatment purposes. If managed and handled properly, it can be considered patient-generated data. There are other types of personal health data that are available from various sources such as hospitals, doctor's offices, clinics, radiology centers or any other caregivers. Aforementioned health documents are deemed as a personal health record (PHR). According to American Health Information Management Association (AHIMA) [3], PHR can be defined as an electronic, lifelong resource of health information

[*]Mohammed Abdulkareem Alyami, Email: aljakman2006@hotmail.com

https://dx.doi.org/10.25046/aj030219

needed by individuals to make health decisions. However, it is not easy to collect all the relevant personal health data because of the fact that they are in different data types, available from different sources, and stored in different media and devices. To overcome such difficulties, it is desirable to have personal health data in one place where users have full control over their own clinical data. In order to be useful, the clinical data should be sharable when needed for diagnosis and treatment. Without proper clinical information (medical history, allergies, current medications, and adverse reactions) medical mistakes could occur when making medical decisions due to insufficient information. Even if a patient has a complete medical history and all the necessary clinical data, if it is not shared properly among caregivers at the time of need, discontinuity in care may occur. In order to meet the needs of such scenario, PHRS should have the following properties: robust and private storage, easy retrieval and maintenance, secure, sharable, and able to handle emergency situations.

There are two types of PHRS: untethered and tethered. Untethered PHR is an independent PHRS where patients have full control over their own personal health records. They can collect, manage, and share their health records. On the other hand, tethered PHRS are linked to a specific healthcare providers' EHR system, where the users typically gain easy access to their own records through secure portals and see their own clinical information such as test results, immunization records, family history, and other relevant information. They can also utilize secure messaging with their collaborating clinicians. The participating patients need to share the cost and the information with their care provider. However, these tethered records may not be complete since the information sources are from one provider only. Despite all the benefits PHRS provide, the adoption rate of PHR by the general public still remains low in the U.S. [4-6]. In our previous work [7], we identified six barriers (usability, ownership, interoperability, privacy and security, portability and motivation) that cause the slow adoption of PHRS. One of the main concerns for not having PHRS is the ownership issue. For example, one incident happened on January 1, 2012. Google Health™ System, one of the biggest PHRS providers, stopped its service and asked their registered customers to move their records to their computers or other PHRS vendors, which exacerbated ownership concerns by the public.

As an attempt to overcome some of the barriers, we propose an untethered PHRS that utilizes personal cloud storage, offers simplicity in organizing various kinds of clinical data by utilizing Dublin Core (DC) metadata, and provides easy access to emergency clinical data to paramedics or clinicians in case of emergency. DC metadata has been successfully applied in many areas, but since it is not specifically designed for clinical data, there are some limitations in its expressive power in the healthcare domain. In this research, we simplified the categorization of clinical data by human body part for easy retrieval of clinical data using DC, so users can manage their own clinical data without in-depth knowledge about clinical information. As a proof of our concept, we developed a system called My Clinical Record System (MCRS) to help users store, organize, retrieve, and share their clinical data with caregivers when needed, including in emergency situations. In an emergency situation, clinicians (e.g. physicians, paramedics, nurses, and others) can access patient's data using their license numbers and the patient's name and date of birth. Emergency information consists of current medication lists, known allergies, and side effects. By having complete medical history, MCRS users may be able to reduce medical errors and improve patients' outcomes. It also ensures continuity of care by sharing personal clinical data among healthcare providers when needed. The remainder of this paper is organized as follows: Section 2 discusses related work and in section 3 we focus on the clinical data's type and format. In Section 4, we discuss how to solve the data organization and retrieval issues using DC metadata to facilitate better and more accurate data retrieval. Section 5 discuses cloud-based storage. In section 6, we discuss the current situation in the healthcare industry (AS-IS). Section 7 uses a scenario as an example. Section 8 discusses the solution domain. In section 9, we introduce MCRS as a proof of concept and finally we conclude our study and discuss our future work in section 10.

2. Literature Review

In [8], Fearon defined Meta-data as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. However, metadata standards have not been employed by many repositories and most of the meta-data was generally descriptive, rather than administrative or for preservation [9]. In [10], Greenberg investigated the ability of resource authors to create acceptable metadata in an organizational web site at the National Institute of Environmental Health Sciences in the U.S. The findings of their study indicate that resource authors can create better quality metadata when working with Dublin Core. In some circumstances, they may be able to create quality metadata better than a metadata professional. In [11], Talha developed their own metadata tool called Metadata Management System (MMS) to facilitate the creation, maintenance and storage of metadata. MMS supports two well-known metadata models, Dublin Core and SCORM 1.2 (IEEE Learning Object Metadata). The authors implemented their metadata tools in the Malaysia Grid for Learning (MyGfL). The results of the study indicate that MMS can substantially improve the discovery, retrieval management and control of web resources and learning objects in their MyGfL portal. In the healthcare domain, meta-data has been utilized as a method for confidentiality tags that indicate data sensitivity levels. This enables patients to give consent to the exchange of some parts of their health records (e.g. the medical diagnosis), while withholding consent for the exchange of other areas (e.g. a mental health counseling session) [12]. This work can help in increasing patients' privacy and allow them with some freedom in exchanging their health records. However, their work focused on EHR, but did not incorporate PHRS, which limited the scope of their study. Other researchers have adopted the ontology approach to quickly search and access relevant and meaningful information among large numbers of CDA documents within healthcare providers' systems (Electronic Health System), which in turn enables semantic interoperability[5, 13]. However, these studies were limited to one health data type (CDA), whereas health records include other data types such as images (e.g. x-rays, scanned documents, ultrasounds, and others), and observed symptoms noted by patients and clinical sensors' data. In [14] Patel developed a system called TrialX, on top of PHR, where patients not only can search by keywords, as in ClinicalTrials.gov, but also by demographics (e.g. age, gender, city and study site). This system enables patients to match their health condition to clinical trials. This system can accelerate and improve the search results among health records. However, the authors did not take into account how to retrieve relevant clinical data since the amount and types of personal clinical data continue to grow, which makes it difficult to find such data. In [15], Appelboom reviewed the literature on smart wearable body sensors and found that these sensors are accurate and have clinical utility, but still are underutilized in the healthcare industry. These devices can be used to monitor physiological, cardiovascular and many other factors of health variables and transmit data either to a personal device or to an online storage site. The smart wearable body sensors are placed on different parts of the user's body based on the purpose of the sensor device. For instance, the physical therapy sensor is placed on the ankle; the cardiopulmonary sensor can be placed on the wrists, fingers, arms or thighs.

In [16], Zhang developed an application to apply meta-data efficiently on clinical trial data. The authors chose Microsoft Excel due to the wealth of built-in features (e.g. spell checking, sorting, filtering, finding, replacing, importing and exporting data capabilities), which contribute to the ease of use, power, and flexibility of the overall meta-data application. They focused on the analysis process in a drug development environment such as adverse clinical events (ACE), Electrocardiogram (ECG), laboratories (LAB), and vital signs (VITAL), where the raw data is stored in the clinical trial database and then the data can be manipulated.

Another study in [17], Teitz developed a website called HealthCyberMap with the goal of mapping Internet health information resources in novel ways for enhanced retrieval and navigation. They used Protégé-2000 to model and populate a qualified DC RDF meta-data base. They also extended the DC elements by adding quality and location elements. Also, the W3C RDFPic project extends the DC schema by adding its own elements such as camera, film, lens and film development date for describing and retrieving digitized photos. In [18], Ekblaw built a system (RedRec) to enable patients to access their medical health records across health providers (e.g. pediatrician, university physician, dentist, employer health plan provider, specialists, and others). Their system applies novel, blockchain smart contracts to create a decentralized content-management system for healthcare data across providers. RedRec governs medical records access while providing the patient with the ability to share, review, and post new records via flexible interface. The raw medical record content is kept securely in providers' existing data storage. However, when the patient wants to retrieve data from their provider's database, their Database Gatekeeper checks authentication. If it is approved, the Gatekeeper retrieves the relevant data for the requester and allows a sync with the local database.

In order to solve interoperability problems in exchanging clinical data, in [19], Dogac proposed archetypes to overcome the interoperability problems. They provided guidelines on how ebXML Registries can be used as an efficient medium for annotating, storing, discovering and retrieving archetypes. They also used ebXMLWeb services to retrieve data from clinical information systems. An archetype is defined as "a reusable, formal expression of a distinct, domain-level concept such as "blood pressure", "physical ex-amination", or "laboratory results", expressed in the form of constraints on data whose instances conform to some reference model"[19]. However, these studies are not comprehensive. Their limitations come from: focus only on one health data type, do not take into account PHR, lack of retrieving relevant clinical data, and overlook emergency access to medical records. Our proposed approach overcomes these limitations. Our study is comprehensive, which covers many different clinical and nonclinical documents such as images (e.g. x-rays, scanned documents, ultrasounds, and others), text (e.g. CDA, CCR, CCD, and others), and observed symptoms noted by patients and their clinical sensors' data. This can organize these various data types in a way that can help in storing and retrieving such data in an efficient way.

3. Clinical Data

In this section, we describe the types, formats, and sources of clinical data.

3.1. Measurements data from portable medical devices, sensors, or mobile application

One way to collect measurement data is through explicit clinical sensors. A clinical sensor is a device that responds to a physical stimulus and transmits a resulting impulse for interpretation or recording. Some sensors are designed to work outside the human body, while others can be implanted within the human body [20]. In this research, we are referring to clinical sensors for homecare settings, such as blood oxygen monitors, thermometers for body temperature, heart rate sensors, blood pressure sensors, and others. In addition to these textual data types, there can be non-textual data generated from sensors such as electrocardiogram measurement devices. The clinical sensors play a major role in healthcare, including early detection of diseases, diagnosis, disease monitoring and treatment monitoring.

Another method to collect measurement data is through mobile apps. For instance, most smartphones (e.g. Android or IOS) offer health and fitness apps that help users monitor their daily activities and health (e.g. track diet and nutrition calories, track vital signs, track fitness progress, share health data with their doctor electronically, and others). The data collected from these applications can be sent as a message or an email attachment to whom the users want to share it with. For interoperability, the collected data needs to be in standard format, such as HL7 CDA or in standard code such as SNOMED-CT.

3.2. Observed Symptoms

Patients sometimes experience particular symptoms (e.g. chest pain, nausea, vomiting, shortness of breath and others). If the patient notices such symptoms, they should be recorded and shared with their physician for proper treatment. If these symptoms are not shared with their physician, due to incomplete information, misdiagnosis could occur. When patients are recording, the observed symptoms should be described in standardized code such as SNOMED-CT. This will allow semantic interoperability, since the same symptoms can be described in multiple ways. Without codified descriptions, there can be discrepancies about the perceptions regarding symptoms between patients, nurses, or physicians [21].

3.3. Images

Most of the medical imaging machines produce standard image format called Digital Imaging and Communications in Medicine (DICOM). DICOM is defined as the international standard for medical images and related information (ISO 12052). There are two types of clinical data images: images that are based on DICOM standard (e.g. x-rays, Computed Tomography (CT), Magnetic Resonance (MR), and ultrasound devices) and scanned documents. The DICOM –format combines images and meta-data that describes the medical imaging procedure. Accessing data in DICOM files becomes as easy as working with TIFF or JPEG images [22]. On the other hand, the scanned documents (e.g. PDF/ JPEG) are difficult to retrieve because the content is not searchable. For example, some physicians write notes on clinical forms while diagnosing their patients and then type them on the computer. They also sometimes scan the notes and upload them to the patient records. Either way is time consuming, difficult to retrieve in a timely manner, and consumes relatively large storage space. In addition, the patient may have more than one doctor or may have been treated by many healthcare providers, which in turn fragments his/her records. So when patients obtain their records, they mostly receive them either printed out or sent as an email attachment. This makes it difficult to retrieve scanned documents because its content cannot be retrieved by computers. To alleviate such issues, we have utilized meta-data to describe such medical documents so computerized retrieval and systematic organization are possible.

3.4. Clinical Document

Clinical documentation (CD) is a computerized record describing a medical treatment, medical trial or clinical test, which can be exchanged among healthcare providers [23]. EHR data may be collected from healthcare providers. There are four types of clinical document formats: Continuity of Care Record (CCR), Clinical Document Architecture (CDA), Continuity of Care Document (CCD), and Consolidated CDA (C-CDA). All of which allow healthcare providers to exchange clinical information summary about a patient. However, CCR was excluded from the 2014 edition of EHR Certification, which is a standard certification criteria for EHRs that was established by The Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health Information Technology (ONC), as a valid way to send summary of care documents. Hence, the content from a CCR was merged into a CDA format and called Continuity of Care Document (CCD). C-CDA includes nine different types of commonly used CDA documents such as CCD, consultation notes, discharge summary, imaging integration, DICOM diagnostic imaging reports, history and physical, operative note, progress note, procedure note, and unstructured documents. Each C-CDA Document Template was designed to satisfy a specific information exchange situation.

In the following subsection, we will discuss the difference between these formats (CCR, CDA, and CCR):

- CCR documents provide a snapshot of treatment and basic patient information such as diagnosis and reason for referral between healthcare providers. It also uses only specified XML code and does not support narrative text (free-text), which hinders physicians from writing notes if needed, and it is not electronically acceptable by all systems.
- CDA documents are a flexible standard which can be read by the human eye or processed by machine due to the use of XML language. It is based on the HL7 Reference Information Model (RIM), and uses HL7 V3 data types. CDA can be transported using different methods such as HL7 V2, HL7 V3, DICOM, MIME-encoded attachments, HTTP, or FTP. CDA documents can include text, images and even multimedia.
- CCD documents are not a complete medical history of the patient but it includes only the information critical to effectively provide continuity of care. Its primary purpose is exchanging patient information between different healthcare providers. It is based on XML standard and can be displayed on a web browser using style sheet. It also allows narrative text which is an advantage over other standards like CCR [24].

When using untethered PHRS, patients are responsible for collecting clinical data from their healthcare providers or from their own patient-generated measurement data and keeping it in their own personal cloud space. For example, CDA, CCR and CCD can be obtained from healthcare providers, X-rays can be obtained from the radiology department, and lab test results can be obtained from the test lab or doctor’s office. Patients can share their health records with their clinicians by either electronically transmitting or granting access to their storage through the PHRS. If electronic sharing is not allowed, the patient may download the file and make hardcopies or store them in a USB, CD, or other mediums for sharing [25].

3.5. Meta-data

There are two different methods of storing meta-data. In the first method, meta-data can be embedded in the data (e.g. in the header of a digital file). The advantages of this option are ensuring that the meta-data will not be lost, eliminating the need for linking data and meta-data, and updating the data and meta-data together. In the second method, meta-data can be stored separately. The advantage of this option is that it can simplify the management of meta-data and can expedite the retrieval of the data [8]. In our approach, we employed the latter method to accelerate the retrieval of clinical data and to enhance expressive power. However, in this method, there can be inconsistencies between meta-data and clinical data. This can occur when transitioning to a new platform, integration between different systems or sharing data across multiple systems [12]. There are many meta-data formats that have been accepted internationally including: Dublin Core (DC), Federal Geographic Data Committee (FGDC), Encoded Archival Description (EAD), and Government Information Locator Service (GILS) [26]. Some of the metadata standards/schemas are generic, while others are domain-specific. For example , SCORM 1.2 (IEEE LOM) is used for educational approaches and rights management; Friend of a Friend (FOAF) is used for people and organizations; Simple Knowledge Organization (SKOS) is used for concept collections; Asset Description Metadata Schema (ADMS) is used for describing semantic interoperability assets; and Dublin Core is used for published material (text, images).

Generic metadata formats, such as Dublin Core, tend to be easy to use and widely adopted. In this study, we adopted DC metadata because it was developed for author-generated metadata, supports resource sharing and interoperability among information systems, has the broadest level of commonality of elements, extensibility, international acceptability and the flexibility it provides for extensions to the basic elements.

DC metadata solves one of the major issues in using meta-data among healthcare systems, which lacks interoperability [7]. Therefore, among these different metadata formats, DC metadata is the most appropriate format that aligns with our approach.

Meta-data benefits personal health record management in many ways. These benefits include the following:

- Consistency in definitions: properly defined tags provide structured information about the clinical data stored.

- Clarification of the relationships: meta-data can be used to clarify the relationships among the clinical items by defining categories and associated relationships within the category. We have defined each tag in DC for clinical purposes. When the data is uploaded or modified, the corresponding meta-data is required to update as well.

3.6. Meta-data Management

Meta-data management ensures that the data is associated with the datasets and utilized efficiently throughout and across organizations. Data governance is needed for successful meta-data implementation so it can provide trustworthy, timely and relevant information to decision makers, as well as personal users. For successful implementation, data governance must be aligned with the intended purposes of the users or organizations. This means that a data governance program starts by specifying its strategy, goals and the scope of its success. An organization needs to define the three data governance pillars including policies, people (and people skills), and processes. Once the above steps are in place, the organization can determine the best tools and technology to implement its data governance initiative [27].

3.7. Dublin Core Meta-data for Clinical Use

The DC Meta-Data Initiative (DCMI), is an open organization supporting innovation in meta-data design and best practices across the meta-data ecology. The DC Meta-data consists of 15 optional elements including: title creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage and rights.

In this study, we defined the usage of DC Meta-data elements for clinical purposes to describe and retrieve clinical data efficiently as shown in Table 1. Some of the meta-data elements - title subject, description, type, data and resource - are required for the integrity of data. These elements must be present for every clinical data item. The optional fields can be skipped, but if it has been filled, the metadata quality will be increased.

Table 1: Meta-data schema for PHR

Entity	Description
DC. Title	The title of the document
DC. Creator	The author of the document
DC. Subject	Subject of the document
DC. Description	Description of the document
DC. Relation	- One of the body parts (Thorax, Abdomen, Heart, extremities, Integumentary, Head, Urinary or Reproductive) - This element is linked to the subject element
DC. Date	Date of meta data creation
DC. Type	Type will be used for lab tests (blood work, urinalysis, fecal sample, nasopharyngeal sample, oropharyngeal sample and others) or images (x-ray, cat scan/ CT, Ultrasound, Magnetic resonance/MR, Scanned Document, Electrocardiogram, EKG/ or ECG, CDA, C-CCD and others)
DC. Format	PDFs, Text, JPEGs , TIFFs, HL7 CDA, and others
DC. Identifier	Optional Document ID
DC. Language	English and other languages
DC. Coverage	Geographical and time-related information
DC. Rights	Copyright and access rights (secured or unsecured)
DC. Source	Data source

4. Healthcare Processes

4.1. AS-IS Healthcare Processes

Some of the issues that the current healthcare industry is having include discontinuity of care and unacceptably high rates of medical mistakes due to unavailable patient medical records at the time of need. Some of the factors that cause these issues are listed below and illustrated in the Figure 1. (the numbers listed below correspond to the numbers in the Figure 1)

- Personal clinical data is difficult to collect and manage because they are located over multiple places such as doctor’s offices, radiology centers, hospitals, or some clinics (1).

- Heterogeneous data types such as text, images, charts, or paper- based documents (2) [8].

- Discontinuity of care due to lack of communication among caregivers. This is caused by distributed and fragmented medical information (3).

- Lack of evidence-based treatment due to limited access to medical records (4).

- Medical errors due to incomplete medical history or access to emergency health information (e.g. allergies, current medication list, side effects, and others) at the time of need (5). According to Sunyaev. [28], most PHRS do not offer built-in emergency access to the record, except through third-party services that are available for HealthVault. For example, Microsoft Health Vault provides users with

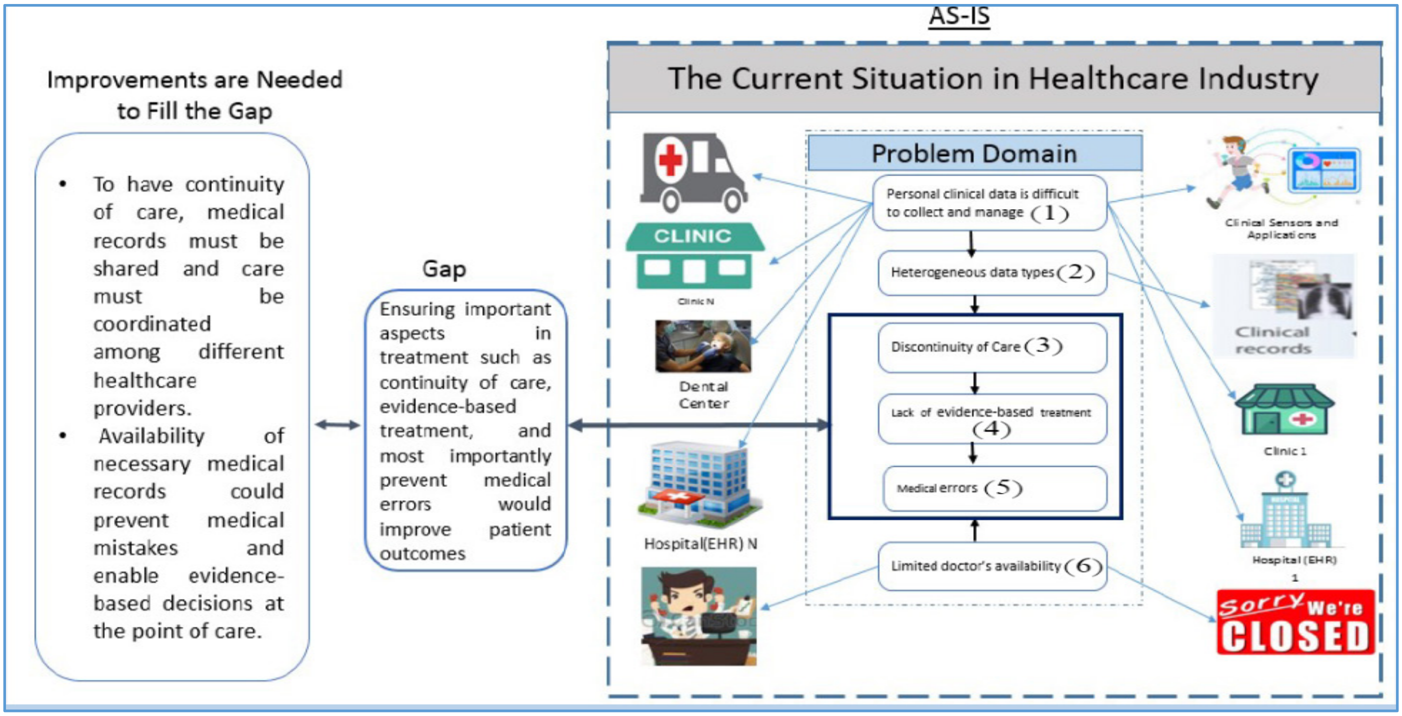

AS-IS
Improvements are Needed to Fill the Gap
• To have continuity of care, medical records must be shared and care must be coordinated among different healthcare providers.
• Availability of necessary medical records could prevent medical mistakes and enable evidence-based decisions at the point of care.
Gap
Ensuring important aspects in treatment such as continuity of care, evidence-based treatment, and most importantly prevent medical errors would improve patient outcomes
The Current Situation in Healthcare Industry
Problem Domain
Personal clinical data is difficult to collect and manage (1)
Heterogeneous data types (2)
Discontinuity of Care (3)
Lack of evidence-based treatment (4)
Medical errors (5)
Limited doctor's availability (6)
CLINIC
Clinic N
Dental Center
Hospital(EHR) N
Clinical Sensors and Applications
Clinical records
Clinic 1
Hospital (EHR) 1
Sorry We're CLOSED

Figure 1: Problem Domain in the Current Healthcare Industry

access codes that can be given to emergency responders and other people they trust to allow access to the emergency information.

- Limited doctor availability. For instance, patients may need to be seen on the weekend or on a holiday when their doctor's office is closed (6).

4.2. Scenario for the Need of PHRS

John was suffering from tiredness and lack of energy for the past 4 weeks. He has several chronic conditions (type 2 diabetes, chronic kidney disease, chronic lower back pain, generalized anxiety disorder, depression, bipolar disorder, dyslipidemia, hypothyroidism, coronary artery disease and congestive heart failure) and is on multiple medications prescribed by his PCP, cardiologist, psychiatrist and pain management doctor. John had recently requested to become one of Dr. Smith's patients. After multiple attempts, Dr. Smith was able to obtain some of his previous medical records. Dr. Smith also was interested in reviewing his previous medical diagnoses and prior/current medications. Unfortunately, there was no integration of the records and medications taken. After several interviews with John and his wife, Dr. Smith was able to determine that the patient was using the same type of long-acting insulin twice a day because one had the generic name and the other the commercial or brand name. The patient was supposed to use this insulin once a day only. This patient's error kept his glucose at very low levels in blood which led to constant tiredness and lack of energy. Once the dose was corrected, the patient felt better. The immediate access to medical and prescription information would have allowed Dr. Smith to identify the error faster and provide him with the ability to take prompt corrective measures.

4.3. Solution Concept (TO-BE) Processes

To have continuity of care, medical records must be shared and care must be coordinated among different healthcare providers. Availability of necessary medical records could help prevent medical mistakes and enable evidence-based decisions at the point of care. It would be convenient to have clinical data stored in the same place for easy sharing and retrieval. Well-managed personal cloud space could outlive the lifetime of PHRS since clinical data is stored independently. In our approach, we separate the clinical data from applications to make the data independent from the application. Also, the users can have alternative applications to access their clinical data. Such independence helps clinical data outlive its applications. Our proposed solution concepts are illustrated in the Figure 2. In the Figure, the clinical data is separated from the application for data independence. The numbers in the Figure 2 correspond to the problems listed in the Figure 1.

5. My Clinical Record System (MCRS): A Proof of Concept

As mentioned in the introduction, there are a number of obstacles in collecting and maintaining personal health data. In an attempt to remove such obstacles, we developed a web-based system called My Clinical Record System (MCRS) that can help users to upload, organize, share, and retrieve relevant health data.

5.1. Overview of MCRS

Some of the main features of MCRS are listed below:

- Users can search their cloud storage not only by keyword, but also by any of the meta-data elements using document finder search features. They also can search by two elements such as subject and date in order to filter data by showing more relevant data. Additionally, they can find a group of records based on a date range they specify.

- MCRS allows users to share their health records with their physicians using the share document feature as shown in the Figure 3.
- MCRS helps users to create meta-data for any documents (textual and non-textual data) and upload them to their cloud storage. For the non-textual data generated from sensors such as electrocardiogram (ECG) measurement devices, the users can scan them or take a picture and then upload them as an image or use a plotted number. Thus, the user will be able to include such data into their PHR and share it with their healthcare providers. MCRS also helps to retrieve those documents easily and can direct the users to its location if more information is needed.
- To overcome the ownership barrier, we separate the clinical data from applications, which will give the users more freedom by not limiting them to one provider or application. Also, their data is saved on their own storage, thus we do not have to store it in our system.
- To overcome the interoperability barrier, we used DC standards to describe any clinical data using our tag definition. The DC meta-data content for doctor visit summary document is shown in the Figure 4 and in the Figure 5.

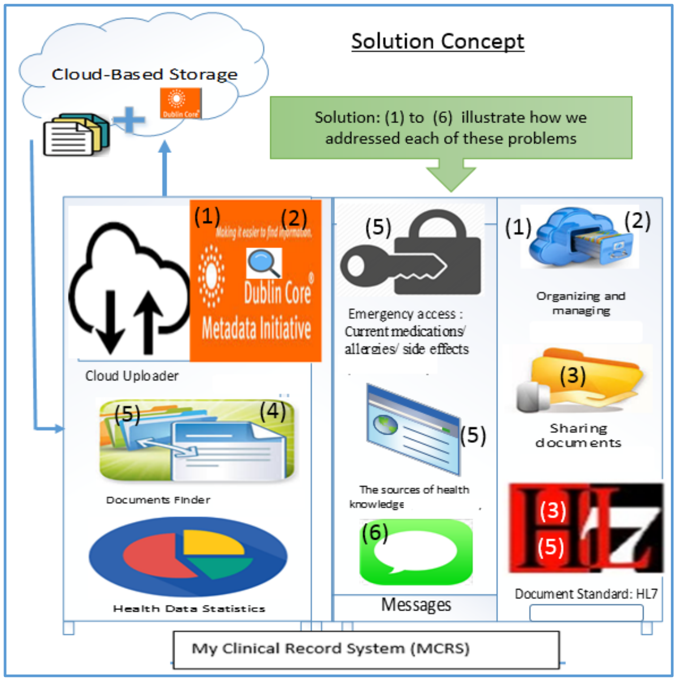
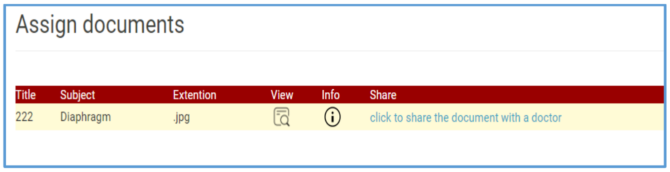

Solution Concept
Cloud-Based Storage
Dublin Core
Solution: (1) to (6) illustrate how we addressed each of these problems
(1)
(2)
Making it easier to find information.
Dublin Core
Metadata Initiative
Cloud Uploader
(5)
(4)
Documents Finder
Health Data Statistics
(5)
Emergency access : Current medications/ allergies/ side effects
(5)
The sources of health knowledge
(6)
Messages
(1)
(2)
Organizing and managing
(3)
Sharing documents
(3)
(5)
Document Standard: HL7
My Clinical Record System (MCRS)

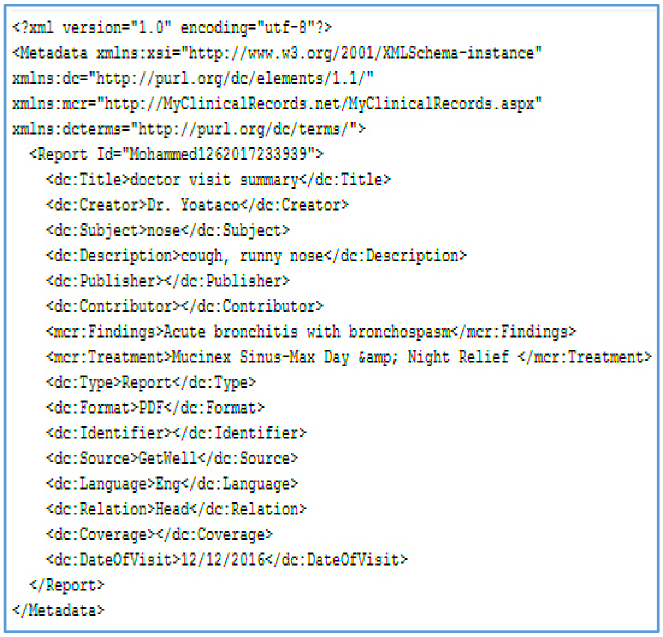
Figure 2: Solution Concept



Assign documents

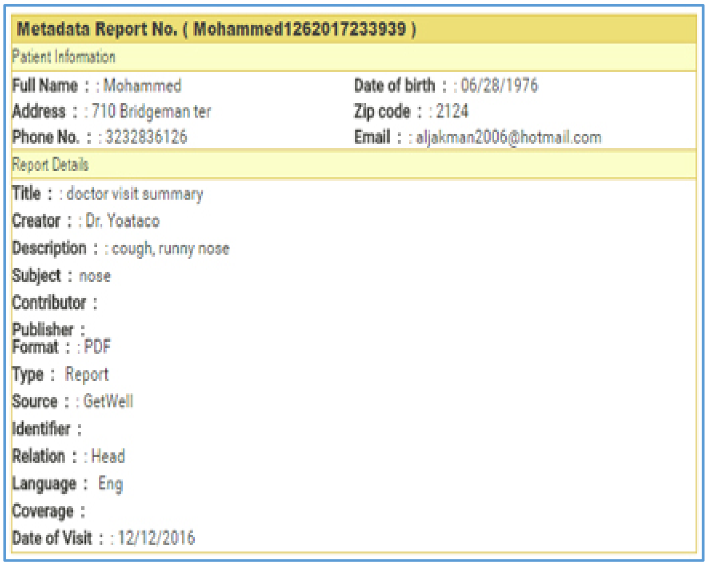
Title	Subject	Extention	View	Info	Share
222	Diaphragm	.jpg	🔍	ⓘ	click to share the document with a doctor

Figure 3: Sharing Health Records with Physician



```
<?xml version="1.0" encoding="utf-8"?>
<Metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:mcr="http://MyClinicalRecords.net/MyClinicalRecords.aspx"
xmlns:dcterms="http://purl.org/dc/terms/">
  <Report Id="Mohammed1262017233939">
    <dc:Title>doctor visit summary</dc:Title>
    <dc:Creator>Dr. Yoataco</dc:Creator>
    <dc:Subject>nose</dc:Subject>
    <dc:Description>cough, runny nose</dc:Description>
    <dc:Publisher></dc:Publisher>
    <dc:Contributor></dc:Contributor>
    <mcr:Findings>Acute bronchitis with bronchospasm</mcr:Findings>
    <mcr:Treatment>Mucinex Sinus-Max Day &amp; Night Relief </mcr:Treatment>
    <dc:Type>Report</dc:Type>
    <dc:Format>PDF</dc:Format>
    <dc:Identifier></dc:Identifier>
    <dc:Source>GetWell</dc:Source>
    <dc:Language>Eng</dc:Language>
    <dc:Relation>Head</dc:Relation>
    <dc:Coverage></dc:Coverage>
    <dc:DateOfVisit>12/12/2016</dc:DateOfVisit>
  </Report>
</Metadata>
```

Figure 4: DC Meta-data Content for Doctor Visit Summary



Metadata Report No. (Mohammed1262017233939)

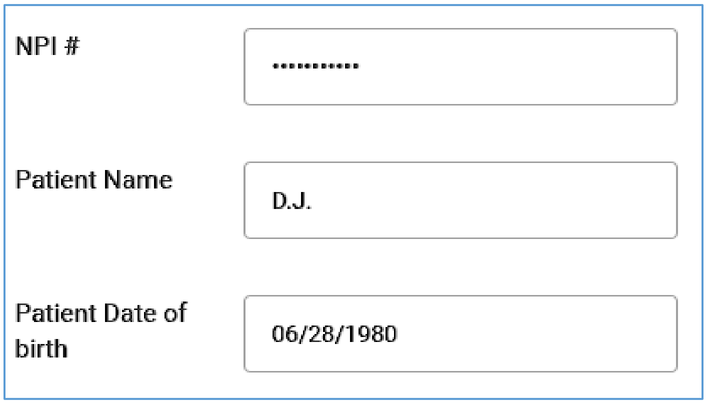
Patient Information

Full Name : : Mohammed	**Date of birth :** : 06/28/1976
Address : : 710 Bridgeman ter	**Zip code :** : 2124
Phone No. : : 3232836126	**Email :** : aljakman2006@hotmail.com

Report Details

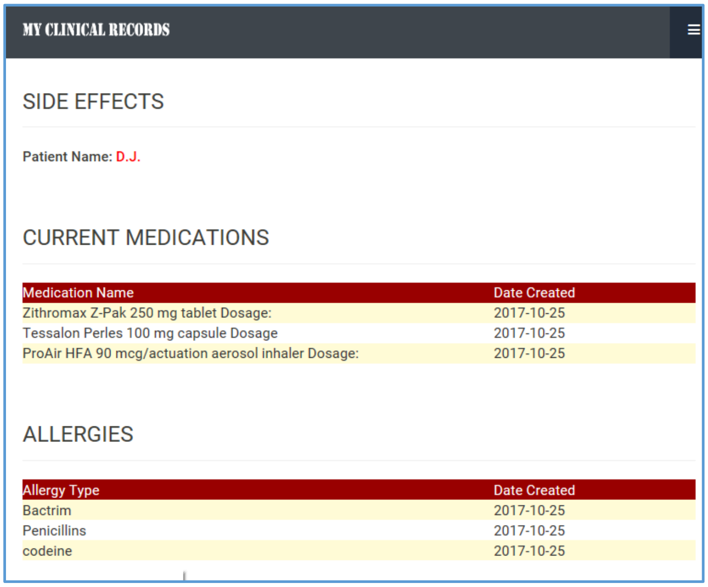
Title : : doctor visit summary
Creator : : Dr. Yoataco
Description : : cough, runny nose
Subject : nose
Contributor :
Publisher :
Format : : PDF
Type : Report
Source : : GetWell
Identifier :
Relation : : Head
Language : Eng
Coverage :
Date of Visit : : 12/12/2016

Figure 5: Visualization of the Meta-data



NPI #	•••••••••••
Patient Name	D.J.
Patient Date of birth	06/28/1980

Figure 6: Emergency Access with Valid License Number



MY CLINICAL RECORDS

SIDE EFFECTS

Patient Name: D.J.

CURRENT MEDICATIONS

Medication Name	Date Created
Zithromax Z-Pak 250 mg tablet Dosage:	2017-10-25
Tessalon Perles 100 mg capsule Dosage	2017-10-25
ProAir HFA 90 mcg/actuation aerosol inhaler Dosage:	2017-10-25

ALLERGIES

Allergy Type	Date Created
Bactrim	2017-10-25
Penicillins	2017-10-25
codeine	2017-10-25

Figure 7: Emergency Information

- The easy access to users' health data and the ability to contribute to their record enhances users' motivation to use PHR.
- MCRS enables emergency clinical data access by emergency crew only with valid license number. We use the National Provider Identifier (NPI), patient name, and date of birth for the emergency medical information access as shown in the Figure 6. MCRS uses the Application Programming Interface (API) that was provided by NPPES NPI Registry in order to verify the NPI. Emergency information contains allergies, current medication list, and side effects. This information is updated regularly by patients as shown in the Figure 7. It also contains any references to the time of the last update.
    - Healthcare providers apply for NPI using the National Plan and Provider Enumeration System (NPPES) [29].
    - NPI can be validated through NPPES NPI Registry.

5.2. Using MCRS

We use patient's Dropbox access token to allow the connection between Dropbox and MCRS, so patients can have their own storage and have the ability to provide access to their storage through MCRS when needed. This allows users to keep their own data without binding to any specific application. MCRS contains no clinical data as they are stored in the patient's cloud storage. Patients need accounts for Dropbox and MCRS separately.

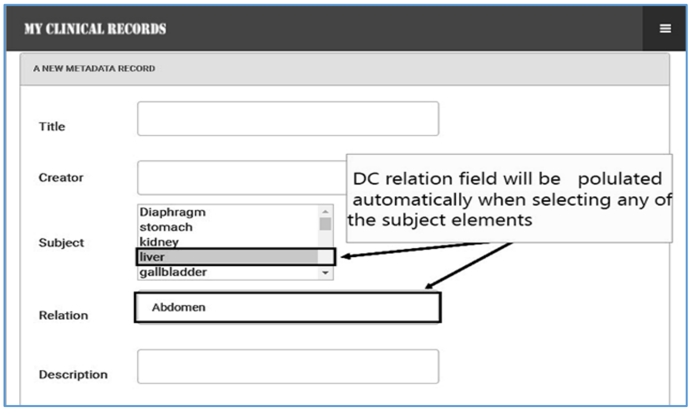
5.3. Personal Cloud Storage

Cloud storage is a place where users can store their data and access it anytime, from anywhere, and from any device via the Internet. It is maintained, managed and operated by cloud storage service providers such as Google, Amazon, or Microsoft Cloud storage services have many advantages such as cost savings, ease of use, ability to share data, accessibility, and sustainability. Personal Cloud Storage (PCS) is getting more popular because of the aforementioned convenience. Any cloud storage service provider may be used (SugarSync, Carbonite, IDrive, Dropbox, Google Drive, and others) – for storing personal clinical data as long as they provide required functionality and security. For testing purposes, we chose Dropbox™ as cloud storage. However, for practical use, secure cloud storage services that are HIPAA complaint can be used for privacy and security purposes, such as Dropbox (Business), Box, Google Drive, Microsoft OneDrive, and Carbonite [30]. The contents of each storage are described by DC meta-data for interoperability. In the case that data has embedded meta-data, we create another layer of meta-data so entire contents can be retrieved through our DC meta-data.

5.4. Managing Health Data

MCRS categorizes clinical data based on human body parts. There are eight categories: abdomen, heart, head, thorax, extremities, integumentary, urinary, and reproductive, as shown in Table 2. Any clinical data will be stored and linked based on these categories using the relation and subject tag elements of the DC meta-data. We kept the categories to a minimum so it can be simple enough to be used by patients. Users can specify the category (the human body part of interest) when searching for relevant clinical data, so it can show only the clinical documents (e.g. doctor visit summary, x-ray, and others) that are related to that part.

When using the relationships between the resource (DC subject) and target resource (DC relation), it is possible to combine the result to a greater scope, e.g. instead of eyes and ears, it can be categorized by head. This can be done by predefining each part of the human body and associating it with its related category in the system. Also, we have constrained the DC subject to a small core set that can be selected from a drop-down menu (all possible parts of the human body) to best describe the subjects (as shown in the Figure 8). When the users select the subject element, the DC relation field will be populated automatically with the associated part of its related category. For example, when a user searches by keyword (e.g. head) and chooses the element (e.g. relation), the search results will be filtered and show only all clinical documents that are relevant to the head (e.g. eyes, ears, brain, mouth, teeth, nose, and chin) as shown in the Figure 9. This is a less time-consuming method to filter the data instead of showing all documents as shown in the Figure 10. Also, the user can filter the search by date if they need to specify a period of time to find clinical documents.



MY CLINICAL RECORDS

A NEW METADATA RECORD

Title

Creator

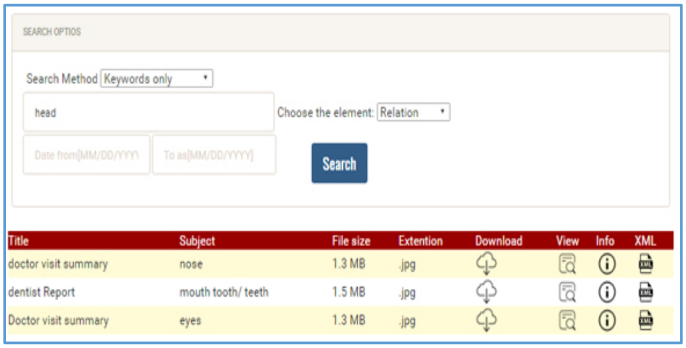
Subject: Diaphragm
stomach
kidney
liver
gallbladder

Relation: Abdomen

Description

DC relation field will be polulated automatically when selecting any of the subject elements

Figure 8: Create DC Meta-data



SEARCH OPTIOS

Search Method: Keywords only

head

Choose the element: Relation

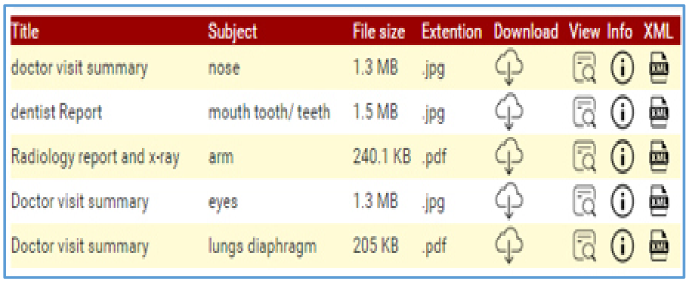
Date from(MM/DD/YYYY

To as(MM/DD/YYYY)

Search

Title	Subject	File size	Extention	Download	View	Info	XML
doctor visit summary	nose	1.3 MB	.jpg	☁	🔍	ⓘ	▤
dentist Report	mouth tooth/ teeth	1.5 MB	.jpg	☁	🔍	ⓘ	▤
Doctor visit summary	eyes	1.3 MB	.jpg	☁	🔍	ⓘ	▤

Figure 9: Example of Retrieving DC Meta-data for only Related Documents



Title	Subject	File size	Extention	Download	View	Info	XML
doctor visit summary	nose	1.3 MB	.jpg	☁	🔍	ⓘ	▤
dentist Report	mouth tooth/ teeth	1.5 MB	.jpg	☁	🔍	ⓘ	▤
Radiology report and x-ray	arm	240.1 KB	.pdf	☁	🔍	ⓘ	▤
Doctor visit summary	eyes	1.3 MB	.jpg	☁	🔍	ⓘ	▤
Doctor visit summary	lungs diaphragm	205 KB	.pdf	☁	🔍	ⓘ	▤

Figure 10: Example of DC Meta-data for all Documents

Table 2: Human Body Categories

Body categories	Body parts
The abdomen	Contains diaphragm, stomach, liver, gallbladder, pancreas, small intestine, large intestine, cava, spleen, and others.
Heart	Contains superior vena cava, pulmonary artery, pulmonary veins, pulmonic valve, tricuspid valve, inferior vena cava, right atrium, right ventricle, left ventricle, aortic valve, mitral valve, left atrium, aorta and others.
Head	Contains eyes, ears, brain, mouth, teeth, nose, chin, spinal cord, tonsil, uvula, gullet, meninges, pharynx and others.
Thorax	Contains lungs, diaphragm/pleura, nasopharynx/oral, cavity, trachea/Larynx, ribs, capillaries, bronchial tube, windpipe/trachea, chest, esophagus and others.
Extremities	Contains arms, elbows, hands, wrists, shoulders, hips/thighs, fingers, thumbs, legs, knees, toe, vertebral column, neck, ankles, breast, back pain, feet and others.
Integumentary	Skin and associated structures such as hair, nails, sweat glands, and oil glands
Urinary	Kidneys, ureters, urinary bladder, and urethra
Reproductive	Gonads (testes or ovaries) and associated organs; in females: uterine tubes, uterus, and vagina; in males: epididymis, ductus deferens, prostate gland, and penis

6. Conclusion

As the medical industry is going through a paradigm shift from clinician-centered to patient-centered, readily available complete personal medical history is becoming crucial. This will help ensure the three major goals in medical industry: evidence-based treatment, continuity of care, and prevention of medical mistakes. In this paper, we proposed an untethered personal health record system to help achieve such goals. Our proposed system, MCRS, provides a method to collect and organize heterogeneous personal health data using DC meta-data. The retrieval of the organized data was completed by the reorganized DC tags, which allowed the organization of clinical data by body parts for easy retrieval. Finally, we allowed access to personal emergency clinical data to emergency crew only by their license number at the time of need. In the future, we plan to expand the usage of clinical data collected from the application to analyze and identify the regional characteristics in health mapping to build better public policy for the nation. Our experiment was limited to one healthcare provider using the clinical data from that provider. Also, the data collection was limited due to the patient privacy regulations in the healthcare industry.

References

[1] This paper is an extension of work originally presented in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017.

[2] Key Developments in the Connected Health Markets https://www.parksassociates.com/whitepapers.

[3] AHIMA e-HIM Personal Health Record Work Group. "Defining the Personal Health Record." Journal of AHIMA 76, no.6 (June 2005): 24-25.

[4] Richard Brandt and Rich Rice, "Building a better PHR paradigm: Lessons from the discontinuation of Google Health™", Health Policy and Technology Volume 3, Issue 3, 2014, Pages 200-207, ISSN 2211-8837, https://doi.org/10.1016/j.hlpt.2014.04.004. (http://www.sciencedirect.com/science/article/pii/S221188371400032X)

[5] Morgan Price, Paule Bellwood, Nicole Kitson, Iryna Davies, Jens Weber and Francis Lau, "Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review". BMC Medical Informatics and Decision Making 15, 32 (2015), https://doi.org/10.1186/s12911-015-0159-1

[6] Price, Margaux M. and Pak, Richard and M{\"u}ller, Hendrik and Stronge, Aideen, "Older adults' perceptions of usefulness of personal health records" "Universal Access in the Information Society 12, 191-204 (2013), doi="10.1007/s10209-012-0275-y

[7] Alyami, M.Aalyami and Song, Yeong-Tae: Removing Barriers in using Personal Health Record Systems. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). DOI: 10.1109/ICIS.2016.7550810

[8] Fearon. James D. and Laitin, David D. "Ethnicity, Insurgency, and Civil War" American Political Science Review volume={97} pages={75–90}, (2003). DOI={10.1017/S0003055403000534

[9] Sweet Lauren E. and Moulaison Heather Lea. " Electronic Health Records Data and Metadata: Challenges for Big Data in the United States" Big Data. January 2014, 1(4): 245-251. https://doi.org/10.1089/big.2013.0023

[10] Greenberg, Jane, Maria Cristina Pattuelli, Bijan Parsia, and W. Davenport Robertson. "Author-generated Dublin Core metadata for web resources: a baseline study in an organization." In International Conference on Dublin Core and Metadata Applications, pp. 38-45. 2001.

[11] Talha, Norhaizan Mat. "Metadata management system (MMS)." In International Conference on Dublin Core and Metadata Applications. 2004.

[12] Haugen, Mary Beth and Herrin, Barry and Slivochka, Sharon and Tolley, Lori McNeil and Warner, Diana and Washington, Lydia,"Rules for handling and maintaining metadata in the EHR" Journal of AHIMA 84, no.5 (May 2013): 50-54. URL: http://europepmc.org/abstract/MED/23755436

[13] P. Dhivya and S. Roobini and A. Sindhuja, "A. Symptoms Based Treatment Based on Personal Health Record Using Cloud Computing" Procedia Computer Science 47, 22-29 (2015), doi = "https://doi.org/10.1016/j.procs.2015.03.179",
url = http://www.sciencedirect.com/science/article/pii/S1877050915004470

[14] Chintan Patel and Karthik Gomadam and Sharib Khan and Vivek Garg, "TrialX: Using semantic technologies to match patients to relevant clinical trials based on their Personal Health Records" Web Semantics: Science, Services and Agents on the World Wide Web, 8, 342-347 (2010) doi = "https://doi.org/10.1016/j.websem.2010.08.004", url = http://www.sciencedirect.com/science/article/pii/S1570826810000636

[15] Appelboom, Geoff and Camacho, Elvis and Abraham, Mickey E. and Bruce, Samuel S. and Dumont, Emmanuel LP et al. "Smart wearable body sensors for patient self-assessment and monitoring" Archives of Public Health 72, 28 (2014). doi="10.1377/hlthaff.2012.1216", url=https://doi.org/10.1377/hlthaff.2012.1216

[16] Zhang J., Chen D.& Wong T. 'Metadata application on clinical trial data in drug development,' Proceedings of the Twenty Eighth Annual SAS® Users Group International Conference, Seattle, WA (2003).

[17] Tal Teitz and Dwayne G. Stupack and Jill M. Lahti, "Halting Neuroblastoma Metastasis by Controlling Integrin-Mediated Death" Cell Cycle 5, 681-685 (2006). doi = {10.4161/cc.5.7.2615}, URL = http://dx.doi.org/10.4161/cc.5.7.2615

[18] Ariel Ekblaw, Asaf Azaria, Thiago Vieira, Andrew Lippman. (2016). MedRec: Medical Data Management on the Blockchain. PubPub, https://www.pubpub.org/pub/medrec version: 57e013615dbf3f3300152554].

[19] Asuman Dogac , Gokce B. Laleci , Yildiray Kabak , Seda Unal , Sam Heard , Thomas Beale , Peter Elkin , Farrukh Najmi , Carl Mattocks , David Webber et al. "Exploiting ebXML registry semantic constructs for handling archetype metadata in healthcare informatics" International Journal of Metadata, Semantics and Ontologies 1, 21-36 (2006). https://doi.org/10.1504/IJMSO.2006.008767

[20] Poeggel, Sven, Daniele Tosi, DineshBabu Duraibabu, Gabriel Leen, Deirdre McGrath, and Elfed Lewis. "Optical fibre pressure sensors in medical applications." Sensors 15, no. 7 (2015): 17115-17148.

[21] Guner C, Akin S, Durna Z. Comparison of the symptoms reported by post-operative patients with cancer and nurses' perception of patient symptoms. Eur J Cancer Care 2014;23:523-30. doi: 10.1111/ecc.12144

[22] Jeff Mather, MathWorks "Accessing data in DICOM files" Technical Articles and Newsletters Published 2002http://www.mathworks.com/company/newsletters/articles/accessing-data-in-dicom-files.html?requestedDomain=www.mathworks.com (accessed October 2016)

[23] Meltzer, Eli O., Daniel L. Hamilos, James A. Hadley, Donald C. Lanza, Bradley F. Marple, Richard A. Nicklas, Allen D. Adinoff et al. "Rhinosinusitis: developing guidance for clinical trials." Journal of Allergy and Clinical Immunology 118, no. 5 (2006): S17-S61.

[24] LaConte, Grace M. "Documentation roles in an electronic health record environment." PhD diss., The College of St. Scholastica, 2011.

[25] T.Brnson, Principles of Health Interopersability HL7 ans SNOMED CT, Health Information Technology Standards, 2012.

[26] Chao-chen Chen, Hsueh-hua Chen, Kuang-hua Chen. "The Design of Metadata Interchange for Chinese Information and Implementation of Metadata Management System" BULLETIN of the American Society for Information Science and Technology Vol. 27, No. 5 June / July 2001, url: https://www.asis.org/Bulletin/Jun-01/chen.html

[27] Informatica Metadata Management for Holistic Data Governance. (July 2013).at https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/metadata-management-data-governance_white-paper_2163.pdf

[28] Sunyaev, Ali, Dmitry Chornyi, Christian Mauro, and Helmut Krcmar. "Evaluation framework for personal health records: Microsoft HealthVault vs. Google Health." In System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1-10. IEEE, 2010.

[29] Bindman AB. Using the National Provider Identifier for health care workforce evaluation. Medicare Medicaid Res Rev. 2013;3(3):1-10.

[30] C. N. Daily, "Top 20 Best Cloud Storage Providers – Reviews and Comparison of the Top Secure Solutions and Services. Find Unlimited Cloud Based Data Storage Services and Options," 2017.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 174-183 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

A Model for Optimising the Deployment of Cloud-hosted Application Components for Guaranteeing Multitenancy Isolation

Laud Charles Ochei*,[1], Christopher Ifeanyichukwu Ejiofor[2]

[1]*School of Computing and Digital Media, Robert Gordon University, United Kingdom*

[2]*Department of Computer Science, University of Port Harcourt, Nigeria*

ARTICLE INFO

Article history:
Received: 25 November, 2017
Accepted: 24 February, 2018
Online: 27 March, 2018

Keywords:
Cloud-hosted application
Multitenancy
Degree of Isolation
Simulated Annealing
Metaheuristic
Optimization model
Component
Cloud deployment

ABSTRACT

Tenants associated with a cloud-hosted application seek to reduce running costs and minimize resource consumption by sharing components and resources. However, despite the benefits, sharing resources can affect tenant's access and overall performance if one tenant abruptly experiences a significant workload, particularly if the application fails to accommodate this sudden increase in workload. In cases where a there is a higher or varying degree of isolation between components, this issue can become severe. This paper aims to present novel solutions for deploying components of a cloud-hosted application with the purpose of guaranteeing the required degree of multitenancy isolation through a mathematical optimization model and metaheuristic algorithm. Research conducted through this paper demonstrates that, when compared, optimal solutions achieved through the model had low variability levels and percent deviation. This paper additionally provides areas of application of our optimization model as well as challenges and recommendations for deploying components associated with varying degrees of isolation.

1. Introduction

Designing and planning component deployment of a cloud-hosted application with multiple tenants demands special consideration of the exact category of components that are to be distributed, the number of components to be shared, and the supporting cloud resources required for component deployment. [1] This is because there are different or varying degrees of multitenancy isolation. For instance, in components providing critical functionality, the degree of isolation is higher compared to components that only require slight re-configuration prior to deployment [2].

A low degree of isolation actively encourages tenants to share resources and components, resulting in lower resource consumption and reduced operating costs, however, there are potential challenges in both security and performance in the instance where one component sees a sudden workload surge. A high degree of isolation tends to deliver less security interference, although there are challenges instigated by high running costs and resource consumption in view that these tenants are not sharing resources [2]. Consequently, the software architect's main challenge is to first identify solutions to the opposing trade-off of high degrees of isolation (including excessive resource consumption issues and high operating costs), versus low degrees of isolation (including performance interference issues).

Motivated by these key challenges, this paper presents a model for the deployment of components which provides exemplary solutions specific to cloud-based applications and aims to do so in a way that secures the segregation of multitenancy. The approach for this research includes creating an optimization model which is mapped to a Multichoice Multidimensional Knapsack Problem (MMKP) before solutions are tested using a metaheuristic. The approach is analysed through comparing the different optimal solutions achieved which then collectively compose an exhaustive search tool to analyse the solutions capacity specifically for minor problem occurance, in its entireity.

This paper and its research questions: "***How can we optimize the way components of a cloud-hosted service is deployed for guaranteeing multitenancy isolation?***". It is possible to guarantee the specific degree of isolation essential between

*Laud Charles Ochei, Aberdeen, United Kingdom. l.c.ochei@rgu.ac.uk

https://dx.doi.org/10.25046/aj030220

tenants, whilst efficiently managing the supporting resources at the same time through component deployment optimization of the cloud-based application.

This paper expands on the previous work conducted in [3]. The core contributions of this article are:
1. Mathematical optimization model providing optimal component deployment solutions appropriate for cloud-based applications to guarantee multitenancy isolation.
2. Mathematical equations inspired by open multiclass queuing network models to determine the average request totals for granting component and resource access.
3. Variants of metaheuristic solutions to deliver optimization model resolutions attributed to simulated annealing.
4. Guidelines and recommendations for component deployment in cloud-hosted applications seeking to guarantee required levels of multitenancy isolation.
5. Application areas of a cloud-hosted service where it is possible for the optimization model to be directly applied to component deployment with the aim to guarantee required degree of multitenancy isolation.

The remainder of this paper is structured as outlined below: Section II focuses on the challenge of identifying and delivering near-optimal component deployment solutions specific to cloud-hosted applications that guarantee the essential levels of multitenancy isolation; Section III presents the Optimisation Model; the following section presents the Metaheuristic Solution; Section V considers the Open Multiclass Queuing Model; Section VI evaluates the results and presents the experimental setup; Section VII discusses the results; Section VIII discusses the model's application areas for the model; and Section IX conclude with recommendations of future work.

2. Optimising the Deployment of Components of Cloud-hosted Application with Guarantee for Multitenancy Isolation

This section examines multitenancy isolation, the conflicting trades-offs in delivering optimal deployment influenced by the varying degrees of multitenancy isolation, and other associated research on cloud resources and optimal allocation of such resources.

2.1. Multitenancy Isolation and Trade-offs for Achieving Varying Degrees of Isolation

In a multitenant architecture (also referred to as *multitenancy*), multiple tenants are able to access a single instance of a cloud service. These tenants have to be isolated when there are changes in workload. In the same way that it is possible to isolate multiple tenants, it is also possible to isolate multiple components of a cloud-hosted application.
In this paper, we define "Multitenancy Isolation" as a way of ensuring that other tenants are not affected by the required performance, stored data volume, and access privileges of one of the tenants. accessing the cloud-hosted application [3] [4]. A high degree of isolation is achieved when there is little or no impact on other tenants when a substantial increase in workload occurs for one of the tenants, and vice versa. The three cloud patterns that describe the varying degrees of multitenancy isolation are:
(i) Dedicated Component: tenants cannot share components; however, a component may be associated with one or more tenants or resources;
(ii) Tenant-isolated Component: tenants can share resources or components, and isolation of these is guaranteed; and
(iii) Shared Component: tenants can share resources or components, but these remain separate from other components.

If components required a high degree of isolation between them, then each tenant requires that each component is duplicated. This can be expensive and also lead to increased resource consumption. On the other hand, there could also be a need for a low degree of isolation which could, in contrast, reduce cost and resource consumption. However, any changes in workload levels that the application cannot cope with risk interference [4]. The question, therefore, is how optimal solutions can be identified to resolve trade-offs when conflicting alternatives arise.

2.2. Related Work on Optimal Deployment and Allocation of Cloud Resources

Research on optimal resource allocation in the cloud is quite significant, however, much fewer studies focus on optimal solutions in relation to component deployment across cloud-based applications in a way which guarantees the required degree of multitenancy isolation. Researchers in [5] and [6] aim to keep cloud architecture costs to a minimum by implementing a multitenant SaaS Model. Other authors [7] concentrated on bettering execution times for SaaS providers whilst reducing resource consumption using evolutionary algorithms opposed to traditional heuristics. A heuristic is defined in [8] for the capacity planning purposes for the SaaS inspired by a utility model. The aim of the utility model was to generate profit increases and so it largely concentrated on business-related aspects of delivering the SaaS application.

It is explained in [9] how optimal configuration can be identified for virtual servers, such as using certain tests to determine the required volumes of memory for application hosting. Optimal component distribution is discussed and analyzed by [10] in relation to virtual servers. Research conducted in [11] is of a similar nature to this paper in that it attempts to reduce costs (using a heuristic search approach is inspired on hill climbing), specifically in relation to the use of VMs from an IaaS provider with limitations around SLAs response time.

The studies noted above predominantly focus on scaling back costs associated with cloud architectural resources. The use of metaheuristics is not considered in these studies in delivering optimal solutions that can guarantee the required degree of multitenancy isolation. Additionally, previous research involving optimization models have operated with one objective; an example is where [11] look to minimize VM operational costs. For this paper's model, a bi-objective case is used (i.e., maximising the required degree of multitenancy isolation and number of requests permitted to access a component). Thereafter, a modern metaheuristic inspired by simulated annealing is used to solve the model.

3. Problem Formalisation and Notation

This section defines the problem and explains the process of mapping it to a Multichoice Multidimensional Knapsack Problem (MMKP).

3.1. Description of the Problem

Assuming a tenant has multiple components associated with the same supporting cloud infrastructure. A team may represent a team or department, a company with a responsibility to design a cloud-based application, its components, and underlying processes. Components varying in size and function has to integrate with their cloud-hosted application to achieve effective deployment in a multitenant style. It is also possible to define component categories based on different features, such as function, for example, processing or storage. Within these categories, different components are likely to have differing degrees of isolation enabling some components to deliver the same function which can hence be accessed and used by multiple tenants, opposed to other components which may be solely allocated to certain tenants or departments.

Every component within an application needs a particular allocation of resources from the cloud infrastructure in order to support the volume of requests received. In instances where one component in the application experiences surges in workload, then it must be considered how the designer can choose components to deliver optimal deployment to effectively respond to the sudden changes in such a manner that: (i) maximises component degrees of isolation through ensuring they behave in the same way as components of other tenants, thus, isolating against one and other; and (ii) maximises the total requests permitted to access and use each components.

3.2 Mapping the Problem to a Multichoice Multidimensional Knapsack Problem (MMKP)

The above mentioned optimal component deployment problem can be closely linked to a 0-1 Multichoice Multidimensional Knapsack Problem (MMKP). An MMKP is a variant of the Knapsack Problem commonly depicted as a member of the NP-hard class of problems. For the purpose of this paper, the problem of focus can be formally defined as:

Definition 1 (Optimal Component Deployment Problem): Consider that there are N groups of components ($C_1,\ldots,C_N$) with each group having a_i ($1 \leq \mathrm{i} \leq N$) components useful for designing (or integrating with) a cloud-hosted application. Each component of the application is affiliated with: (i) the degree of isolation that is required between components (I_{ij}); (ii) the rate at which requests arrive to the component λ_{ij}; (iii) the service demand of resources required to support the component D_{ij}; (iii) the average request totals permitted to access the component Q_{ij} and (iv) resources for supporting the component, $r_{ij} = r_{ij}^1 ,r_{ij}^2 ,\ldots,r_{ij}^n$. For the cloud to properly support all components, a certain volume of resources are required; the total number of resources needed can be calculated as $R = (R^1,R^2,\ldots,R^n)$.

The aim of an MMKP is to choose one component present in each category for deployment to the cloud in a manner that ensures that if one component sees sudden increases in load, then the: (a) the degree of isolation of other components is maximized; (b) the total requests permitted to access the component and application is maximised without using more resources than are actually available.

Definition 1 identifies two objectives within the problem. An aggregation function is used to convert the multi-objective problem into a single-objective because of merging the two objective functions merge (i.e., g1=degree of isolation, and g2=number of requests) into one single objective function (i.e., g=optimal function) in a linear way. Because the optimal function of this study is linear, a *priori single weight* strategy is employed to aid defining the weight vector selected based on the decision maker's individual preferences [12]. This paper also adopts the approach discussed in [12] for computing the absolute percentage difference (see section 7.1), the target solutions used to compare against the optimal solution (see section 7.2), and the use of the number of optimal function evaluations as an alternative to measuring the computational effort of the metaheuristic.

Therefore, the purpose is redefined as follows: to deliver a near-optimal solution for component deployment to the cloud-based application that also fulfils system requirements and achieves the best value possible for optimal function, G.

Definition 2 (Optimal Function): For a cloud-hosted application architect, the main issues impacting the optimal deployment of components are changes to workload, which can be expressed as:

$$\sum_{i=1}^{N} \sum_{j \in C_i}^{n} g_{ij} . a_{ij}$$

Subject to:

$$\sum_{i=1}^{N} \sum_{j \in C_i}^{n} r_{ij}^{\alpha} . a_{ij} \leq R^{\alpha} (\alpha = 1, 2, \ldots, N) \qquad (1)$$

$$\sum_{j \in C_i}^{N} a_{ij} = 1$$

$a_{ij} \in 0,1$ (i = 1, 2 ,…, N), j $\in C_i$

where (i) a_{ij} is fixed at 1 if component j is chosen from group C_i and 0 otherwise; (ii) g_{ij} is determined by a weighted calculation of parameters involving the degree of isolation, average requests permitted to access a component, and penalty for constraint violations.

$$g_{ij} = (w1 \times I_{ij}) + (w2 \times Q_{ij}) - (w3 \times P_{ij}) \qquad (2)$$

Specific weight values are allocated to w1, w2, and w3; namely 100, 1 and 0.1 respectively. The allocation of weights is done using a method that provides preference to the required degree of isolation. The penalty, P_{ij}, imposed for components that surmount resource cap is expressed as:

$$P_{ij} = \sum_{i=0}^{n} max\left\{0, \left(\frac{R^k - R^k_{max}}{R^k_{max}}\right)\right\}^2 \qquad (3)$$

For every component (***g***), the degree of isolation, I_{ij}, is assigned either 1, 2, or 3 indicating either shared, tenant-isolated or dedicated components, respectively. The sum described as: $r_{ij} = r^1_{ij}, r^2_{ij}, \dots, r^n_{ij}$ refers to the resource consumption in group C_i. for each individual application component j. Total resource consumption r^{α}_{ij} for all application components needs to be lower than the total available resources in the cloud infrastructure $R = R^{\alpha}, (\alpha = 1, \dots, m)$.

It is presumed that the service demands at the CPU, RAM, Disk I/O, and the supporting bandwidth of each component can be identified and/or measured readily by the SaaS supplier or customer. This assumption enables us to calculate the number of requests, Q_{ij} that may be permitted access for each component through analysis of an open multiclass QN Model [13]. The following section expands further on the open multiclass network.

4. Queuing Network (QN) Model

Queueing network modelling is one modelling approach through which the computer system is depicted as a network of queues that can be solved in an analytic fashion. In its most basic form, a network of queues is an assembly of service centers representative of system resources, and customers representative of business activity, such as transactions [13]. Service centers are basically supporting resources for the components, such as CPU, RAM, disk and bandwidth.

Assumptions: For the purpose of this paper, the following component assumptions are made:

(i) components cannot support other applications or alternative system requirements, and is therefore exclusively deployed to one cloud-application;

(ii) component arrival rates are separate to the main system state and so component requests may have significantly varied behaviours.

(iii) it is possible to identify and effortlessly measure the service demands at the CPU, RAM, Disk, and Bandwidth supporting each component by both the SaaS provider and/or customer.

(iv) sufficient resource is available to support each component during changes to workload, particularly during significant surges of new incoming requests. Ensuring sufficient resource means that there are no overloads during peak times where all components are operating.

The assumptions noted above allow the study to utilise an open multiclass queuing network (QN) model for the purpose of calculating average requests permitted to reach the component, whilst simultaneously ensuring the required degree of isolation, as well as system requirements. The magnitude and intensity of workload volume in an open multiclass QN is determined by request arrival rates. The arrival rate is not typically reliant on the system state, and so is not reliant on the volume of other tenants in the system either [13].

Definition 3 (Open Multiclass Queuing Network Model): Assuming there is a total of N classes, where every class *c* is an open class with arrival rate λ_c. The arrival rates are symbolised as a vector by $\vec{\lambda} = (\lambda_1, \lambda_2, \dots \lambda_N)$. The use of each component in class *c* at the center *k* is given by:

$$U_{c.k}(\vec{\lambda}) = \lambda_c D_{c.k} \qquad (4)$$

To solve the QN model, assumptions are made, such as that a component stands for a single open class system hosting four service centers otherwise referred to as supporting resources, such as CPU, RAM, disk capacity and bandwidth. At any one service center (e.g., CPU), the average request totals for a specific component is:

$$Q_{c,k}(\vec{\lambda}) = \frac{U_{c.k}(\vec{\lambda})}{1 - \sum_{i=1}^{N} U_{i.k}(\vec{\lambda})} \qquad (5)$$

Consequently, to determine the average amount of requests accessing the particular component, the length of the queue of all requests reaching all service centers (i.e., components' supporting resources such as CPU, RAM, disk capacity and bandwidth) would be totaled.

$$Q_c(\vec{\lambda}) = \sum_{i=0}^{n} Q_{c,k}\ \vec{\lambda} \qquad (6)$$

5. Metaheuristic Search

The optimisation problem as explained in the section before is an NP-hard problem renowned for its feasible search capacity and exponential growth [14]. The number of potential and feasible solutions that may achieve optimal component deployment and solve the problem can be determined using this equation:

$$\left\{\binom{n}{r}\right\}^N \qquad (7)$$

The above equation, Equation 4, signifies the different ways one or more (**r**) components can be chosen from each group (comprising of **n** components) from a pool of numerous (N) groups of components, for the purpose of creating and integrating them into a cloud-hosted application upon receipt of updates or changes to workload by the component. Thus, to manage such changes to workloads, the specific number of different ways that one component can be selected (i.e., r=1) from each of the 20 different groups (i.e., N=20), comprising of 10 items per group (i.e., n=10), approximately 10.24 x 10^{12} possible solutions can be identified. Contingent on the number of changes to workload and also the regularity of these, a cloud-hosted service quite large in size could experience a much greater volume of possible solutions.

Therefore, to obtain an optimal solution for the identified optimisation problem, it is essential to use an efficient metaheuristic. In addition, this should be done in real-time with the SaaS customer or cloud architect. Two versions of a simulated annealing algorithm are implied: (i) *SAGreedy*, incorporates greedy principles in conjunction with a simulated annealing algorithm; (ii) *SARandom*, employs randomly propagated solutions in conjunction with a simulated annealing algorithm. Both of these versions can be effectively utilised to achieve near-optimal solutions for component deployment. Additionally, an algorithm was developed for this study to generate an extensive search of the full solution area for a small problem size. Algorithm 1 includes the algorithm for *SA(Greedy)*. However, SA(Random) only needs a minor change to this algorithm, which will be described further in the following section. An extensive breakdown of Algorithm 1 can be viewed below:

Algorithm 1 SA(Greedy) Algorithm

```
1:   SA (Greedy) (mmkpFile, N)
2:   Randomly generated N solutions
3:   Initial temperature fixed to $T_0$ to st. dev. of all
     optimal solutions
4:   Create greedySoln $a^1$ with optimal value $g(a^1)$
5:   optimalSoln = $g(a^1)$
6:   bestSoln = $g(a^1)$
7:   for I = 1, N do
8:   Create neighbour soln $a^2$ with optimal value
     $g(a^2)$
9:   Mutate the soln $a^2$ to improve it
10:  if $a^1 < a^2$ then
11:      bestSoln = $a^2$
12:  else
13:      if random[0,1) < $\exp(-(g(a^2) - g(a^1))/T)$ then
14:          $a^2$ = bestSoln
15:      end if
16:        end if
17:      $T_{i+1} = 0.9 * T_i$
18:  end for
19:  optimalSoln = bestSoln
20:  Return (optimalSoln)
```

5.1. The SAGreedy for Near-optimal Solution

The first algorithm is a combination of simulation annealing and greedy algorithm which is used to obtain a near-optimal solution for the optimisation problem modelled as an MMKP. First, the algorithm extracts the key details from the MMKP problem instance before populating the encompassing variables (i.e., collections of different dimensions storing isolation values of isolation; average request totals; and the resource consumption of components). A basic linear cooling schedule is used, where $T_{i+1} = 0.9Ti$. The method for prescribing and fixing the preliminary temperature T_0 will be to randomly generate an optimal solution whose number is equivalent to the total number of groups (n) in the problem instance, multiplied by the number of iterations (N) used in the experimental settings before running the simulated annealing element of the algorithm.

When the problem instance and/or the total iterations is low, the magnitude of optimal solutions created may be limited by the number of groups (n) in the problem instance, multiplied by total iterations (N) used in the experimental settings. Next, the initial temperature T_0 is determined for the standard deviation of all optimal solutions (Line 2-3) through random generation. The algorithm then uses the greedy solution as the preliminary solution (Line 4) which is assumed as the best current solution. The simulated annealing process enhances the greedy solution further providing a near-optimal solution for cloud component deployment.

The execution of the algorithm in its most basic form for the instance C(4,5,4) is explained as follows: let us imagine that the total number of iterations is 100, 400 (i.e., 4 groups x 100 iterations) optimal solutions are randomly generated before calculating the standard deviation for all solutions. Assuming a value of 50.56, T_0 is identified as 50. It is also assumed that the algorithm creates a foremost greedy solution with $g(a^1) = 2940.12$, before a current random solution with $g(a^2)$ = 2956.55. The solution a^2 will substitute a^1 with probability, P =exp(-16.43/50)=0.72, because $g(a^2) > g(a^1)$. In lines 14 to 16, a random number (rand) is generated between 0 and 1; if rand <0.72, a^2 replaces a^1 and we proceed with a^2. Alternatively, the study continues with a^1. Now, the temperature T is reduced providing $T_1 = 45$ (Line 17). Iterations continue until N (that is, the identified number of iterations set to enable the algorithm to function), is reached, thus the search converges with a high probability of near-optimal solution.

5.2. The SA(Random) for Optimal Solutions

Considering the SA(Random) metaheuristic version, a solution is also randomly generated before being encompassed within the simulated annealing process to provide a preliminary solution. It can be seen in Line 4 that rather than creating a greedy solution, a random solution is created. An optimal solution representative of a set of components with the highest total isolation value and number of requests permitted to reach the component access is then output by the algorithm. Every time a variance in workload is experiences, the optimal solution alters to respond to this.

6. Evaluation

We describe in the section how each instance was generated as well as the process, procedure and set up of the experiment.

6.1. Instance Generation

Reflective of different capacities and sizes, a number of problem instances were randomly generated. Instances were divided into two categories determined by those cited frequently in current literature: (i) OR benchmark Library [15] and other standard MMKP benchmarks, and (ii) the new irregular benchmarks used in [16]. All these benchmarks were used for single objective problems. This study edited and modified this benchmark to fit into a multi-objective case through assigning each component with one of two profit values: isolation values and average number of requests [17].

The values of the MMKP the instance, were produced as follows: (i) random generation of isolation values in the interval [1-3]; (ii) values of component consumption of CPU, RAM, disk and bandwidth (i.e., the weights) were generated in the interval [1-9]; (iii) individual component resource limits (i.e., knapsack capacities for CPU, RAM, disk and bandwidth) were created by halving the maximum resource consumption possible (see Equation 7).

$$c_k = \frac{1}{2} \times m \times R \tag{8}$$

An identical principle has been employed to create instances for OR Benchmark Library, as well as for instances used in [18] [19]. This research considers the total resources/constraints as four (4) for each group, which reflects the minimal resource requirement to deploy a component to the cloud. The notation for each instance is: C(n,r,m), representing the number of groups, the amount of components in each group, and resource totals.

6.2. Experimental Setup and Procedure

For consistency, all experiments were set up and operated using Windows 8.1 on a SAMSUNG Laptop with an Intel(R) CORE(TM) i7-3630QM at 2.40GHZ, 8GB memory and 1TB swap space on the hard disk. Table I outlines the experimental parameters. The algorithm was tested using different sized instances of different densities. In relation to large instances, it was not possible to conduct an exhaustive search due to a lack of memory resource on the machine used.

As a result of this limitation, the MMKP instance was implemented first, C(4,5,4), to provide a benchmark for analysis to enable algorithm comparison.

Table 1. Parameters used in the Experiments

Parameters	**Value**
Isolation Value	[1,2,3]
No. of Requests	[0,10]
Resource consumption	[0,10]
No. of Iterations	N=100 (except Table 4)
No. of Random Changes	5
Temperature	T_0 = st.dev of N randomly generated solns.
Linear Cooling Schedule	$T_{i+1}=0.9Ti$

7. Results

Section 7 discusses the experiment results.

7.1 Comparison of the Obtained Solutions with the Optimal Solutions

The results delivered by algorithms SA(Greedy) and SA(Random) were initially compared with the optimal solutions generated by an exhaustive search for a small problem instance in the entire solution space (i.e., C(4,5,4)). Table 2 and Table 3 portray the findings. The instance id used is noted in Column 1 of Table 2. The second, third and fourth columns respectively highlight the optimal function variables as (FV/IV/RV), representing the optimal function value, isolation value, and number of permitted requests, for Optimal, SA(Random) and SA(Greedy) algorithms. The first and second columns of Table 3 depict a proportion of the optimal values for SA(Random) and SA(Greedy) algorithms, respectively. The final two columns note the absolute percentage difference, indicative of the solution quality, for SA(Random) and SA(Greedy) algorithms, which is measured as follows:

$$\frac{|f(s) - f(s^*)|}{f(s^*)} \tag{9}$$

where s is the obtained solution and s* is the optimal solution generated by the exhaustive search.

It is clear that SA(Greedy) and SA(Random) provide very similar results. Solutions identified for SA(Random) are almost 100% close most of the time to their optimal solution, and over 83% in a smaller proportion of cases whilst in just one occurrence it was below 66%. SA(Greedy) also delivered nearly 100% close solutions created to the optimal solution in a considerable number of cases, and more than 83% in others. Overall, SA(Greedy) delivered better results than SA(Random) when considering percent deviation from the optimal solution.

7.2 Comparison of the Obtained Solutions to a Target Solution

It was not possible to run instances greater than C(4,5,4) because of hardware limitations (i.e., CPU and RAM).

Consequently, the results were compared to a target solution. The target solution for percent deviation and performance rate was determined as (n x max(I) x w1) and ((n x max(I) x w1) + (0.5 x (n x max(Q) x w2))), respectively. So, for instance C(150,20,4), the target solution for computing percent deviation sit at 45,000.

Table III, IV, and V demonstrates average solution behaviour: (i) on a significant selection of varied instances using the same parameters; (ii) over various runs on the same instance (with differing quantities of optimal evaluation); and (iii) over different runs on the same instance. The robustness of solutions in relation to their behaviour on varying types of instances using the same parameters was measured. Table III shows this measure and that solutions are strong when considering average deviation of solution behaviour for both the SA(Greedy) and SA(Random), as implied through their low variability scores.

The average percent deviation and standard deviation (of the percent deviations), for SA(Greedy) is marginally greater than SA(Random) as a result of the significant absolute difference between some solutions and the reference solution. For example, the percent deviations of SA(Greedy) for the instances C(100,20,4) and C150,20,4) are higher than SA(Random). The results show how SA(Random) performs much better than SA(Greedy) in reference to small instances up to C(80,20,4).

Table IV compares solution quality with optimal function evaluations. It can be determined that the overall solution quality is good when both algorithms are tested on large instances. Once again, the standard deviation for SA(Greedy) is notably lower than SA(Random) in addition to great percent deviation stability. Table V highlighted that SA(Greedy) is stronger than SA(Random) evidenced by the average optimal values and low solution variability. The performance rate (PR) was computed by determining the reference solution as a function of the quantity of optimal evaluations. The PR of SA(Greedy) is marginally higher than SA(Random).

Figure I depict the relationship of solution quality relating to optimal values (i.e., fitness value) and the volume of optimal function evaluations. It can be seen from the diagram that SA(Greedy) benefited a little from the preliminary greedy solution more than SA(Random) when optimal function evaluations are few. Nonetheless, solution quality for both algorithms are better as iterations increase. Once 100 optimal evaluations are reached, the optimal solution stabilizes, but thereafter fails to show any further noticeable improvement.

Figure 2 portrays correlations between solution quality relating to percent deviation and the total function evaluations. In line with expectations, SA(Random) reported a smaller percent deviation than SA(Greedy) in the majority of results, particularly in instances where function evaluations are few. An explanation for this could be the low function evaluation total used in the study. However, percent deviation for SA(Greedy) showed greater stability despite being greater than SA(Greedy)'s results.

Table 2. Comparison of SA(Greedy) and SA(Random) with the Optimal Solution

Inst-id	Optimal (FV/IV/RV)	SA(R) (FV/IV/RV)	SA(Greedy) (FV/IV/RV)
I1	1213.93/12/24	1218/12/18	1218/12/18
I2	1213.97/12/14	1208.99/12/9	1109/11/9
I3	1222.99/12/23	1120/11/20	1120/11/20
I4	1119.98/11/20	1023/10/23	1019/10/19
I5	1219.99/12/20	1017/10/17	1017/10/17
I6	1229.92/12/30	1020.96/10/21	1022.99/10/23
I7	1224.90/12/25	1018/10/18	1018/10/18
I8	1228.96/12/29	822/8/22	1224.99/12/29
I9	1021.97/10/22	912/9/12	912/9/12
I10	1236/12/36	1236/12/36	1236/12/36

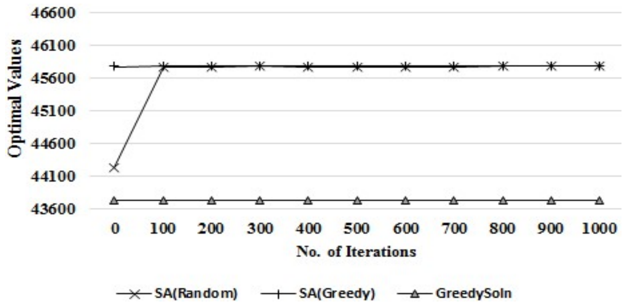


Figure 1. Relationship between Optimal Values and Function Evaluations.

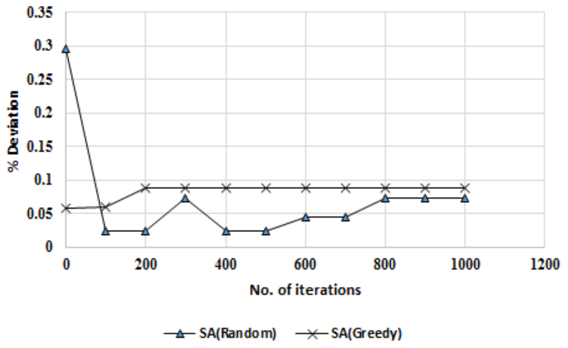


Figure 2. Relationship between Percent Deviation and Function evaluations.

8. Discussion

In Section 8 the implication of the results is discussed further and recommendations for component deployment of a cloud-hosted application that guarantees multitenancy isolation are considered and presented.

8.1 Quality of the Solutions

Solution quality was measured encompassing percent deviation from either the optimal or reference solution. Tables II, III and IV note solutions generated by SA(Greedy) which demonstrate a low percent deviation. The results also indicate that SA(Greedy) performs efficiently for large instances. However, Table III clearly evidences SA(Random) outperforming SA(Greedy) on small instances, that is, from C(5,20,4) to C(80,20,4).

Table 3. Computation of a Fraction of the Obtained Solutions with the Optimal Solution.

Inst-id	SA-R/Opt	SA-G/Opt	%D (SA-R)	%D (SA-G)
I1	0.995	0.995	0.48	0.48
I2	0.996	0.914	0.41	8.65
I3	0.916	0.916	8.42	8.42
I4	0.913	0.910	8.66	9.02
I5	0.834	0.834	16.64	16.64
I6	0.830	0.832	16.99	16.82
I7	0.831	0.831	16.89	16.89
I8	0.669	0.997	33.11	0.32
I9	0.892	0.892	10.76	10.76
I10	1	1	0	0
ST.DEV	0.097	0.064	9.73	6.43

Table 4. Comparison of the Quality of Solution with Instance size

Inst (n=20)	SA-R	SA-G	SA-R	SA-G
C(5)	1528/15/28	1517/15/17	1.87	1.13
C(10)	3058/30/58	3057/30/57	1.93	1.9
C(20)	6082/60/82	6002/59/102	1.37	0.03
C(40)	12182/120/182	12182/120/182	1.52	1.52
C(60)	18317/180/317	18317/180/317	1.76	1.76
C(80)	24388/240/388	24380/240/380	1.62	1.58
C(100)	30252/298/452	30505/300/505	0.84	1.68
C(150)	45648/449/748	45793/450/793	1.44	1.76
Avg. %D			1.54	1.42
STD of Avg. %D			0.33	0.57

8.2. Computational Effort and Performance Rate of Solutions

Hardware limitations of the computer used for the study restricted evaluation and full consideration of computational effort for metaheuristics. On the other hand, it was still possible to compute algorithm performance rates as a function of optimal evaluation totals. Typically, the number of optimal evaluations is considered a computational effort indicator that is independent to the utilised computer system. Table V indicates that the performance rate of SA(Greedy) is marginally higher than SA(Random) when we put into consideration the total reference solutions achieved.

Table 5. Comparing Solution Quality vs. No. of Optimal Evaluations

No. of Iteration	SA(R)	SA(G)	%D (SA-R)	%D (SA-G)
1	44242.1	45776.8	3.30	0.06
100	45760.7	45776.9	0.02	0.06
200	45760.7	45790	0.02	0.09
300	45783	45790	0.07	0.09
400	45760.7	45790	0.02	0.09
500	45760.7	45790	0.02	0.09
600	45770.1	45790	0.04	0.09
700	45770.1	45790	0.04	0.09
800	45783	45790	0.07	0.09
900	45783	45790	0.07	0.09
1000	45783	45790	0.07	0.09
Worst	44242.1	45776.8	0.02	0.06
Best	45783	45790	3.30	0.09
Average	45632.46	45787.61	0.34	0.08
STD	439.77	5.07	0.93	0.01

8.3. Robustness of the Solutions

Low variability was seen from the SA(Greedy) algorithm particularly when run with large instances, therefore suggesting it delivers stronger solutions than SA(Random) making SA(Greedy) the most suitable algorithm for real-time use in a fast-paced environment experiencing regular changes to workload levels. Figure 1 demonstrates instability in SA(Random) for a small amount of evaluations, although following 1,000 evaluations it was seen to converge eventually at the optimal solution. In contrast, the SA(Greedy) showed steady improvement as evaluations increased.

8.4. Required Degree of Isolation

The study's optimisation model assumes that each component for deployment (or group of components) is linked to certain degree of isolation. By mapping the problem to a Multichoice Multidimensional Knapsack Problem linking each component to a profit values: either isolation value; or number of requests permitted to access a component, this was achieved. As a result of this, it was possible to monitor each component separately and independently, responding to each of their unique demands. Where limitations are faced, such as through cost, time or effort used to tag each component, a different algorithm could enable us to perform this function dynamically. In our research prior to this [20], an algorithm capable of dynamically learning the properties of current components in a repository was developed with this information then used to associate each component with the required degree of isolation.

Table 6. Robustness of Solutions over Different Runs on the Same Instance.

Runs	SA(R)	SA(G)	%D SA-R	%D SA-G
1	45767	45658	1.70	1.46
2	45793	45744	1.76	1.65
3	45663	45793	1.47	1.76
4	45460	45658	1.02	1.46
5	45567	45658	1.26	1.46
6	45562	45793	1.25	1.76
7	45569	45658	1.26	1.46
8	45567	45658	1.26	1.46
9	45682	45658	1.52	1.46
10	45663	45767	1.47	1.70
Worst	45460	45658	1.02	1.46
Best	45793	45793	1.76	1.76
Avg	45629	45705	1.40	1.57
STD	97.67	58.40	0.22	0.13
Perf. Rate	2.0E-04	3.0E-04		

9. Application Areas for Utilizing the Optimization Model

In this section, we discuss the various areas where our optimization model can be applied to the deployment components of a cloud-hosted service for guaranteeing the required degree of multitenancy isolation.

9.1. Optimal Allocation in a resource constrained environment

Our optimization model can be used to optimize the allocation of resources especially in a resource constrained environment. or where there are frequent changes in workload. This can be achieved by integrating our model into a load balancer/manager. Many cloud providers have auto-scaling programs (e.g., Amazon Auto scaling) for scaling applications deployed on their cloud infrastructure, usually based on per-defined scaling rules. However, these scaling programs do not have a functionality to provide for guaranteeing the isolation of tenants (or components) associated with the deployed service. It is the responsibility of the customer to implement such a functionality for individual service deployed to the cloud. In an environment where there are frequent workload changes, there would be a high possibility of performance interference. In such a situation, our model can be used to select the optimal configuration for deploying a component that maximizes the number of request that can be allowed to access the component while at the same maximizing the degree of isolation between tenants (or components).

9.2. Monitoring Runtime Information of Components

Another area where our model can be very useful is in monitoring the runtime information of components. Many providers offer monitoring information on network availability and utilization of components deployed on their cloud infrastructure based on some pre-defined configuration rules. However, none of this information can assure that the component is functioning efficiently and guarantees the required degree of multitenancy isolation on the application level. It is the responsibility of the customer to extract and interpret these values, adjust the configuration, and thus provide optimal configuration that guarantees the required degree of multitenancy isolation.

Our model can be implemented in the form of a simple web service-based application to monitor the service and automatically change the rules, for example, based on previous experiences, user input, or once the average utilization of resources exceeds a defined threshold. This application can be either be deployed separately or integrated into different cloud-hosted services for monitoring the health status of a cloud service.

9.2. Managing the Provisioning and Decommissioning of Components

The provisioning and decommissioning of components or functionality offered to customers by many cloud providers is through the configuration of pre-defined rules. For example, a rule can state that once an average utilization of a system resource (e.g., RAM, disk space) exceeds a defined threshold then a component shall be started. When runtime information of components is extracted, and made available as stated earlier, they can be used to make important decisions concerning the provisioning of required components and decommissioning of unused components.

10. Conclusion and Future Work

Within this research, to provide a further contribution to the current literature on multitenancy isolation and optimized component deployment, we have developed an optimisation model alongside a metaheuristic solution largely inspired by simulated annealing for the purpose of delivering near-optimal component deployment solutions specific to cloud-based applications that guarantee multitenancy isolation. First, an optimisation problem was formulated to capture the implementation of the required degree of isolation between components. Next, the problem was mapped against an MMKP to enable early resolution through the use of a metaheuristic.

Results show greater consistency and dependability from SA(Greedy), particularly when functioning on large instances, whereas SA(Random) operates sufficiently on small instances. When considering algorithm strength, SA(Greedy) portrayed low variability and thus generates higher quality solutions in dynamic environments experiencing regular changes in workload levels. SA(random) appears to have greater sensitivity and instability to small deviations of input instances compared to SA(Greedy), more so in relation to large instances.

We intend on testing the MMKP problem instances further using various metaheuristic types (e.g., genetic and estimated distribution algorithms) and combinations (e.g., simulated annealing merged with genetic algorithm) with the aim of identifying the most efficient metaheuristic to generate optimal solutions to operate in a variety of cloud deployment situations. For future research, a decision support system will be developed capable of creating or integrating an elastic load balancer for runtime information monitoring in relation to single specific components in order to deliver near-optimal component deployment solutions specific to responses required from cloud-hosted applications to changes in workload. This would be of immense value to cloud designers and SaaS customers for decision making around the provision and decommission of components.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was supported by the Tertiary Education Trust Fund (TETFUND), Nigeria and IDEAS Research Institute, Robert Gordon University, UK.

References

[1] C. Fehling, F. Leymann, R. Retter, W. Schupeck, and P. Arbitter, *Cloud Computing Patterns*. Springer, 2014.

[2] R. Krebs, C. Momm, and S. Kounev, “Architectural concerns in multi-tenant saas applications.” *CLOSER*, vol. 12, pp. 426–431, 2012.

[3] L. C. Ochei, A. Petrovski, and J. Bass, “Optimizing the Deployment of Cloud-hosted Application Components for Guaranteeing Multitenancy Isolation,” 2016 International Conference on Information Society (i-Society 2016).

[4] L. C. Ochei, J. Bass, and A. Petrovski, “Implementing the required degree of multitenancy isolation: A case study of cloud-hosted bug tracking system,” in *13th IEEE International Conference on Services Computing (SCC 2016)*. IEEE, 2016.

[5] F. Shaikh and D. Patil, “Multi-tenant e-commerce based on saas model to minimize it cost,” in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on*. IEEE, 2014, pp. 1–4.

[6] D. Westermann and C. Momm, “Using software performance curves for dependable and cost-efficient service hosting,” in *Proceedings of the 2nd International Workshop on the Quality of Service-Oriented Software Systems*. ACM, 2010, p. 3.

[7] Z. I. M. Yusoh and M. Tang, “Composite saas placement and resource optimization in cloud computing using evolutionary algorithms,” in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*. IEEE, 2012, pp. 590–597.

[8] D. Candeia, R. A. Santos, and R. Lopes, “Business-driven long-term capacity planning for saas applications,” *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 290– 303, 2015.

[9] M. L. Abbott and M. T. Fisher, *The art of scalability: Scalable web architecture, processes, and organizations for the modern enterprise*. Pearson Education, 2009.

[10] F. Leymann, C. Fehling, R. Mietzner, A. Nowak, and S. Dustdar, “Moving applications to the cloud: an approach based on application model enrichment,” *International Journal of Cooperative Information Systems*, vol. 20, no. 03, pp. 307–356, 2011.

[11] A. Aldhalaan and D. A. Menascé, “Near-optimal allocation of vms from iaas providers by saas providers,” in *Cloud and Autonomic Computing (ICCAC), 2015 International Conference on*. IEEE, 2015, pp. 228–231.

[12] E.-G. Talbi, *Metaheuristics: from design to implementation*. John Wiley & Sons, 2009, vol. 74.

[13] D. Menasce, V. Almeida, and D. Lawrence, *Performance by design: capacity planning by example*. Prentice Hall, 2004.

[14] F. Rothlauf, *Design of modern heuristics: principles and application*. Springer Science & Business Media, 2011.

[15] J. E. Beasley, “Or-library: distributing test problems by electronic mail,” *Journal of the operational research society*, vol. 41, no. 11, pp. 1069–1072, 1990.

[16] Z. Eckart and L. Marco. Test problems and test data for multiobjective optimizers. Computer Engineering (TIK) ETH Zurich. [Online]. Available: http://www.tik.ee.ethz.ch/sop/.../testProblemSuite/

[17] E. Zitzler and L. Thiele, “Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach,” *IEEE transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[18] R. Parra-Hernandez and N. J. Dimopoulos, “A new heuristic for solving the multichoice multidimensional knapsack problem,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, no. 5, pp. 708–717, 2005.

[19] N. Cherfi and M. Hifi, “A column generation method for the multiple-choice multi-dimensional knapsack problem,” *Computational Optimization and Applications*, vol. 46, no. 1, pp. 51–73, 2010.

[20] L. C. Ochei, A. Petrovski, and J. Bass, “An approach for achieving the required degree of multitenancy isolation for components of a cloud-hosted application,” in *4th International IBM Cloud Academy Conference (ICACON 2016)*, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 184-189 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Predicting Smoking Status Using Machine Learning Algorithms and Statistical Analysis

Charles Frank[1], Asmail Habach[2], Raed Seetan[*,1], Abdullah Wahbeh[1]

[1] *Computer Science Department, Slippery Rock University, 16057, USA*

[2] *Mathematics Department, Slippery Rock University, 16057, US*

A R T I C L E I N F O

Article history:
Received: 23 February, 2018
Accepted: 11 March, 2018
Online: 27 March, 2018

Keywords:
Machine Learning
Smoker Status Prediction
Healthcare

A B S T R A C T

Smoking has been proven to negatively affect health in a multitude of ways. As of 2009, smoking has been considered the leading cause of preventable morbidity and mortality in the United States, continuing to plague the country's overall health. This study aims to investigate the viability and effectiveness of some machine learning algorithms for predicting the smoking status of patients based on their blood tests and vital readings results. The analysis of this study is divided into two parts: In part 1, we use One-way ANOVA analysis with SAS tool to show the statistically significant difference in blood test readings between smokers and non-smokers. The results show that the difference in INR, which measures the effectiveness of anticoagulants, was significant in favor of non-smokers which further confirms the health risks associated with smoking. In part 2, we use five machine learning algorithms: Naïve Bayes, MLP, Logistic regression classifier, J48 and Decision Table to predict the smoking status of patients. To compare the effectiveness of these algorithms we use: Precision, Recall, F-measure and Accuracy measures. The results show that the Logistic algorithm outperformed the four other algorithms with Precision, Recall, F-Measure, and Accuracy of 83%, 83.4%, 83.2%, 83.44%, respectively.

1. Introduction

As of 2009, smoking has been considered the leading cause of preventable morbidity and mortality in the United States, continuing to plague the country's overall health [1]. Patients admitted to a hospital are often asked their smoking status upon admission, but a simple yes/no answer can be misleading. Patients who answer no can previously be smokers, or have recently quit smoking. The 'no' responses also do not consider their household member's smoking status, which can lead to continued exposure to secondhand smoke. Lastly, a 'no' response could still experience tobacco exposure through other forms, such as chewing tobacco. This study aims to use machine learning algorithms to predict a patient's smoking status based on medical data collected during their stay at a medical center. In the future, these predictive models may be useful for evaluating a patient's smoking status who is unable to speak.

Smoking has been proven to negatively affect your health in a multitude of ways. Smoking and secondhand smoke can magnify current harmful health conditions, and has been linked as the cause for others. Smoking and secondhand smoke often trigger asthma attacks for persons suffering from Asthma, and almost every case of Buerger's disease has been linked to some form of tobacco exposure. Various forms of cancer are caused by smoking, secondhand smoke, and other tobacco products [2]. In addition to being deemed the cause of certain cancers, most commonly known for causing lung and gum cancer, smoking and secondhand smoke also prevents the human body from fighting against cancer. Gum disease is often caused by chewing tobacco products, but continuing to smoke after gum damage can inhibit the body from repairing itself, including the gums. Smoking, secondhand smoke, and tobacco products are included in creating and preventing the recovery of the following additional diseases or health conditions: chronic obstructive pulmonary disease (COPD), diabetes, heart disease, stroke, HIV, mental health conditions such as depression and anxiety, pregnancy, and vision loss or blindness [3].

The objective of this study is to gain insights on smoking by exploring and studying patient's summary information after hospital admission. Using predictive machine learning models, a better understanding of tobacco's effect on a patient's health status can be obtained if the models produce valuable results. Using such models and existing patients' related information, understanding how vitals and patient data reflect the use of tobacco and smoking, could help medical professionals have better understanding of the

[*] Corresponding Author: Raed Seetan, Computer Science Department, Slippery Rock University, 16057, USA, Email: raed.seetan@sru.edu Tel: 724-738-2940

https://dx.doi.org/10.25046/aj030221

smokers population, which in turn can help better treat and handle patients with previous or current tobacco use more effectively.

Machine learning techniques are being applied to a growing number of domains including the healthcare industry. The fields of machine learning and statistics are closely related, but different in terms a number of terminologies, emphasis, and focus. In this work, machine learning is used to predict the smoking status of patients using several classification algorithms. Such algorithms include Multilayer Perceptron, Bayes Naïve, Logistic Regression, J48, and Decision Tree. The algorithms are used with the objective of predicting a patient's smoking status based on vitals. To determine if smoking has negative effects on vitals, One-way ANOVA analysis with SAS tool will be used repeatedly to determine whether different blood test readings from the patients are statistically different between smokers and non-smokers. The dataset used in this study was obtained from a community hospital in the Greater Pittsburgh Area [4]. The data set consists of 40,000 patients as well as 33 attributes.

The remainder of this paper is structured as follows: Section 2 discusses related work and the used dataset. Section 3 presents analytic methods and results. Section 4 discussed the results provides further recommendations; and Section 5 concludes the study.

2. Related Work

The I2b2 is a national center for Biomedical Computing based at Partners HealthCare System in Boston Massachusetts [5]. I2b2 announced an open smoking classification task using discharge summaries. Data was obtained from a hospital (covered outpatient, emergency room, inpatient domains). The smoking status of each discharge summary was evaluated based on a number of criteria. Every patient was classified as "smoker", "non-smoker", or "unknown". If a patient is a smoker, and temporal hints are presented, then smokers can be classified as "past smoker" or "current smoker." Summaries without temporal hints remained classified as "smoker".

Uzuner et. al. utilized the i2b2 NLP challenge smoking classification task to determine the smoking status of patients based on their discharge records [6]. Micro-average and macro-averaged precision, recall, and F-measure were metrics used to evaluate performance in the study. A total of 11 teams with 23 different submissions used a variety of predictive models to identify smoking status through the challenge with 12 submissions scoring F-measures above 0.84. Results showed that when a decision is made on the patent smoking status based on the explicitly stated information in medical discharge summaries, human annotators agreed with each other more than 80% of the time. In addition, the results showed that the discharge summaries express smoking status using a limited number of key textual features, and that many of the effective smoking status identifiers benefit from these features.

McCormick et. al., also utilized the i2b2 NLP challenge smoking classification task using several predictive models on patient's data to classify a patient's smoker status [7]. A classifier relying on semantic features from an unmodified version of MedLEE (a clinical NLP engine) was compared to another classifier which relied on lexical features. The classifiers were compared to the performance of rule based symbolic classifiers. The supervised classifier trained by MedLEE stacked up with the top performing classifier in the i2b2 NLP Challenge with micro-averaged precision of 0.90, recall of 0.89, and F-measure of 0.89.

Dumortier et. al. studied a number of machine learning approaches to use situational features associated urges to smoke during a quit attempt in order to accurately classify high-urge states. The authors used a number of classifiers including Bayes, discriminant analysis, and decision tree learning methods. Data was collected from over 300 participants. Sensitivity, specificity, accuracy and precision measures were used to evaluate the performance of the selected classifiers. Results showed that algorithms based on feature selection achieved high classification rates with only few features. The classification tree method (accuracy = 86%) outperformed the naive Bayes and discriminant analysis methods. Results also suggest that machine learning can be helpful for dealing with smoking cessation matters and to predict smoking urges [8].

3. Data Analysis and Results

The analysis is divided into two parts. In part 1, we use One-way ANOVA analysis with SAS tool to show the statistically significant difference in blood test readings between smokers and non-smokers. In part 2, we use five machine learning algorithms - Naïve Bayes, MLP, Logistic regression classifier, J48 and Decision Table - to predict the smoking status of patients. To compare the effectiveness of these algorithms we use four metrics, namely Precision, Recall, F-measure and Accuracy measures.

3.1. Statistical Analysis using ANOVA Test

In this work, One-way ANOVA analysis with SAS tool [9] was used repeatedly to determine whether different blood test readings from the patients are statistically different between smokers and non-smokers. So, our hypothesis are as follows.

Null Hypothesis (H0): There is no statistical difference in blood test readings between smokers and non-smokers.

Alternative Hypothesis (H1): There is a statistical difference in blood test readings between smokers and non-smokers.

The One-way ANOVA test, based on a 0.05 significance level, and the decision rule will be based on the p-value from the SAS outputs. If the p-value is less than 0.05, the null hypothesis is rejected and the alternative hypothesis is accepted. On the other hand, if the p-value is greater than 0.05, the null hypothesis is accepted. The analysis will be repeated for all blood tests, each of which is listed in Table 1 along with a brief description of its significance.

The results in Table 2 show that there is a significant statistical difference between smokers and non-smokers when it comes to three blood tests: INR, HB, and HCT. To investigate whether these differences were in favor of smokers or non-smokers, descriptive analysis was used (Figures 1, 2, and 3) to show the distribution of each blood test between smokers and non-smokers.

Figure 1 shows that non-smokers have higher values of INR than smokers. According to Mayo Clinic [10], an INR range of 2.0 to 3.0 is generally an effectiveness of anticoagulants. This shows that non-smokers have a more effective therapeutic range than smokers.

Table 1: Lab value definitions

Blood Test	Significance
INR	Measures the effectiveness of the anticoagulants
Platelets	Involved in clotting
Glucose	Main source of energy and sugar
RBC	Red blood cells: carry oxygen and waste products
HB	Hemoglobin: Important enzyme in the RBCs
HCT	Hematocrit: measures the %RBC in the blood
RDW	Red blood cell distribution width

Table 2: Consolidated statistical results

Vital Reading	P-Value	Decision
INR	<0.0001	Reject the Null Hypothesis
Platelets	0.2935	Accept the Null Hypothesis
Glucose	0.1559	Accept the Null Hypothesis
RBC	0.0882	Accept the Null Hypothesis
HB	0.0005	Reject the Null Hypothesis
HCT	0.0022	Reject the Null Hypothesis
RDW	03509	Accept the Null Hypothesis

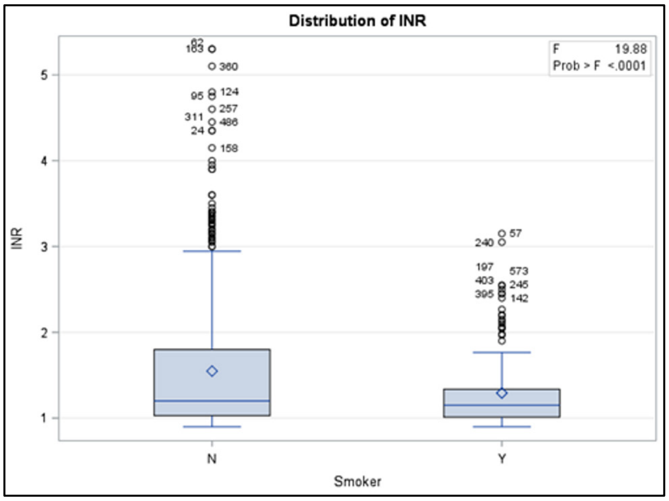


Figure 1: Distribution of INR blood test results between smokers and non-smokers

Figure 2 shows that non-smokers have lower values of HB than smokers. According to Mayo Clinic, an HB range between 12.0 and 17.5 is considered normal. This shows that although the readings of HB blood tests were statistically different between non-smokers and smokers, the difference was in general within the normal range.

Figure 3 shows that non-smokers have lower values of HCT than smokers. According to Mayo Clinic, an HB range between 37.0 and 52.0 is considered normal. This shows that although the readings of HCT blood tests were statistically different between non-smokers and smokers, the difference was in general within the normal range.

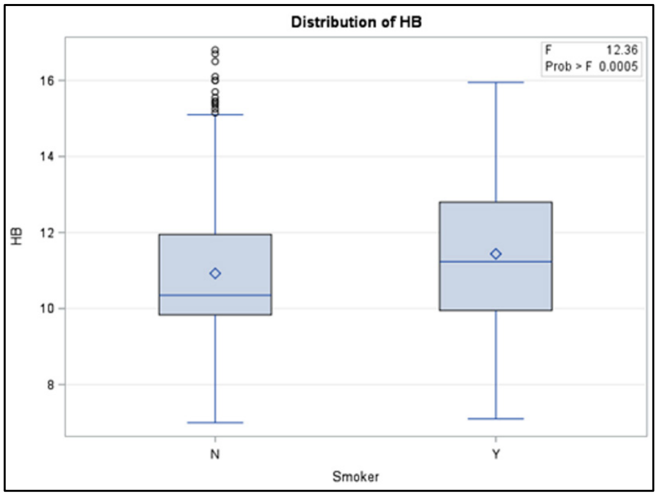


Figure 2: Distribution of HB blood test results between smokers and non-smokers

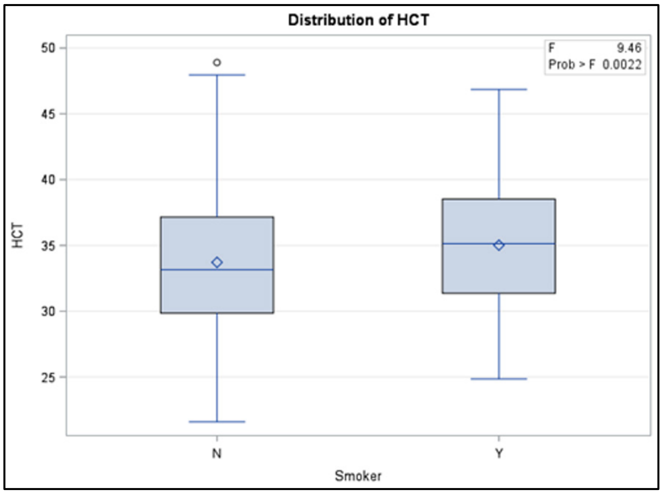


Figure 3: Distribution of HCT blood test results between smokers and non-smokers

3.2. Classification Analysis using Machine Learning

In this work, the Waikato Environment for Knowledge Analysis (Weka) (https://www.cs.waikato.ac.nz/ml/weka/) will be utilized to analyze the dataset [11]. The machine learning models utilized in this study include five classification algorithms, namely, Naïve Bayes, Multilayer Perceptron, Logistic, J48, and Decision Table.

3.2.1. Classifiers Description

Table 3 provides a summary about the classification algorithms characteristics and features. Naive Bayes is a popular versatile algorithm based on Bayes' Theorem, from the English mathematician Thomas Bayes. Bayes' Theorem provides the relationship between the probability of two events and the conditional probabilities of those events. The Naïve Bayes Classifier assumes that the presence of one feature of a class is not related to the presence or absence of another. Naïve Bayes classifier is a well-known algorithm because of its reputation for computational efficiency and overall predicative performance [12].

Table 3: Summary of classifiers characteristics and feature

Algorithm	Characteristics and Feature
Naïve Bayes	Computationally efficient, independence assumptions between the features, needs less training data, works with continuous and discrete data.
MLP	Many perceptrons organized into layers, ANN models are trained but not programmed, consist of three layers: input layer, hidden layer, and output layer.
Logistic	Multinomial logistic regression model with a ridge estimator
J48	Creates a binary tree, selects the most discriminatory features, and comprehensibility
Decision Table	Groups class instances based on rules, easy to understand, provides good performance

Multilayer Perceptron (MLP) is an Artificial Neural Network (ANN) model that maps sets of input data onto sets of suitable output data. ANN models are trained, not programmed. This means that the model takes a training set of data and applies what it has learned to a new set of data (the test data). The MLP ANN model is similar to a logistic regression classifier, with three layers: input layer, hidden layer, and output layer. The hidden layer exits to create space where the input data can be linearly separated. More hidden layers may be used for added benefit and performance, but MLP is used because of its overall performance [13].

The Logistic algorithm is a classifier for building and using multinomial logistic regression model with a ridge estimator to classify data. The version implemented using Weka states that it is slightly modified from the normal Logistic regression model, mainly to handle instance weights [14].

The J48 algorithm is a popular implementation of the C4.5 decision tree algorithm. Decision tree models are predictive machine learning models that determine the output value based on the attributes of input data. Each node of a decision tree signifies each attribute of the input data. The J48 model creates a decision tree that identifies the attribute of the training set that discriminates instances most clearly. Instances that have no ambiguity are terminated and assigned an obtained value, while other cases look for an attribute with the most information gain. When the decision tree is complete, and values are assigned to their respective attributes, target values of a new instance are predictively assigned [15].

Lastly, the Decision Table algorithm utilizes a simple decision table to classify data. Decision tables are best described to programmers as an if-then-else statement, and less complicated as a flow chart. A decision table groups class instances based on rules. These rules sort through instances and their attributes and classify each instance based on those rules. Decision tables are often easier to understand than other algorithm models while providing necessary performance [16].

Each model was run with a 66% split, using 66% of input as the training data and 34% as the test data. All algorithms are implemented through Weka after preprocessing, Table 4.

Table 4: Weka Schema

Algorithm	Weka Schema Attribute
Naïve Bayes	weka.classifiers.bayes.NaiveBayes
MLP	weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 5 -E 20 -H a
Logistic	weka.classifiers.functions.Logistic -R 1.0E-8 - M -1 -num-decimal-places 4
J48	weka.classifiers.trees.J48 -C 0.25 -M 2
Decision Table	weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"

3.2.2. Preprocessing

A sample of the large dataset was used for analysis due to available resources. A few samples were created prior to this study for previous work. The sample contains 534 total patients, with 311 non-smokers and 87 smokers. This remains relatively consistent with the overall ratio of smokers to non-smokers in the full dataset. The smoker attribute contained 136 missing values, accounting for 25% of the patients in the sample. To account for missing values, we utilized a Weka filter called "ReplaceMissingValues". This filter replaces missing values through the selected attribute with modes and means of the values in the training set.

After addressing missing values for the class attribute, oversampling was applied to add additional data for analysis. Oversampling was added to try and alter the ratio of smokers to non-smokers closer to the original dataset. The SMOTE algorithm was used through Weka and applied three times, bringing the total instances to about 1000 patients. All preprocessing is done using Weka as listed in Table 5.

Table 5: Weka Filters

Algorithm	Type	Weka Attribute
SMOTE	Filter	weka.filters.supervised.instance.SMOTE
ReplaceMissing Value	Filter	weka.fliters.unsupervised.attribute.-ReplaceMissingValue

3.2.3. Means of Analysis

To evaluate the performance of the machine learning model's, four different measures are used, namely: Precision, Recall, F-measure, and Accuracy, shown in equations 1-4. Precision shows the percent of positive marked instances that truly are positive. Recall is the percentage of positive instances that are correctly identified. Recall is also referred to as sensitivity. F-measure or F-score is a measure of accuracy, that considers the harmonic mean of precision and recall. Accuracy is simply the amount of correctly classified instances from an algorithm.

$$Precision = \frac{TP}{TP + FN} \tag{1}$$

$$Recall = \frac{TN}{TN + FP} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (3)$$

$$\mathrm{F} - measure = 2 \,.\, \frac{Precision\ .Recall}{Precision + Recall} \qquad (4)$$

Figures 4 shows the performance of the five algorithms using the Precision measure. Results show that the J48 and Logistic achieved the highest precision with 83%, followed by MLP, Decision Tree, and Naïve Bayes with Precision values of 81%, 80.5%, and 77.8% respectively.

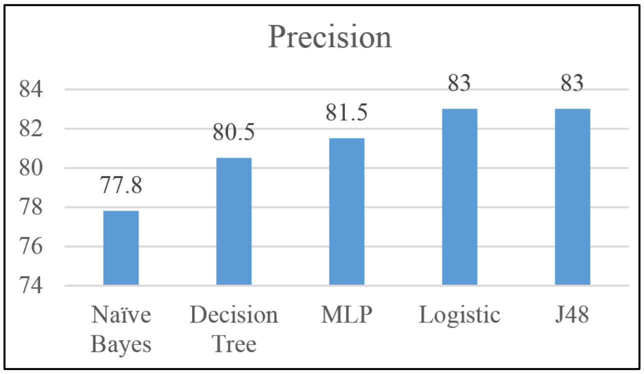


Figure 4: Precision for Naïve Bayes, MLP, Logistic, J48, and Decision Tree results

Figures 5 shows the performance of the five algorithms using the Recall measure. Results show that the J48 achieved the highest Recall with 83.4%, followed by Logistic, MLP, Decision Tree, and Naïve Bayes with Recall values of 83.1%, 81.8%, 81.1% and 77.8% respectively.

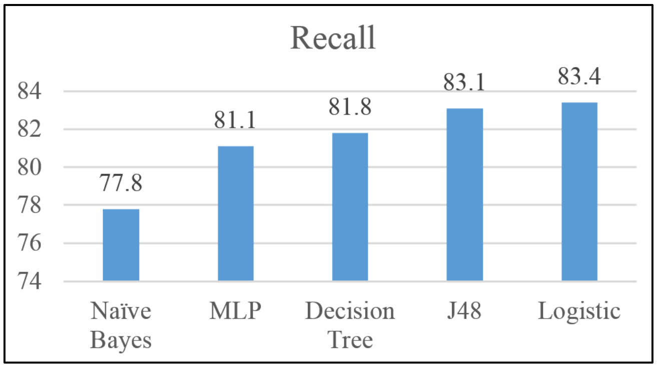


Figure 5: Recall for Naïve Bayes, MLP, Logistic, J48, and Decision Tree results

Figures 6 shows the performance of the five algorithms using the F-Measure. Results show that the Logistic achieved the highest F-Measure with 83.2%, followed by J48, MLP, Decision Tree, and Naïve Bayes with Recall values of 83.1%, 81.3%, 81% and 78.8% respectively.

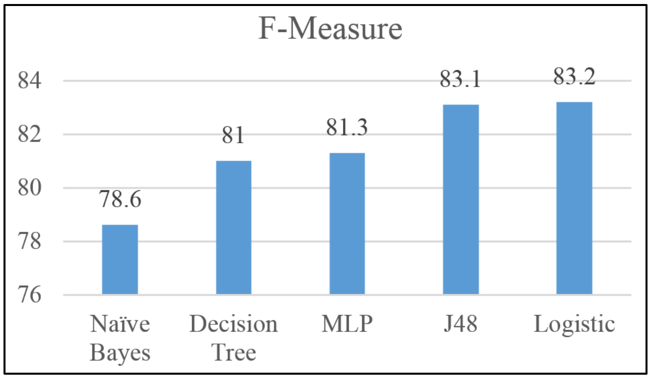
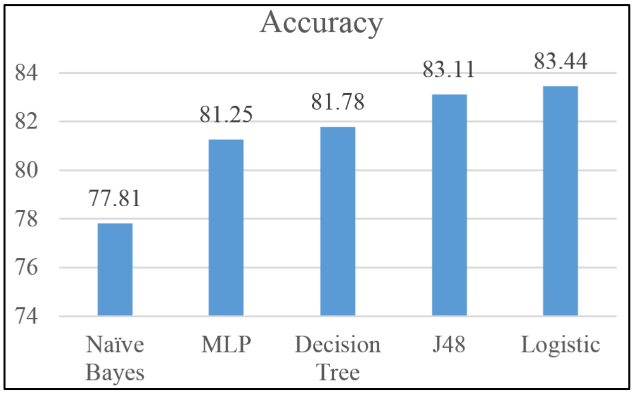


Figure 6: F-measure for Naïve Bayes, MLP, Logistic, J48, and Decision Tree results



Overall, the results show an indication that the five algorithms are relatively reliable when it comes to predicting the smoking status of patients. Logistic algorithm outperformed the four other algorithms with Precision (83%), Recall (83.4%), F-Measure 83.2%, and Accuracy (83.44%).

The study addressed the potential of machine learning algorithms to predict the status of smoking among a smoker population. Results showed the potential of such algorithms to predict the smoking states with accuracy level of 83.44%. However, this study has few limitations. There are several items that could be addressed to further this study and improve outcomes, beginning with data preprocessing. Several other methods are available to handle missing values in the dataset. In this study, the ReplaceMissingValues filter was applied in Weka to handle missing values. The Weka Filter replaces null values with means and modes from the training set. Using a method such as replacing null values with moving averages could produce results that are more realistic to the actual smoking status of patients. Other methods of handling missing values could also be explored.

Another improvement could be to increase the size of the sample dataset. In this study, a sample of 534 patients was used and the SMOTE model was applied in preprocessing. Using a larger sample set could also bring the results closer to what would be expected when applying these tests to larger sets of patients. In most cases, there will be more than 534 patient entries to analyze.

So, learning the results of these tests on larger real-world sets could further prove the value of these tests.

Lastly, other models may outperform those tested in this study. While five algorithms were tested, and showed wholesome results, others may provide better marks. Clustering models may be a point of interest as those tested in this study are all classifier models. There is an abundant amount of classifier models out there and their results are worth testing.

5. Conclusion

This study showed that five machine learning models can be used reliably to determine the smoking status of patients given blood tests and vital readings attributes. These algorithms are Naïve Bayes, MLP, Logistic, J46 and Decision Tree. Logistic algorithm outperformed the other four algorithms with precision, recall, F-Measure, and accuracy of 83%, 83.4%, 83.2%, 83.44%, respectively.

Using One-way ANOVA analysis with SAS tool, the study also confirmed that there is a significant statistical difference between smokers and non-smokers when it comes to three blood tests: INR, HB, and HCT. The difference was within the normal range with HB and HCT, but it was in favor of non-smokers with INR which measures the effectiveness of anticoagulants. In the future, the models could be implemented in hospital systems to identify patients who do not specify smoking status. Also, the findings from SAS confirms the negative health effects of being a smoker.

References

[1] Centers for Disease Control and Prevention: https://www.cdc.gov/tobacco/campaign/tips/diseases/?gclid=CjwKEAjw5_vHBRCBtt2NqqCDjiESJABD5rCJcbOfOo7pywRlcabSxkzh0VIifcvYI05u-hQ9SsI9RRoCDZfw_wcB

[2] The Mayo Clinic (2017). Retrieved May 01, 2017, from http://www.mayoclinic.org/

[3] S., Dube, A., McClave, C., James, R., Caraballo, , R., Kaufmann, & T., Pechacek, (2010, September 10). Vital Signs: Current Cigarette Smoking Among Adults Aged =18 Years United States, 2009. Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5935a3.htm

[4] M., Keyes, C., Frank, A., Habach and R. Seetan. Artificial Neural Network Predictability: Patients' Susceptibility to Hospital Acquired Venous Thromboembolism. The 32th Annual Conference of the Pennsylvania Association of Computer and Information Science Educators (PACISE), At Edinboro University of Pennsylvania. March 31st and April 1st, 2017.

[5] Partners Healthcare. (2017). Informatics for Integrating Biology & the Bedside. Retrieved May 23, 2017, from i2b2: https://www.i2b2.org/NLP/DataSets/Main.php

[6] O., Uzuner, I., Goldstein, Y., Luo, & I., Kohane, (2008). Identifying Patient Smoking Status from Medical Discharge Records. J Am Med Inform Assoc, 15(1), 14-24. doi:10.1197/jamia.m2408

[7] P., McCormick, N., Elhadad & P., Stetson (2008). Use of Semantic Features to Classify Patient Smoking Status. Retrieved March 5, 2017, from Columbia.edu: http://people.dbmi.columbia.edu/noemie/papers/amia08_patrick.pdf

[8] A., Dumortier, E., Beckjord, S., Shiffman,, & E., Sejdić (2016). Classifying smoking urges via machine learning. Computer methods and programs in biomedicine, 137, 203-213.

[9] SAS Institute Inc., SAS 9.4 Help and Documentation, Cary, NC: SAS Institute Inc., 2017.

[10] The Mayo Clinic (2017). Retrieved May 01, 2017, from http://www.mayoclinic.org/

[11] E., Frank, M., Hall, and I., Witten (2016). The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[12] E., Frank, & R., Bouckaert, (2006, September). Naive bayes for text classification with unbalanced classes. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 503-510). Springer, Berlin, Heidelberg.

[13] W., Gardner, and S. Dorling. "Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences." Atmos. Environ. 32, no. 14-15 (1998): 2627-2636.

[14] W., Hosmer, S., Lemeshow, and R., Sturdivant. Applied logistic regression. Vol. 398. John Wiley & Sons, 2013.

[15] D., Dietrich, B., Heller, and B., Yang. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. (2015). Wiley.

[16] H., Lu, & H., Liu, (2000). Decision tables: Scalable classification exploring RDBMS capabilities. In Proceedings of the 26th International Conference on Very Large Data Bases, VLDB'00 (p. 373).

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No.2, 190-194 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Performance Analysis of NLMS Channel Estimation for AMC-COFDM System

Assia Hamidane*, Daoud Berkani

Ecole Nationale Polytechnique, Department of Electronic, 16200, Algeria

ARTICLE INFO

Article history:
Received: 28 February, 2018
Accepted: 12 March, 2018
Online: 27 March, 2018

Keywords:
AMC
COFDM
CSI
NLMS
SE

ABSTRACT

In this paper we focus on the adaptive modulation and coding (AMC) techniques, makes use of the channel state information (CSI) to improve the spectral efficiency (SE) in wireless communication systems. To achieve higher data rates and lower bit error rate (BER's) channel coding can be carried out in OFDM, called COFDM. The NLMS channel estimator is used. The simulation analysis presented includes comparisons of SE and throughput vs. SNR for different proposed schemes.

1. Introduction

Intensive research interests have been in adaptive transmission techniques over wireless channels to efficiently support multimedia services, Internet access, 4G and future 5G mobile communications. One important issue is the ability to combat inter-symbol interference (ISI) in the wideband transmission over multipath fading channels.

Multicarrier modulation technique, such as orthogonal frequency division multiplex (OFDM), appears to be a promising solution to this problem. It is expended in various broadcast technologies like wireless local area network (WLAN), worldwide interoperability for microwave access (WiMAX), long term evolution (LTE), digital video broadcasting (DVB) and digital audio broadcasting (DAB) [1]. OFDM is a multicarrier system which employs a prominent number of close spaced carriers that are regulated with low data rate. In order to avoid mutual interference, the signals are made as orthogonal to each other. Coded Orthogonal Frequency Division Modulation (COFDM) is a kind of OFDM where error correction coding is comprised into the signal [2]. This system is able to attain excellent performance on frequency selective channels because the merged benefits of multicarrier modulation and coding. The multicarrier transmission reaches better results if incorporated with additive techniques in order to achieve higher efficiency in terms of throughput and ASE. The AMC-COFDM systems allows to choose the most appropriate Modulation and Coding Scheme (MCS) depending on the quality of the received signal and the channel conditions.

The aim of this paper is to develop an adaptive technique that improve the COFDM system performances: keep the error probability below a specified threshold or maximize the ASE. The MCS selection is based on the amount of received packets with errors, having as target that of minimizing that number and reducing the implementation complexity by using NLMS algorithm.

The rest of this paper is organized as follows. Section 2 gives the COFDM description where we show how the COFDM is better than OFDM. In Section 3, we present the details of AMC system and explain NLMS algorithm for channel estimation. Simulation results have been discussed in Section 4 and finally, conclusions are provided in the last section.

2. COFDM System Description

The Coded Orthogonal Frequency Division Multiplexing (COFDM) system comprises of three main elements. These are Guard interval/cyclic prefix (CP), Channel coding/interleaving and IFFT. These technical aspects make the system tolerant to multipath fading and ISI [3]. Description of this transceiver system is listed:

The digital bits are randomly generated, convolutional coding is applied with different coding rate schemes (R=1/2, 2/3 or 3/4).

*Assia Hamidane, +213 560 768 303, assia.hamidane@g.enp.edu.dz

https://dx.doi.org/10.25046/aj030222

The coded data is then interleaved to eliminate the effect of selective fading. The data is thereafter modulated using different proposed schemes (M-PSK, M-QAM). A serial-to-parallel conversion is performed. Thereafter, an IFFT is executed and guard time is added prior to parallel-to-serial conversion for making OFDM symbols tolerant to ISI. The data is thereafter digital-to-analog converted (DAC) and transmitted over the Rayleigh fading channel. At the receiver level, time and frequency synchronization is performed. The GI is removed and serial-to-parallel conversion is performed, and then an FFT is executed. The channel is estimated for each sub-carrier and the channel frequency selectivity is compensated for (equalized). After a final parallel-to-serial conversion, the data can then be demodulated and decoded. In order to achieve better performance, the receiver has to know the channel effect.

3. AMC-COFDM Proposed System

Figure 1 illustrates the block scheme of structure of the proposed AMC-COFDM system, exploiting her potentialities taking that the change of modulation level and coding rate according to the CSI, for each symbol. The block diagram represents the whole system model. Basically, the proposed system is divided in two main sections namely transmitter and receiver:

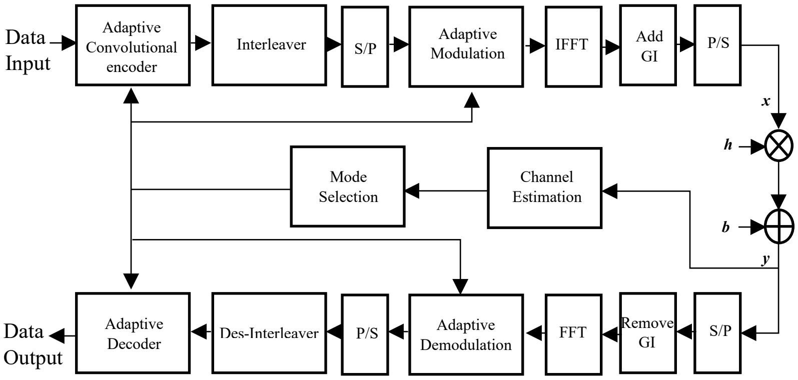


Figure 1. The proposed AMC-COFDM transceiver block diagram.

At the transmitter level, the digital bits are randomly generated and encoded using an adaptive conventional coding technique, where three coding rates are proposed (1/2, 3/4, 2/3) with a constraint length of 3 and a polynomial generator [7,5]8. The sequence at the encoder output is permuted. The employed interleaver should have the capability of breaking up low-weight input sequences so that the permuted sequences have large distances between the 1's. Thus the incoming serial coded sequence is divided into low-rate data streams that are simultaneously transmitted, where each of the low-rate data streams is mapped using an adaptive modulation technique. There are four available modulation types for modeling the data onto sub-carriers: BPSK, QPSK, 16-QAM and 64-QAM used with gray coding in the constellation map. Depending on the feedback information, from the receiver, the mapping scheme and coding rate are adjusted.

The next stage consists to perform invers fast Fourier transform (IFFT), insert a guard area (GI) by the addition of cyclic prefix (CP). thus combating the inter-symbol interference (ISI) and inter-carrier interference (ICI) introduced by the multipath fading channel. As a last step Before the transmission through channel, a parallel to serial conversion is made.

As shown in the diagram block, at the reception level the inverse procedure is performed, where the serial to parallel conversion, the CP remove, and FFT are executed. To obtain the original data bits, the reverse process (demapping, de-interleaving, Viterbi decoding) is executed.

Besides, the NLMS channel estimator method is used. The knowledge on the Channel Impulse Response (CIR) provided by this method to the detector, is transmitted also through feedback channel. The basic concept of this channel estimation algorithm is to know sequence of bits, which is reoccurred in every transmission burst. The transferred symbols going through the channel transmission, can be regarded as a circular convolution between the CIR and the transmitted data block. In the receiver side, the channel coefficients are normally unknown. It demands to be efficiently estimated to maintain a low computational complexity.

3.1. Least Mean Square (LMS) Algorithm

The Least Mean Square (LMS) algorithm was first developed in 1959 by Widrow and Hoff through their studies of pattern recognition [4]. From then it became one of the most widely used algorithms in adaptive filtering. The LMS is an adaptive filter algorithm known as stochastic gradient-based algorithm since it utilizes the gradient vector of the filter tap weights to converge on the optimal wiener solution [5,6]. It is well known by his computational simplicity and implementation facility. This simplicity that made it the benchmark against which all other adaptive filtering algorithms are judged. With each iteration of our algorithm, the tap weights of the adaptive filter are updated according to the following relation:

$$\mathbf{h}(n+1)=\mathbf{h}(n)+2\mu\, e(n)\,\mathbf{x}(n) \tag{1}$$

Where $\mathrm{x}(n)$ represents the input vector of time delayed input values:

$$\mathbf{x}(n)=\left[\mathrm{x}(\mathrm{n})\,\mathrm{x}(\mathrm{n}-1)\,\mathrm{x}(\mathrm{n}-2)\ldots\ \mathrm{x}(\mathrm{n}-\mathrm{N}+1)\right]^{T} \tag{2}$$

The vector $\mathbf{h}(n)=\left[\mathrm{h}(\mathrm{n})\,\mathrm{h}_1(\mathrm{n})\,\mathrm{h}_2(\mathrm{n})\ldots\ \mathrm{h}_{\mathrm{N}-1}(n)\right]^{T}$ is the coefficients of adaptive FIR filter tap weight vector at time n . The parameter μ is step size (a small positive constant) which controls the influence of the updating factor. The choice of a suitable value for μ is imperative to the performance of the LMS algorithm, if it is too small the time the adaptive filter takes to converge on the optimal solution will be too long; if μ is too large the adaptive filter becomes unstable and its output diverges [7,8].

The three distinct steps required for each LMS algorithm iteration are given as follows:

- First, the output of the FIR filter $y(n)$ is calculated by following equation:

$$y(n)=\sum_{i=0}^{N-1} h(n)x(n-i)=\mathbf{h}^{T}(n)\mathbf{x}(n) \tag{3}$$

- Then, we calculate the error estimation value using the following relation:

$$\mathrm{e}(n)=\mathrm{d}(n)-\mathrm{y}(n) \qquad (4)$$

- Thereafter, the tap weights of the FIR vector are updated using the relation (1) in preparation for the next iteration.

The primary trouble of the LMS algorithm is its sensitivity to the scaling of its input signals. This causes it very hard to select h that guarantees stability of the algorithm. Also, among its disadvantages is having a fixed step size parameter for every iteration [9]. This requires an understanding of the statistics of the input signal prior to commencing the adaptive filtering operation. In practice this is rarely achievable.

3.2. Normalized Least Mean Square (NLMS)Algorithm

The normalized least mean square algorithm (NLMS) is a variant of the LMS algorithm that resolves these problems by annealing with the power of the input signal and calculating maximum step size value, it is proportional to the inverse of the total expected energy of the instantaneous values of the coefficients of the input vector $\mathbf{x}(n)$. [i.e. Step size=1/dot product (input vector, input vector)].

The recursion formula for the NLMS algorithm is given by the following equation:

$$\mathbf{h}(n+1)=\mathbf{h}(n)+\left[\mathbf{x}(n)^{T}\,\mathbf{x}(n)\right]^{-1} e(n)\,\mathbf{x}(n) \qquad (5)$$

- Implementation of the NLMS algorithm:

We implement the NLMS algorithm in MatLab. It gives a great stability with unknown signals since the step size parameter is selected based on the current input values. The NLMS algorithm practical implementation is similar to that of the LMS algorithm. Each iteration of this algorithm requires these following steps:

first, we calculate the output of the adaptive filter using the following relation:

$$y(n)=\sum_{i=0}^{N-1} h(n)x(n-i) = \mathbf{h}^{T}(n)\mathbf{x}(n) \qquad (6)$$

Then, we calculate the error estimation value using the relation (4).

Thereafter, the step size value for the input vector is determined by:

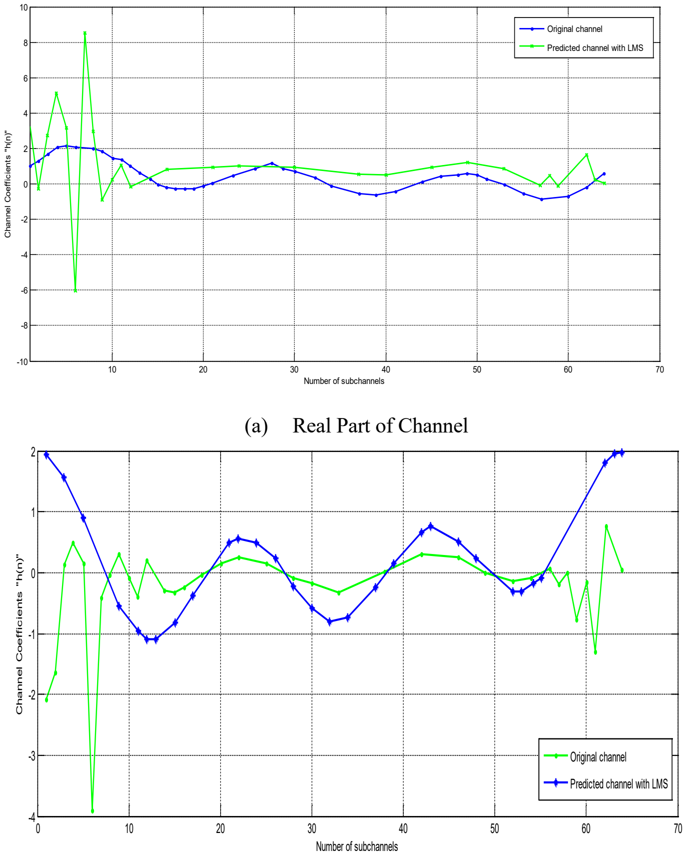
$$\mu(n)=\frac{1}{\mathbf{x}(n)^{T}\,\mathbf{x}(n)} \qquad (7)$$

And then, the filter tap weights are updated using the relation (5) [i.e.

$\mathbf{h}(n+1)=\mathbf{h}(n)+\left[\mathbf{x}(n)^{T}\,\mathbf{x}(n)\right]^{-1} e(n)\,\mathbf{x}(n)$] in preparation for the next iteration.

3.3. Analysis of Simulated Algorithms

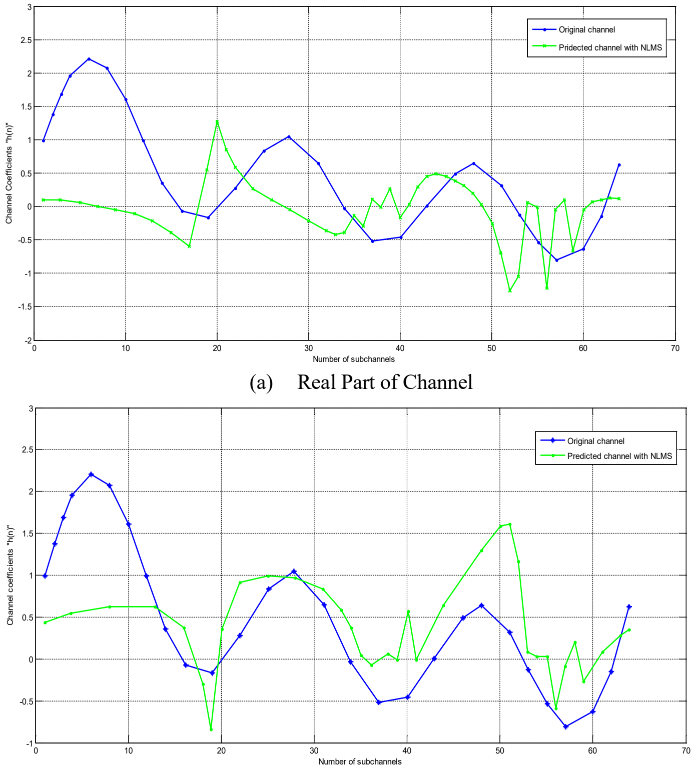
In this section, we will examine the performance of LMS algorithm in comparison with the NLMS algorithm. The adaptive filter is a 1025th order FIR filter. The step size was set to 0.0005.



(a) Real Part of Channel

(b) Imaginary Part of Channel

Figure 2. Channel estimation using the LMS algorithm.



(a) Real Part of Channel

(b) Imaginary Part of Channel

Figure 3. Channel estimation using the NLMS algorithm.

Figure 2 shows the original and predicted values of channel coefficients with LMS algorithm, but the tracking is not efficient. It does not predict the coefficients exactly. So, NLMS algorithm is adopted to have better tracking compared to LMS. It is shown in figure 3.

The LMS algorithm is the most popular adaptive algorithm because of its simplicity. However, the LMS algorithm suffers from slow and data dependent convergence behavior. The NLMS algorithm, an equally simple, but more robust variant of the LMS algorithm, exhibits a better balance between simplicity and performance than the LMS algorithm. Due to its good properties the NLMS has been largely used in real-time applications.

4. Simulation Results

This paper considers a NLMS-AMC-COFDM binary transmission system. The Simulation model was implemented in MatLab. The system parameters used in simulations are given in Table 1.

Table 1. Parameters definition.

Parameter	Values
Number of sub channels	1024
Number of pilots	128
GI	256
Number of subcarriers	896
Channel length	16
Modulation schemes	BPSK, QPSK, 16QAM, 64QAM

In this section throughput vs. SNR performance is compared for various combination "modulation - code rate" schemes (BPSK1/2, QPSK 3/4, 16QAM 3/4 and 64QAM 2/3) with NLMS-Coded OFDM. In adaptive modulation and coding (AMC) technique modulation scheme and convolution code rate are altered according to the variation in the communication channel. The transmitter will choose the appropriate combination scheme is depend upon the SNR threshold.

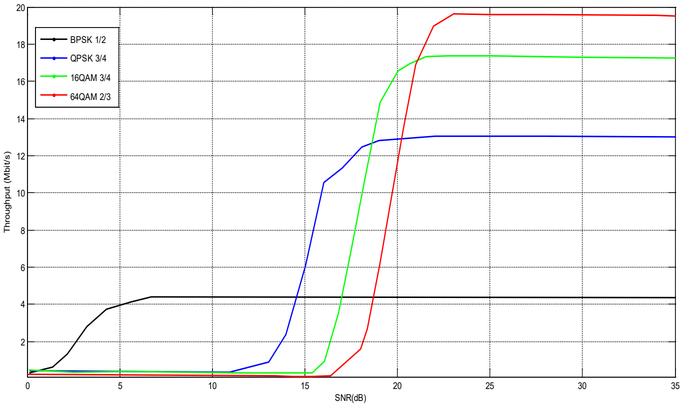


Figure 4. Throughput of AMC schemes for NLMS-COFDM system.

Figure 4 presents the throughput versus SNR graphs for different proposed combination schemes selected by channel estimator using the COFDM scenario. We observe that 64QAM with code rate 2/3 offers a significant performance gain of 3dB to 4dB. As we see here, at a given SNR NLMS-COFDM can provide a significant increase in throughput when combined with suitable link adaptation, since higher throughput modes can be used at much lower values of SNR.

AMC increases the robustness of the system. In this way, the spectral efficiency of NLMS-COFDM and NLMS-AMC - COFDM are exploited in figure 5 to improve the performance. At lower SNR values BPSK is preferred because it gives lower BER values compared to remaining techniques. As modulation order increases, SE also increases for a given SNR value.

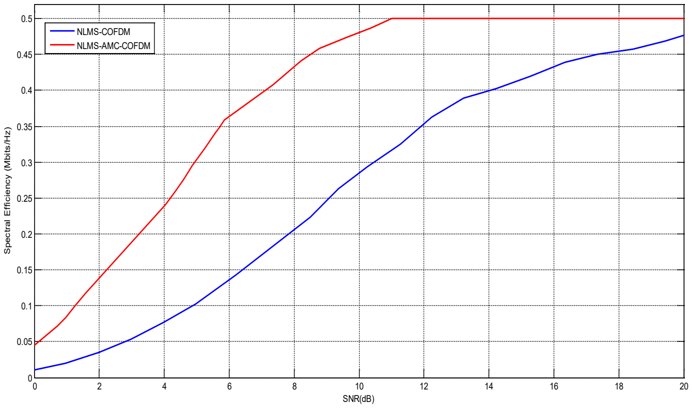


Figure 5. Spectral Efficiency of AMC-NLMS-COFDM.

5. Conclusion

This paper exploits the Coded OFDM performance in terms of SE and throughput for different modulation and code rate schemes over a fading channel. COFDM is powerful modulation technique to achieve higher bit rate and to eliminate ISI. So, the COFDM has chosen for high speed digital communications of the next generation. This paper discusses the importance of Adaptive Modulation Coding technique. NLMS channel estimator is considered because of its low complexity. The most suitable MCS is chosen at the transmitter on the basis of channel estimator in terms of BER.

References

[1] S. K. Chronopoulos, G. Tatsis, V. Raptis and P. Kosta- rakis, "Enhanced PAPR in OFDM without Deteriorating BER Performance" International Journal of Communications, Network and System Sciences, 4 (3), 164-169, 2011. https://doi.org/10.4236/ijcns.2011.43020

[2] Chen, H.M. and Chen, W.C. and Chung, C.D. "Spectrally precoded OFDM and OFDMA with cyclic prefix and unconstrained guard ratios", Wireless Commun., IEEE Trans., **10**(5), 416-1427, 2011. https://doi.org/10.1109/TWC.2010.041211.091801

[3] Fantacci, R. and Marabissi, D. and Tarchi, D. and Habib, I., "Adaptive modulation and coding techniques for OFDMA systems", Wireless Commun., IEEE Trans., **8**(9), 4876-4883, 2009. https://doi.org/10.1109/TWC.2009.090253

[4] Soria, E.; Calpe, J.; Chambers, J.; Martinez, M.; Camps, G.; Guerrero, J.D.M.; "A novel approach to introducing adaptive filters based on the LMS algorithm and its variants", IEEE Trans., **47**(1), 127-133, Feb 2008. https://doi.org/10.1109/TE.2003.822632

[5] Rachana. N ; Pradeep. K ; Poonam. B;" An approach to implement LMS and NLMS adaptive noise cancellation algorithm in frequency domain", IEEE Confluence The Next Generation Information Technology Summit

(Confluence), Noida, India, 2014. https://doi.org/10.1109/CONFLUENCE.2014.6949268

[6] Lee, K.A.; Gan,W.S; "Improving convergence of the NLMS algorithm using constrained subband updates", Sig. Proc. Let. IEEE, **11**(9), 736-739, 2004. https://doi.org/10.1109/LSP.2004.833445

[7] R. G. Soumya ; N. Naveen ; M. J. Lal ; "Application of Adaptive Filter Using Adaptive Line Enhancer Techniques", IEEE Advances in Computing and Communications (ICACC), Cochin, India, 2013. https://doi.org/10.1109/ICACC.2013.39

[8] Minchao. L ; Xiaoli Xi ; "A New Variable Step-Size NLMS Adaptive Filtering Algorithm", IEEE Information Technology and Applications (ITA), Chengdu, China, 2014. https://doi.org/10.1109/ITA.2013.62

[9] Sajjad A. G.; Muhammad F. S.;" System identification using LMS, NLMS and RLS", IEEE Research and Development (SCOReD), Putrajaya, Malaysia, 2013. https://doi.org/10.1109/SCOReD.2013.7002542

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 195-205 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

A decision-making-approach for the purchasing organizational structure in Moroccan health care system

Kaoutar Jenoui[*], Abdellah Abouabdellah

Industrial Engineering laboratory, Team of Modeling, Optimization of Industrial Systems and Logistics (MOSIL) ENSA, University Ibn Tofail, Kenitra, Morocco.

ARTICLE INFO

Article history:
Received: 06 November, 2017
Accepted: 28 January, 2018
Online: 31 March, 2018

Keywords:
Healthcare organization
Centralization strategy
Hidden quality costs
Organizational structure

ABSTRACT

Excellence in hospitals supply management results in better quality, best prices, and good deliveries. One of the questions that come up as healthcare organization to capture the economies of scale in purchasing prices and process costs is whether their purchasing activities should be centralized or decentralized. In most cases, centralization strategy usually gives good supplier's service with a lower cost, but the consideration of supplier's cost in the hospital sector are mainly limited to visible ones. The high levels of hidden quality costs generated by suppliers and their unknown presence have serious consequences on the decisions made by the managers. However, the existence of this kind of costs has not been considered yet. Therefore, the main objective of this paper is to propose a decision-making-approach, integrating a new method of measuring supplier's hidden quality costs, in order to help managers to choose the appropriate purchasing organizational structure in the hospital sector.

1. Introduction

In Morocco, the ministry of health has undertaken several actions to make medicines and medical devices available and accessible to the population. It has a national list of medicines and essential medical devices, according to the last revision which was carried out at the end of 2011. The pharmaceutical products covered by this list benefit from an annual budget allocated by the ministry of health, to ensure their availability at the level of public hospitals and basic health care facilities. This budget increased by 67% between 2002 and 2012, reaching the sum of 1.6 billion Dirham [1-3]. Despite efforts made by the ministry, the access to medicines and medical devices in hospitals remains insufficient [2] [4]. The main factor that reduces the availability of medicines is a failure to achieve a good supplier's quality. Experiences have shown that it is possible to improve access to these products by making the best use of resources and streamlining management processes. Indeed, it is necessary to highlight the organizational purchasing structure, and improve the supplier's quality [5-7]. In this perspective, it is mandatory to study the relationship that exists between non-quality, costs and organizational structure. In fact, a change in an organizational structure impacts the entire system including customers, suppliers, and competitors. In the healthcare system, the suppliers are considered as the main stakeholders contributing to the improvement of the hospital's performance. The demand continues to support the growing supplies, which has changed the behavior of organizations to achieve good quality. This change includes both simple corrections and more complex changes for suppliers, at the level of their attitudes and their performances. The particularity of healthcare facilities is that the patient receives its service in real time. Though, the impact of the structure on the supplier's performance affects the quality of care and purchaser's satisfaction, this dissatisfaction could be estimated by a cost, which is considered as a hidden cost. Theses hidden quality costs may increase or decrease depending on the choice of the organizational structure.

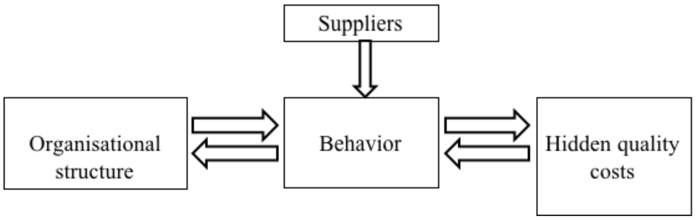


Figure 1. Relationship that exists between organizational structure and hidden costs

In light of the importance of hospital logistics, we propose in this paper to consider the supplier's hidden quality costs as one of the

[*] JENOUI Kaoutar, Email : k.jenoui@gmail.com

https://dx.doi.org/10.25046/aj030223

principle criteria to be based on for deciding the right purchasing strategy. This paper is an extension of a work originally presented in Logistiqua'17 conference [1]. It's organized as follows: Section II presents some definitions of the hospital sector, followed by section III that describes the problematic, and section IV that provides a literature review. In section V a deciding-making-approach is proposed and validated through a practical case study in section VI. Finally, section VII concludes the paper and states future work.

2. Hospital pharmacy presentation

This section provides important definitions that describe hospital pharmacy services, as well as the requirements for providing pharmaceutical care.

2.1. The hospital, hospital pharmacy and SEGMA hospital definition

The hospital is a healthcare unit composed of specialized scientific equipment, and certified healthcare professionals, working as a team for the common purpose of providing medical services to the population. While the hospital pharmacy defined as a department in a hospital, which is responsible for the supply of medical products, and headed by professionally competent and qualified pharmacist who directly supervises and ensures compounding and distribution of medication to in and out patients [8].

The status SEGMA stands for "State service managed in an autonomous way" and is defined as the services of the state whose expenses operations are executed by a head of the department attached to the superior accountant of the kingdom. In 2004, there was 38.4% of the total number of services dedicated to hospital care with SEGMA status. Currently 63% of the number of hospitals obtains the SEGMA status [9].

2.2. Goals of hospital pharmacy

The hospital pharmacy aims to provide safe and effective pharmaceutical care and services to users and health care facilities, through a global professional and ethical standards. Pharmaceutical services components are responsible for three main missions. Firstly, it's responsible for the procurement, distribution and control of pharmaceutical products, secondly, the evaluation and dissemination of comprehensive information about medical products and their use to the institution's employees and patients. And finally, it's responsible for the supplier's monitoring, evaluation, and assurance of the quality of drugs [8].

2.3. Hospital's pharmacy in Centralized and decentralized structure

Figure 2 shows the logistics flows for centralized and decentralized structure. The centralized structure aims to make orders for many hospitals independently when there is no grouping of commons orders. While decentralized structure is considered to refer to a practice model in which a pharmacist is responsible for orders purchasing and distribution services, including order validation and possibly order entry [5][8].

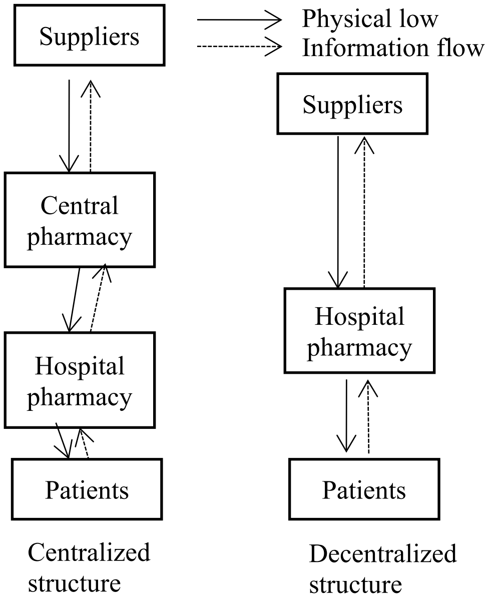


Figure 2. The information and physical flow in both structures

3. Problematic description

In 24 years, the Moroccan health care system switched many times between centralized and decentralized structure [9]. (See table 1)

In face of all these changes and developments of the organizational structure, Moroccan hospitals still suffer from inadequate quality, the insufficient availability of products and the increased costs [3].In decentralization structure, centers of storage and distribution cost up to one billion Dirhams per year, and operated by 200-250 people [1-4].Also, storage and distribution of medicines in Morocco cost the ministry of health over 30 million Dirhams per year. While in centralization structure, the rate of obsolescence of medicines decreased by 8% at the end of 2002, and the purchase prices of medicines through the centralized system in 2001 were lower than the prices paid in the decentralized system. This drop in prices allowed buying 50% more products. All these data show the supplier's bad quality and the inappropriate choice of the organizational structure, which indicates the importance of estimating poor quality costs generated by suppliers, also called hidden costs. Therefore, in this paper we propose a decision-making-approach, integrating a new method of measuring supplier's hidden quality costs, in order to help the managers to choose the appropriate purchasing organizational structure in the hospital sector.

4. Literature review

To discuss the literature related to this subject, this review will cover three principle topics, and we will start by a brief review of each of them:

4.1. Organizational structure

Decision-making can be performed in a centralized or a decentralized structure. In a centralized one, there exists a central responsibility for decision-making, whereas in a decentralized structure the individual entities can make their own decisions. In practice, each approach has its advantages and disadvantages. Most commonly, the strategic decisions are usually made centrally while operation decisions are decentralized. The performance of each approach depends on specific environment and particular decisions [9] [10]. Many researchers studied the effects of centralization and decentralization on the multi-item replenishment problem in a two-echelon supply chain. Chen et al. proposed both centralized and decentralized decision models and proved the optimal properties of both models to minimize costs [11] [12]. They found that it is beneficial to adopt centralized control and proposed a mechanism to coordinate the decentralized system so that each player in the chain benefit from it. Behdani et al. studied disruptions in a multi-plant company and considered alternative policies for coping with them. To model this complex system, they used an agent-based simulation model [13].

Table1. A brief history of the health care supply system

1980	Decentralization	With many difficulties faced by the central pharmacy, there has been an introduction of direct delivery system
1985-1986	Centralization	Worsening problems due to the small size of premises and storage spaces, implies return to centralization. There was also a need of the implementation of a new unit, and the improvement of the central pharmacy program, followed by a progressive implementation of semi-autonomous management for the regional hospitals which directly procure themselves.
1994	Decentralization	Establishment of the procurement division, reporting directly to the general secretariat of the ministry.
1995	Centralization	Commissioning of the storage unit to centralize procurement for medicines.
1997	Decentralization	Decentralization purchasing for SEGMA hospitals
2001	Centralization	Centralized purchasing (procurement, storage, distribution) for SEGMA regional hospitals by the procurement division.
2003	Decentralization	In view of difficulties of regular supply, it is envisaged to decentralize the supply of SEGMA hospitals.

However, no one has considered the hidden costs generated in each strategy. In our study, we will focus on this objective in a particular way.

Hidden quality cost:

In traditional systems, quality losses occur when the product deviates beyond the specification limits, and becomes unacceptable [14]. Taguchi proposed a narrower view of characteristic acceptability to indicate that any deviation from a characteristic's target value results in a loss, and that a higher quality measurement results in minimal variation from the target value [21-22]. Other authors analyze different aspects related to hidden quality cost in the construction business, and highlight the following examples of hidden quality costs: schedule delays, litigation and claims, loss of reputation and the subsequent impact on future business opportunities, loss of schedule and productivity, low operational efficiency, work inactivity from waiting or idle time, etc. [15-19]. Taguchi loss functions have been recently used for non-manufacturing applications. They were implemented to evaluate product quality as an aid to the selection of suppliers, been based on quantitative quality characteristics. These products are used for evaluation, comparison, and ranking process. There are several methods for the estimation of hidden costs [23-25]. However, none of them considers the quantification of hidden costs associated to qualitative criteria that presents subjectivity and uncertainty. The question which arises while measuring the cost associated to qualitative criteria is how to determine the target value, and the particular value specification that characterize the performance of each element.

4.2. Multi attributes decision making technique

There are many techniques developed for the supplier evaluation problem. Some of these techniques are categorical method, weighted point method [26], matrix approach [27], vendor performance matrix approach [28], vendor profile analysis (VPA) [29], analytic hierarchy process (AHP) [30-31], analytic network process (ANP) [32], mathematical programming [33-34] and multiple objective programming (MOP) [35-37]. However, most of these methods are more adapted to precise data. They don't have enough influence on factors such as imprecision preferences, qualitative criteria and incomplete information. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method based on the concept that the chosen alternative should have the shortest geometric distance from the positive ideal solution, and the longest geometric distance from the negative ideal solution [7]. Therefore, TOSPIS method is more appropriate to overcome problems in estimating hidden costs associated to qualitative characteristics. To sum up, the main contribution of this paper is the development of a method for the quantification of hidden quality costs based on Taguchi loss function and TOPSIS method, included in a decision-making-approach. We also compare the behavior of medicines suppliers under the centralized and decentralized strategies based on hidden quality costs. The suitability of the two strategies under two different scenarios is explored through a real case study in a Moroccan hospital.

5. The proposed deciding-making-approach

Based on the review of previous works and field interview with suppliers in Morocco, the organizational structure applied to purchase medical products for hospitals, helps to evolve the supplier's chain performance to flourish and vice versa. As indicated in figure 3, the organizational structure in the healthcare supply system is a critical component for any supplier's system improvement, and the decision to centralize or decentralize purchasing should be based on studies that focus on the actual supplier's hidden quality costs for each structure.

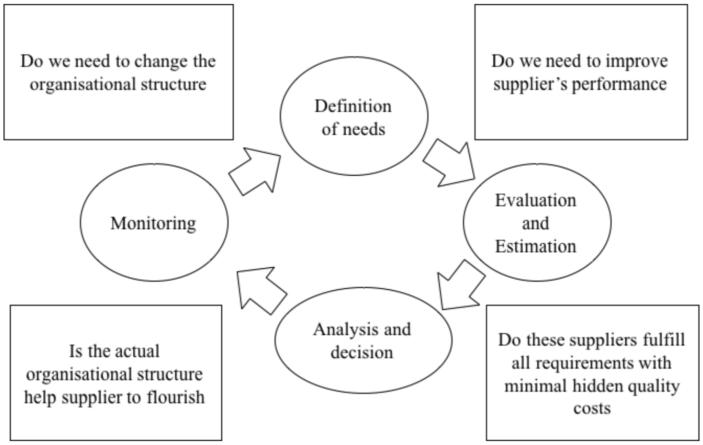

Do we need to change the organisational structure
Definition of needs
Do we need to improve supplier's performance
Monitoring
Evaluation and Estimation
Analysis and decision
Is the actual organisational structure help supplier to flourish
Do these suppliers fulfill all requirements with minimal hidden quality costs

Figure 3. Deciding-making-approach lifecycle

5.1. Definition of needs

As indicated in the figure, targets of change include improving effectiveness at two principle levels: Supplier's performance and organizational structure. There is a strong relationship between these two elements that helps ensure solid communication. Therefore, in this phase, we should:

- Define the needs regarding supplier's specifications, functional and technical requirements.
- Recognize the problems inherent in managing the actual organizational structure and obstacles that overcome.
- Discover whether the actual resources and capabilities help increase hospital's ability to create value.

5.2. Evaluation and estimation of supplier's hidden quality costs

In this phase, the full capabilities of suppliers are being gauged against defined requirements. We believe that the supplier's quality loss occurs when his performance deviates from a target, the higher the deviation is, the worst the service is rendered, and the higher the purchaser is dissatisfied. This dissatisfaction could be estimated by a cost. According to Taguchi, as indicated in figure 4, the variation of this cost is represented by a quadratic curve. The curve is centered on the target value, which represents the best performance, and determining this best value is not usually a simple task.

Three types of loss functions have been presented in the Taguchi loss function [19], nominal the better (NTB), larger the better (LTB); and smaller the better (STB). Firstly, It is necessary to determine the criteria to be based on to estimate the supplier's hidden cost, the target value, then the upper and lower specification limits.

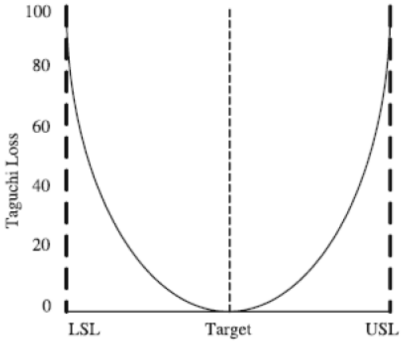


Figure 4. Two-sided loss function with specification preference

5.2.1. Supplier's hidden costs for quantitative quality characteristic

In this work, we adopt the form NTB for the quantitative quality characteristics which means that the nominal value is the best value. The loss function is given by Eq. 1.

$$L\ (y_i) = k\ (T\text{-}y_i)^2 \qquad (1)$$

Where k is the loss coefficient, whose value is constant depending on the cost at the specification limits and the width of the specification, y_i is a particular value specification, that presents the value of the product characteristic, and L(y_i) is the loss associated to y_i [19].

5.2.2. Supplier's hidden costs for qualitative quality characteristic

For qualitative criteria, we use a one-sided minimum (smaller-is-better) loss function of the form:

$$L\ (y_i) = k.(y_i)^2 \qquad (2)$$

where the meanings of L(y_i), y_i and k are the same as in equation 1. As indicated in figure 5, our philosophy aims to consider that the target value is presented as the best supplier's profile, which is characterized by the highest performance level at all quality attributes, and where the loss is equal to zero. Any supplier's distance from the ideal profile generates a cost; the smaller the distance, the lower the hidden cost is. This cost is estimated using TOPSIS method. This method measures this distance based on the supplier's performance level. It hypothesizes two artificial alternatives: The Ideal alternative that has the best level for all attributes considered, and the negative ideal alternative that has the worst attribute values [28]. TOPSIS allows determining the alternative that is the closest to the ideal solution and farthest from negative ideal alternative.

Let assumes that we have m alternatives, and n criteria. We have the score of each option with respect to each criterion. The supplier's distance from the ideal profile is calculated according to the following steps:

Let x_{ij} be the score of option i with respect to criterion j. We have a matrix X = (x_{ij}) with dimension m×n. Let J be the set of benefit attributes or criteria (LTB), and let J' be the set of negative attributes or criteria (SMB).

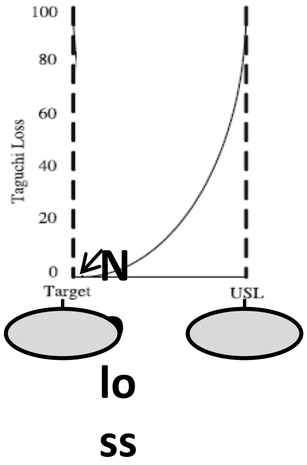


Figure 5. Loss function in case of qualitative criterion

Step 1: We construct normalized decision matrix

This step transforms various attribute dimensions into non-dimensional attributes, which allows comparisons across criteria. Scores or data are normalized as follows:

$$r_{ij} = x_{ij} / (\Sigma\, x^2_{ij}) \quad \text{for } i = 1, \dots, m;\ j = 1, \dots, n \tag{3}$$

Step 2: We construct the weighted normalized decision matrix.

In this step, we assume that we have a set of weights for each criteria w_j for j = 1,…,n. We multiply each column of the normalized decision matrix by its associated weight. An element of the new matrix is:

$$v_{ij} = w_j r_{ij} \tag{4}$$

Step 3: We determine the ideal and negative ideal solutions.

Ideal solution:

$$A^* = \{v_1^*, \dots, v_n^*\}, \tag{5}$$

Where $v_j^* = \{ \max (v_{ij}) \text{ if } j \in J ;\ \min (v_{ij}) \text{ if } j \in J' \}$

Negative ideal solution:

$$A' = \{ v_1', \dots, v_n' \}, \tag{6}$$

Where $v' = \{ \min (v_{ij}) \text{ if } j \in J ;\ \max (v_{ij}) \text{ if } j \in J' \}$

Step 4: We calculate the separation measures for each alternative.

The separation from the ideal alternative is:

$$S_i^* = [\Sigma\, (v_j^* - v_{ij})^2]^{½} \qquad i = 1, \dots, m \tag{7}$$

Similarly, the separation from the negative ideal alternative is:

$$S'_i = [\, \Sigma\, (v_j' - v_{ij})^2]^{½} \qquad i = 1, \dots, m \tag{8}$$

Step 5: We calculate the relative closeness to ideal solution y_i*

$$y_i^* = S_i^* / (S_i^* + S'_i),\ \ 0 < y_i^* < 1 \tag{9}$$

The supplier's distance from the ideal profile determines yi, which allows calculating the loss occurred by equation (2).

5.3. Analysis and decision

There are significant costs associated with organizational structure management and supplier's performance. The choice of centralization or decentralization structure may reduce supplier's flexibility and therefore makes it difficult to be responsive to changing purchasing management. Furthermore, the structure chosen changes the distance and the number of touch points between hospitals entry and suppliers. Poor planning and oversight in such a highly complex demand leads to unnecessary expenses. For example, when many "hands" are involved, the right products in the right mix and volumes may not be in the right places at the right time. Therefore, this method allows us to evaluate our satisfaction with supplier's quality in the current structure, and to be taken into account before taking any decision of change.

5.4. Supplier's Monitoring

The movement of the healthcare system organization from one form to another often requires change within the organization itself. Therefore, the organization shall operate procedures for approval and continued monitoring of all its suppliers whose products or services may affect product safety. Monitoring a supplier's performance will provide an indication of changes within the suppliers' systems and procedures that could lead to potential issues. The results of evaluations and follow-up actions shall be recorded, for future decisions.

6. Practical case study: Military hospital of Rabat

We choose in this study the Mohamed V Military Hospital of Rabat/Morocco. This hospital carries the SEGMA status, and has been subject to the change of the structure from centralization to decentralization many times from the year 2000. The objective of this study is to report the experience of this hospital in the involvement of different types of structure in the management of medical products, materials and methods from 2001 to 2003. Taking into account the total number of suppliers who participated during this period, as shown in figure 6, we find that 47% of these suppliers participated during the decentralization period, 36% of them participated during centralization, and only 17% participated in both strategies. Our study will be focused on the estimation of hidden quality costs of suppliers, who have been involved in the purchase of medical products, in both strategies: centralization in 2001, and decentralization in 2003, in order to study the impact of organizational structure on supplier's behavior.

6.1. Definition of needs

As a result of the team efforts, and as recent data are confidential, we accessed to the suppliers' evaluation data in 2001 (centralization), and those in 2003 (decentralization). We highlight that old data will not have a significant impact on our approach. In 2001, the hospital suffered from the unavailability of medical products, and the poor quality delivered by the suppliers, which was evaluated on several bases (see figure 7). The question which arises when changing the organizational structure from centralization to decentralization is whether this change will help suppliers to improve their performance or to get it worse. Therefore, in this section we will estimate the hidden quality cost behind the non-quality provided by suppliers in both strategies.

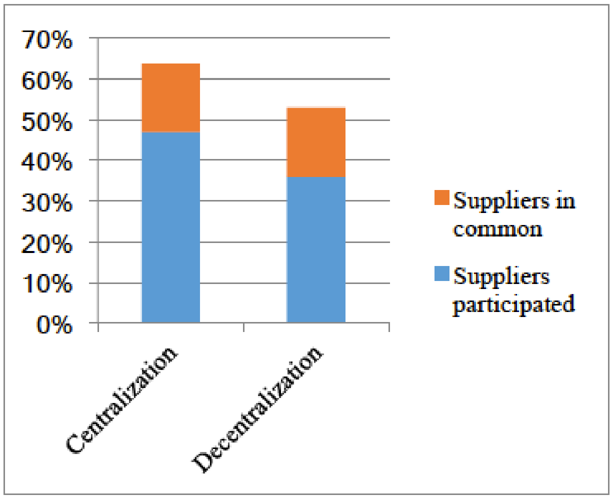


Figure 6. Suppliers participated in the period 2001-2003

6.2. Evaluation and estimation

6.2.1. Principle criteria for supplier's hidden quality costs

For the purpose of estimating supplier's hidden quality costs in the case of this hospital, we select three principle criteria: the delivery, the validity period as a single quantitative criterion, and quality as a multiple qualitative criterion, which is subdivided into five sub attributes. The objective is to estimate the supplier's hidden quality costs in seven cases as presented in the table 2.

6.2.2. Estimation of supplier's hidden quality costs

To apply the existing Taguchi loss function for quantitative quality characteristic, and the developed method for qualitative quality characteristic, we develop a purchasing model, in which we assume that the purchaser only knows the prices that might have different impacts on product quality and delivery. We estimate the loss associated to 7 suppliers according to these seven criteria.

- Quantitative quality characteristic

To calculate the loss associated to the quantitative criteria, we first need to specify the upper specification limit and lower specification limit for each criterion. For the case of this hospital, they are determined as follow:

- The maximum acceptable delivery and the minimum acceptable delivery are 7 days.
- The minimum period of validity is 18 months.

The average price of the market is 80.

As shown in table 3, we assume that the effect of price on delivery and validity period follows an exponential function, where delta price term shows the difference between the supplier price and the minimum price showed in the market. Then the estimation of the loss occurred in both structures, is determined in table 4 and 5.

- Qualitative quality characteristic

To estimate the loss associated to the quality criterion for each supplier, we suppose that: The maximum deviation from the ideal profile is 40%

- For the purchaser's orientation, the weight for each sub-criterion in the quality criterion is 0.2

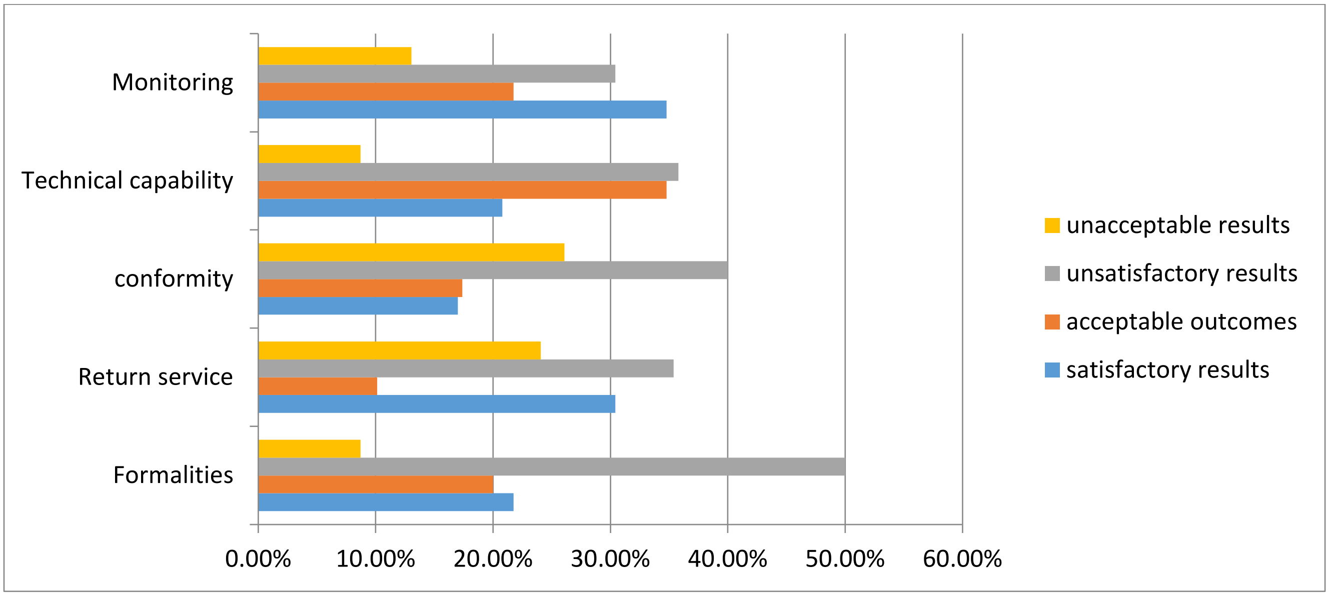


Figure 7. Supplier's evaluation criteria

Table 2. Criteria for supplier's hidden quality costs and their determination

Quantitative quality characteristic	1. Delivery		Costs to catch up the purchase unavailability as consequence of the supplier's delay or to pay extra fees to store the command for later use as a result of an order delivered in advance.
	2. Period of validity		Costs generated by expired medications due to suppliers that do not respect the validity period.
Qualitative quality characteristic	3.Quality	3.1. Monitoring	Costs to unlock the logistics problem in the hospital's upstream supply chain, generated by supplier's bad manufacturing practices, which include criteria for personnel, facilities, equipment, materials, manufacturing operations, labeling, packaging, quality control and, stability testing. These costs imply a supplier's bad conformity.
		3.2. Technical capability	Costs to maximize the return potential of the hospital's pharmacy generated by suppliers who have a bad experience in pharmaceutical returns service, and also to generate the minimum return credit for the hospital's outdated pharmaceuticals.
		3.3. Conformity	Costs to avoid unavailability of medicines in the pharmacy, and to monitor and check the supplier ongoing process of reviewing the degree to which programmed activities are completed and objectives are being met.
		3.4. Return service	Costs related to small errors of medications structure which can cause patients death. Hospital pharmacy policy requires providers to establish a continuous, systematic, and criteria-based evaluation system, such as the form and dose of drug use, that will help ensure the appropriate use of drugs.
		3.5.Formalities	Costs to assure continuity of hospital's supply and reliability of medicines quality, due to a supplier's bad technical capability. Hospital's pharmacy analysis confirms the reasonableness of the type and amount of resources proposed by the supplier. This analysis covers the proposed types and quantities of materials, labor, processes, special tooling, facilities, the reasonableness of scrap and spoilage, and other factors set forth in the supplier's proposal.

Table 3. Functions used to describe the effects of price on delivery and loss coefficients values.

	Delivery in advance	Late delivery	Period of validity in advance
Impact of price	High	Medium	Medium
Price function	$e^{\Delta price/-300}$	$e^{\Delta price/-200}$	$e^{\Delta price/-100}$
K	333.3	222.22	111.1

First, we have to determine for each supplier the independent variable y_i in equation 2. For this purpose, we rate each attribute of each alternative in the form of a matrix. To study the impact of these matrices values on final results, we assume that we have 2 experts; each one of them is assigned with the task of evaluating each supplier against each sub-criterion in each strategy. Then we determine the ideal solution, the ideal negative solution, and consequently the relative closeness to the ideal solution according to each expert (See table 6).

To get the values of k, we first need to specify the upper specification limit for each measure. As mentioned earlier, a loss of 100% will take place if a supplier image is 40% far away from the ideal profile. The other numerical measure that we need to model is how price affect quality characteristics. The expert might have different priorities regarding these criteria. For instance, one expert might be interested in high quality, while another in on-time delivery. For this reason, we assume that price effect is considered as medium or high in both centralized and decentralized structure. We use an exponential function to model the price effect for all experts.

Substituting these values in equation 2, we obtain the following loss coefficient values k, as shown in Table 7.

Table 4. Loss function results for delivery criterion

	Suppliers	y (days)	Loss L(y)
Centralization	S2	-8	21331.2
	S4	-5	8332.5
	S3	-9	26997.3
	S1	0	0
	S5	10	22220
	S6	7	10887.8
	S7	14	43551.2
Decentralization	S6	-9	26997.3
	S4	-4	5332.8
	S2	-3	2999.7
	S3	9	17998.2
	S5	10	22220
	S1	12	31996.8
	S7	14	43551.2

Table5. Loss function results for period of validity criterion

	Centralization		Decentralization	
	y (Month)	L(y)	y (Month)	L(y)
S1	2,8	871.024	2	444.4
S2	4,5	2249.775	3,4	1284.316
S3	2,3	587.719	5,2	3004.144
S4	3,6	1439.856	3,6	1439.856
S6	4,7	2454.199	3,1	1067.671
S5	5,9	3867.391	6,5	4693.975
S7	6,9	5289.471	6,7	4987.279

Table 6. Particular value specification results for each structure

	Centralization		Decentralization	
	Expert 1	Expert 2	Expert 1	Expert 2
S1	0.95299848	0.98239346	0.7435345	0.794534
S2	0.79439686	0.69798985	0.5469879	0.537699
S3	0.59252325	0.540469002	0.2435345	0.265476
S4	0.80358862	0.79417179	0.6967687	0.7176987
S5	0.99745556	0.96301608	0.2276879	0.19756789
S6	0.87174095	0.85333792	0.1632234	0.1624908
S7	0.52251934	0.55605416	0.2987698	0.283423

Table7. Functions used to describe the effects of price on quality and loss coefficients values

	Price function	k
Medium	$e^{\Delta price/-200}$	2222,22
High	$e^{\Delta price/-300}$	3333,3

After determining the y values for each supplier in each strategy, and the loss coefficients values k, we calculate the loss occurred in each case as presented in table 8.

Table8. The supplier's hidden quality loss values according to quality criterion

				Medium	High
Centralization	Expert 1		y	L(y)	L(y)
		S1	0,95299848	2018,2156	3027,3234
		S2	0,79439686	1402,35569	2103,53354
		S3	0,59252325	780,178424	1170,26764
		S4	0,80358862	1434,99603	2152,49404
		S5	0,99745556	2210,90588	3316,35882
		S6	0,87174095	1688,72152	2533,08228
		S7	0,52251934	606,719401	910,079101
	Expert 2			Medium	High
			y	L(y)	L(y)
		S1	0,98239346	2144,63835	3216,95753
		S2	0,69798985	1082,63324	1623,94986
		S3	0,60469002	812,547475	1218,82121
		S4	0,79417179	1401,56117	2102,34175
		S5	0,96301608	2060,86821	3091,30232
		S6	0,85333792	1618,17405	2427,26108
		S7	0,55605416	687,09586	1030,64379
Decentralization	Expert 1			Medium	High
			y	L1(y)	L2(y)
		S1	0,7435345	1228,52894	1842,79341
		S2	0,5469879	664,872824	997,309236
		S3	0,2435345	131,796577	197,694865
		S4	0,6967687	1078,84837	1618,27255
		S5	0,3876879	334,0009	501,001349
		S6	0,1632234	59,203582	88,805373
		S7	0,2987698	198,361113	297,541669
	Expert 2			Medium	High
			y	L1(y)	L2(y)
		S1	0,794534	1402,83992	2104,25988
		S2	0,537699	642,482941	963,724411
		S3	0,265476	156,615115	234,922673
		S4	0,7176987	1144,63616	1716,95424
		S5	0,356789	282,882483	424,323725
		S6	0,198769	87,7971562	131,695734
		S7	0,283423	178,506208	267,759312

6.3. Analysis and decision

We assume that the effect of price on delivery and validity period is the same in both centralization and decentralization strategies. However, there is a big change at the level of the supplier's hidden costs. The figure 8 indicates that loss results for the delivery criterion are much greater in the case of decentralization. The purpose of having several suppliers allows to benefit from the specific superiorities of each supplier in his specialty and to spread their risks. However, a purchase in small quantity leads to less favorable price and costs relationship. This decreases the chance of negotiating effectively the delivery deadlines, and does not help to put pressure on suppliers to meet the delivery date. The figure 9 shows that for most of suppliers, the cost relative to period of validity is high in the case of centralization. In fact, most of suppliers have medicines on stock whose period of validity have been already started, which constitutes a loss to the organization as long as they are stored. In one hand, the supplier tries to get the maximum profit, by lowering the unit purchase price to encourage the buyer. In other hand, the buyer is mainly based on maximizing the cost criterion, and does not pay enough attention to the period of validity. As a result, the buyer procures to optimize purchasing costs; but, in return the stocks in the hospital pharmacy hides a huge loss.

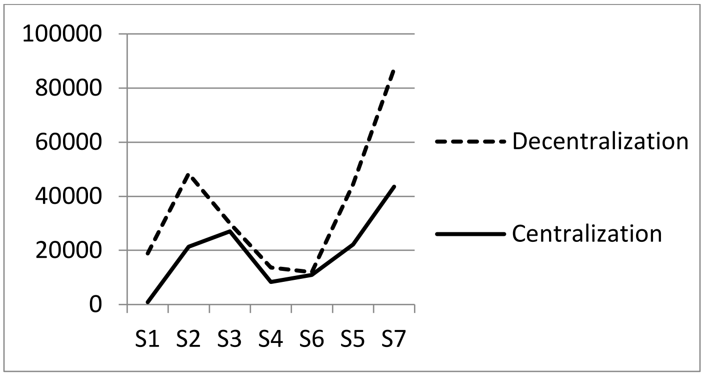


Figure 8. Loss results for delivery criterion in both structures

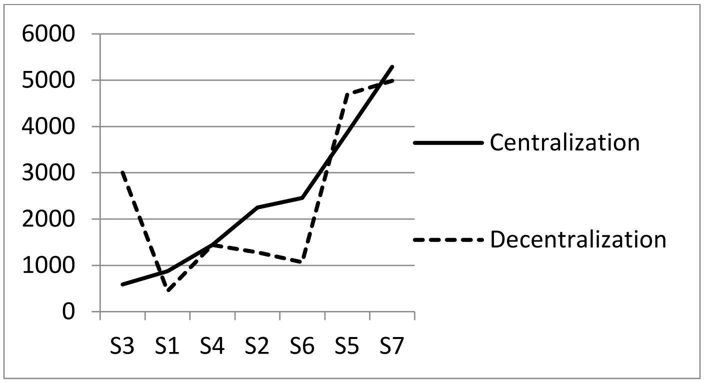


Figure 9. Loss results for period of validity criterion in both structures

In figure 10, the results determining the closest supplier to the best profile differ from one expert to another. In centralization strategy, the first expert indicates that S1 is the closest one, the second expert marks S3, while both experts' opinion answers for the decentralization strategy coincide at the same level for S6, which explains that experts' experiences bring different points of view on suppliers' evaluation.

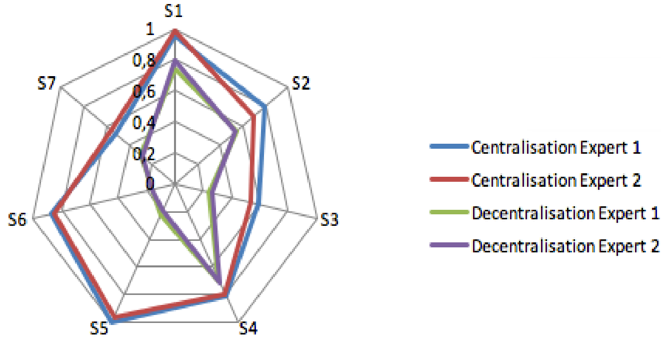


Figure 10. Expert's evaluation results

There are two parameters that might impact the supplier's loss function results: the value y, and the loss coefficient k. In one hand, the value y is influenced by the experts' evaluation (figure 11 shows that although both curves are almost symmetric for each strategy and follow the same trend, the maximum difference value is up to 5,6% in centralization strategy). In other hand, the loss coefficient is influenced by the determination of the impact of price on quality criterion, which is different from an organizational structure to another. As indicated in Figure 12, an increase in the loss coefficient results in an increase of the loss function, whose variation presented in an offset curve with a huge difference that comes to a maximum of 40%. Relatively speaking, the difference between experts' evaluations in centralization and decentralization strategy doesn't have a big impact on the variation of supplier's loss function in both structures which shows the effectiveness of our method.

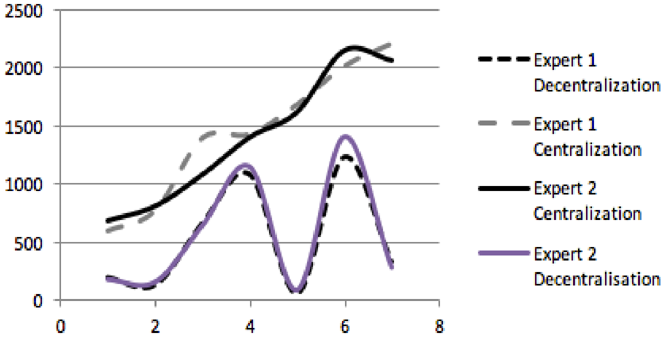


Figure 11. Variation of loss function in the case of medium impact of price on quality in both strategies

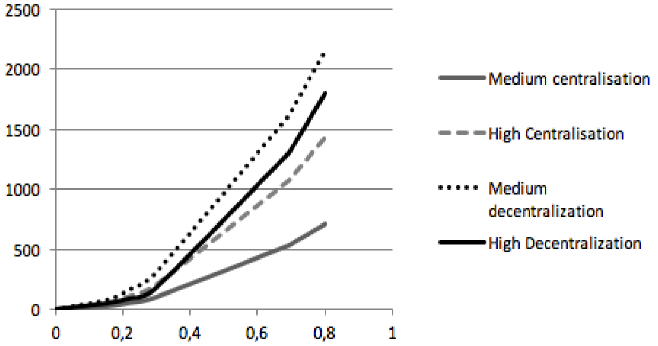


Figure 12. Variation of loss function in the case of first expert in both strategies

As shown in table 9, supplier's hidden quality costs are higher in the case of centralization, and this is due to:

- Centralization avoids sharing the responsibilities that can lead to coordination difficulties, and promotes decision-making consistent with the defined strategy.
- The centralized purchasing team may lack expertise on certain products, and their distance from user expectations may cause dissatisfaction.
- An error on a purchase will have multiple consequences, to the extent that the volumes purchased are more important, whether in terms of quality or delivery.
- Communication problems may occur between the various hospitals and the centralized purchasing site.
- Increased risk of dependence on suppliers as there is a risk of having too many purchases from few suppliers.

Table 9. Supplier's hidden quality costs in both organizational structures

	Centralization	Decentralization	Gain (C - D)/ C
S1	2889,2396	19226,7289	0,13190022
S2	22733,5557	27662,1728	0,03414751
S3	27777,4784	3131,49658	-0,1707576
S4	9767,49603	6411,64837	-0,0232507
S5	13098,7059	1401,6719	-0,0810419
S6	23908,7215	22279,2036	-0,0112899
S7	44157,9194	43749,56111	-0,00282927

7. Conclusion

The purchasing organizational structure is considerably changed according to many factors. However, this change doesn't take into account its impact on supplier's hidden quality costs. In this study, a deciding making approach is elaborated for the hospital sector. Regarding to this new approach, a new method of measuring supplier's hidden quality costs is developed. The idea is to use improved Taguchi quality loss function in order to measure the loss associated to supplier's bad quality and help the manager to take an appropriate decision regarding the purchasing structure. Based on this work, we conclude that an erroneous value of the loss coefficient can result in misallocation of the loss function value in centralization and decentralization strategy. Therefore, future research needs to address this coefficient, which represents the impact of price on quality, differs from a structure to another and which is considered as an important component of quality loss function.

References

[1] K. Jenoui, A. Abouabdellah, "Estimating supplier's hidden quality costs with Taguchi quality loss function and Topsis method". In 10th International Colloquium on Logistics and Supply Chain Management, Rabat, Morocco, 2017. https://doi.org/ 10.1109/LOGISTIQUA.2017.7962881

[2] A. Abouabdellah, A. Cherkaoui, "Decision Support System for Predicting the degree of a cancer patient's empowerment". Journal of Theoretical and Applied Information Technology, 60(3), 517-523, 2014.

[3] A. Marie, C. Giuliani, A. Abouabdellah, A. Cherkaoui, "The empowerment of patients factoring: Reporting to a holonic approach". In 9th Conference on service systems and service management, Troyes, France, 2006.

[4] D. Serrou, A. Abouabdellah, "Logistics in the hospital: Methodology for measuring performance". ARPN journal of engineering and applied sciences, 11(5), 250-256, 2016.

[5] D. Serrou, A. Abouabdellah, "Study grouping pharmacies impact on the performance of the hospital supply chain". In 6th International conference on industrial engineering and systems management, Seville, Spain, 2015.

[6] D. Serrou, A. Abouabdellah, "Study of Improved fiscal performance of hospital supply chain pharmacies by centralizing". In 45th International conference on computers & industrial engineering, Metz, France, 2015.

[7] WN. Pi, C. Low, Int J AdvManuf, "Supplier Evaluation and Selection Using Taguchi Loss Functions". The International Journal of Advanced Manufacturing Technology, 5 (1), 1123-1130, 2005. https://doi.org/10.1007/s00170-003-1975-5

[8] P. Trouiller, Technical support for improving the management and organization of hospital pharmacies, 2013.

[9] J. Dumoulin, Analyse du système d'approvisionnement pharmaceutique au Maroc : l'expérience de regroupement des achats depuis 2001, 2004.

[10] Saharidis, G.K.D., Kouikoglou, V.S., Dallery, Y. "Centralized and decentralized control polices for a two-stage stochastic supply chain with subcontracting". International Journal of Production Economics. 11(6), 117-126, 2009. https://doi.org/10.1016/j.ijpe.2008.10.001

[11] Chen, J.M, Chen, T.-H. "The multi-item replenishment problem in a two-echelon supply chain: the effect of centralization versus decentralization". Computers & Operations Research. 32(10), 3191-3207, 2005. https://doi.org/10.1016/j.cor.2004.05.007

[12] Chang, M.H, Harrington, J.E., Jr. "Centralization vs. Decentralization in a Multi-Unit Organization: A Computational Model of a Retail Chain as a Multi-Agent Adaptive System". Management Science. 46(10), 1427-1440, 2000. https://doi.org/10.1287/mnsc.46.11.1427.12085

[13] Behdani, B.Lukszo, Z.Adhitya, A.Srinivasan, "Decentralized vs. centralized management of abnormal situations in a multi-plant enterprise using an agent-based approach, in Computer Aided Chemical Engineering; 28(C), 1219-1224, 2010. https://doi.org/10.1016/S1570-7946(10)28204-4

[14] Hall, Ma, Tomkins, "A cost of quality analysis of a building project: Towards a complete methodology for design and build". Construction Management and Economics Journal. 19 (7), 727-740, 2001.

[15] A. Kazaz, M. Birgonul, "The evidence of poor quality in high rise and medium rise housing units: A case study of mass housing projects in Turkey". Building and environment journal. 40 (11), 1548-1556, 2005. https://doi.org/10.1016/j.buildenv.2004.11.023

[16] P. Love, "Auditing the indirect consequences of rework in construction: A case based approach". Management and audit journal. 17 (3), 138-146, 2002. https://doi.org/10.1108/02686900210419921

[17] P. Love, H. Li, "Knocking down walls and building bridges overcoming the problems associated with quality certification". Construction management and economics journal. 18(3), 321–332, 1999.

[18] P. Love, D. Edwards, "Determinants of rework in building construction projects". Engineering Construction & Architectural Management. 11(4), 259–274, 2004.

[19] R. Kethley, T. A. Waller, "Improving Customer Service in the Real Estate Industry: A Property Selection Model Using Taguchi Loss Functions". Total Quality Management. 13(6), 739-748, 2002.

[20] C. Quigley, C. McNamara, "Evaluating Product Quality: An Application of the Taguchi Quality Loss Concept". International Journal of Purchasing Materials. 28(3), 619-628, 1992.

[21] R. Kethley, T. A. Waller, "Improving Customer Service in the Real Estate Industry: A Property Selection Model Using Taguchi Loss Functions". Total Quality Management. 13(6), 739-748, 2002. https://doi.org/10.1080/0954412022000010109.

[22] H. C. Li, "Quality Loss Functions for the Measurement of Service Quality". The International Journal of Advanced Manufacturing Technology, 2(5), 29-37, 2003. https://doi.org/ 10.1007/s001700300004.

[23] R. Kethley, B. Waller, T. Festervand, "Improving customer service in the real estate industry: a property selection model using Taguchi loss functions". Total Quality Management, 13(6), 739–748, 2002. https://doi.org/ 10.1080/0954412022000010109

[24] M. W. Kim, W. Liao, "Estimating hidden quality costs with quality loss functions". Accounting Horizons. 8(1), 8–18, 1994.

[25] P. Kotler, "Strategies for marketing service excellence". The 1st Global

Conf, Management Centre, Europe, Brussels, 1991.

[26] S. K. Krishnan, “Increasing the visibility of hidden failure cost”. Measuring Business Excellence, 10(4), 77–101, 2006.

[27] E. Timmerman, “An approach to vendor performance evaluation”. Journal of Purchasing and Supply Management, 1(7), 27-32, 1986.

[28] R. E. Gregory, “Source selection: a matrix approach”. Journal of Purchasing and Materials Management. 22(2), 24–29, 1986.

[29] W. R. Soukup, “Supplier selection strategies”. Journal of Purchasing and Materials Management. 28(2), 7-12, 1987.

[30] K. Thompson, “Vendor prole analysis”. Journal of Purchasing and Materials Management. 26 (1), 11-18, 1990.

[31] G. Barbarosoglu, T. Yazgac, “An application of the analytic hierarchy process to the supplier selection problem”. Production and Inventory Management Journal, 38(1), 14–21, 1997.

[32] R. Nydick, R. P. Hill, “Using the Analytic Hierarchy Process to structure the supplier selection procedure”. International Journal of Purchasing and Materials Management, 25(2), 31–36, 1992.

[33] J. Sarkis, S. Talluri, “A model for strategic supplier selection”. In 9th International IPSERA Conference, London, UK, 2000.

[34] S. S. Chaudhry, F. G. Forst, J. L. Zydiak, “Vendor selection with price breaks”. European Journal of Operational Research, 16(6), 52–66, 1993.

[35] E. C. Rosenthal, J. L. Zydiak, S. S. Chaudhry, “Vendor selection with bundling”. Decision Sciences, 26 (1), 35- 48, 1995.

[36] F. P. Bua, W. M. Jackson, “A goal programming model for purchase planning”. Journal of Purchasing and Materials Management, 19(3), 27–34, 1983.

[37] C. A.Weber, L. M. Ellram, “Supplier selection using multi objective programming: a decision support system approach”. International Journal of Physical Distribution and Logistics Management. 23(2), 3-14, 1992.

[38] K. Jenoui, A. Abouabdellah, "System of multisourcing supplier's selection and evaluation in the hospital sector integrating the criteria: Total cost, Gap time, Risk performance". ARPN Journal of Engineering and Applied Sciences. 11(17), 10433-37, 2015.

[39] K. Jenoui, A. Abouabdellah, "Implementation of a decision support system heuristic for selecting suppliers in the hospital sector". In 6th International conference on industrial engineering and systems management, Seville, Spain, 2015.

[40] K. Jenoui, A. Abouabdellah, " Single or multiple sourcing strategy: a mathematical model for decision making in the hospital sector". In 11th International conference on intelligent systems: theories and applications, Mohamedia, Morocco, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 206-210 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

The impact of Big Data on the Android Mobile Platform for Natural Disaster Situations

Zijadin Krasniqi[*,1], Adriana Gjonaj[2]

[1]*Department of Information Technology, Tax Administration of Kosovo.*

[2]*Department of Informatics, Mathematics, and Statistics, European University of Tirana, Albania.*

ARTICLE INFO

Article history:
Received: 10 January ,2018
Accepted: 06 March ,2018
Online: 31 March ,2018

Keywords:
Android
OCEMA
Big Data
DWH
Natural Disasters

ABSTRACT

This application developing for the project OCEMA comes as result of architecture building in Android, then the development of a professional modeling in Talend Open Studio for Big Data, which enabled the integration of data from many data sources. One of its uses is the quick identification of people found in areas affected by natural disasters. The application identifies the persons who do not have an ID card, or another identification document, by using identification through fingerprint or personal number. OCEMA Application has access to all the agencies involved in natural disasters managing, such as ISK, HMIK, MIK, FSK and CRA. This application is developed to connect to web applications as well, such as applications, which gather real time information on earthquakes and weather in the world.

The study of literature and the actual work with these systems has shown some important components of success for DWH systems, DataMart and mobile application development.

1. Introduction

In the year 2002 alone, disasters affected a staggering 680 million people worldwide. Natural disasters include bring earthquakes, flooding, droughts, heat waves, cold spells and events that destroy countries, businesses and individual property and well-being.
The goal of this study is to investigate the adoption of cell phones as preparedness efforts during the natural disaster seasons in a developing country. Since Kosovo is a country vulnerable to natural disasters, the present study aims to investigate the adoption of cell phones for natural disaster preparedness [1]

OCEMA applications initially are secured by a username and password, and then divided in 5 buttons, such as: ISK, CRA, HMIK, MIK and FSK. This application enables access to real time information, such as in the case of the agency of civil register, offering two possibilities to identify the victims of natural disasters.

It uses the identification through personal number, but in cases where there is no ID card, the identification can be done using the fingerprints. Part of this application is access to institutions such as the Institution of Seismology, the Medical Institution, the Hydro meteorological Institution and the Firefighter’s Service. These institutions have updated information. Beside this, the OCEMA application has the possibility to access important web pages in order to monitor climate and seismic conditions in world level.

2. Mobile Platforms

The four most common smart phone operating systems, by market share in the third quarter of 2013 (IDC 2013) are:
•Android by Google, with a distinct dominance at 81 %,
•iOS, on Apple's iPhone, at 12.9 %,
•Windows Phone by Microsoft at 3.6 %
•and BlackBerry at 1.7 %.

There are many significant differences between these platforms. Some are visible to the users such as availability of specific features or design of the user interface (UI), and some are invisible to the users but affect the developers of mobile applications.

2.1. Android architecture

Starting from the bottom we have Linux Kernel, Android is built up on the Linux Kernel. Linux is already being used extensively from so many years and its kernel had received so many security patches. Linux Kernel provides basic system functionality like

[*]Zijadin Krasniqi, Tax Administration of Kosovo, zijadinkrasniqi@hotmail.com

https://dx.doi.org/10.25046/aj030224

process management, memory management, device management like camera, keypad, display etc.

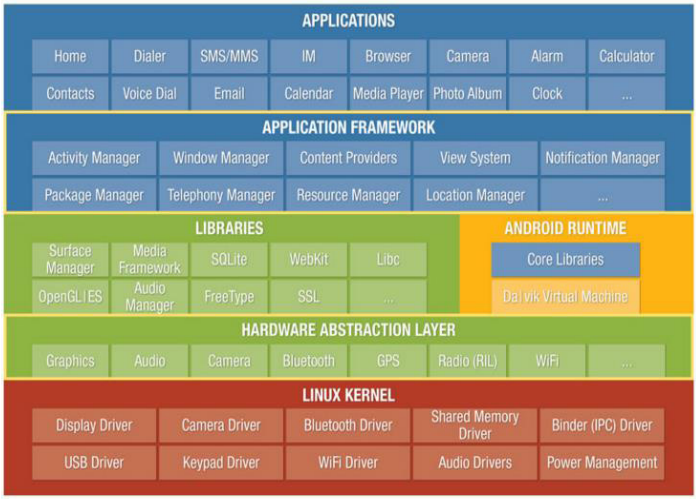


Figure 1: Android Architecture [2]

As the base for a mobile computing environment, the Linux kernel provides Android with several key security features, including:

- A user-based permissions model
- Process isolation
- Extensible mechanism for secure IPC
- The ability to remove unnecessary and potentially insecure parts of the kernel.

As a multiuser operating system, a fundamental security objective of the Linux kernel is to isolate user resources from one another. The Linux security philosophy is to protect user resources from one another [2]

Taking into consideration the Android as a target platform for application comes as the result of reviewing various works using Android technology.

The Android SDK allows application development with great ease. There are many inbuilt features and tools in an Android device, which can be integrated and programmed to be used as and when required from within the application [3]

3. Case study: SQLite_OCEMA database integration can be implemented in Android mobile devices

The operational Centre of the Emergency Management Agency requires analysis in daily, weekly and monthly bases. What is most important is the capability of real time data analysis. Therefore, this increases the need for data collection and data integration in a single database. The scope is to be able to analyze possible problems faced during emergency situations and natural hazards and then determine the highest levels of entities from which we can collect data.

The following data form the bases of the conceptual model of OCEMA database, which we are going to build.

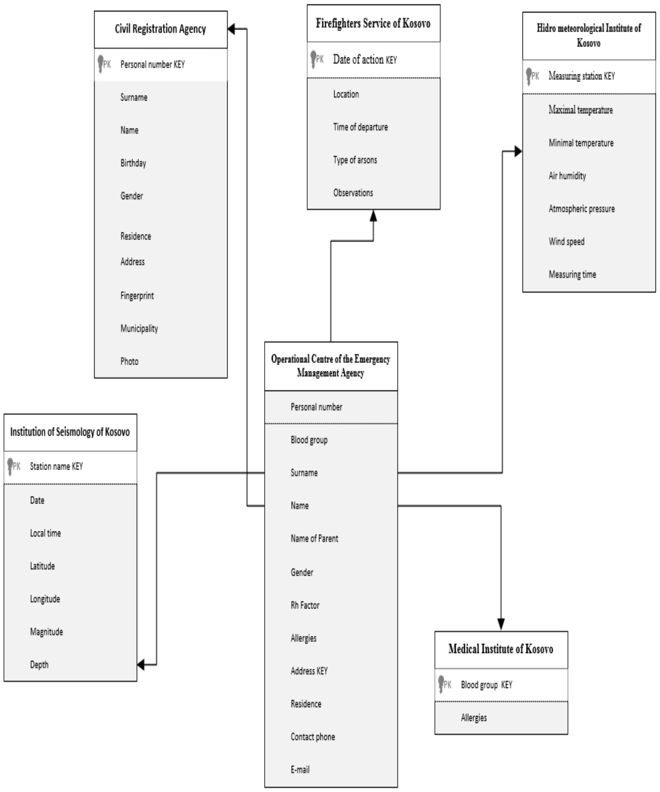


Figure 2: The Entity Relationship diagram of OCEMA database

We start with the description of the entities used in the OCEMA database.

The entity named Medical Institute of Kosovo is composed of the following data: Blood group for the patients and the eventual allergies present in the population. Entry data from MIK are an important information source at national level as far as the health of the population is concerned.

The entity named Civil Registry Agency includes several fields involving records on entry data regarding the civil registration at a national level, among them including the following characteristics: personal number, surname, name, date of birth, residence, address, fingerprints, district and photo.

All these attributes of the CRA entity include the population's records. In our study, the principal searching attributes on individuals will be the personal number and the fingerprints.
The entity named Institute of Seismology of Kosovo includes the following attributes: name of the weather station from where we get the real time data on seismic waves and information from the instruments of online monitoring of seismic waves.

Then in this entity we include: the date of seismic waves, local time, latitude, longitude, magnitude and seismic depth.
The data entity generated by the Hydro meteorological Institute of Kosovo includes a very significant dynamics of available data aiming the comprehensive real time monitoring in order to prevent flooding from rainfall, snow and also extreme draught caused by extreme maximal temperatures.

This entity includes the following attributes: the weather measuring station, maximal temperature, minimal temperature, air humidity, atmospheric pressure, wind speed, measuring time. It is very important to keep under control the continuous monitoring and observation of many factors, which affect arsons, such as the case of arsons involving homes, forests and pastures.

The entity of Firefighters Service is composed of the following attributes: date of action, location, time of departure, type of arsons, observations. All of these offer a very good opportunity to get information from this very important institution.

4. Data integration in SQLite_OCEMA through Talend Open Studio for Big Data

We have chosen SQLite as the main database, where we will integrate the tables from Oracle SQL, such as: CRA, HMIK, MIK, ISK and FSK. Most of the enterprise companies use Oracle in their applications in order to load data. In the case of data loading in SQLite_OCEMA, we have built the following job in Talend Open Studio for Big Data, as shown in the Figure below:

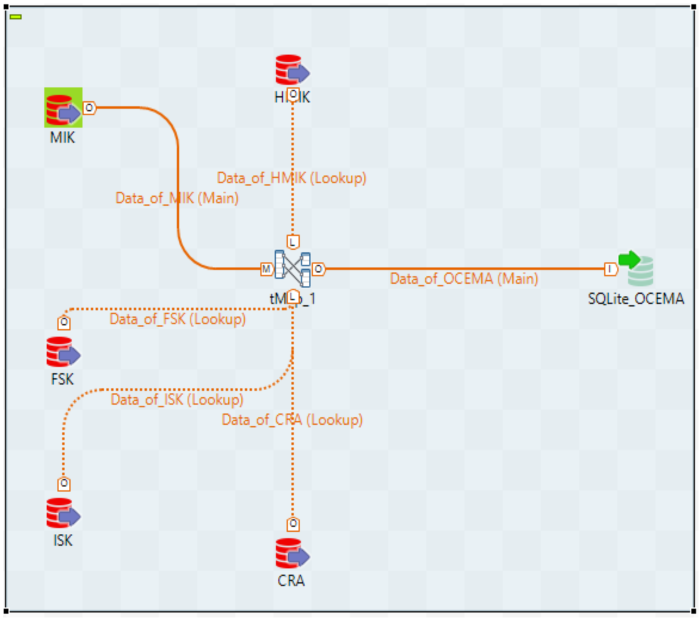


Figure 3: Talend's Job for data loading from the ORACLE SQL tables in SQLite_OCEMA

These heterogeneous data sources, HMIK, MIK, FSK, ISK and CRA, then will be stored in SQLite_OCEMA databases, as shown in the TOS schema.

An important step in ensuring the job developing with ETL is the successful connection between TOS and SQLite_OCEMA database.

4.1. Projecting the SQLite _OCEMA database

Unlike most RDBMS products, SQLite does not have client/server architecture. Most large-scale database systems have a large server package that makes up the database engine.

The database server often consists of multiple processes that work in concert to manage client connection, I/O files, caches, query optimization and query processing. A database instance typically consists of a large number of files organized into one or more directory trees on the server's file system.

In order to access the database, all the files must be present and correct. This can make it somewhat difficult to move or reliably back up a database instance. To access the database, client software libraries are typically provided by the database vendor. These libraries must be integrated into any client application that wishes to access the database server.

These client libraries provide APIs to find and connect to the database server, as well as set up and execute database queries and commands. The Figure 4 below shows how everything fits together in a typical client/server RDBMS.

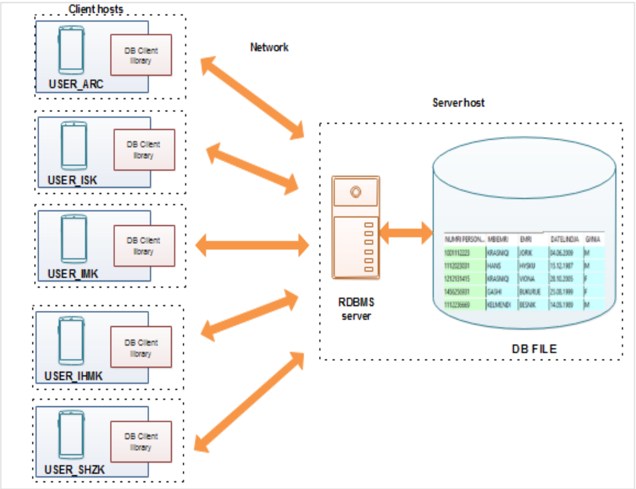


Figure 4: Client/server RDBMS architecture [4]

In contrast, SQLite does not have a separate server. The entire database engine is integrated into whatever application needs to access a database.

The only shared resource among applications is the single database file as it sits on disk. If we need to move or back up the database, we can simply copy the file. Below, in the Figure 5, we will see the SQLite architecture.

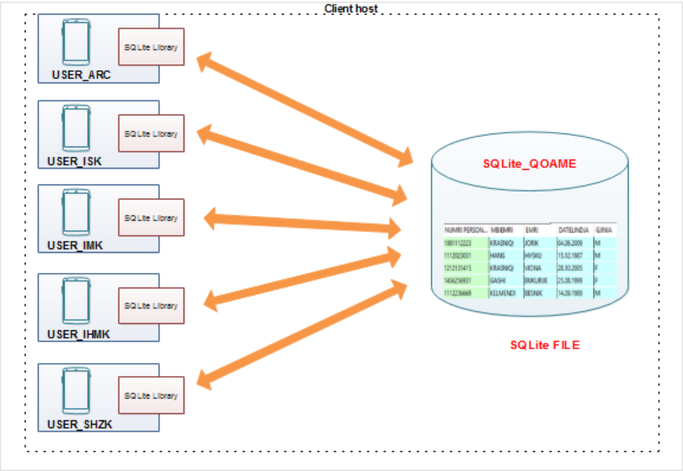


Figure 5: The SQLite server architecture [4]

By eliminating the server, a significant amount of complexity is removed. This simplifies the software components and nearly eliminates the need for advanced operating support. Unlike a traditional RDBMS server that requires advanced multitasking and high performance inter-process communication, SQLite requires little more than the ability to read and write to some type of storage.SQLite is designed to be integrated directly into an executable. This eliminated the need for an external library and simplifies distribution and installation. Removing external dependencies also removes most versioning issues.

If the SQLite code is built right into the application, we never need to worry about linking to the correct version of a client library or that the client library is version-compatible with the database server. Eliminating the server imposes some restrictions. SQLite is designed to address localize storage needs, such as a web server accessing the local database. This means it isn't well suited for situations where multiple client machines need to access a centralized database. That situation is more representative of client/server architecture and is better serviced by a database system that uses the same architecture. [4]

4.2. SQLite_OCEMA database-the case of android application for CRA

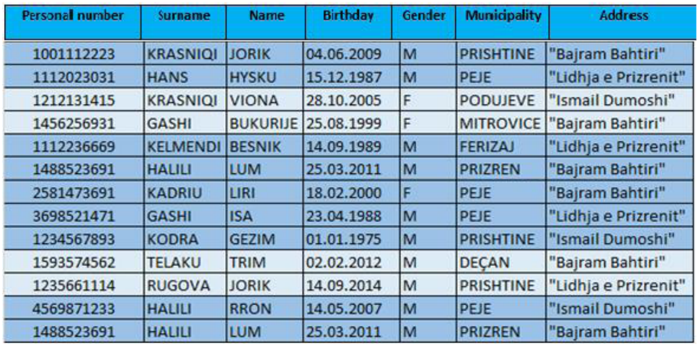
Android uses the SQLite database system, which is an open-source, stand-alone SQL database, widely used by many popular applications.

SQLite is a lightweight transactional database engine that occupies a small amount of disk storage and memory, thus, it is a perfect choice for creating databases on many mobile operating systems such as Android and iOS [2]

The Android platform offers full support for SQLite databases. The created database will be accessible by name from each of the classes inside the application, not those from outside. In our case, when we are searching the data from the Civil Registration Agency in case of emergencies and natural disasters, we have the possibility to search for missing people during lifesaving operations and also in the warning period of natural hazards.

Thus, we use the personal number identification method to find people struck there where the natural disaster happened. If we take the Civil Registration Agency table of the SQLite_OCEMA database as in the following Table 1:

Table 1: Civil Registration Agency data table in the SQLite_OCEMA database.



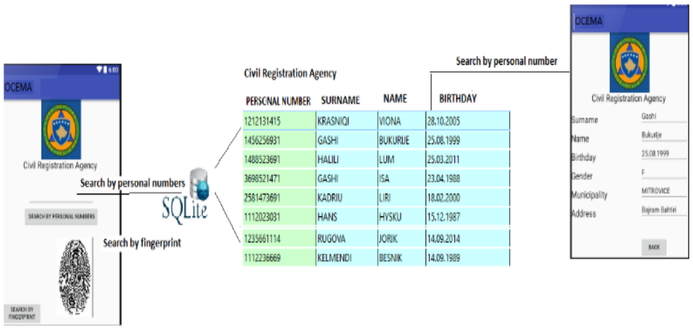
Personal number	Surname	Name	Birthday	Gender	Municipality	Address
1001112223	KRASNIQI	JORIK	04.06.2009	M	PRISHTINE	"Bajram Bahtiri"
1112023031	HANS	HYSKU	15.12.1987	M	PEJE	"Lidhja e Prizrenit"
1212131415	KRASNIQI	VIONA	28.10.2005	F	PODUJEVE	"Ismail Dumoshi"
1456256931	GASHI	BUKURIJE	25.08.1999	F	MITROVICE	"Bajram Bahtiri"
1112236669	KELMENDI	BESNIK	14.09.1989	M	FERIZAJ	"Lidhja e Prizrenit"
1488523691	HALILI	LUM	25.03.2011	M	PRIZREN	"Bajram Bahtiri"
2581473691	KADRIU	LIRI	18.02.2000	F	PEJE	"Bajram Bahtiri"
3698521471	GASHI	ISA	23.04.1988	M	PEJE	"Lidhja e Prizrenit"
1234567893	KODRA	GEZIM	01.01.1975	M	PRISHTINE	"Ismail Dumoshi"
1593574562	TELAKU	TRIM	02.02.2012	M	DEÇAN	"Bajram Bahtiri"
1235661114	RUGOVA	JORIK	14.09.2014	M	PRISHTINE	"Lidhja e Prizrenit"
4569871233	HALILI	RRON	14.05.2007	M	PEJE	"Ismail Dumoshi"
1488523691	HALILI	LUM	25.03.2011	M	PRIZREN	"Bajram Bahtiri"

5. Searching the SQLite_OCEMA data using the personal number

If we set as criteria the searching of data using the personal number method from the SQLite _OCEMA database and we use the recommended methods to create a new SQLite database.

When a database has been successfully opened, the SQLite Open Helper will cache it, so you can (and should) use these methods each time you query or perform a transaction on the database, rather than caching the open database within your application.

A call to getWritableDatabase can fail due to disk space or permission issues, so it's good practice to fall back to getReadableDatabase method for database queries if necessary. In most cases this method will provide the same, cached writeable database instance as getWritableDatabase [5]. An example is shown in the Figure 6.

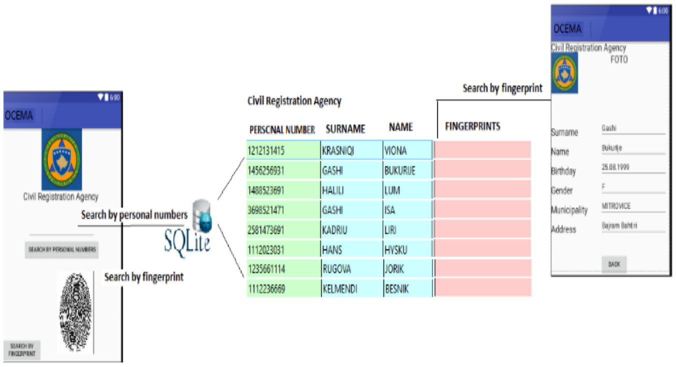

OCEMA
Civil Registration Agency
Search by personal numbers
SEARCH BY PERSONAL NUMBERS
Search by fingerprint
SEARCH BY FINGERPRINT
SQLite
Civil Registration Agency
Search by personal number
OCEMA
Civil Registration Agency
BACK

PERSCNAL NUMBER	SURNAME	NAME	BIRTHDAY
1212131415	KRASNIQI	VIONA	28.10.2005
1456256931	GASHI	BUKURIJE	25.08.1999
1488523691	HALILI	LUM	25.03.2011
3698521471	GASHI	ISA	23.04.1988
2581473691	KADRIU	LIRI	18.02.2000
1112023031	HANS	HYSKU	15.12.1987
1235661114	RUGOVA	JORIK	14.09.2014
1112236669	KELMENDI	BESNIK	14.09.1989

Surname	Gashi
Name	Bukurije
Birthday	25.08.1999
Gender	F
Municipality	MITROVICE
Address	Bajram Bahtiri

Figure 6: Searching by personal number in SQLite_OCEMA database [6]

This is a part of the application development. In this part, we search for people struck by natural disasters and emergency situations by using the fingerprint method, as shown in the Figure.


OCEMA
Civil Registration Agency
Search by personal numbers
SEARCH BY PERSONAL NUMBERS
Search by fingerprint
SEARCH BY FINGERPRINT
SQLite
Civil Registration Agency
Search by fingerprint
OCEMA
Civil Registration Agency
FOTO
BACK

PERSCNAL NUMBER	SURNAME	NAME	FINGERPRINTS
1212131415	KRASNIQI	VIONA	
1456256931	GASHI	BUKURIJE	
1488523691	HALILI	LUM	
3698521471	GASHI	ISA	
2581473691	KADRIU	LIRI	
1112023031	HANS	HYSKU	
1235661114	RUGOVA	JORIK	
1112236669	KELMENDI	BESNIK	

Surname	Gashi
Name	Bukurije
Birthday	25.08.1999
Gender	F
Municipality	MITROVICE
Address	Bajram Bahtiri

Figure 7: Searching by using the fingerprints in SQLite_OCEMA database.[6]

This result is displayed every time the user launches the matching activity from our proposed application, by comparing the initial fingerprint with the ones stored in database.

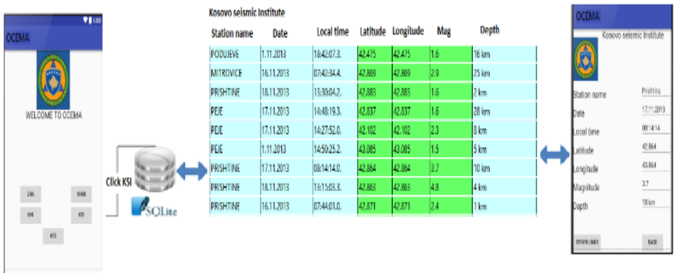
One utility of the proposed application consists in quick identification of the people struck in areas hit by natural hazards or emergency operations, in the quick identification of those people who do not have an ID card or another document of identification, by using the fingerprints identification method.

Also, the application could be deployed in the access systems of the public institutions like schools, libraries, hospitals and so on[7]

6. SQLite_OCEMA database- the case of android application for ISK

The Institution of Seismology has a large number of weather measuring stations across the territory of Kosovo. These stations transmit their data to the institute on daily, weekly and monthly bases. Upon creating and building the SQLite_OCEMA database, these data can be reported in a shorter time, better to say in real time across all the weather measuring stations.

By clicking on the ISK (Institute of Seismology of Kosovo) button in the OCEMA basic application, we gain access to search the latest updated ISK data, as shown below in the Figure 8.



Kosovo seismic Institute

Station name	Date	Local time	Latitude	Longitude	Mag	Depth
PODUJEVE	1.11.2013	18:42:07.3.	42.475	42.475	1.6	16 km
MITROVICE	16.11.2013	07:42:34.4.	42.869	42.869	2.9	25 km
PRISHTINE	18.11.2013	13:30:04.2.	42.883	42.883	1.6	2 km
PEJE	17.11.2013	14:48:19.3.	42.837	42.837	1.6	28 km
PEJE	17.11.2013	14:27:52.0.	42.102	42.102	2.3	8 km
PEJE	1.11.2013	14:50:25.2.	43.085	43.085	1.5	5 km
PRISHTINE	17.11.2013	08:14:14.0.	42.864	42.864	3.7	10 km
PRISHTINE	18.11.2013	13:15:03.3.	42.863	42.863	4.8	4 km
PRISHTINE	16.11.2013	07:44:01.0.	42.871	42.871	2.4	1 km

OCEMA

Kosovo seismic Institute

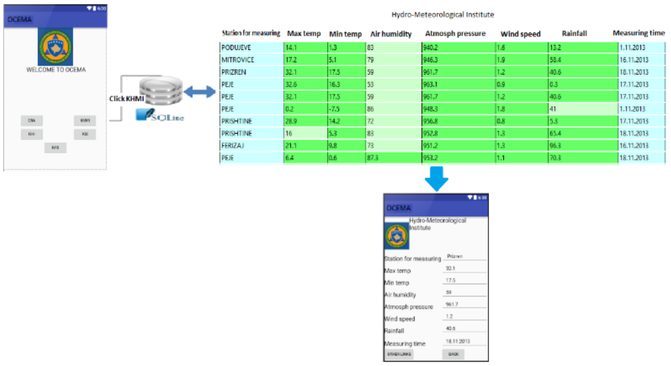
Station name Prishtina
Date 17.11.2013
Local time 08:14:14
Latitude 42.864
Longitude 43.864
Magnitude 3.7
Depth 10 km

Figure 8: Access to the data of Institute of Seismology of Kosovo. [6]

After the creation of the Institution of Seismology of Kosovo table in the SQLite_OCEMA database, we can click on the ISK button to search the latest updated data of this institution, including magnitude of seismic waves, seismic depth, geographical position, latitude-longitude, date and time of the seismic wave.

7. SQLite_OCEMA database, the case of Android application for HMIK

The Hydro meteorological Institute of Kosovo has a large number of weather measuring stations across the territory of Kosovo. These stations transmit their data to the institute on daily, weekly and monthly bases. Upon creating and building the SQLite_OCEMA database, these data can be reported in a shorter time, better to say in real time across all the stations, including data from the river flows, data from sensors placed in locations where there is more snow.



Hydro-Meteorological Institute

Station for measuring	Max temp	Min temp	Air humidity	Atmosph pressure	Wind speed	Rainfall	Measuring time
PODUJEVE	14.1	1.3	83	940.2	1.6	13.2	1.11.2013
MITROVICE	17.2	5.1	79	946.3	1.9	58.4	16.11.2013
PRIZREN	32.1	17.5	59	961.7	1.2	40.6	18.11.2013
PEJE	32.6	16.3	53	963.1	0.9	0.3	17.11.2013
PEJE	32.1	17.5	59	961.7	1.2	40.6	17.11.2013
PEJE	0.2	-7.5	86	948.3	1.8	41	1.11.2013
PRISHTINE	28.9	14.2	72	956.8	0.8	5.3	17.11.2013
PRISHTINE	16	5.3	83	952.8	1.3	65.4	18.11.2013
FERIZAJ	21.1	9.8	73	951.2	1.3	96.3	16.11.2013
PEJE	6.4	0.6	87.3	953.2	1.1	70.3	18.11.2013

OCEMA

Hydro-Meteorological Institute

Station for measuring Prizren
Max temp 32.1
Min temp 17.5
Air humidity 59
Atmosph pressure 961.7
Wind speed 1.2
Rainfall 40.6
Measuring time 18.11.2013

Figure 9: Data from SQLite database for the HMIK case [6]

Thanks to this application we have been able to get detailed information from all the weather measuring stations in the territory of Kosovo, which provide a representation of warnings in case of flooding from atmospheric precipitations, wind speed, air humidity and the maximum and minimum temperatures.

8. Conclusion

Mobile technology has shown a considerable impact on society over the past few years. Gradually, this technology has become part of every person's life and is the easiest mode to deliver information. The development of application OCEMA for the doctoral project comes as a result of building architecture in Android Platform, then a professional model in Talend Open Studio for Big Data, which enabled the data integration from multiple data sources. Then we have developed the applications in Android Platform using the Java programming language. We have made it possible to reach near real time representation of the sources.

It is widely known that during natural disasters and emergencies, we do not have all the information available, but in those moments, everything is important, above all the information related to the human factor. EMA should be aware of these kinds of situations and precede the capacity building in systems communication and developing applications for the management of emergencies and natural disasters.

References

[1] A. Lawson, L.Willoughby, "Adoption of cell phones as preparedness efforts for natural disasters", Issues in Information Systems, International Association for Computer Information Systems, USA, 2012, Vol. 13, pp 11-20 http://iacis.org/iis/2012/79_iis_2012_11-20.pdf

[2] A.Agrawal "Android Application Security Part 2-Understanding Android Operating System", 2015, http://manifestsecurity.com/android-application-security-part-2

[3] T. Tejassvi, "An Android application to support flash flood disaster response management in India" M.sc Thesis, University of Twente -Netherlands, 2014.

[4] J. Kreibich, Using SQLite, Published by O'Reilly Media, 2010, USA.

[5] R. Meier, Professional Android™ 4 Application Development, John Wiley & Sons, Inc., 2012, Canada.

[6] EMERGENCY MANAGEMENT AGENCY-LOGO (2015) https://ame.rks-gov.net/en/Home/PgrID/494/PageID/4.

[7] O.Dospinescu,I. Lîsîi ,"The Recognition of Fingerprints on Mobile Applications – an Android Case Study" , IBIMA Publishing , Romania,2016 http://ibimapublishing.com/journals/JEERBE/2016/813264/813264.html

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 220-230 (2018)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Fuel Cell/ Super-capacitor power management system assessment and Lifetime Cost study in a 500kVA UPS

Imen Ben Amira*, Abdessattar Guermazi

University of Sfax, Electrical Department, Sfax Engineering National School, P.O. Box 1173, 3038 Sfax, Tunisia

ARTICLE INFO

Article history:
Received: 26 January,2018
Accepted: 03 March ,2018
Online: 31 March ,2018

Keywords:
Fuel cell
Super-capacitor
Lifetime
Cost
Degradation
Uninterruptible Power Supply.

ABSTRACT

A 500 KVA Uninterruptible power supply (UPS) using Fuel Cells (FC) and super-capacitors (SCs) was studied with the worst case of 10 minutes and eight hours of interruption per day. A power management system was established to control the FC and the SCs in order to extract the hybridization benefits with a comparison between a Proton exchange membrane FC (PEMFC) working alone and another combined with SCs. Moreover, possible FC degradations were discussed. The start/stop cycling, the high-power loads and load changes degradations were taken into consideration in order to estimate the FC lifetime span using a prediction formula. Besides, the FC costs were studied to estimate the best average cost. Finally, the SCs filter constant time and their charging currents were revealed.

ABREVIATION

UPS	Uninterruptible power supply
SC, SCs	Super-capacitor, super-capacitors
FC	Fuel cell
N_{s_sc}	Number of SCs series elements
N_{p_sc}	Number of parallel SCs pack branches
N_{s_fc}	Number of FC series elements
N_{p_fc}	Number of parallel FC branches
P_{sc}	Super-capacitors pack power
P_{fc}	Fuel cell power
P_{load}	Load power
U_{sc}	Super-capacitors pack voltage
U_{fc}	Fuel cell voltage
UM	Maximum SCs pack voltage
Um	Minimum SCs pack voltage
SoH	State of Health
i_{ch}	Super-capacitors charging current
PEMFC	Fuel cell with Proton Exchange Membrane
EoL	End of Life
Tau	Filter constant time
€	Euro
EIS	Electrochemical Impedance Spectroscopy

1. Introduction

UPS have reached mature levels in term of electrical power quality provided to sensitive loads despite power disturbances and outages. It is available on the market in a wide range depending on their ratings; from 300VA to provide back up for a computer to several megawatts for an entire building.

Most backup power systems technologies involve engine-generator sets and lead-acid battery either with SCs systems or without such systems. Although the engine-generator sets are reliable and last for long runtime, they depend on the availability of fuel, require excessive maintenance and cause a lot of pollution. Besides, batteries are unsuitable for long runtime; they last for limited life span, are very sensitive to temperature and have got

* Imen Ben Amira, Email: imenbenamiraenis@gmail.com

https://dx.doi.org/10.25046/aj030226

power fluctuations which lead to lifetime and capacity uncertainty but they involve low capital investment.

Further improvement FC technology will be a real boost for UPS for low and high-power applications. They are emerging as a valuable alternative to engine generator sets and lead-acid batteries in backup power systems. FC have high specific energy and reliability and present no pollution to the environment. PEMFC are the most common thanks to their relatively fast response and low operating temperature.

The FC lifetime can vary significantly from an application to another, ranging from 4.000 h approximately for intermittent operation to 40.000 h for stationary applications [1]. But the most common drawback of FC is their high cost which is an important criterion for any project.

The FC durability is defined by the end of life (EoL) concept related to the inability to guarantee a minimum performance or to ensure basic operations. Thus, the FC EoL can be defined by the observation of a failure or a certain loss in performances. In case of failure, the user needs a complex and unexpected maintenance in order to bring the system back to its normal operational conditions. Failure is the most undesired situation; it should be prevented in advance by detecting these events. On the other hand, the loss in performances leads to a degraded mode implying reduction of power capacity or overconsumption. It has been proved that the aging effect causes the cell degradation in performances, which means that an old stack needs a higher amount of hydrogen than a new one to provide the same power [2]. In fact, the FC loss in voltage is related to the stack power reduction and so its inability to accomplish some technical requirements.

FC need an auxiliary power source to overcome fast transients such as SCs. In fact, UPS systems depending on the FC combined with the SCs are not that extensively investigated. Thanks to hybridization more advanced UPS systems are expected. This will allow FC functionalities enhancing the global system life expectancy and minimizing stresses caused by load currents [3].

Today, SCs can provide the required energy in a short time despite their low energy density. For high voltage applications, SCs cells can be connected in series and deliver higher power of a similar-sized battery [4]. However, SCs voltage drops sharply while discharging.

Different combination topologies of FC/SCs have been studied. Among these combinations the parallel topology using two choppers seems to be the most efficient and flexible [5]. Fuel Cells and SCs can be directly connected to the DC bus without a converter which allows the DC bus voltage to be maintained around the desired value. Meanwhile, the high charging and discharging cycles cause high power losses of the bidirectional converter and reduces the power efficiency of the entire system [6]. Besides, the load power is filtered during the UPS autonomous operation by a low-pass filter. The SCs absorb high power demands and prevent the FC to face them. The FC supply the smoothed power and can recharge the SCs when power demand is low [5].

In [7], the author estimated the FC lifetime relying on the Electrochemical Impedance Spectroscopy (EIS) measurements which consequently could help predict the FC maintenance planning.

In [8], the author gave a mathematical equation of expected FC lifetime taking into account the on/off cycles, load changes, high power load, idling, and air pollution factor. They proved that the FC reached its end when its voltage decreased by 10% at a constant current.

In [9], the author uses a Stochastic Dynamic Programming in order to decrease the total cost of the FC and increase its lifetime by 14% while increasing the fuel consumption by only 3.5%. This programming reduced significantly the FC transient load.

In [10], the author used two energy management strategies by a dual-stack FC system in order to improve the fuel economy of a vehicle and make the driving range longer. They kept the FC's on/off frequency at a reasonable range.

In [11], the author studied an analysis of life-cycle cost of five trams and took into account the initial infrastructure cost, the tram' power, the mode of operation and the replacement cost. They proved that using a FC hybrid tram cost 16.6% to 19.1% lower than a tram using an electric network.

The objective of this paper was to extract the best amortized lifetime cost of the hybrid system, considering power dissipation and the FC and SCs cost based on the amount of energy supplied. The cost of the hybrid system for the UPS was compared to that of an independent FC. The FC/SCs hybridization in backup systems for UPS was considered using an optimal energy management control in order to improve the lifetime of the hybrid system and reduce its overall cost.

2. Presentation of the UPS

2.1. UPS specification

The studied UPS is a 500 kVA rated power which main specifications are:

- Backup time: T = 10 min;
- Nominal FC power: P_N = 480 kW;
- DC-bus voltage: V_{dc} = 400 V;
- Power factor: FP = 0.9;
- FC nominal voltage: Ufc_{nom} = 650 V.
- Nominal SCs pack voltage: U_{SC} = 300 V.
- Total efficiency: η = 95%;

The studied UPS is a system with an Online/Double-Conversion. It interferes in grid failures before the generators start up. The maintenance has to be applied each time the FC falls up; the study was limited to twenty cycles of failure. In order to estimate the cost, the first cycle of autonomy was considered. The worst case of autonomy has been chosen in this work with two long interruptions per day: 10 minutes of interruption and 8 hours one.

2.2. *Sizing and Modeling Supercpacitors*

The SCs must deliver all the power rating P_N = 480 kW in 10 s, with a delivered energy around 4.8 MJ. The simplified circuit is shown in figure 1.

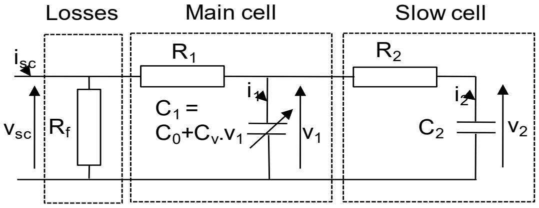


Figure 1: Super-capacitor simplified circuit.

The energy Esc stored in the SCs at the voltage U_{SC} is written as [5]:

$$E_{sc}=\frac{1}{2}C_{eq}U_{sc}^{2}=\frac{1}{2}\frac{N_{p\_sc}}{N_{s\_sc}}C_{sc}U_{sc}^{2} \qquad (1)$$

Where: C_{eq} is the equivalent capacity of the SCs, Np_{sc} are the parallel branches of the SCs Ns_{sc} are the series connections of SCs and Csc is the SC capacitance.

For maximum SCs voltage U_M = 300 V, we can recover 75% of the total energy initially stored if the voltage after discharge is in the range of U_m = 150 V. The efficiency coefficient k is equal to 0.9 in the worst case [12]. The energy delivered by the SCs is expressed as:

$$P_N\Delta t=k.\frac{1}{2}(C_{eq}U_{M}^{2}-\frac{1}{2}C_{eq}U_{m}^{2}) \qquad (2)$$

The calculation gives the equivalent capacitance C_{eq}=158 F.

The SCs studied in this paper are the Maxwell/BCAP3000 type, rated 3000 F, 2.7 V. The number of components in series $Ns_{sc}$is determined by the initial voltage of the pack in the charged state (300 V). We finally get Ns_{sc} = 112 and Np_{sc} = 6.

In order to model the SC, we refer to the simplified model with two branches that describes faithfully the electrical behavior of an SC element. This model was developed by Bonert and Zubieta [13]. It contains two parts in which the capacity of the main branch is nonlinear and varies depending on the voltage as shown in figure 1.

The main capacitance C_1 consists of a constant capacity C0 (F) and a parameter denoted by C_v (in F/V) and it is written as: $C_1 = C_0 + C_V * V_1$, where C_v is a constant and V_1 is the voltage across C_1. The equivalent circuit parameters are detailed in [5].

- R_1 = 0.360 mΩ;
- C_0= 2100 F;
- C_v = 623 F/V;
- R_2 = 1.92 Ω;
- C_2 = 172 F.

The SCs voltage is described as following:

$$U_{sc}=V_{sc}\times N_{s\text{-}sc}=N_{s\text{-}sc}\times(V_1+R_1*i_{sc}) \qquad (3)$$

$$U_{sc}=N_{s\text{-}sc}\times(V_1+R_1*\frac{I_{sc}}{N_{p\text{-}sc}}) \qquad (4)$$

Where U_{sc} and Isc are the super-capacitors bank voltage and current respectively, v_{sc} and i_{sc} are the elementary SC voltage and current respectively.

The voltage v_2 at the terminals of the capacity C_2 is given by:

$$V_2=\frac{1}{C_2}\int i_2dt=\frac{1}{C_2}\int\frac{1}{R_2}(V_1-V_2)dt \qquad (5)$$

The current i_1 is expressed as:

$$i_1=\frac{dQ_1}{dt}=C_1\frac{dV_1}{dt}=(C_0+C_vV_1)\frac{dV_1}{dt} \qquad (6)$$

Where:

$$Q_1=C_0V_1+\frac{1}{2}C_vV_1^{2} \qquad (7)$$

From this, we have:

$$V_1=\frac{-C_0+\sqrt{C_0^{2}+2C_vQ_1}}{C_v} \qquad (8)$$

Taking into account the variation of the SC capacitance as a voltage function, the energy E_{sc} of the SC is given by:

$$E_{sc}=N_{s\text{-}sc}*N_{p\text{-}sc}*(\frac{1}{2}C_0V_{sc}^{2}+\frac{1}{3}C_vV_{sc}^{3}) \qquad (9)$$

The SCs pack sizing meets the energy of 4.8 MJ specified initially. Furthermore, in [5] the simulated SCs circuit was validated by an experimental charge/discharge test at constant currents which showed good agreement with the developed model.

2.3. *Fuel cell Sizing and Modeling:*

The studied FC is a PEMFC simulated by MATLAB/SIMULINK using the datasheet of Nexa™ power module. The FC model should be rescaled to make it fit for UPS applications and suitable for the power demand. The Nexa™ FC stack provides1.2 kW of net output power and a voltage ranging from 43V to 26 V at full load.

The equations for calculating the variation of the voltage can be found in several works [14]. This model is described using a combination of basic laws and empirical models. The FC voltage is described as follows:

$$V_{cell} = E_{fc} - \Delta V_{fc} \qquad (10)$$

Where: E_{fc} is the open circuit voltage and ΔV_{fc} the sum of activation losses (due to the start-up electrochemical reactions at the cathode), ohmic losses (caused by the resistance imposed by the bipolar plates and the electrodes) and concentration losses (by the variation of the reactants concentrations). We can write E_{fc} according to [14] using the cell temperature T_{fc}, the partial pressure of hydrogen PH_2 and oxygen PO_2 as:

$$E_{fc} = 1.229 - 8.5*10^{-4}(T_{fc} - 298) + 4.308*10^{-5}*T_{fc}*Ln(P_{H2}*P_{O2}{}^{0.5}) \qquad (11)$$

ΔV_{fc}Can be defined as:

$$\Delta V_{fc} = A_{fc}Ln(\frac{i_{fc}}{j_0}) + r_{fc}j_{fc} + B_{fc}Ln(1 - \frac{i_{fc}}{j_{lim-fc}}) \qquad (12)$$

Where A_{fc} and j_0 are the losses activation parameter, $\mathrm{B_{fc}}$ is the modeling constant (V), $\mathrm{j_{fc}}$ is the current density of the stack (A/cm2), r_{fc} are the ohmic losses parameter and $j_{\lim -fc}$ is the maximum current density of the FC (A/cm2).

The total electric power of the stack is calculated by the following equation:

$$P_{fc} = N_{s-fc} \times V_{fc} \times i_{fc} \qquad (13)$$

Where V_{fc} and i_{fc} are the voltage and current of the FC and Ns_{fc} is the series fuel cells number.

The FC parameters are summarized in Table 1. The results were verified by comparing the operating conditions with the datasheet provided by the manufacturer and given in Figure 2.

There is an agreement with the datasheet polarization curve and the polarization curve of the FC potential despite minor differences due to the estimation of some parameters affecting the activation regions. This difference is tolerable considering the neglect of some physical processes like water flooding at the cathode and anode drying.

From this comparison, we can say that the developed model is accurate and can be used to simulate the FC performances.

The consumption model uses an optimization algorithm that transforms the electric power and FC auxiliaries into a hydrogen consumption rate. The instantaneous hydrogen consumption rate depends on i_{fc} and can be expressed as: [15]

Table 1: Fuel Cell parameters

A_{fc}	$120cm^2$	P_{H2}	2 atm
B_{fc}	0.027	P_{O2}	$1.9\ atm$
T_{fc}	60^0	j_{lim-fc}	$2A/cm^2$
N_{s_fc}	18	r_{fc}	$0.5\Omega cm^2$
N_{p_fc}	18	j_0	$2\,\mu A/cm^2$

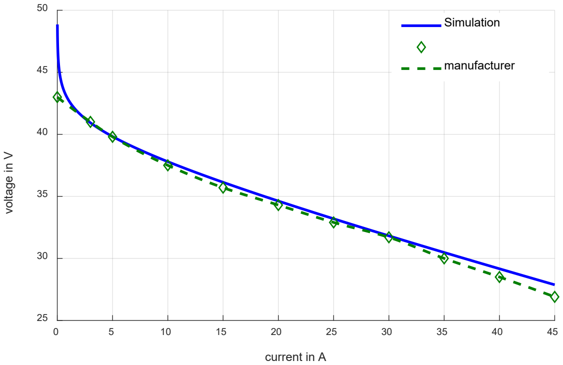


Figure 2: Polarization curves of the FC according to the proposed model and the datasheet

$$cons_{H2} = \frac{0.0337 * N_{s-fc} * i_{fc} + 0.0112}{60} \ \text{(g/s)} \qquad (14)$$

The lifecycle cost can be approximated since the power source degradation in a backup system can be predicted. This is a priority objective of many researchers who try to make it as minimum as possible. Indeed, the degradation is not caused perforce by operation conditions but mainly by aging.

The FC couldn't compete with the internal combustion engine with a price over 47€/kW at the transport domain [16, 17]. In fact the lifecycle cost depends on the FC cost, the fuel efficiency, the method of hydrogen production, the production capacity and the social cost [18]. Besides at stationary domain the FC cost is around 3000€/kW [19]. However, when the power needed increases, the FC cost decreases.

Few studies focused on the PEMFC lifetime and cost. If the researcher is seeking for prolonging the FC lifetime, he will always face cost problems. Thus to reduce the cost, the lifetime will be shortened. In fact the FC cost could be reduced by using small stacks at a higher power; however this will reduce the efficiency and increase the degradation and decrease the system lifetime [9].

The total FC cost (C_{tot}) takes into account the FC stack manufacture cost (C_{stack}), the hydrogen consumption cost (C_{H2}) and the maintenance cost (C_{maint}).

$$C_{tot} = C_{stack} + \int_0^{600} C_{H2} dt + C_{ma\,\mathrm{int}} \tag{15}$$

In another hand, In order to estimate the FC lifetime, one should determine the FC degradations and its EOL. In fact, the FC EOL indicates a loss of about 20% of the active surface area, and the voltage value drops by 10% compared to what it was for a total surface area (120 cm^2) [20, 8].

Lifetime can be expressed as:

$$T_f = \frac{\Delta_V}{r_d} \tag{16}$$

Where: r_d presents the FC performance decay rate and Δ_V is the decreased value of FC performance from the beginning of its lifetime to its end.

We can express r_d as follows:

$$r_d = n_1 V_1 + n_2 V_2 + t_1 V'_1 \tag{17}$$

Where n_1 is the average start-stop cycles per hour, V_1 is the start-stop cycle degradation value, n_2 is the average load change cycles per hour, V_2 is the load change cycles degradation value, t_1 is the average high power load operation time per hour, V'_1 is the high power load degradation value.

Generally, lifetime loss can be expressed as:

$$T_f = \frac{\Delta_V}{\sum n_i V_i + \sum t_j V'_j} \tag{18}$$

Where n_i is the operating cycles of condition i per hour, V_i is the average degradation value in condition i, t_j is the average operating time of condition j per hour and V_j is the average degradation value in condition j.

This formula can be used for any FC at any condition. But in our case with the Nexa module, zero output voltage degradation was witnessed under dynamic load test conditions according to the manufacturer. Therefore, we focused on on/off cycling and the high power load degradations.

3. Fuel cell/Super-capacitor combination

3.1. The UPS system design

The power sharing control is elaborated in order to benefit the SCs rapid charge and discharge ability and reduce the FC stress caused by instantaneous power load demands. The main goal of this combination is that the SCs support the power transients and to sleek the FC high-power demands.

The structure shown in figure 3 contains two ideal DC/DC converters which are supposed with no losses. These converters are two types two quadrants, reversible current for the SC and unidirectional for the FC to avoid current return, and have a bus voltage around 400 V.

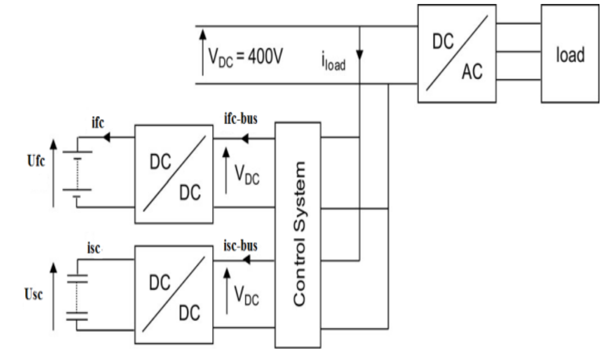


Figure 3: Topology of the controlled FC/SC combination with control system

The main goal of the management policy is to drive the FC only in its high efficiency domain, away from its high and low power regions. Thus, control strategies maintain the SOC of the SCs within a specified range. It mitigates the FC stress and maximizes its life span since it uses the energy stored in the SCs to absorb high power peaks.

The UPS block diagram respects the topology of an Online/Double-Conversion system. The backup autonomy considers a10 minutes of interruption and another one of eight 8 hours.

The hybrid control system generates two current reference signals ifc_{bus} for the FC and isc_{bus} for the SCs with the following constraints: FC current must be limited to a maximum value in order to match the reactant delivery rate and the usage rate and a minimum value. Besides, the SC pack must have a maximum value of 300V and a minimum value of 150V to indicate the charge/discharge cycle limits.

At the beginning, sudden power variations are diverted to the SCs thanks to a low-pass filter which is activated each load change. Then this filter is canceled and the FC supplies the full load power Pload. After that, at a certain limit of the load power P_{lim} , the FC supplies the load and the SCs with energy. When fully charged, the SCs pack can intervene in the same way in case of another power demand and the filter activated again. When the SCs reach 75% of discharge, their voltage drops until 150 V which is the minimum limit set. The SCs pack role is inhibited and the FC handles up the full power.

The equation of the FC power P_{fc} is given by:

$$P_{fc} = P_{load}[1 - \exp(-\frac{1}{T_{au}})] + P_{IV} \tag{19}$$

Where: P_{load} is the power load cycle, P_{IV} is the starting FC power and T_{au} is the adjustable constant of the low-pass filter.

The chart in figure 4 explains the power distribution principle between the FC and the SCs.

The power management is developed and gives the optimal sizing of the FC and SCs according to the critical load profile which corresponds to the typical load power with large power impulses as illustrated in figure 5. It is delivered by an UPS relying on batteries in information technology (IT).

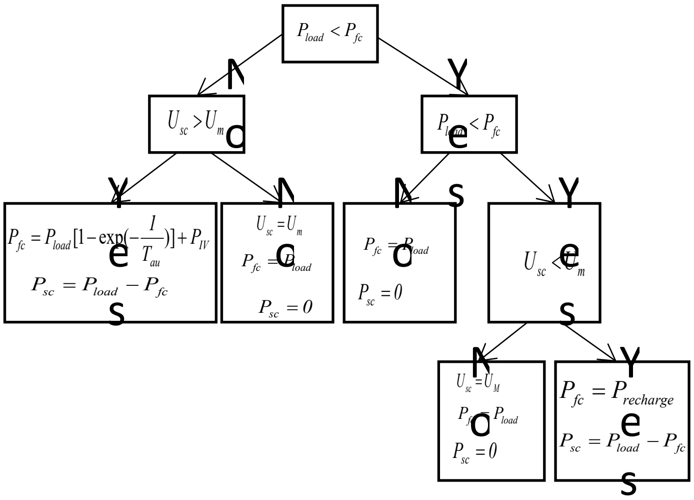


Figure 4: FC/SCs combination system chart

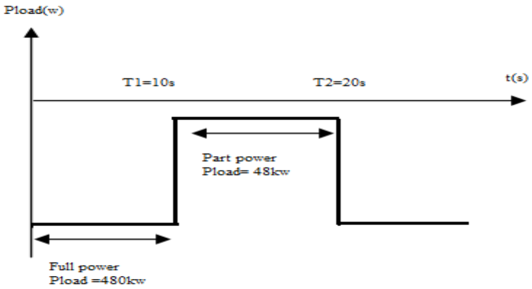


Figure 5: Load cycle profile

The energy management system has to be appropriate in order to obtain the necessary size of the FC and the SCs. The two major parameters of the power management system are the SCs current charging (ich) and the filter time (Tau).

The SCs current charge i_{ch} was varied (100A, 200A et 400A) for Tau=2 and the load charge variation of figure 5 in order to extract the SCs power waveforms as shown in figure 6. This test emphasized on the SCs rest time ($P_{sc} = 0$) which decreased while increasing i_{ch}, thus the FC power varied too (since P_{sc} and P_{fc} are complementary) and so is its lifetime and the hydrogen consumption. For i_{ch} = 400A the SCs are charged rapidly via the FC in order to interfere in case another load variation occurs. However, at low charge currents the SCs are not totally charged to face the next power charge variation. Consequently the optimum i_{ch} is 400A.

The second test aims to vary the filter constant Tau (1, 2, 3 et 10) at 400A of i_{ch} as indicted in figures 7 and 8 which present the SCs (P_{sc}) power and the FC power (P_{fc}) variations simultaneously. In fact, the waveforms of theses storage elements vary for each Tau value. The filter dynamic could also affect the FC and SCs lifetime; the optimal solution seems to opt for a Tau = 2s.

3.2. Simulation results:

The combination of the FC and SCs has to last for the whole backup time needed by our UPS. For the rest of simulation the ich=400A and Tau=2 are adopted.

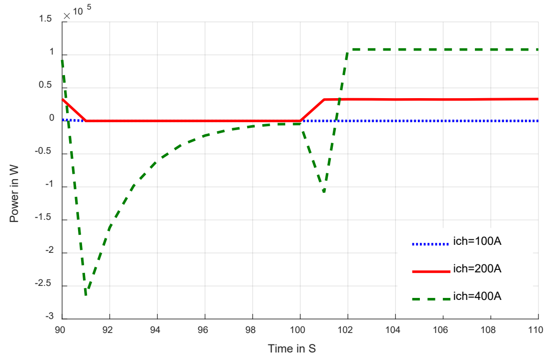


Figure 6: PSC for different values of Tau and current charge of the SCs

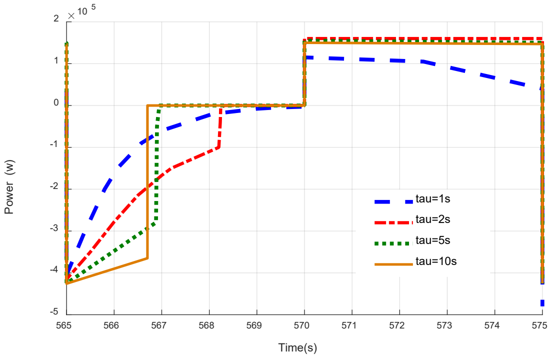


Figure 7: SCs power evolution for different values of Tau versus Pload

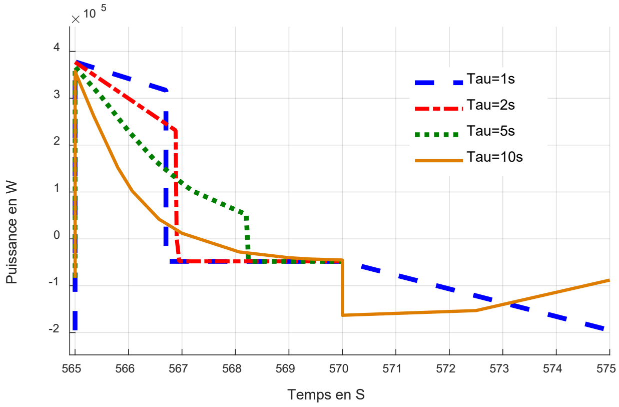


Figure 8: SCs power evolution for different values of Tau versus Pload

Figure 9 shows the FC voltage difference when it performs alone and when it is accompanied with the SCs. Figure 10 gives the SCs voltage and proves the optimum choice of Tau while reaching 150V for just seconds before being charged up to 319V. As for figures 11 and 12, they display the load power Pload versus the FC power and versus SCs power respectively obtained. Furthermore, figure 13 shows the FC current (i_{fc}) when the system relies on only the FC and when the system is hybrid in order to calculate the high power load operation period (while the FC current increases 46A). Figure 14 depicts the instantaneous consumption of hydrogen with and without SCs which follows the FC current shape.

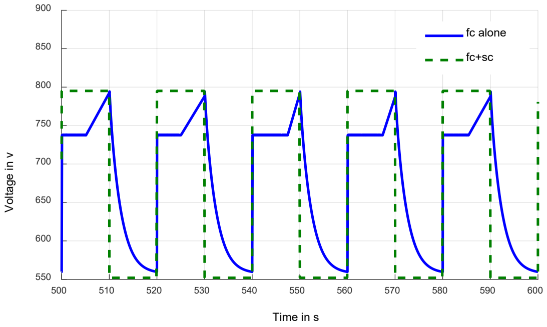


Figure 9: FC voltage

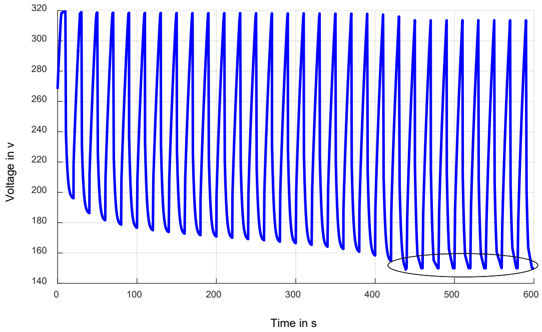


Figure 10: SCs voltage

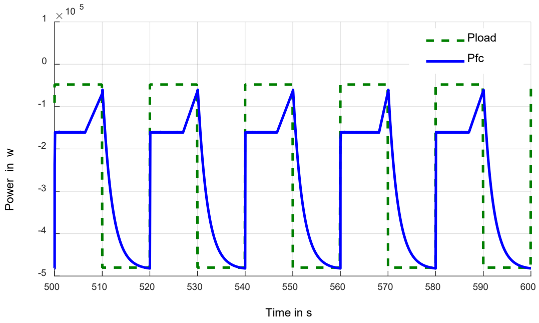


Figure 11: FC power versus load power with control system

At 550 s in figures 12 and 13, the FC is subjected to a sudden power demand by the load. The low-pass filter is applied in order to divert power variation to the SCs. At 558.83s, the FC supplies the full power and then the filter is off. Then at 560s the FC supplies the load and the SCs (in charging cycle) with energy.

At 570 s, another power demand arrives, the filter is activated again and the SCs meet this sudden power requirement until 578.83 s when the filter is shut down and the FC continues supplying the full power load: the same principle is repeated with each load power demand. We notice that the SCs recharge process stops when their voltage reaches its maximum value (319V); this avoids their deterioration.

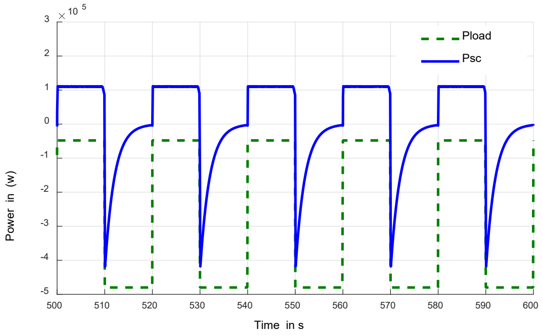


Figure 12: SCs power versus load power with control system

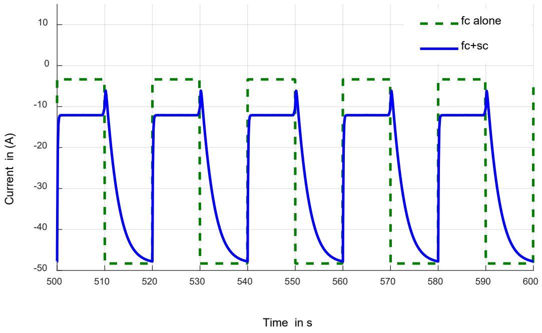


Figure 13: Single FC pack current

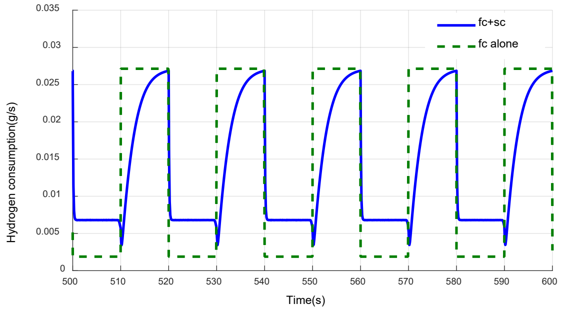


Figure 14: Hydrogen instantaneous consumption with control system

As it’s clear, the SCs make a significant effect on U_{fc} since they ensure the instant power changes only for a short period due to the quick drop of its voltage and the FC fulfills the rest of the supplied power until the 10 min are over.

In UPS applications, the FC is so efficient but the air supply compressor, cooling pump and radiator could affect its performance. Thus the air supply control strategy and resultant transients have a significant impact on the UPS efficiency and economy.

3.3. Gain in FC RMS Current and gain in FC Energy Losses:

In order to evaluate the FC/SCs association performance and to verify the good choice of Tau as well as the number of branches of SCs in parallel, two important criteria have been tested; the RMS current in the FC as well as its energy losses.

Certainly the more the number of SCs increases and Tau is greater the less the FC is solicited since SCs provide most of the load power for a longer time. For this, these quantities were calculated after each cycle and divided by the SCs cost.

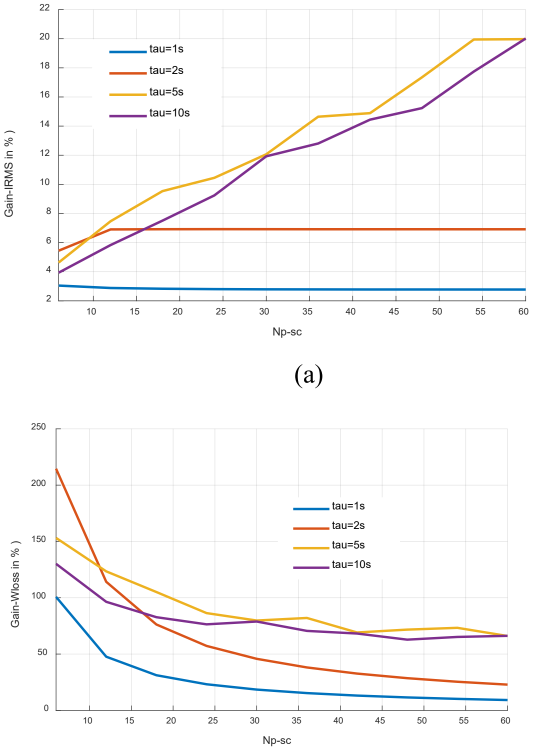
The internal resistance of the FC can give an approximate image of ohmic losses thanks to the RMS FC current (I_{rms}) which is compared to the FC current when it acts without SCs pack ($I_{rms-ref}$) [5]. The chemical processes are neglected.

The gain in FC RMS current is expressed as:

$$Gain_{IRMS}(\%) = \frac{I_{RMS\_ref} - I_{RMS}}{I_{RMS\_ref}} * 100 \qquad (20)$$

Figure 15 (a) shows the RMS current gain of the FC while varying the constant of the filter Tau (1, 2, 5, and 10) and the number of parallel branches of SCs (6, 12, 18, 24, 30, 36, 42, 48, 54 and 60). It increases as the number of parallel SCs branches the filter constant increase until an upper limit after what there is no enhancement.

The gain in RMS current is divided by the cost of the SCs pack as shown in figure 15 (b). It shows a maximum value about 214 % per € of gain/cost at Tau=2s and Np_{sc} =6 branches.



(a)

(b)

Figure 15: Gain in RMS current versus Tau and N_{p_sc}

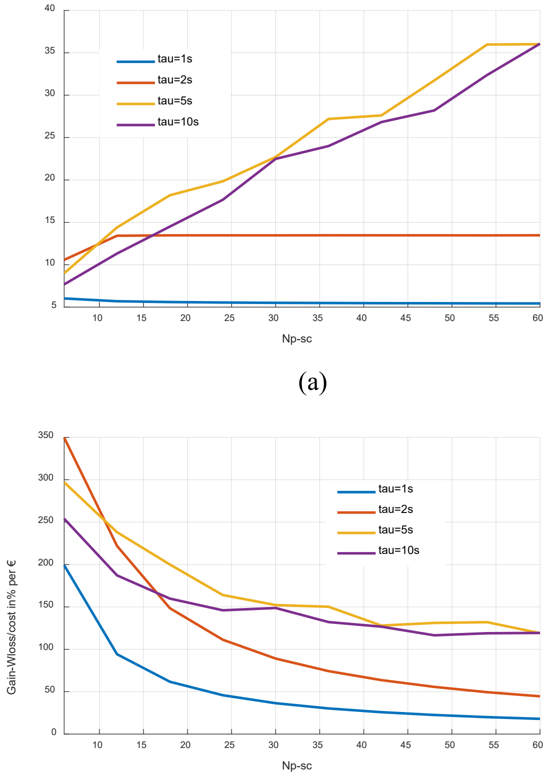
In another hand, reducing the energy losses for any storage element is the goal of any researcher in order to reduce its direct effect on its longevity. These losses were calculated for the FC by a comparison between FC losses (W_{loss}) when the system is hybrid and the FC losses while it is without SCs pack ($W_{loss-ref}$).

The gain is obtained as:

$$Gain_{Wlosses}(\%) = \frac{W_{loss-ref} - W_{loss}}{W_{loss-ref}} * 100 \qquad (21)$$

Figure 16 (a) and (b) shows gain in FC energy losses and the gain per cost with the same conditions of Tau and Np_{sc} for the gain in RMS current.

The gain in FC evolves as increasing Tau and Np_{sc} until no enhancement, but the gain per cost reaches 349.72 % per € for Tau = 2s and Np_{sc}= 6 branches.



(a)

(b)

Figure 16: Gain in FC losses versus Tau and N_{p_sc}

4. Lifetime cost evaluation

Some measurements may be available in order to evaluate the SOH of the FC and indicate its EOL.

EIS measurements are either given by the manufacturer or measured before the using the stack. However, these measurements cannot be performed all the lifetime of an embedded system [21].

The voltage decreases with time as presented in figure 17 which indicates the evolution of degradation mechanisms inside the

PEMFC; that's why some papers use the voltage and the power to evaluate the degradation rate of the system [22].

In order to evaluate the SOH of the FC, two thresholds can be defined. The first one is a threshold to conform a mission which decides if the FC can asset a given mission (the FC is not out of use if we operate in degraded mode). The second one is the definitive EoL; the FC is not able to deliver the power in safe conditions (loss of 10% of voltage).

Figure 17 shows that the voltage drops from 795V for a new stack to 791V for a stack at 90% of SoH and to 786V for a stack at EoL. The FC won't perform when the profile contain current variations if the stack is aging. It's clear that's there are differences between voltage at Beginning of Life (BOL) and voltage after a loss of 20% of membrane.

The Beginning of degradations can be declared when the system starts up. After only few hours the power decreases continuously. Degradations could be a result of an increase in resistance of the FC [23].

Figure 18 proves that aging FC consumes more Hydrogen than a new FC stacks; the more the FC is aged, the more the cost grows.

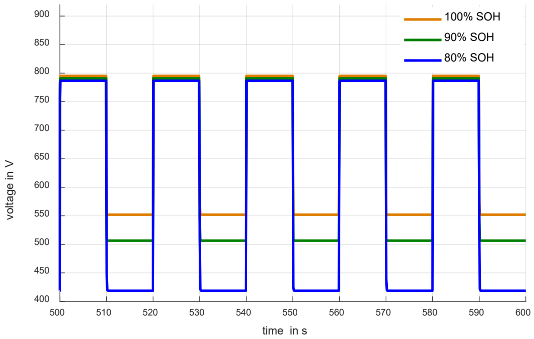


Figure 17: FC voltage from the BOL to its EOL

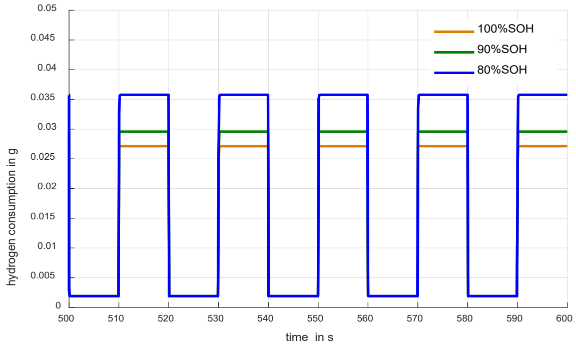


Figure 18. Hydrogen instantaneous consumption from the BOL to its EOL

The lifetime is calculated with (16) taking into account the high power load, the load change and the start/stop cycles in this paper. The FC degradation depends on the way the system operates.

Two interruptions occur per day, otherwise two start/stop cycles for 8h and 10 min, thus we can deduce n_1 the average start-stop cycles per hour. Furthermore, the load power period is 20s; for 10 min the load power makes 30 cycles and for 8h it makes 1440 cycles thus we can calculate the average load change cycles per hour n_2 .

Besides, the high load power degradation is defined when the FC operates at full power with a current that exceeds 46A. However when the system operates with a current less than 35A the Nexa module presents no degradation.

The degradation values are given in table 2. The high load power period is obtained by calculating the periods while the FC current exceeds 46A in seconds. After that, theses periods are expressed according to the backup autonomy in min/h; they are about 9.3min/h when the storage system is hybrid and about 30 min/h when the FC acts alone. In [24], the author gave FC degradation causes, the interest of use of auxiliary system for the FC lifetime and calculated the cost of the SCs and FC pack [24].

Table 2: PEM Fuel cells voltage degradation rates [15]

Operating conditions voltage	Degradation rate	Load spectrum	Values
Start-stop	1.1 mV/cycle	n_1	0.245 cycles/h
Load change	0.4185µv/cycle	n_2	180 cycles/h
High load power	0.54 mV/h	t_1	9.3 min/h (for Tau=2)

The study was limited to twenty cycles of FC failure. In order to estimate the cost, the first cycle of maintenance was considered.

The cost of the FC is calculated with equation (15). The PEMFC stack manufacture cost is estimated by 100 000€ [25] thus the first maintenance cycle FC stack cost is 5K€, the hydrogen cost is about 5€/kg and the maintenance cost is estimated by 20 000€.

A good modeling minimizes the SCs and FCs size. Consequently, the mass and cost of the UPS will be reduced. However, the FC hydrogen consumption and the power variations increased compared to the full design. Subsequently, there is a compromise between reducing the SC size and the FC lifetime [24].

Figures 19 and 20 are drawn in order to prove the optimum choice of Tau with a less impact on the FC lifetime and cost. Tau =2 provides 1510.28h of autonomy and costs only 14.15€ for 10 minutes of autonomy.

The cost of the SCs is assumed to be 0.01€/Farad [5]. The total cost of the SCs is 112*6*30=20k€. Their lifetime is about 1 million cycles of charge/ discharge. The charge cycle lasts 10s and the discharge cycle lasts 10s too which means the complete cycle duration is 20s. For the two long interruptions with 10 minutes and 8 hours, the SCs operate 29400s per day otherwise 1470 cycles per day. During 1510.28 h of the FC lifetime, the SCs operate

92504.65 cycles. Consequently, the SCs cost for the first maintenance cycle is 2K€.

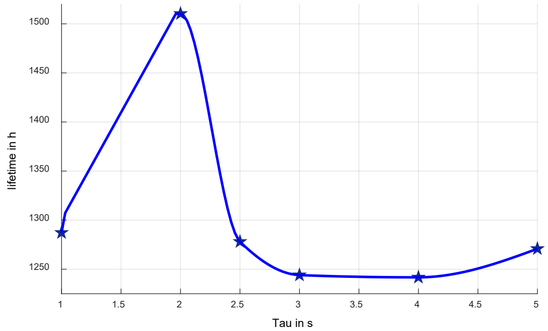


Figure 19: FC lifetime for different values of Tau

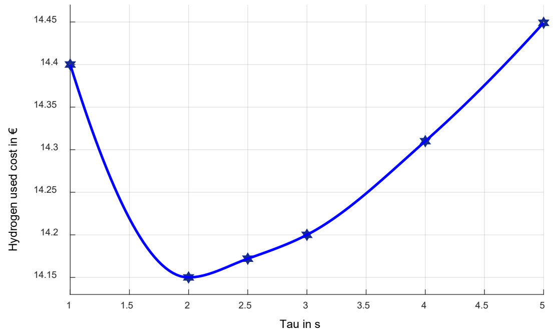


Figure 20: Hydrogen cost in 10 minutes for different values of Tau

The energy is product of the mean load power charge ((480KVA + 48KVA)/2) and the FC lifetime duration and it is expressed by KWh.

The comparison of using hybrid storage system relying on FC/SCs and using only the FC is summarized in table 3.

Table 3: The first maintenance cycle FC' lifetime cost.

	PAC alone	controlled FC/SC system
lifetime	1117.3 h	1510.28 h
The H2cost	97.57 k€	129.63 k€
The FC stack cost	5 k€	5 k€
The maintenance cost	20 k€	20k€
The SCs cost	-	2 K€
Energy	294 967.2 KWh	398 713.92KWh
The FC cost	0.4155 €/KWh	0.3928 €/KWh

Clearly, using SCs increases the FC lifetime since it lasts only 1117.3h when they are not used and also decreases the total FC cost per KWh.

The Gain in FC lifetime is not the only issue in the UPS. The SCs lifetime should also be considered which explains the fact that the power management decreases the SCs power and cost. Therefore, a compromise between all performances has to be made in order to decrease the cost of the whole UPS system and get a better lifetime.

5. Conclusion

In this paper a 500-kVA rated UPS was presented. The system included a FC standalone architecture and another where it is coupled with SCs and simulated on Matlab-Simulink. Simulations show that FC/SCs hybridization could reduce significantly the FC peak current requirement and so the possibility of downsizing the potential of the FC system. In fact, downsizing the FC can reduce the system peak efficiency but improve the average system efficiency. FC degradations were presented while taking into account high load power, load change and start/stop degradations in order to estimate its lifetime by a quick evaluation method. The expression of the formula is suitable for all applications. The cell voltage decreases of about 10% at a constant current at the EOL; this percentage is not definite. In fact, to evaluate the future SoH of the system it is crucial to know if the FC remains in a healthy mode or it is already degraded. Furthermore, an evaluation method of the FC economic lifetime was determined. It was based on cost ratio and optimal benefits. Clearly, the fuel consumption increases in degraded modes and with FCs use and aging.

Acknowledgment

This work is a contribution of the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique of Tunisia. The authors would like to acknowledge the conference paper "Fuel cell/supercapacitors combination in Uninterruptible Power Supply (UPS)" Published in: Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2015, 16th International Conference published by IEEE.

References

[1] S. D. Knights, K. M. Colbow, J. St-Pierre, and D. P. Wilkinson, Aging mechanisms and lifetime of PEFC and DMFC, J. Power Sources, vol. 127, no. 12, pp. 127–134, Mar. 2004.

[2] R. N. d. Fonseca, Optimization of the Sizing and Energy Management Strategy for a Hybrid Fuel Cell Vehicle Including Fuel Cell Dynamics and Durability Constraints, Ph.D. dissertation, The National Institute of Polytechnique in Toulouse, 2013

[3] Hredzak, V. G. Agelidis, and G. D. Demetriades, A Low Complexity Control System for a Hybrid DC Power Source Based on Ultracapacitor–Lead–Acid Battery Configuration, IEEE Trans. Power Electron., vol. 29, no. 6, pp. 2882–2891, Jun. 2014.

[4] P. Thounthong, S. Raël, and B. Davat, Energy management of fuel cell/battery/supercapacitor hybrid power source for vehicle applications, J. Power Sources, vol. 193, pp. 376–385, 2009.

[5] A. Lahyani, P. Venet, A. Guermazi, and A. Troudi, Battery/Supercapacitors Combination in Uninterruptible Power Supply (UPS), IEEE Trans. Power Electron., vol. 28, no. 4, pp. 1509-1522, April 2013.

[6] J. S. Yoo, J. Y. Choi, M. K. Yang, H. S. Cho, and W. Y. Choi, High Efficiency Power Conversion System for Battery-Ultracapacitor Hybrid Energy Storages, IEEE Trans. Power Electron., vol. 978, pp. 2830–2835, March 2013.

[7] R. Onanena, L. Oukhellou, D. Candusso, A. Same, D.Hissel,P. Aknin, Estimation of fuel cell operating time for predictive maintenance strategies, Int. J. of hydrogen energy, vol. 35,no. 15, pp. 8022–8031, 2011

[8] P. Pei, Q. Chang, T. Tang, A quick evaluating method for automotive fuel cell lifetime, Int. J. of hydrogen energy, vol. 33, pp. 3829–3836, 2008

[9] T. Fletcher, R. Thring, M. Watkinson, An Energy Management Strategy to concurrently optimise fuel consumption & PEM fuel cell lifetime in a hybrid vehicle, Int. J. of Hydrogen Energy, vol. 157, 2016

[10] X. Han, F. Li, T. Zhang, T. Zhang, K. Song, Economic energy management strategy design and simulation for a dual-stack fuel cell electric vehicle, Int. J. of hydrogen energy, 2017

[11] W. Zhang, J. Li, ,L. Xu, M. Ouyanga, Y. Liu, Q. Hand, K. Li, Comparison study on life-cycle costs of different trams 1 powered by fuel cell systems and others, Int. J. of hydrogen energy, vol. 41, no. 38, pp. 16577–16591, 2016

[12] W. Choi, J. W. Howze, and P. Enjeti, Fuel-cell powered uninterruptible power supply systems: Design considerations, J. Power Sources, vol. 157, pp. 311–317, 2006.

[13] L. Zubieta and R. Bonert, Characterization of double-layer capacitors for power electronics applications, IEEE Trans. Industry Applications, vol. 36, no. 1, pp. 199–205, Jan./Feb. 2000.

[14] I. Lachhab and L. Krichen, An improved energy management strategy for FC/UC hybrid electric vehicles propelled by motor-wheels, Int. J. of hydrogen energy, vol. 39, pp. 571–581, 2014.

[15] NexaTM Power Module User's Manual, MAN5100078, Ballard

[16] T. Cao, H. Lin , L. Chen, Y. He and W. Tao, Numerical investigation of the coupled water and thermal management in PEM fuel cell, J. Applied Energy, vol. 112, pp. 1115, 2013

[17] Y. Sun, J. Ogden and M. Delucchi, Societal lifetime cost of hydrogen fuel cell vehicles, Int. J. of Hydrogen Energy, vol. 35, no. 21, pp. 11932, 2010

[18] J. Lee, M. Yoo, K. Cha, TW. Lim and T. Hur, Life cycle cost analysis to examine the economical feasibility of hydrogen as an alternative fuel, Int. J. of Hydrogen Energy, vol. 34, no. 10, pp.4243, 2009

[19] The Fuel cell and Hydrogen annual review, the 4th energy wave, 2016

[20] H. Chen, P. Pei, M. Song, Lifetime prediction and the economic lifetime of Proton Exchange Membrane fuel cells, J. Applied Energy, vol. 142, pp. 154-163, 2015

[21] M.Jouin, M. Bressel, S. Morando, R. Gouriveau, D. Hissel, M. C. Péra, N. Zerhouni, S. Jemei, M. Hilairet and B. O. Bouamama, "Estimating the end-of-life of PEM fuel cells : guidelines and metrics", J. Applied Energy, vol. 76, May 2016

[22] X. Zhang, Y. Rui, Z. Tong, X. Sichuan, S. Yong, N. Huaisheng, "The characteristics of voltage degradation of a proton exchange membrane fuel cell under a road operating environment", J. Hydrogen Energy, vol. 39, no. 17, pp. 9420 -9429, 2014

[23] D. Bezmalinovic, B. Simic and F. Barbir, "Characterization of PEM fuel cell degradation by polarization change Curves" , J. Power Sources vol. 294, pp. 82-87, 2015

[24] I.B. Amira, A. Lahyani, A. Guermazi, Fuel Cell/Supercapacitors combination in uninterruptible power supply(UPS), The 16 int. Conf. Of Sciences et Techniques d'Automatique STA'2015

[25] A. Chauvin, Contribution à l'optimisation globale pour le dimensionnement et la gestion d'énergie de véhicules hybrides électriques basée sur une approche combinatoire, Ph.D. dissertation, The national Institute of applied sciences in Lyon, 2015.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 231-237 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

An Aggregation Model for Energy Resources Management and Market Negotiations

Omid Abrishambaf, Pedro Faria*, João Spínola, Zita Vale

GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering – Polytechnic of Porto (ISEP/IPP), Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal

ARTICLE INFO

Article history:
Received: 04 December,2017
Accepted: 27 January,2018
Online: 06 April,2018

Keywords:
Demand Response
Aggregator
Distributed Generation
Smart grid
K-Mean Clustering

ABSTRACT

Currently the use of distributed energy resources, especially renewable generation, and demand response programs are widely discussed in scientific contexts, since they are a reality in nowadays electricity markets and distribution networks. In order to benefit from these concepts, an efficient energy management system is needed to prevent energy wasting and increase profits. In this paper, an optimization based aggregation model is presented for distributed energy resources and demand response program management. This aggregation model allows different types of customers to participate in electricity market through several tariffs based demand response programs. The optimization algorithm is a mixed-integer linear problem, which focuses on minimizing operational costs of the aggregator. Moreover, the aggregation process has been done via K-Means clustering algorithm, which obtains the aggregated costs and energy of resources for remuneration. By this way, the aggregator is aware of energy available and minimum selling price in order to participate in the market with profit. A realistic low voltage distribution network has been proposed as a case study in order to test and validate the proposed methodology. This distribution network consists of 25 distributed generation units, including photovoltaic, wind and biomass generation, and 20 consumers, including residential, commercial, and industrial buildings.

1. Introduction

The present paper is an extension of work originally proposed in *2017 IEEE Manchester PowerTech* [1]. The generation variation in Distributed Renewable Energy Resources (DRER) is a topic of introduction in a lot of research works, since they have a key role in nowadays power system [2]. By appropriate management on the consumption in demand side, energy efficiency and optimal energy usage should be addressed. Curtailment Service Provider (CSP), Virtual Power Player (VPP), and aggregator are entities that can provide reliable solutions for the management of consumption and generation resources, since these can be aggregated and represented as a unique resource in electricity markets [3-5].

In this context, an aggregator is responsible to optimally manage a certain number of resources in a region, and aggregate them as one resource. This simplifies the process of energy negotiation in electricity markets [6]. Moreover, if other players, such as Balance Responsible Parties (BRPs), exist in the network, the role of aggregator would be more efficient and important [7].

Nowadays, there are several European countries that employ the aggregator concept for electricity consumers [8]. As an example, France is one of these countries that accepted aggregated loads in every ancillary service program, and BRPs and aggregators have been reorganized based on [9], [10]:

- Performing electricity market negotiations, to calculate compensation costs by aggregator for BRP;
- Aggregator has no direct interaction with BRP, however, it establishes contract with an electricity supplier in order to have flexibility services.

In fact, an aggregator is accountable not only for DRERs, but also is responsible for Demand Response (DR) programs [10]. According to the Federal Energy Regulatory Commission (FERC) [11], DR program is referred as "Changes in electric use by demand-side resources from their normal consumption patterns in response to changes in the price of electricity, or to

*Corresponding Author: Pedro Faria, GECAD Research Group Polytechnic of Porto (ISEP/IPP), Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal, Email: pnsfaria@gmail.com

https://dx.doi.org/10.25046/aj030227

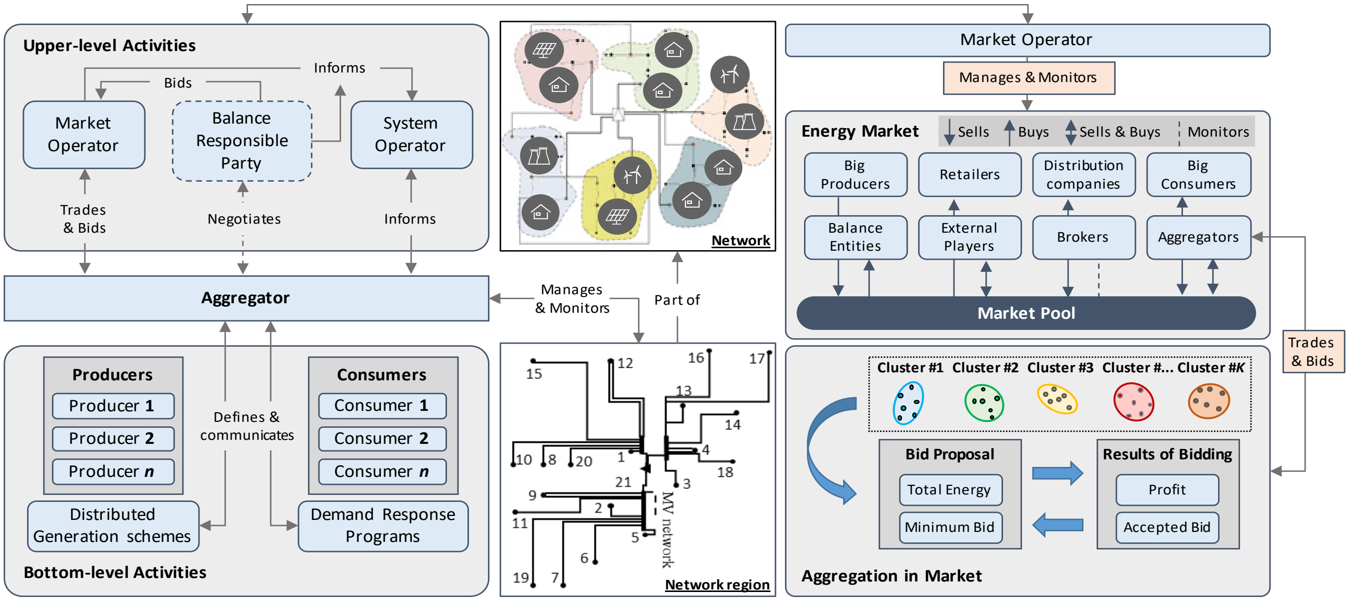


Figure 1. Overal architecture of the aggregation model.

incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized". The role of an aggregator in terms of DR programs is to gather all electricity consumer who can participate in DR programs, and present them as one. Therefore, it can be considered as a flexible player [12]. For this purpose, the aggregator can establish bidirectional contracts with end-users for DR programs to manage consumption resources, and consequently, to have flexibility in electricity market negotiations. In order to manage the generation of end-users, which are considered as prosumer (a consumer who is able to produce electricity), the aggregator can play the role of VPP, as [13]-[14] demonstrated before. It is clear that the generation capacity of these prosumers is not significant, thus, the network management would be difficult for system operators. Therefore, the need of a third party, namely an aggregator, is evident to gather all these small-scale consumption and generation resources, and participate in electricity market.

This paper represents an optimization based aggregation model for DRERs and DR programs managements, which enables small and medium resources to have active participation in the electricity markets. The aggregator controls demand-side customers by providing them several tariffs based DR programs, which brings flexibility in the electricity market negotiations. Moreover, this aggregator model gathers energy of resources and aggregated costs to be aware of available energy and minimum selling cost for defining remunerations, and also participate in the market with profit.

The rest of the paper is organized as follow. Section 2 details the aggregator model architecture considered for the aggregation. Section 3 describes the mathematical formulation considered for the optimization problem and aggregation process. Section 4 explains a case study that will test and validate the proposed method, and its results are expressed in Section 5. Finally, main conclusions of the work are proposed in Section 6.

2. Aggregator Model Architecture

This section focuses on how the presented aggregation model performs scheduling, aggregating and remuneration. The overall view of the presented model is illustrated by Figure 1. In this aggregation model, the consumption and generation resources are classified in several groups, where the output of the aggregation process will be the energy and cost of each group. As one can see in Figure 1 and also proposed in [15], the functionality of the aggregator is categorized in two sections of upper-level and bottom-level. In the upper level, the aggregator negotiates with players, such as market operator, BRP, and system operator; however, in the bottom level, it deals with demand-side users, namely small and medium scale consumers and producers.

The aggregator performs the scheduling process relying on external suppliers, Distributed Generation (DG) especially renewable resources, and DR programs. The customers who can execute DR programs would be able to establish contract with aggregator in three programs: load shifting, load reduction, and load curtailment. The load shifting model has been adapted from [16], and in this aggregation model it is considered as a free DR program. Load reduction and curtailment are the programs that aggregator takes them into account for scheduling and participating in the market. The aggregator considers a linear cost function for all external suppliers, DGs, load reduction and load curtailment. In this model, the aggregation process is done by K-Means Clustering algorithm by respect to the scheduled energy and its costs. In the aggregation process, only the resources that have been selected form the scheduling, are considered, and the rest that have no interaction in scheduling process, will not be considered. The aggregator categorizes the resources in several groups, and specifies a remuneration for each group, which called group tariff. This means the remuneration process should be calculated after the aggregation. The resources that are classified in a group, will be remunerated with same price. For this reason, the maximum price available in each group will be selected for group tariff. Therefore, the cheapest resource in the group will be motivated to participate in aggregation, since the group tariff is greater than the price initially defined, and also the most expensive resource will be satisfied, since the group tariff is as same as the price that it proposed.

In this way, the aggregator is able to participate in the market with a bid for each group. In each bid, the aggregator deliberates the gathered energy from the resources and also the group tariff as the minimum rate. The energy in each group is related to the aggregation of scheduled resources of that related group, therefore, the aggregator can easily manage its activities. On the other hand, the aggregator will be able to have negotiation in the market by biding the available energy of each group with a certain price, where this price should be greater or equal to the group tariff for the aggregator to gain profits or at least obtain the amount expended for the resources.

3. Optimization problem

The mathematical formulation regarding the presented aggregation model, especially resource scheduling, will be presented in this section. The optimization problem developed for the aggregator scheduling contains several continuous and discrete variables, therefore, the problem is considered as a mixed-integer linear problem (MILP). The objective function considered for the aggregation model is to minimize its Operational Cost (OC) and is shown by (1). It should be noted that in this model it is supposed the technical verification of the network is the obligation of the network operator, and the aggregator is not responsible for this matter.

$$MinOC = \sum_{s=1}^{S} P^{Sup}_{(s,t)} \cdot C^{Sup}_{(s,t)} + \sum_{p=1}^{P} P^{DG}_{(p,t)} \cdot C^{DG}_{(p,t)} + \sum_{c=1}^{Cs}\left[\begin{array}{l} P^{Red}_{(c,t)} \cdot C^{Red}_{(c,t)} + P^{Cut}_{(c,t)} \cdot C^{Cut}_{(c,t)} \\ +\sum_{d=1}^{T} P^{Shift}_{(c,t,d)} \cdot C^{Shift}_{(c,t,d)} \end{array}\right] \quad \forall t \in \{1,...,T\} \tag{1}$$

In this objective function, $P^{Sup}_{(s,t)}$ is purchased energy from external supplier, $P^{DG}_{(p,t)}$ denotes the attained energy from DG, $P^{Red}_{(c,t)}$ stands for DR load reduction, $P^{Cut}_{(c,t)}$ is for DR load curtailment, and $P^{Shift}_{(c,t,d)}$ represents DR load shifting.

There are several constraints that should be considered in the objective function. The first constraint stands for load balance, as (2) shows. In this equation, $P^{Load}_{(c,t)}$ presents the required demand of consumers.

Also, the technical limitations of all resources available in the proposed methodology should be considered. Therefore, (3) represents the generation limitations of external supplier in term of minimum and maximum ($P^{Sup\min}_{(s,t)}$, $P^{Sup\max}_{(s,t)}$), and (4) considers DG limitations ($P^{DG\min}_{(p,t)}, P^{DG\max}_{(p,t)}$).

$$\sum_{c=1}^{Cs}\left[\begin{array}{l} P^{Load}_{(c,t)} - P^{Red}_{(c,t)} - P^{Cut}_{(c,t)} \\ -\sum_{d=1}^{T}\left(P^{Shift}_{(c,t,d)} - P^{Shift}_{(c,d,t)}\right) \end{array}\right] -\sum_{s=1}^{S} P^{Sup}_{(s,t)} - \sum_{p=1}^{P} P^{DG}_{(p,t)} = 0 \quad \forall t \in \{1,...,T\} \tag{2}$$

$$P^{Sup\min}_{(s,t)} \le P^{Sup}_{(s,t)} \le P^{Sup\max}_{(s,t)} \quad \forall s \in \{1,...,S\}, \forall t \in \{1,...,T\} \tag{3}$$

$$P^{DG\min}_{(p,t)} \le P^{DG}_{(p,t)} \le P^{DG\max}_{(p,t)} \quad \forall p \in \{1,...,P\}, \forall t \in \{1,...,T\} \tag{4}$$

DR technical limitations, including load reduction, curtailment, shifting, are presented by (5)-(8).

$$P^{Red\min}_{(c,t)} \le P^{Red}_{(c,t)} \le P^{Red\max}_{(c,t)} \tag{5}$$

$$P^{Cut\min}_{(c,t)} \le P^{Cut}_{(c,t)} \le P^{Cut\max}_{(c,t)} \tag{6}$$

$$P^{Cut}_{(c,t)} = P^{Cut\max}_{(c,t)} \cdot X^{Cut}_{(c,t)} \quad X^{Cut}_{(c,t)} \in \{0,1\} \quad \forall c \in \{1,...,C\}, \forall t \in \{1,...,T\} \tag{7}$$

$$P^{Shift\min}_{(c,t,d)} \le P^{Shift}_{(c,t,d)} \le P^{Shift\max}_{(c,t,d)} \tag{8}$$

Although load shifting may not be pleasant for end-users, it is an appropriate and practical tool for aggregator. Load shifting process may limit consumers use of devices in a certain period, however, it enables the aggregator to manage the consumption based on the offered generation capacity. For this purpose, the limitations of maximum energy that will be shifted out from a period ($P^{Shift\_out}_{(c,t)}$), and enters to another period ($P^{Shift\_in}_{(c,t)}$) are proposed in (9) and (10).

$$\sum_{d=1}^{T} P^{Shift}_{(c,t,d)} \le P^{Shift\_out}_{(c,t)} \tag{9}$$

$$\sum_{d=1}^{T} P^{Shift}_{(c,d,t)} \le P^{Shift\_in}_{(c,t)} \quad \forall c \in \{1,...,C\}, \forall t,d \in \{1,...,T\} \tag{10}$$

Moreover, (11) demonstrates the constraint regarding the groups tariff and their remuneration, which is the maximum price of group. The groups are separated based on the type of available resources (DG or DR).

$$G^{DG}_{(k,t)} = max(C^{DG}_{(p,t)}) \quad G^{DR}_{(k,t)} = max(C^{Red}_{(c,t)}, C^{Cut}_{(c,t)}), \forall c \in \{1,...,C\} \quad \forall p \in \{1,...,P\}, \forall k \in \{1,...,K\}, \forall t \in \{1,...,T\} \tag{11}$$

As a summary, the mathematical formulation for resources scheduling and their remuneration performed by the aggregator have been explained in this part. The methodology presented in this section will be employed in a case study in the next section.

4. Case Study

In order to examine the model represented in this paper, a case study is proposed. For this purpose, an low voltage distribution network of a university campus, in Porto, Portugal, is considered for the aggregator, which has been adapted from [17]. This distribution network is shown in the bottom of Figure 1 (Network region) and is considered as a part of main network. The network consists of underground electrical lines with 21 buses, where a MV/LV transformer in BUS #21, connects the campus network to the main network.

For this case study, we considered that there are 20 consumers, and 26 producers in the network. The consumers include 8 Residential (RE) buildings, 10 Commercial buildings in three scales of small (C-S), medium (C-M), and large (C-L), and 2 Industrial (IN) units, which are classified based on average daily consumption. Moreover, producers consist of renewable resources including 20 Photovoltaic (PV) units and 4 wind generators, 1 biomass, and external suppliers. The generation and consumption profiles of whole network considered for day-ahead scheduling in a winter day are shown on Figure 2.

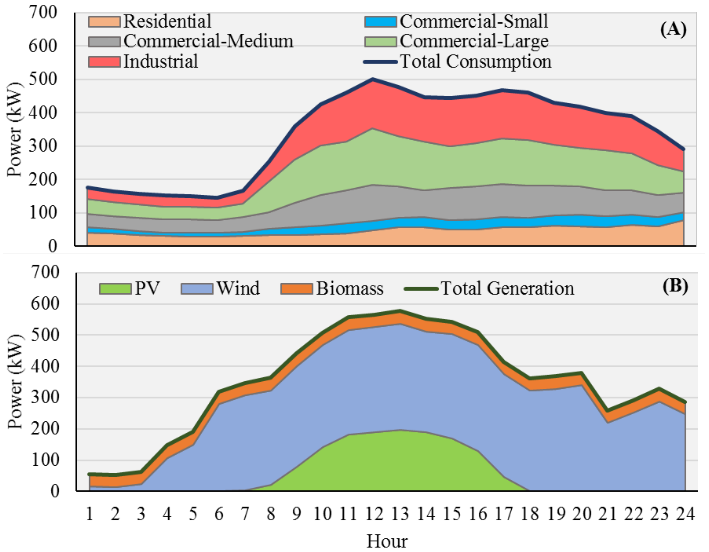


Figure 2. Day-ahead profiles of the network considered for case study: (A) Consumption, (B) Production.

As it can be seen in Figure 2 – (A), large commercial buildings and industrial units have occupied a huge part of consumption, and peak periods start from period #10 to #23. In this case study, it is presumed that the biomass production, and external suppliers have maximum capacity of 40 and 500 kW respectively. The external suppliers profile is not illustrated in Figure 2 – (B), since it is out of scope of figure and is a constant value during all periods. Moreover, it is considered that all producers would be able to contribute in the aggregation process, except external suppliers. Additionally, as you can see in Figure 2 – (B), since a winter day selected for the case study, PV producers have no significant generation, therefore, the aggregator should rely on wind, biomass, external suppliers and DR programs to prevent purchasing energy from the market. However, by comparing both parts of Figure 2, it is obvious that there are some periods that aggregator has more generation than consumption, therefore, it would be able to sell energy to the market and gain profits.

Regarding DR programs, Figure 3 demonstrates linear costs considered for each consumer based on its type. These costs are for load reduction and load curtailment, where 20% of the initial consumption is considered as maximum load reduction, and 15% for maximum load curtailment.

Furthermore, the linear costs considered for energy resources are shown on Figure 4. Each point in Figure 4 is the individual cost of each resource, where resource #1 to #20 are all PV, #21 to #24 are wind generators, #25 is biomass unit, and #26 illustrates external suppliers. It is should be mentioned that the costs demonstrated in Figure 4, are constant in all periods.

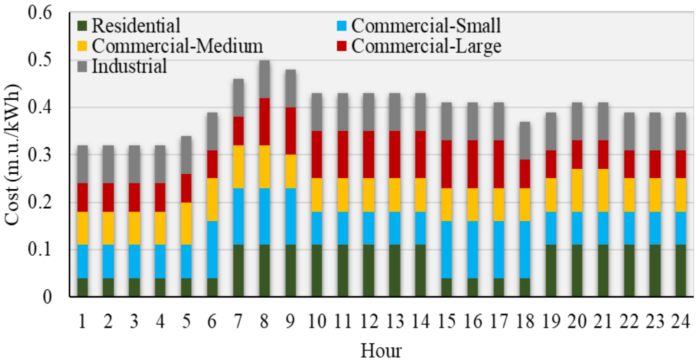


Figure 3. DR program costs for consumers.

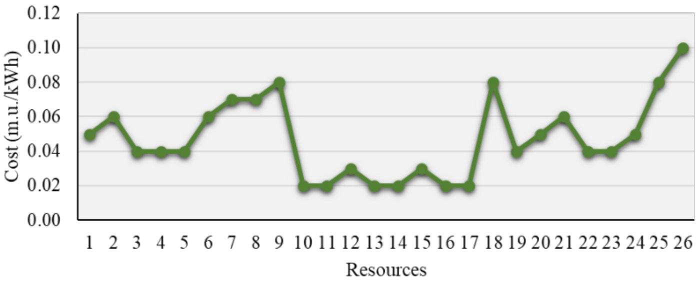


Figure 4. Individual cost for each energy resource.

Additionally, Figure 5 represents the day-ahead market prices considered for the aggregator in order to participate in market negotiations.

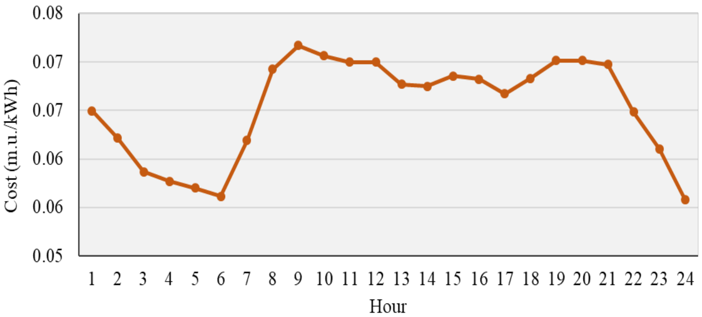


Figure 5. MIBEL market price for Portugues section in a winter day.

These prices are for a winter day in 2017 and have been adapted from Portuguese sector of Iberian Electricity Markets (MIBEL) [18]. In order to model the participation of the aggregator in the electricity markets, a market place should be taken into account, to guaranty its contribution in the competition. For this purpose, a market pool is an appropriate solution to ensure that third parties, such as aggregator, would be able to present energy bids.

5. Results

This section concerns the aggregation and scheduling results of the case study presented in the previous section. The optimization problem of aggregation and scheduling presented in this paper, has been solved through TOMLAB [19]. Additionally, the market negotiation results are represented, which shows how the aggregator utilizes these results for providing a bid. In the case study, we considered that the aggregator meets a drop from external suppliers in first four periods that can supply only 10% of their capacity. The reason of this lack of energy is considered as a fault or any other causes in the external suppliers. Figure 6 shows the network consumption before and after the scheduling of aggregator.

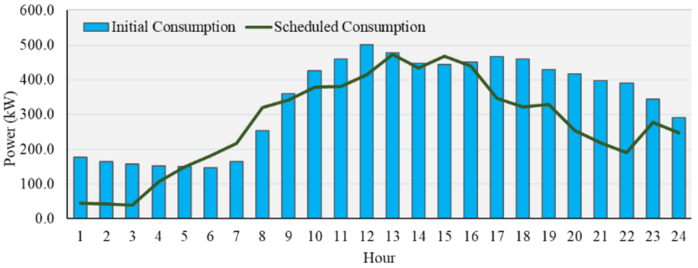


Figure 6. Total consumption of the network.

The scheduling results shown in Figure 6 are based on DG and available energy during each period. Additionally, there are several periods that scheduled consumption profile are greater or smaller than the initial profile. This is due to the utilization of DR programs by aggregator. With this in mind, Figure 7 illustrates more information regarding the generation and DR scheduling.

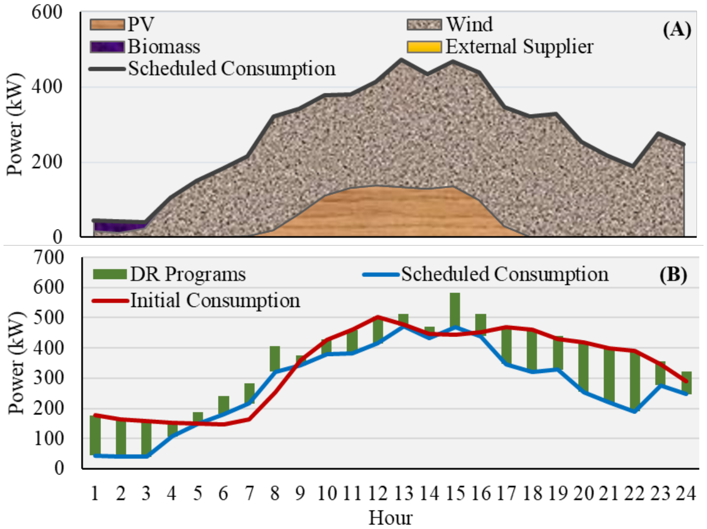


Figure 7. Detailed scheduling results of aggregator: (A) Generation scheduling, (B) DR scheduling.

As one can see in Figure 7 – (A), since the DG suppliers are considered as cheapest resources comparing with external suppliers, the aggregator utilizes the available DG energy, especially PV and wind, to supply the demand, and in the first four periods, it employs biomass generation to supply the loads. In other words, the aggregator reduced the consumption to the available DG energy in order to prevent purchasing energy from market for minimizing the costs. This means, in the periods that the DG generation is not adequate for the demand, aggregator applies DR programs to regulate the difference between the consumption and generation, as illustrated in Figure 7 – (B). The DR programs that aggregator employed to balance the network for each single period, are shown on Figure 8. The utilized DR programs include load reduction, load curtailment, and load shifting.

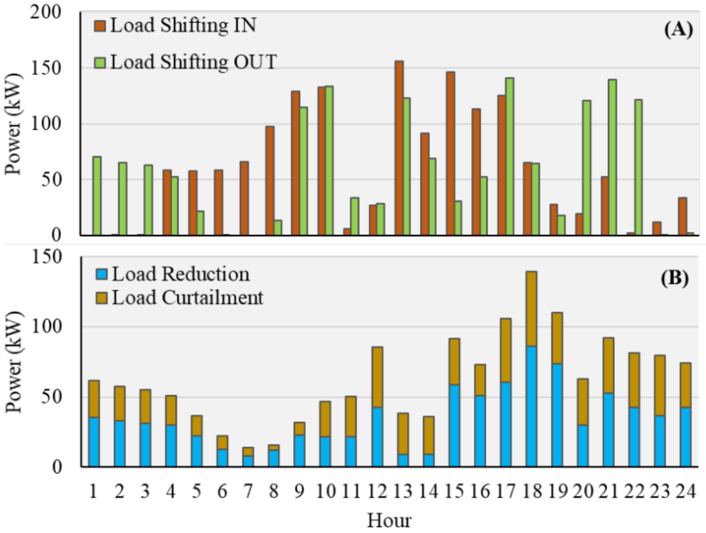


Figure 8. DR programs used by aggregator for network balancing: (A) Load shifting, (B) Load reduction and curtailment.

The incoming and outcoming consumption of each period during load shifting are shown on Figure 8 – (A), which occurred during low generation periods, and shifted to high generation periods. The load shifting enables the aggregator to manage the consumption and shift it to desired periods to prevent purchasing energy from the market, since it is more expensive comparing with DG resources.

Additionally, Table 1 shows the results of aggregation and remuneration processes for period number 12.

Table 1. Remuneration and aggregation results for a single period.

Group		1	2	3	4	5
DG	PV (kW)	0	36.23	19.64	37.54	43.86
	Wind (kW)	250.58	0	0	0	26.88
	Biomass (kW)	0	0	0	0	0
	tariff (m.u./kWh)	0.05	0.02	0.05	0.06	0.06
DR	Residential (kW)	0	0	0	0	0
	Commercial Small (kW)	0	0	0	0	10.09
	Commercial Medium (kW)	0	0	11.07	20.05	7.11
	Commercial Large (kW)	0	0	0	0	0
	Industrial (kW)	26.21	10.94	0	0	0
	tariff (m.u./kWh)	0.06	0.06	0.05	0.05	0.05
Total (kW)		500.18				

In Table 1, the total energy as well as the number of resources in each group have been calculated by aggregation computation, however, the group tariff has been indicated by remuneration calculation. Moreover, in order to calculate the profit of the aggregator after paying all resources, including DG and incentives for DR participation, (12) is proposed. This profit is the monetary benefit that aggregator gained after its operations.

$$Profit = C_{(t)}^{mcp}\cdot\left[\sum_{p\in k}^{P} P_{(p,t)}^{DG} + \sum_{c\in k}^{C}(P_{(c,t)}^{Red}+P_{(c,t)}^{Cut})\right]$$
$$-\left[G_{(k,t)}^{DG}\cdot\sum_{p\in k}^{P}P_{(p,t)}^{DG} + G_{(k,t)}^{DR}\cdot\sum_{c\in k}^{C}(P_{(c,t)}^{Red}+P_{(c,t)}^{Cut})\right] \qquad (12)$$
$$\forall k\in\{1,...,K\},\ \forall t\in\{1,...,T\}$$

In (12), the $C_{(t)}^{mcp}$ denotes market clearing price, which is considered in this case study is equal to the market prices provided in Figure 5. The classification of the resources in the several groups enables the aggregator to provide lower group tariffs, comparing with the situation that all resources are in the same group. It is true that with classification of resources in several groups, high group tariff will be still remained, however, the chance of aggregator to reach some group tariff with lower rates will be increased. The financial profit gained by aggregator during period number 12, is shown on Table 2. In this single period, the aggregator has total energy of 500.18 kW, which has incoming of 35 monetary unit from the energy that sold to the market. However, it also paid 22.84 monetary unit for all resources, including DG units and DR incentives, and in total, 12.17 monetary unit will be the final profit of aggregator during period number 12.

Table 2. Gained profit by aggregator during market negotiations for one period.

Parameter	Value
Costs paid to all resources (m.u.)	22.84
Market clearing price (m.u./kWh)	0,0700
Income from market sell (m.u.)	35.00
Total aggregator profit (m.u.)	12.17

The profit of aggregator shown on Table 2, is for a single period (considered as one hour in a day in this case study), and even with a few number of consumers and generators, it could gain profit from market negotiations. This means that if the aggregator is responsible for a larger network He will be able to aggregate more energy capacity for clustering, and therefore, with great participation in market, which leads to obtain a satisfying amount of financial benefits. However, this profitability depends on the capabilities and offers of aggregator in market negotiations and existing competitions. Figure 9 demonstrates the financial results concerning the participation of aggregator in the electricity market for all periods of case study. These results are obtained after the scheduling and remuneration processes. It should be noted that only the resources that participated in these processes, are considered. The costs of each period in Figure 9 follows the same process represented in Table 2, which the gained profit is a subtract of costs paid to all resources and the income from market participation.

The last results of this section are related to a comparison that shows the impact of load shifting method for aggregator. For this purpose, it is considered that the aggregator is not capable to employ load shifting during scheduling process. The scheduling results, without load shifting, are illustrated in Figure 10. The results shown in Figure 10 (without load shifting) can be compared with the scheduling results demonstrated in Figure 7 (with load shifting).

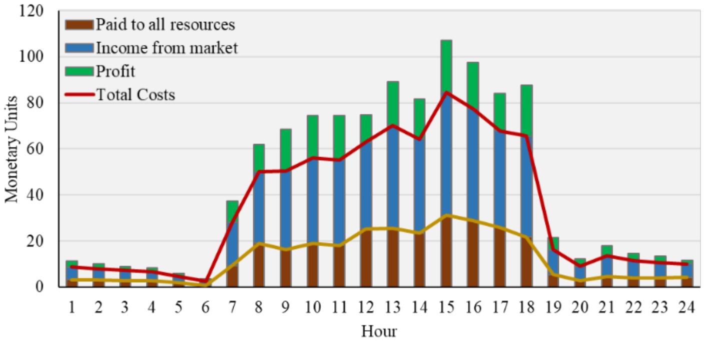


Figure 9. Detailed aggregator costs after scheduling and remuneration process for all periods.

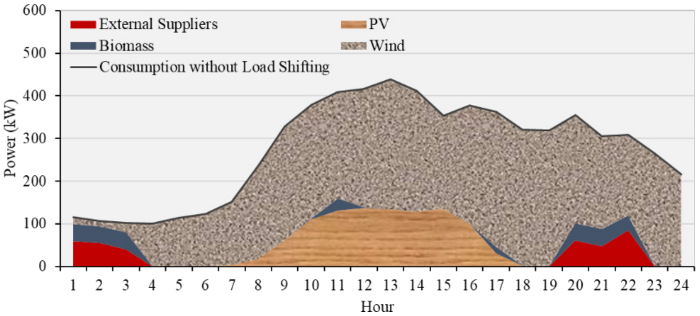


Figure 10. DG scheduling results without load shifting.

As one can see in Figure 10, in some periods the aggregator not only utilizes all available DG resources to supply the demand, but also, it is forced to use energy from external suppliers to feed all demand. By this way, since the electricity price of external suppliers are more expensive than the DG resources, the total costs of aggregator will be increased, and therefore, the gained profit will be decreased. However, as Figure 7 demonstrated, if the aggregator utilized load shifting scenario, and shift the load from the moments that there is no adequate DG energy, to the periods with high DG energy, its operational costs will be reduced, and obtained financial benefits will be increased.

6. Conclusions

This paper presented an aggregator model for distributed energy resource and demand response program management. The presented model considered the resources able to participate in electricity market negotiations through the aggregator. The aggregator has capability of demand-side flexibility by establishing several demand response contracts with consumers.

The main focus of the paper was given to a business model that aggregator utilized it to gather energy of resources and their costs, to define a fair remuneration tariff for all resources, as well as an affordable price for market participation. By this way, the aggregator guarantees that the small-scale resources, including distributed generation and demand response programs, will participate in the electricity market, and therefore, getting profits.

The results of case study demonstrate that the aggregator model is able to perform an optimal scheduling for distributed resources, in order to minimize the operational costs of the aggregator. This is done through implementing several DR programs. The final outcomes of aggregation and remuneration processes validated the proposed method, and proved that the aggregator can gain financial benefits from market negotiations, even by paying a fair tariff to all available resources.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 641794 (project DREAM-GO). This work also received funding from the following projects: NETEFFICITY Project (ANI | P2020 – 18015); and from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013.

References

[1] J. Spinola, P. Faria, Z. Vale, "Model for the integration of distributed energy resources in energy markets by an aggregator", in 2017 IEEE Manchester PowerTech, Manchester UK, 2017. https://doi.org/10.1109/PTC.2017.7981234

[2] S. M. Nosratabadi, R.-A. Hooshmand, E. Gholipour, "A comprehensive review on microgrid and virtual power plant concepts employed for distributed energy resources scheduling in power systems," Renew. Sustain. Energy Rev., **67**, 341–363, 2017. https://doi.org/10.1016/j.rser.2016.09.025

[3] C. Battistelli, A. J. Conejo, "Optimal management of the automatic generation control service in smart user grids including electric vehicles and distributed resources," Electr. Power Syst. Res., **111**, 22–31, 2014. https://doi.org/10.1016/j.epsr.2014.01.008

[4] A. Roos, S. Ø. Ottesen, T. F. Bolkesjø, "Modeling Consumer Flexibility of an Aggregator Participating in the Wholesale Power Market and the Regulation Capacity Market," Energy Procedia, **58**, 79–86, 2014. https://doi.org/10.1016/j.egypro.2014.10.412

[5] O. Abrishambaf, P. Faria, L. Gomes, J. Spínola, Z. Vale J. Corchado, "Implementation of a Real-Time Microgrid Simulation Platform Based on Centralized and Distributed Management", Energies, **10**(6), 806-820, 2017. http://dx.doi.org/10.3390/en10060806

[6] D. J. Vergados, I. Mamounakis, P. Makris, E. Varvarigos, "Prosumer clustering into virtual microgrids for cost reduction in renewable energy trading markets," Sustain. Energy, Grids Networks, **7**, 90–103, 2016. https://doi.org/10.1016/j.segan.2016.06.002

[7] S. Rahnama, S. E. Shafiei, J. Stoustrup, H. Rasmussen, J. Bendtsen, "Evaluation of Aggregators for Integration of Large-scale Consumers in Smart Grid," IFAC Proc. Vol., **47**(3), 1879–1885, 2014. https://doi.org/10.3182/20140824-6-ZA-1003.00601

[8] EG3 Report - Smart Grid Task Force, "Regulatory Recommendations for the Deployment of Flexibility," 2015.

[9] Smart Energy Demand Coalition, "Mapping Demand Response in Europe Today 2015," 2015.

[10] Smart Energy Demand Coalition, "Enabling independent aggregation in the European electricity markets," 2015.

[11] Federal Energy Regulatory Commission, "Assessment of Demand Response & Advanced Metering," 2011.

[12] Smart Grid Task Force, "Regulatory Recommendations for the Deployment of Flexibility," 2015.

[13] S. Rahmani-Dabbagh, M. K. Sheikh-El-Eslami, "A profit sharing scheme for distributed energy resources integrated into a virtual power plant," Appl. Energy, **184**, 313–328, 2016. https://doi.org/10.1016/j.apenergy.2016.10.022

[14] P. Faria, J. Spínola, Z. Vale, "Aggregation and Remuneration of Electricity Consumers and Producers for the Definition of Demand-Response Programs," IEEE Trans. on Industrial Informatics, **12(**3), 952–961, 2016. https://doi.org/10.1109/TII.2016.2541542

[15] N. Mahmoudi, E. Heydarian-Forushani, M. Shafie-khah, T. K. Saha, M. E. H. Golshan, P. Siano, "A bottom-up approach for demand response aggregators' participation in electricity markets," Electr. Power Syst. Res., **143**, 121–129, 2017. https://doi.org/10.1016/j.epsr.2016.08.038

[16] P. Faria, Z. Vale, J. Baptista, "Constrained consumption shifting management in the distributed energy resources scheduling considering demand response," Energy Convers. Manag., **93**, 309–320, 2015. https://doi.org/10.1016/j.enconman.2015.01.028

[17] M. Silva, F. Fernandes, H. Morais, S. Ramos, Z. Vale, "Hour-ahead energy resource management in university campus microgrid," 2015 IEEE Eindhoven PowerTech, Eindhoven Netherlands, 2015. https://doi.org/10.1109/PTC.2015.7232449.

[18] MIBEL Electricity Market. Available online: http://www.omip.pt (accessed on 30 Nov. 2017)

[19] TOMLAB Optimization. Available online: https://tomopt.com/tomlab/ (accessed on 30 Nov. 2017)

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 238-251 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

An Overview of Data Center Metrics and a Novel Approach for a New Family of Metrics

Moises Levy*, Daniel Raviv

Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, 33431, FL, USA

ARTICLE INFO

Article history:
Received: 13 February, 2018
Accepted: 04 April, 2018
Online: 11 April, 2018

Keywords:
Data center
Mission critical
Metric
Measurement
Performance
Efficiency
Productivity
Sustainability
Operations
Infrastructure
Management
Risk
Site

ABSTRACT

Data centers' mission critical nature, significant power consumption, and increasing reliance on them for digital information, have created an urgent need to monitor and adequately manage these facilities. Metrics are a key part of this effort as their indicators raise flags that lead to optimization of resource utilization. A thorough review of existing data center metrics presented in this paper shows that while existing metrics are valuable, they overlook important aspects. New metrics should enable a holistic understanding of the data center behavior. This paper proposes a novel framework using a multidimensional approach for a new family of data center metrics. Performance is examined across four different sub-dimensions: productivity, efficiency, sustainability, and operations. Risk associated with each of those sub-dimensions is contemplated. External risks are introduced, namely site risk, as another dimension of the metrics, and makes reference to a methodology that explains how it is calculated. Results from metrics across all sub-dimensions can be normalized to the same scale and incorporated in one graph, which simplifies visualization and reporting. The new family of data center metrics can help to standardize a process that evolves into a best practice to help evaluate data centers, to compare them to each other, and to improve the decision-making process.

1. Introduction

The ongoing significant and increasing reliance on digital information has led data centers to play a key role in guaranteeing that information is constantly available for users, and adequately stored. A novel framework for data center metrics using a multidimensional approach was introduced in a paper originally presented at the 15th LACCEI International Multi-Conference for Engineering, Education, and Technology: Global Partnerships for Development and Engineering Education in 2017 [1], for which this work is an extension.

Data centers are energy intensive complexes, and this sector is expected to grow substantially. These circumstances have prompted the desire and need to make data centers more efficient and sustainable, while at the same time ensuring reliability and availability. Efforts undertaken to pursue this goal include new legislation, the development of standards and best practices to follow when designing, building and operating a data center, and metrics to monitor performance and find areas of improvement.

Being that data centers are a growing field subject to new and evolving technologies, current practices ignore important pieces of information. Existing data center metrics are very specific, and fail to take into consideration the holistic performance of the data center. In addition, there is an imminent need for metrics to incorporate an assessment of the risk to which the data center is exposed.

The proposed concept addresses concerns from the recent United States Data Center Usage Report (June 2016) [2], which communicates the need to expand research on data center performance metrics that better capture efficiency, in order to identify and understand areas of improvement. The main motivation of this paper is to consolidate existing metrics and current practices, explain areas of improvement, and finally propose a novel multidimensional approach for data center metrics incorporating productivity, efficiency, sustainability, and operations, as well as measurements of all the different risks associated to the data center. This paper adds technical and scientific support to existing theoretical and practical work that has mostly been carried out from outside academia. The ultimate goal

*Corresponding Author: Moises Levy. Florida Atlantic University, USA
Email: mlevy2015@fau.edu

https://dx.doi.org/10.25046/aj030228

of this work is to help standardize a process that eventually becomes a best practice to rate data centers.

2. Background

A data center can be defined as a dedicated facility with all the resources required for storage, processing and sharing digital information, and its support areas. Data centers comprise the required infrastructure (e.g., power distribution, environmental control systems, telecommunications, security, fire protection, and automation) and information technology (IT) equipment (including servers, storage and network/communication equipment). Data centers are very dynamic; equipment can be upgraded frequently, new equipment may be added, obsolete equipment may be removed, and old and new systems may be in use simultaneously. Several threats can cause failures in a data center, including technical issues and human errors. The cost of downtime depends on the industry, and could reach thousands of dollars per minute. Recent reports show that the average cost associated with data center downtime for an unplanned outage is approximately $ 9,000 per minute, an increase of about 60% from 2010 to 2016 [3].

According to standards, best practices, and user requirements, the infrastructure of data centers must comply with stringent technical requirements that guarantee reliability, availability, and security, as they highly correlate with cost and efficiency. An infrastructure with high reliability and availability must have system redundancy, which makes it more expensive, and probably less efficient [4]. In this context, reliability is the ability of the system to perform its functions under stated conditions for a specified period of time, whereas availability refers to the degree to which a system is operational when it is required for use. Redundancy is the multiplication of data center components used to enhance reliability, since they may malfunction at a certain point due to maintenance, upgrade, replacement, or failure [5].

Data centers can consume 40 times more energy than conventional office buildings. IT equipment by itself can consume 1100 W/m^2 [6], and its high concentration in a data center results in higher power densities, compared to conventional office buildings, where energy consumption ranges between 30 and 110 W/m^2. In addition, the energy consumption profile of data centers is very different from conventional office buildings, since IT equipment power consumption represents more than 50% of the total consumption in data centers. Figure 1 shows an example of energy consumption breakdown [7].

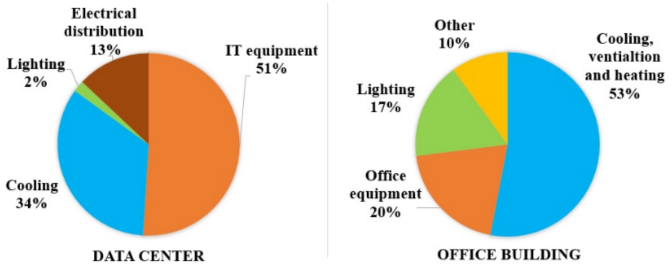


Figure 1: Energy consumption profile in a data center and office building

The United States is home to approximately 3 million data centers, representing one data center for every 100 people [8]. Data center electricity consumption increased approximately 24% from 2005-2010, 4% from 2010-2014, and is expected to grow 4% from 2014-2020 [2]. In 2014, data centers in the United States consumed around 70 billion kWh, which is 1.8% of total electricity consumption, and are predicted to consume around 73 billion kWh in 2020 [2]. Emerging technologies and energy management strategies may decrease the projected energy consumption.

Environmental impact of data centers varies depending on the energy sources used and the total heat generated. The differences between the lowest and the highest greenhouse gas (GHG) emissions associated with each energy source is approximately a factor of 200 [9]. For the period from 2002 to 2020, the emissions associated with the IT sector and the data center sector are estimated to grow by 180% and 240% respectively, considering business as usual [10]. This growth rate is much faster compared to the 30% increase in total emissions from all sources, again considering business as usual. Conversely, the IT sector has contributed to a reduction in emissions in other sectors, since IT is an enabling infrastructure for the global economy. For 1 kWh consumed by the IT sector in the United States, other 10 kWh are saved in other sectors due to the increase in economic productivity and energy efficiency [11]. The growing ubiquity of IT driven technologies has revolutionized and optimized the relation between efficiency and productivity, and energy consumption across every sector of the economy.

The deregulation of telecommunications in the United States through the Telecommunications Act of 1996 promoted the creation of a number of standards and best practices related to telecommunications, and more recently to data centers. Standards provide the most complete and reliable guidance to design or assess a data center, from national/ local codes (required) to performance standards (optional). Due to their mission critical tasks and elevated costs, data centers must be designed, built and operated in compliance with those standards to ensure basic performance and efficiency. Standards, however, do not translate into best practices when the objective is attaining the highest possible reliability and availability, under the best performance. Optional standards and best practices have contributed to achieving this goal.

Data centers standards and best practices evolve continually adapting to emerging needs, and addressing new issues and challenges. The *ASHRAE Technical Committee 9.9* has created a set of guidelines regarding the optimal and allowable range of temperature and humidity set points for data centers [12] [13] [14]. The *ANSI/ASHRAE standard 90.4 (Energy Standard for Data Centers)* establishes the minimum threshold for data center energy efficient design, construction, operation maintenance, and utilization of renewable energy resources [15]. The *Singapore Standard SS 564 (Green Data Centers)* addresses planning, building, operation and metrics of green data centers [16]. Standards contribute to classify data centers based on their reliability and infrastructure redundancy, such as the *ANSI/BICSI-002* (class F0 to F4) [5], the *ANSI/TIA-942* (rating 1 to 4) [17] and the *Data Center Site Infrastructure Standard* from Uptime Institute (tier levels 1 to 4) [18] [19].

Likewise, there is significant concern among legislators about how efficiently an IT facility uses energy. Governments have also imposed regulations on data centers depending on the nature of the business The *Energy Efficiency Improvement Act of 2014* (H.R. 2126) demands federal data centers to implement energy efficiency standards. This is motivated by the fact that federal data centers energy consumption represents 10% of all data centers in the United States. This bill encourages federal data centers to improve energy efficiency and develop best practices to reduce energy consumption [20]. According to the *Data Center Optimization Initiative* (DCOI), federal data centers must reduce their power usage efficiency below a specified threshold, unless they are scheduled to be shutdown, as part of the *Federal Data Center Consolidation Initiative* (FDCCI). In addition, federal data centers must replace manual collection and reporting with automated infrastructure management tools by the end of 2018, and must address different metric targets including energy metering, power usage effectiveness, virtualization, server utilization and facility monitoring [21].

Organizations that have contributed to the creation of standards, white papers, best practices and other documents related to the Data Center industry are: the International Organization for Standardization (ISO), the National Institute of Standards and Technology (NIST), the Institute of Electrical and Electronics Engineers (IEEE), the International Electrotechnical Commission (IEC), the National Fire Protection Association (NFPA), the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), the U.S. Department of Energy (DOE), the U.S. Environmental Protection Agency (EPA), the U.S. Green Building Council (USBGC), the Telecommunication Industry Association (TIA), the Building Industry Consulting Service International (BICSI), the Uptime Institute, The Green Grid (TGG), the Association for Computer Operations Management (AFCOM), the International Computer Room Expert Association (ICREA), the International Data Center Authority (IDCA), the European Commission (EU Code of Conduct for Data Centres), the British Computer Society (BCS), the Japan's Green IT Promotion Council (GIPC), the Japan Data Center Council (JDCC), and the Singapore Standards Council among others.

3. Overview of Data Center Metrics

Metrics are measures of quantitative assessment that allow comparisons or tracking of performance, efficiency, productivity, progress or other parameters over time. Through different metrics data centers can be evaluated in comparison to goals established, or to similar data centers. Variations or inconsistencies in measurements can produce a false result for a metric, which is why it is very important to standardize metrics. Much of the current metrics, standards and legislation on data centers is focused towards energy efficiency, as this has proven to be a challenge given the rapid growth of the sector and its energy intensive nature.

One of the most widely used energy efficiency metrics is *Power Usage Effectiveness* (PUE), introduced in 2007 [22]. Figure 2 shows the energy flow in a data center. PUE is calculated as the ratio of energy used in a facility to energy delivered to IT equipment [23]. Similarly, *Data Center infrastructure Efficiency* (DCiE) is defined as the reciprocal of PUE [24] with a value ranging between 0 and 1 to make the metric easier to understand in terms of efficiency. Organizations such as the EPA have selected PUE as the metric to analyze energy performance in data centers, and define Source PUE as the ratio of total facility energy used to UPS energy [25], [26]. The main limitation of PUE is that it only measures the efficiency of the building infrastructure supporting a given data center, but it indicates nothing about the efficiency of IT equipment, or operational efficiency, or risk involved [2] [27] [28].

Multiple metrics have been developed to measure other aspects of efficiency. *Data Center Size metric* (e.g., 'mini', 'small', 'medium', 'large', 'massive', and 'mega' data centers) is based on the number of racks and the physical compute area of the data center [29]. Other classifications based on the size of the data center have been proposed such as 'hyperscale', 'service provider', 'internal', 'server room', and 'server closet' [2]. The rack *Density metric* (e.g., low, medium, extreme, and high) considers the measured peak power consumption on every rack (density for rack) and across the compute space [29]. *Data Center Density* (DCD) is defined as the total power consumption of all equipment divided by the area. *Compute Power Efficiency* (CPE) is estimated as the IT equipment utilization multiplied by the IT equipment power consumption and divided by the total facility power consumption [30].

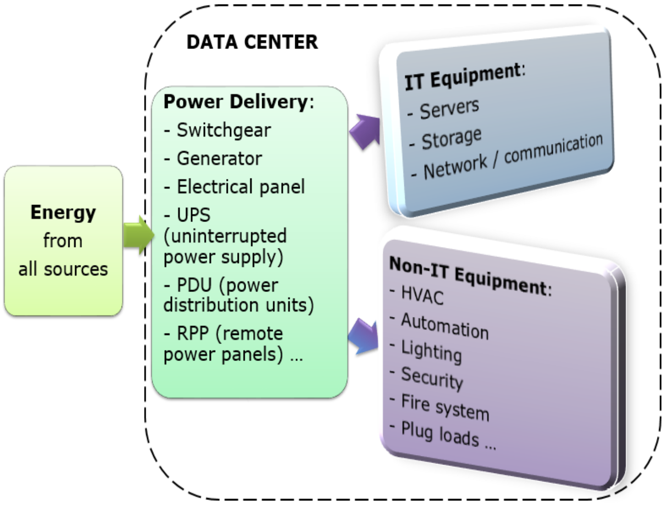


Figure 2: Data center energy flow diagram

Metrics have been proposed to measure server efficiency and performance for computer servers and storage. The metric *FLOPS* (floating-point operations per second) *per Watt* measures performance per unit of power [31], and is used to rank the most energy efficient supercomputers' speed. Standard Performance Evaluation Corporation (SPEC) has contributed with proposals such as *SPEC Power and Performance Benchmark Methodology*, which are techniques recommended for integrating performance and energy measures in a single benchmark; *SPECpower_ssj2008*, a general-purpose computer server energy-efficiency measure, or power versus utilization; *SPECvirt_sc2013* for consolidation and virtualization; *SPEComp2012* for highly parallel complex computer calculations; *SPECweb2009* for web applications [32]. The Transaction Processing Performance Council (TPC) uses benchmarks with workloads that are specific to database and data management such as: *TPC-Energy*, which focuses on energy benchmarks (e.g. *TPC-C* and *TPC-E* for online transaction processing, *TPC-H* and *TPC-DS* for business intelligence or data warehouse applications, *TPC-VMS* for virtualized environment) [32]. VMWare proposed *VMmark* to measure energy consumption, performance and scalability of virtualization platforms [32]. The Storage Networking Industry Association (SNIA) Emerald Program and the Storage Performance Council (SPC) have contributed to establishing measurements for storage performance and the power consumption associated with the workloads [32]. The *Space Wattage and Performance* metric (SWaP = Performance/ (Space x Power)) [33] incorporates the height in rack units (space), the power consumption (measured during actual benchmark runs or taken from technical documentation), and the performance (measured by industry standard benchmarks such as SPEC).

For cooling and ventilation system efficiency, metrics have also been proposed. *HVAC effectiveness* is measured through the ratio of IT equipment energy consumption to HVAC system energy consumption. *Airflow Efficiency* shows the total fan power needed by unit of airflow (total fan power in Watts / total fan airflow in cfm). *Cooling System Efficiency* is calculated as the ratio of average cooling system power consumption divided by the average data center cooling load [26]. *Cooling System Sizing Factor* shows the ratio between the installed cooling capacity and the peak cooling load. *Air Economizer Utilization Factor* and *Water Economizer Utilization Factor* measure usage at full capacity over a year in percentage terms. *Air temperature* metric measures the difference between the supply and return air temperature in the data center. *Relative humidity* metric measures the difference between the return and supply air humidity in the data center [2]. The *Rack Cooling Index* (RCI) gauges cooling efficiency for IT equipment cabinets compared to the IT equipment intake temperature [34]. The *Return Temperature Index* (RTI) gauges the performance of air management systems [35].

The *Coefficient of Performance of the Ensemble* (COP) gauges the ratio of the total heat load to power consumption of the cooling system [36]. The Cooling Capacity Factor (CCF) estimates the utilization of the cooling capacity, by dividing total rated cooling capacity by the UPS output multiplied by 110% [37]. The *Energy Efficiency Ratio* (EER) is the ratio of the cooling capacity to power input at 95 ºF, and the *Seasonal Energy Efficiency Ratio* (EER) is the ratio of the total heat removed during the annual cooling season by total energy consumed in the same season [38]. The *Sensible Coefficient of Performance* (SCOP) is defined as the ratio of net sensible cooling capacity divided by total power required to produce that cooling (excluding reheat and humidifier) n consistent units [39] [40]. This metric was chosen to computer room air conditioner due to the unique nature and operation of data center facilities.

For electrical equipment different metrics have been proposed [2]. *UPS load factor* gauges the relation between the peak value and the nominal capacity. *UPS system efficiency* shows the relation of the output power and the input power. *Lighting density* shows the lighting power consumption per area.

Data center sustainability metrics have also been introduced. The *Green Energy Coefficient* (GEC) measures the percentage of total energy sourced from alternative energy sources, such as solar, wind, or geothermal plants, and encourages the use of renewable energy [41]. *Carbon Usage Effectiveness* (CUE) measures total CO_2 emissions in relation to IT equipment energy consumption [42]. *Water Usage Effectiveness* (WUE) measures total water usage in relation to IT equipment energy consumption [43]. *Energy Reuse Effectiveness* (ERE) gauges how energy is reused outside of the data center. It is calculated as the total energy minus reuse energy divided by IT equipment energy, or the *Energy Reuse Factor* (ERF) calculated as the reuse energy divided by the total energy [44]. The *Electronic Disposal Efficiency* (EDE) measures how responsible the discarded electronic and electrical equipment are managed, as the ratio of equipment disposed of through known responsible entities and the total of equipment disposed [45].

Other metrics quantify energy consumption related to environmental sustainability. The following categories of metrics are defined: IT strategy, IT hardware asset utilization, IT energy and power efficient hardware deployment, and site physical infrastructure overhead [46]. For those metrics different factors are defined, such as, the *Site-Infrastructure Power Overhead Multiplier* (SI-POM) estimated as the power consumption at the utility meter divided by the total power consumption at the plug of all IT equipment; the *IT Hardware Power Overhead Multiplier* (H-POM), estimated as the ratio of the AC hardware load at the plug and the DC hardware compute load, showing the IT equipment efficiency; the *Deployed Hardware Utilization Ratio* (DH-UR), estimated as the number of servers running live applications divided by the total servers deployed, or as the ratio of terabytes of storage holding data and the total of terabytes of storage deployed; the *Deployed Hardware Utilization Efficiency* (DH-UE) measured as the ratio of minimum number of servers required for peak compute load and the total number of servers deployed, which shows the possibilities of virtualization; and free cooling [46]. The *Free Cooling* metric estimates potential savings using outside air; and the *Energy Save* metric calculates the amount of money, energy, or carbon emission savings that accrue if IT equipment hibernates while it is not in use [46].

After recognizing the need for performance metrics that better capture the efficiency of a given data center, different entities have proposed metrics that measure the functionality of the data center (e.g. amount of computations it performs) and relate that to energy utilization. An example of these are metrics to track useful work produced at a data center compared to power or energy consumed producing it. “Useful work” is defined as the tasks executed in a period of time, each one having a specific weight related to its importance.

Different productivity metrics have also been proposed. Although they are often specific to each user’s activity, these metrics provide a framework for comparison. The *Data Center Performance Efficiency* metric is the ratio of useful work to total facility power [30]. The *Data Center Energy Productivity* (DCeP) metric is defined as the amount of useful work produced, divided by the total data center energy consumed producing it [47]; The *Data Center Compute Efficiency* (DCcE) metric gauges the efficiency of compute resources, intended to find areas of improvement [48]; The *Data Center Storage Productivity* (DCsP) metric expresses the ratio of useful storage system work to energy consumed [49].

Fixed and proportional overhead metrics help analysts and managers understand how energy and cost influences the use of IT equipment. The fixed portion of energy consumption considers use when all IT equipment are unused. The variable part takes into account IT equipment load [50]. The *Data Center Fixed to Variable Energy Ratio* (DC-FVER) metric, defined as the fixed energy divided by the variable energy plus one (1 + fixed energy / variable energy). It shows the inefficiencies through the wasted energy not delivering 'useful work'. The metric reflects the proportion of energy consumption that is variable, considering IT equipment, software and infrastructure [51]. Metrics related to energy-proportional computing, based on computing systems consuming energy in proportion to the work performed have been proposed. The *Idle-to-Peak power Ratio* (IPR) metric is defined as the ratio system's idle consumption with no utilization over the full utilization power consumption [52]. The *Linear Deviation Ratio* (LDR) metric, shows how linear power is, compared to utilization curve [52].

Other metrics have been also proposed. The *Digital Service Efficiency* (DSE) metric shows the productivity and efficiency of the infrastructure through performance, cost, environmental impact, and revenue. Performance is measured by transactions (buy or sell) per energy, per user, per server and per time. Cost is measured by amount of money per energy, per transaction and per server. Environmental impact is estimated in metric tons of carbon dioxide per energy and per transaction. Revenue is estimated per transaction and per user [53]. The *Availability, Capacity and Efficiency* (ACE) performance assessment factors in the availability of IT equipment during failures, the physical capacity available, and how efficient the cool air delivery to IT equipment is [54].

Metrics that combine measurements of efficiency and productivity have been also proposed. The *Corporate Average Datacenter Efficiency* (CADE), estimated as the IT equipment efficiency factored by the facility efficiency (CADE = IT equipment efficiency × facility efficiency = IT equipment asset utilization × IT equipment energy efficiency × Site asset utilization × Site energy utilization) [55]. The *Data Center Energy Efficiency and Productivity* (DC-EEP) index results from multiplying the *IT Productivity per Embedded Watt* (IT-PEW) and the *Site Infrastructure Energy Efficiency ratio* (SI-EER) [56].The IT organization is mainly responsible for the IT-PEW. The SI-EER is measured dividing the power required for the whole data center by the conditioned power delivered to the IT equipment.

The *Datacenter Performance per Energy* (DPPE) considers four sub-metrics: *IT Equipment Utilization* (ITEU), *IT Equipment Energy Efficiency* (ITEE), *Power Usage Effectiveness* (PUE) and *Green Energy Coefficient* (GEC). ITEU is the ratio of total measured power of IT equipment to total rated power of IT equipment. ITEE is the ratio of total rated capacity of IT equipment and total rated energy consumption of IT equipment. PUE is the total energy consumption of the data center divided by the total energy consumption of IT equipment, which promotes energy saving of facilities. GEC is the green energy divided by the total energy consumption of the data center, which promotes the use of green energy. Then the metric is defined as DPPE = ITEU × ITEE × 1/PUE × 1/ (1-GEC) [57].

The *Performance Indicator* (PI) metric was introduced to visualize the data center cooling performance in terms of the balance of the following metrics: thermal conformance, thermal resilience and energy efficiency [58] [59]. IT thermal conformance indicates the proportion of IT equipment operating within recommended inlet air temperatures ranges during normal operation. IT thermal resilience shows if there is any equipment at risk of overheating in case redundant cooling units are not operating due to a failure or planned maintenance. Energy efficiency is measured through the PUE ratio, and it indicates how the facility is operated compared to pre-established energy efficiency ratings.

Different measurements for operational performance have been proposed recently. The methodology *Engineering Operations Ratio* [60] shows the systems operational performance related to its design, as the operational effectiveness for each component (e.g., designed PUE related to actual PUE). The *Data Center Performance Index* [61], takes into account three categories: availability, efficiency and environmental. Availability is measured through the number of incidents or the time of loss of service; performance is gauged through *Power Usage Effectiveness* (PUE) and *Water Usage Effectiveness* (WUE); and environmental is assessed using the Greenhouse Gas (GHG) emissions. Possible indexes are A, B, C, D or not qualifying.

In addition, there have been different proposals to incorporate probability and risk to data center key indicators. The *Class* metric indicates the probability of failure of a data center in the next 12 months, and the associated risk is the Class multiplied by the consequences [62]. It is based on the standard IEEE-3006.7 [63], where *Class* (also called unreliability) is defined as one minus reliability. The *Data Center Risk Index* ranks different countries related to the probability of the factors that affect operations of the facility. Those factors include energy (cost and security), telecommunications (bandwidth), sustainability (alternative energy), water availability, natural disasters, ease of doing business, taxes, political stability, and GDP per capita [64]. The index is mainly designed to contribute on decisions based on the risk profile of the country, although it does not take into account specific business requirements or the fact that some risks can be mitigated.

The authors of this paper have previously proposed a *Data Center Site Risk* metric [65], to help evaluate data center sites and compare them to each other, or to compare different scenarios where the data center operates. The methodology for the *Data Center Site Risk* metric of a specific location is simplified in four steps: the first one is to identify threats and vulnerabilities; the second is to quantify the probability of occurrence of the events; the third one is to estimate potential consequences or impact of the events; and the last one to calculate the total risk level considering weights.

Existing data center metrics reviewed in this paper are listed in Table A.1 (Appendix A), classified by type and main promoter.

Each metric listed uses its own definition for terms such as efficiency, productivity, performance, risk, among others, which must be taken into account when doing comparisons. In addition, there has been academic research in specific data center metrics and risk, such as a modified PUE metric using power demand [66], PUE for a CCHP Natural gas or Biogas Fuelled Architecture [67], PUE for application layers [68], performance metrics for communication systems [69], load dependent energy efficiency metrics [70], workload power efficiency metric [71], airflow and temperature risk [72], power distribution systems risk [73], quality of service and resource provisioning in cloud computing [74][75][76], life cycle cost using a risk analysis model [77], risk for cloud data center overbooking [78], risk management for virtual machines consolidation [79], risk management under smart grid environment [80], and risk for data center operations in deregulated electricity markets [81].

These joint efforts have significantly improved efficiency on the data center infrastructure so that energy consumption has started to flatten out over time [2].

4. Multidimensional Approach for New Family of Data Center Metrics

Existing metrics fail to incorporate important aspects such as the risk involved in processes and operations, for a holistic understanding of the data center behavior. This being the case, comparisons between data center scores with the purpose of evaluating areas of improvement is not an easy task. Furthermore, currently there is no metric that examines performance and risk simultaneously. A data center may have high performance indicators, with a high risk of failure. Research must therefore be refocused to incorporate risks, management and performance. Having this information may work as an early warning system so that mitigation strategies and actions are undertaken on such mission critical facilities.

A new family of metrics can help understand the performance of new and existing data centers, including their associated risk. The proposed novel data center multidimensional scorecard effectively combines performance and risk. Performance is inspected across four different sub-dimensions: productivity, efficiency, sustainability and operations. Risk associated with each of those sub-dimensions is contemplated. External risks are also considered independently of performance, namely site risk.

Figure 3 shows a diagram of the proposed data center multidimensional metric. It is important to highlight that correlation can exist between the different elements of the scorecard; however, for the sake of the explanation this proposal has been simplified by assuming there is no correlation between different performance sub-dimensions. Measurements through different mechanisms are explained below.

4.1 Productivity

Productivity gives a sense of work accomplished and can be estimated through different indicators, such as the ratio of useful work completed to energy usage, or useful work completed to the cost of the data center.

Useful work can be understood as the sum of weighted tasks carried out in a period, such as transactions, amount of information processed, or units of production. The weight of each task is allocated depending on its importance. A normalization factor should be considered to allow the addition of different tasks.

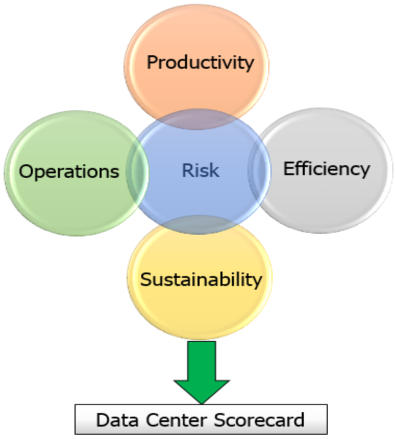


Figure 3: Data center multidimensional metric

Table 1 presents general concepts for these productivity measurements.

Table 1. Productivity measurements

Productivity	Concept
Useful work	- Sum of weighted tasks carried out in a period. - Useful work per energy consumption. - Useful work per physical space. - Useful work to cost of the data center.
Downtime	- Actual downtime, in terms of length, frequency, and recovery time. - A separate measurement within this category will calculate the impact of downtime on productivity, measured as the 'useful work' that was not carried out as well as other indirect tangible and intangible costs due to this failure. To obtain the data a process needs to be established where downtime data (date, time and duration) is sent to this system. To calculate the impact on productivity a scale could be defined based on previous reports of data center outages costs [3].
Quality of service	Quality of service measurements compared to pre-established values. These measurements can include variables in the time domain (e.g., maximum or average waiting time, congestion detection, latency), or scheduling and availability of resources.

Obtaining data to estimate these metrics requires a process for measuring 'useful work', and costs for the specific data center. Cost includes capital expenditures (fixed assets and infrastructure equipment) and operating expenses (energy, human resources, maintenance, insurance, taxes, among other). When including monetary values, and comparing these metrics across time, all future values of money need to be brought to present value so the comparison is consistent. Once processes are in place, calculations of these metrics can be performed in real-time automatically.

4.2 Efficiency

Efficiency has been given substantial attention due to the high energy consumption of the data center sector. Many initiatives have emerged to measure efficiency. Key indicators show how energy efficient site infrastructure, IT equipment, environmental control systems, and other systems are. Power consumption and utilization data can be directly collected from various equipment elements. It can be measured through different energy efficiency measurements briefly explained in Table 2.

Table 2. Efficiency measurements

Efficiency	Concept
Site infrastructure	The ratio of the energy delivered to IT equipment total to total energy used by the data center The value becomes higher if is more efficient. Promotes energy management in facility.
IT equipment utilization	The ratio of total measured power of IT equipment to total rated power of IT equipment. The value becomes higher if is more efficient. The lowest value if all IT equipment are unused. Promotes efficient operation of IT equipment.
IT equipment efficiency	The ratio of the total potential capacity of IT equipment and the total energy consumption of IT equipment. The value becomes higher with more efficient IT equipment. Promotes the procurement of efficient IT equipment, with higher processing capacity per energy.
Physical space utilization	Physical space used divided by total physical space. Energy consumption of all equipment divided by total physical space. It can also be measured in each cabinet (rack unit or area). Promotes efficient planning of physical space.

4.3 Sustainability

Sustainability can be defined as development that addresses current needs without jeopardizing future generations' capabilities to satisfy their own needs [82].

The sustainability of a data center can be measured in different ways, such as calculating the ratio of green energy sources to total energy, estimating the carbon footprint, or the water usage. In addition, an evaluation may be conducted on how environmentally friendly the associated processes, materials, and components are. Table 3 presents general concepts for these sustainability measurements.

Table 3. Sustainability measurements

Sustainability	Concept
Carbon footprint	Each energy source can be assigned a different value of carbon footprint. Use of existing and widely known methods to evaluate the greenhouse gas emissions (e.g., carbon dioxide, methane, nitrous oxide, fluorinated gases). This estimation can be automated.
Green energy sources	The ratio of green energy consumption to total energy consumption. The data can be obtained automatically from real-time measurements.
Water usage	The ratio of total water used to energy consumption of IT equipment. The data can be obtained automatically from real-time measurements.
Environmentally friendly	How environmentally friendly processes, materials and components are. The information is collected by conducting an analysis or audit of processes. It must be updated if a process changes.

4.4 Operations

Operations measurements gauge how well managed a data center is. This must include an analysis of operations, including site infrastructure, IT equipment, maintenance, human resources training, and security systems, among other factors. Audits of systems and processes are necessary to gather the required data. This data should include factors such as documentation, planning, human resources activities and training, status and quality of maintenance, service level agreement, and security. Table 4 shows general concepts for the operations measurements recommended.

Table 4. Operations measurements

Operations	Concept
Documentation	Procedures and policies for data center should be formally documented. All information should be available in digital format.
Planning	Effective planning is desired to reduce downtime. Planning for maintenance, new components, relocations, upgrades, replacements, and life cycle evaluations are needed.
Organization and human resources.	Organizations with an integrated approach are desired, including interactions between different departments and reporting chain. Personnel with the required qualifications and technical training is needed to properly operate the facility.
Maintenance	Preventive, reactive and deferred maintenance programs are required. Consider the equipment manufacturer or vendor recommendations. Predictive maintenance and failure analysis programs should be included. Maintenance management systems are desired to track the status, frequency and quality of the related activities.
Service level agreement	The commitment that prevails with the service provider, including the quality, availability and responsibilities for the service.
Security	Electronic and physical security. Evaluated and assessed against pre-established scales.

Different organizations have recognized the importance to standardize the operation and management of data centers, and have contributed to standards addressing this issue. They can be used as guidelines for quantifying the relatively subjective variables that comprise this sub-dimension. The "Data Center Site Infrastructure Tier Standard: Operational Sustainability" from Uptime Institute [83] includes management and operations. The "Data Centre Operations Standard" from EPI addresses operations and maintenance requirements [84]. BICSI is currently developing the new Data Center Operations standard (BICSI-009) to be used as a reference for operations and maintenance after a data center is built.

4.5 Risk

Data center performance cannot be completely evaluated if the risks that may impact it are not considered. Optimization must involve risk, defined as potential threats that, if materialized, could negatively impact the performance of the data center.

The new family of data center metrics intends not only to measure performance for each process area, but also to associate it with its level of risk. That way, the user may implement actions to achieve the optimum performance and later adjust that performance to a tolerable level of risk, which may again deviate the metrics from their optimum performance. In the long term, if variables remain unchanged, this model will lead to a stable equilibrium.

Risk of these sub-dimensions of performance, as well as the external risk which is independent of performance, are also measured through the use of metrics. They can be described as a causal system, where output depends on present and past inputs. An important strategy to reduce probability of failure is redundancy of resources, but this component may affect performance and costs [4].

4.5.1 Risk associated to performance

Table 5 presents general concepts for estimating the risk of the four identified sub-dimensions of performance.

Table 5. Risk related to performance

Risk	Concept
Productivity risk	Assessed as the downtime probability of occurrence times its impact. The probability is estimated using present and past data of downtime. Impact of downtime on productivity is calculated (e.g., the work that was not carried out with the required quality of service, as well as other indirect tangible and intangible costs associated with a failure). The impact would consider the cost of downtime [3].
Efficiency risk	Estimated with the ratio of processing utilization, IT equipment, physical space utilization, and IT equipment energy utilization, to their respective total capacities. Considers projected growth. When utilization is close to or at capacity, there is no room for growth, which means the risk that future projections will not be met is high. This directly influences performance.
Sustainability risk	Considers historic behavior of the different green energy sources, the percentage composition of each source, and their probability of failure.
Operations risk	Assessed by the operational risk, including documentation, planning, organization and human resources, maintenance, service level of agreement, and security. Analysis of historical data in order to estimate probability of failure due to improper operation in the areas identified, and its impact on performance.

4.5.2 External risk

Performance risks are not the only risks involved in the data center. There are other major risk factors that are external to the actual operation of the data center that must be considered in this analysis, namely site risk.

The authors of this paper have previously proposed a *data center site risk metric* [65] [85], which is a component of the comprehensive family of new data center metrics proposed in this paper. The methodology of the *site risk metric* helps identify potential threats and vulnerabilities that are divided into four main categories: 'utilities', 'natural hazards and environment', 'transportation and adjacent properties', and 'other'. The allocation of weights among each category is based on the significance of the impact of these factors on the data center operation [85].

The methodology quantifies the probability of occurrence of the events according to five pre-established levels of likelihood, and estimates potential consequences of each event using five pre-established levels of impact. It calculates the total risk level associated to the data center location by multiplying the probability of occurrence by the consequences or impact of each threat. That product is then multiplied by the respective weight.

Through this analysis, the different threats and vulnerabilities can be prioritized depending on the values of the probability of occurrence, impact, and the assigned weight. Understanding risk concentration by category can add value when analyzing mitigation strategies. This methodology provides a good sense of what the different risks and potential threats and vulnerabilities are for a data center site.

The *site risk metric* score is summarized into a time dependent function for a specific time instant [85]:

Data Center Site Risk Score =

$$\sum_{i=1}^{tc} \sum_{j=1}^{k_i} RL_{(i,j)} * W2_{(i,j)}$$

Where:

i: Threat categories (i = 1,…,t)

tc: Total number of threat categories (tc = 4)
(i=1: Services, i=2: Natural disasters, i=3: Transportation and adjacent properties, i=4: Other).

j: Specific threat (j=1,… k_i).

k_i: Number of threats for each category
(e.g., k_1=5, k_2=10, k_3=10, k_4=6).

$RL_{(i,j)}$: Risk level for the specific threat.

$W2_{(i,j)}$: Adjusted weight for the specific threat.

The risk level (RL): $RL_{(i,j)} = PO_{(i,j)} * I_{(i,j)}$

Where:

$PO_{(i,j)}$: Probability of occurrence of the specific threat.

$I_{(i,j)}$: Impact of the specific threat.

The adjusted weight (W2) for the threat: $W2_{(i,j)} = W_i * W1_{(i,j)}$

Where:

W_i: Weight of the specific category.

$W1_{(i,j)}$: Weight of the specific subcategory.

The data center site risk metric varies from 1 to 25, where 1 is the lowest level and 25 is the highest level of risk achievable for a specific site. The user must determine an acceptable level of risk, so that if the final score lies above that level, that particular location is not recommended.

4.6 New family of Data Center Metrics

The proposed new family of data center metrics can be summarized into a time dependent function, to be further developed. For a specific time instant, regardless of the correlation between different parameters, the data center score can be defined as a function of different metrics, risks and weights:

Data Center Score =

= f (P_1, P_2, P_3, P_4, R_1, R_2, R_3, R_4, R_E, W_1, W_2, W_3, W_4)

= f (P_i, R_i, R_E, W_i) with i=1,…,4

Where: P_i: Performance.

R_i: Risk of process.

R_E: External risk.

W_i: Weight of each category.

And sub-indexes: 1: Productivity.

2: Efficiency.

3: Sustainability.

4: Operations.

To the best possible extent, each value and weight assigned to each key indicator must be backed with enough support such as research or facts that lead to such conclusions. The outcome is a scorecard that assists in finding areas of improvement, which should be strategically addressed. The quality of information used before assigning each value is very important. Equally weighted data center scores are desired, since a data center must be productive, efficient, well managed and sustainable.

The new family of data center metrics scorecard can be taken as a decision-making trigger. It involves both technical and non-technical aspects, as failures and risks may not only be due to technical issues but also to non-technical ones such as human error. The metrics should measure parameters and processes. Given some premises, a data center may be ideal at a certain point in time, but when conditions change, that same data center may not be optimal. Depending on the final score calculated with the value of the dimensions, the scorecard would rank each data center on a scale to be defined, which allows for tangible comparisons between different data centers, or 'before and after' on the same data center. The multidimensional metric can also be transformed through different operators as a composite metric with just one value. Furthermore, the proposed metric has the possibility to incorporate new performance and risk measurements in the future, preventing it from becoming outdated.

4.7 Visualization tool

To confirm cross-comparability all the different indicators can be normalized. Each key indicator for performance can be presented in a scale interpreted in such a way that a higher value implies a more positive outcome, so minimum and maximum values correspond to the worst and best possible expected outcomes. Conversely, each risk value can be presented in a scale interpreted as a higher value implies a higher level of risk, and a more undesirable scenario. Figure 4 shows an example with indicators selected arbitrarily for illustrative purposes.

The graph (Figure 4) shows that the data center has a high productivity level, followed by efficiency, but its operations and sustainability indicators show greater room for improvement. Levels of risk associated to each category and site risk are low. Risk tolerance depends on the user, but working with this example, if we hypothetically set a maximum tolerable risk of 25%, actions would need to be implemented to reduce the risk related to productivity.

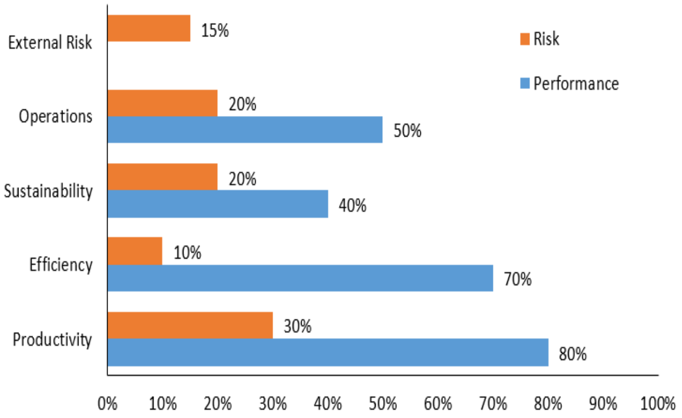


Figure 4: Data center multidimensional metric

Spider graphs can be generated to allow straightforward visual comparisons and trade-off analysis between different scenarios. This is very helpful when simulating or forecasting different strategies. It also enables clear reporting to stakeholders. Figure 5 shows an example with two different data centers, at different times that can involve specific strategies implemented.

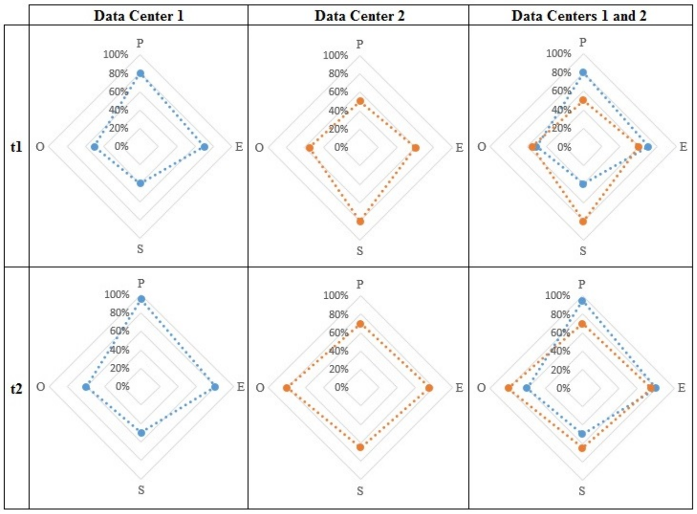


Figure 5: Data center multidimensional performance metric comparison (P: Productivity, E: Efficiency, O: Operations, S: Sustainability)

Edges of diamonds show measurement of the four dimensions of performance: productivity (P), efficiency (E), sustainability (S) and operations (O). The larger the diamond, the better the performance. It can be intuitively seen that in the scenario of time 1 (t1), data center 1 is more productive and efficient, but less sustainable and not so well operated, compared to data center 2. Over time (from t1 to t2), data center 1 has improved all its performance components, but data center 2 has worsened the sustainability indicator and improved all other values. Risk can be analyzed in a similar spider graph, but for simplicity, it was not included in Figure 5.

4.8 Automation

Equipment needs permanent monitoring and maintenance to assure proper and efficient performance. Measurements should be automatic when possible as the automated metrics receive information directly from the different systems that quantify parameters. Real-time collection of relevant data is required for reliable metrics. Obtaining it is not a trivial task, especially in existing data centers that lack adequate instrumentation to collect the data [22], or if the related process cannot be easily automated. This underscores the need for new approaches to data center monitoring and management systems [86] [87]. Gathering the right data and understanding its nature is more important than simply collecting more data [88].

Real-time data can be gathered and updated automatically through a monitoring and management system [87] for parameters such as power consumption, temperature, humidity, air flow, differential air pressure, closure, motion, vibration, and IT equipment resource utilization. Improvements by some equipment manufacturers include the ability to directly access measurements of power, temperature, airflow, and resource utilization for each device. These measurements may include parameters such as the air inlet temperature, airflow, outlet temperature, power utilization, CPU utilization, memory utilization, and I/O utilization. Platform level telemetry can transform data center infrastructure management, allowing direct data access from the equipment processor [89]. Other parameters, such as some aspects of sustainability, operations and external risks, are more difficult to automate, since they require audits, human observation, evaluation, analysis, and the need to be periodically updated. All required data must be entered, automatically or manually, into a system in order to estimate all metrics and to visualize results clearly for decision-makers to undertake adequate actions.

5. Conclusions

The novel family of data center metrics as described in the paper provides a comprehensive view of the data center, using a multidimensional approach to combine performance (including productivity, efficiency, sustainability and operations) and risks (associated to performance and site risk). Using this approach, areas of improvement around which to create a strategy can be detected.

Given the mission critical nature of data centers, metrics must provide a holistic, yet quantitative, understanding of the data center behavior, in order to improve the utilization of all the resources involved. When issues identified by the metrics are addressed, processes can be optimized or moved closer to their desired point based on the vision for the data center.

Actions undertaken will impact the metric results in real time. When variables are re-measured, the result of the metrics should improve. As a result, new strategies may lead to modification of the overall metric as related to performance and risk values. This does not mean attempting to achieve the optimal performance. It is different from existing metrics in that it does not limit itself to measuring a value whose optimization can be automated, instead, this is only part of its scope. It aggregates measurements from multiple aspects to provide a global and objective vision to decide the extent to which optimization is desired. To recognize trends or predict future behavior, tools such as predictive analysis and machine learning are required. Machine learning considers algorithms that can learn from and make predictions on data.

The new family of data center metrics may contribute to standardize a process that eventually becomes a best practice. It may help to assess data centers, to compare them to each other, to make comparisons between different scenarios, and to provide a ranking of how the data center behaves. The outcome is a scorecard that will constitute a strong basis for decision-making.

6. Future Research

Further research needs to be conducted to assess and validate the new family of data center metrics proposed, based on a multidimensional approach and including performance and risk, in order to understand which parameters are most reasonable to use for each specific purpose.

Studying the correlation between the different metric scores and their associated parameters and risks, simulations of different scenarios can help to visualize how a change in parameters impacts risk, and likewise, how a change in risk factors affects metric results. Since there are currently no solutions available that include all proposed factors, new dynamic models and simulation tools are needed to validate and calibrate the metric. This would assist in implementing numerous data center strategies to improve metrics, for which a theoretical approach must be conducted. By tracking the proposed metric, the data center stakeholders will better understand data center performance and risk in a multidimensional view.

References

[1] M. Levy and D. Raviv, “A Novel Framework for Data Center Metrics using a Multidimensional Approach,” *15th LACCEI International Multi-Conference for Engineering, Education, and Technology: Global Partnerships for Development and Engineering Education*. Boca Raton, FL, 2017.

[2] Ernest Orlando Lawrence Berkeley National Laboratory, “United States Data Center Energy Usage Report. LBNL-1005775,” 2016.

[3] Ponemon Institute LLC, “Cost of Data Center Outages. Data Center Performance Benchmark Series,” 2016.

[4] L. Minas and B. Ellison, *Energy Efficiency for Information Technology: How to Reduce Power Consumption in Servers and Data Centers*. Intel Press, 2009.

[5] “Data Center Design and Implementation Best Practices,” ANSI/BICSI 002, 2014.

[6] S. Greenberg, E. Mills, and B. Tschudi, “Best Practices for Data Centers : Lessons Learned from Benchmarking 22 Data Centers,” *2006 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA, pp. 76–87, 2006.

[7] D. Bouley, “Estimating a Data Center’s Electrical Carbon Footprint. White paper #66,” *APC by Schneider Electric*. 2012.

[8] U.S. Department of Energy. Office of Energy Efficiency & Renewable Energy, “10 Facts to Know About Data Centers,” 2014. [Online]. Available: http://energy.gov/eere/articles/10-facts-know-about-data-centers.

[9] I. Ahmad and S. Ranka, “Chapter 43: Overview of Data Centers Energy Efficiency Evolution,” in *Handbook of Energy-Aware and Green Computing*, vol. 2, CRC Press, 2012, pp. 983–1027.

[10] The Climate Group - Global eSustainability Initiative (GeSI), "SMART 2020: Enabling the low carbon economy in the information age," 2008.

[11] J. A. Laitner and K. Ehrhardt-Martinez, "Information and Communication Technologies: The Power of Productivity," American Council for an Energy-Efficient Economy, Washington, DC, 2008.

[12] ASHRAE Technical Committee 9.9, "2011 Thermal Guidelines for Data Processing Environments – Expanded Data Center Classes and Usage Guidance," 2011.

[13] ASHRAE Technical Committee 9.9, "Data Center Networking Equipment – Issues and Best Practices," 2012.

[14] ASHRAE Technical Committee 9.9, "Data Center Storage Equipment – Thermal Guidelines , Issues , and Best Practices," pp. 1–85, 2015.

[15] "Energy Standard for Data Centers," ANSI/ASHRAE 90.4, 2016.

[16] "Green data centres – Part 1: Energy and environmental management systems," Singapore Standards SS 564, 2013.

[17] "Telecommunications Infrastructure Standard for Data Centers," ANSI/TIA-942-B, 2017.

[18] P. Turner, J. H. Seader, V. Renaud, and K. G. Brill, "Tier Classifications Define Site Infrastructure Performance," Uptime Institute, 2006.

[19] "Data Center Site Infrastructure Tier Standard: Topology," Uptime Institute, 2018.

[20] B. Howard, "H.R. 2126: The Energy Efficiency Improvement Act," 2014. [Online]. Available: http://www.usgbc.org/articles/hr-2126-energy-efficiency-improvement-act.

[21] T. Scott, "Data Center Optimization Initiative (DCOI). Memorandum," 2016.

[22] The Green Grid, "PUE^{TM}: A Comprehensive Examination of the Metric. White paper # 49," 2012.

[23] The Green Grid, "Green Grid Metrics: Describing Data Center Power Efficiency. Technical Committee White Paper," 2007.

[24] The Green Grid, "Green Grid Data Center Power Efficiency Metrics: PUE and DCIE. White paper # 6," 2008.

[25] U.S. Environmental Protection Agency, "Energy Star for Data Centers Scheduled Portfolio Manager Release on June 7 , 2010," 2010.

[26] W. Lintner, B. Tschudi, and O. VanGeet, "Best Practices Guide for Energy-Efficient Data Center Design," U.S. Department of Energy, 2011.

[27] N. Horner and I. Azevedo, "Power usage effectiveness in data centers: overloaded and underachieving," *The Electricity Journal*, vol. 29, no. 4. Elsevier Inc., pp. 61–69, 2016.

[28] V. Avelar, D. Azevedo, and A. French, *PUE^{TM}: A Comprehensive Examination of the Metric*. Atlanta, GA: ASHRAE Datacom Series, 2012.

[29] M. Andrea, "Data Center Standards. Data Center Size and Density. White paper DCISE-001." The Strategic Directions Group Pty Ltd, 2014.

[30] P. P. Ray, "The green grid saga - a green initiative to data centers: a review," *Indian J. Comput. Sci. Eng.*, vol. 1, no. 4, pp. 333–339, 2010.

[31] S. Sharma, C. H. Hsu, and W. C. Feng, "Making a Case for a Green500 List," *20th IEEE International Parallel & Distributed Processing Symposium (IPDPS) Workshop on High-Performance, Power-Aware Computing (HP-PAC)*. Rhodes Island, Greece, 2006.

[32] K.-D. Lange and K. Huppler, *Server Efficiency: Metrics for Computer Servers and Storage*. Atlanta, GA: ASHRAE Datacom Series, 2015.

[33] Sun Microsystems Inc, "Sun Sparc ® Enterprise T5120 and T5220 Server Architecture. White Paper," 2007.

[34] M. K. Herrlin, "Rack Cooling Effectiveness in Data Centers and Telecom Central Offices: The Rack Cooling Index (RCI)," *ASHRAE Trans.*, vol. 111, no. 2, pp. 725–731, 2005.

[35] M. K. Herrlin, "Airflow and Cooling Performance of Data Centers: Two Performance Metrics," *ASHRAE Trans.*, vol. 114, no. 2, pp. 182–187, 2008.

[36] C. D. Patel, R. K. Sharma, C. E. Bash, and M. Beitelmal, "Energy Flow in the Information Technology Stack : Coefficient of Performance of the Ensemble and its Impact on the Total Cost of Ownership." Hewlett-Packard Development Company, 2006.

[37] K. G. Brill and L. Strong, "Cooling Capacity Factor (CCF) Reveals Stranded Capacity and Data Center Cost Savings. White paper." Upsite Technologies, Inc., 2013.

[38] "2017 Standard for Performance Rating of Unitary Air-conditioning & Air-source Heat Pump Equipment," Air Conditioning Heating and Refrigeration Institute AHRI 210/240, 2017.

[39] "2016 Standard for Performance Rating of Computer and Data Processing Room Air Conditioners," Air Conditioning Heating and Refrigeration Institute AHRI 1360, 2016.

[40] "Method of Testing for Rating Computer and Data Processing Room Unitary Air Conditioners," ANSI/ASHRAE Standard 127-2012, 2012.

[41] The Green Grid et al., "Harmonzing Global Metrics for Data Centers Energy Efficiency," 2012.

[42] The Green Grid, "Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric. White paper # 32," 2010.

[43] The Green Grid, "Water Usage Effectiveness (WUE): A Green Grid Data Center Sustainability Metric. White paper # 35," 2011.

[44] The Green Grid, "ERE: A Metric for Measuring the Benefit of Reuse Energy From a Data Center. White paper # 29," 2010.

[45] The Green Grid, "Electronics disposal efficiency (EDE): an IT recycling metric for enterprises and data centers. White paper #53," 2012.

[46] J. R. Stanley, K. G. Brill, and J. Koomey, "Four Metrics Define Data Center Greenness," Uptime Institute, 2007.

[47] The Green Grid, "A framework for data center energy productivity. White paper # 13," 2008.

[48] The Green Grid, "The green grid data center compute efficiency metric: DCcE. White paper # 34," 2010.

[49] The Green Grid, "The Green Grid Data Center Storage Productivity Metrics (DCsP): Application of Storage System Productivity Operational Metrics. White paper #58," 2014.

[50] L. Newcombe, "Data centre energy efficiency metrics: Existing and proposed metrics to provide effective understanding and reporting of data centre energy. White Paper," BCS Data Centre Specialist Group, 2010.

[51] L. Newcombe, Z. Limbuwala, P. Latham, and V. Smith, "Data centre Fixed to Variable Energy Ratio metric DC - FVER: An alternative to useful work metrics which focuses operators on eliminating fixed energy consumption. White Paper," BCS Data Centre Specialist Group, 2012.

[52] I. Ahmad and S. Ranka, "Chapter 44: Evaluating Performance, Power, and Cooling in High-Performance Computing (HPC) Data Centers," in *Handbook of Energy-Aware and Green Computing*, vol. 2, CRC Press, 2012, pp. 1029–1049.

[53] eBay Inc, "Digital Service Efficiency - Solution Brief. White paper," San Jose, CA, 2013.

[54] D. King and S. Davies, "ACE Performance Assessment - Empowering the Data Center Operator. White paper," Future Facilities, 2014.

[55] J. Kaplan, W. Forrest, and N. Kindler, "Revolutionizing data center energy efficiency." McKinsey & Company, pp. 1–13, 2008.

[56] K. G. Brill, "Data Center Energy Efficiency and Productivity," Uptime Institute, 2007.

[57] Green IT Promotion Council, "Concept of New Metrics for Data Center Energy Efficiency - Introduction of Datacenter Performance per Energy [DPPE]," Japan, 2010.

[58] The Green Grid, "New Metric To Assess and Visualize Data Center Cooling," 2016. [Online]. Available: http://www.thegreengrid.org/~/media/Press Releases/The Green Grid_Press Release_Performance Indicator New metric_vFINAL.pdf.

[59] The Green Grid, "The Performance Indicator Assessing and Visualizing Data Center Cooling Performance. White Paper," 2016.

[60] Infrastructure Masons, "Engineering operations ratio. White paper." 2016.

[61] Infrastructure Masons, "Data Center Performance Index. White paper." 2017.

[62] R. Miller, "Class: New data center metric targets probability and risk," 2015. [Online]. Available: http://datacenterfrontier.com/class-new-data-center-metric-targets-probability-of-failure/.

[63] "IEEE Recommended Practice for Determining the Reliability of 7x24 Continuous Power Systems in Industrial and Commercial Facilities," IEEE Std 3006.7, 2013.

[64] Cushman & Wakefield, "Data Centre Risk Index," 2016.

[65] M. Levy and D. Raviv, "A framework for data center site risk metric," *IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. New York City, NY, pp. 9–15, 2017.

[66] E. Jaureguialzo, "PUE: The Green Grid metric for evaluating the energy efficiency in DC (Data Center). Measurement method using the power demand," *2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC)*. Amsterdam, 2011.

[67] F. De Angelis and U. Grasselli, "The next generation green data center a modified power usage effectiveness metric proposal for a CCHP natural gas or biogas fuelled architecture," *2015 IEEE 15th International Conference on Environment and Electrical Engineering, (EEEIC)*. Rome, pp. 907–911, 2015.

[68] R. Zhou, Y. Shi, and C. Zhu, "AxPUE: Application level metrics for power usage effectiveness in data centers," *2013 IEEE International Conference on Big Data*. Silicon Valley, CA, pp. 110–117, 2013.

[69] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Y. Zomaya, "Performance Metrics for Data Center Communication Systems," *2015 IEEE 8th*

International Conference on Cloud Computing. New York, NY, pp. 98–105, 2015.

[70] D. Schlitt and W. Nebel, "Load dependent data center energy efficiency metric based on component models," *2012 International Conference on Energy Aware Computing*. Guzelyurt, Cyprus, 2012.

[71] T. Wilde *et al.*, "DWPE, a new data center energy-efficiency metric bridging the gap between infrastructure and workload," *2014 International Conference on High Performance Computing & Simulation (HPCS)*. Bologna, pp. 893–901, 2014.

[72] M. Seymour and S. Ikemoto, "Design and management of data center effectiveness, risks and costs," *Annual IEEE Semiconductor Thermal Measurement and Management Symposium*. San Jose, CA, pp. 64–68, 2012.

[73] M. Wiboonrat, "Risk anatomy of data center power distribution systems," *2008 IEEE International Conference on Sustainable Energy Technologies, ICSET 2008*. Singapore, Singapore, pp. 674–679, 2008.

[74] J. Liu, Y. Zhang, Y. Zhou, D. Zhang, and H. Liu, "Aggressive resource provisioning for ensuring QoS in virtualized environments," *IEEE Transactions on Cloud Computing*, vol. 3, no. 2. pp. 119–131, 2015.

[75] S. Parida, S. C. Nayak, and C. Tripathy, "Truthful Resource Allocation Detection Mechanism for Cloud Computing," *The Third International Symposium on Women in Computing and Informatics (WCI '15)*. New York, NY, pp. 487–491, 2015.

[76] S. C. Nayak and C. Tripathy, "Deadline sensitive lease scheduling in cloud computing environment using AHP," *J. King Saud Univ. - Comput. Inf. Sci.*, 2016.

[77] M. Wiboonrat, "Life cycle cost analysis of data center project," *2014 Ninth International Conference on Ecological Vehicles and Renewable Energies (EVER)*. Monte-Carlo, Monaco, 2014.

[78] L. Tomas and J. Tordsson, "An autonomic approach to risk-aware data center overbooking," in *IEEE Transactions on Cloud Computing*, vol. 2, no. 3, 2014, pp. 292–305.

[79] X. Jin, F. Zhang, S. Hu, and Z. Liu, "Risk management for virtual machines consolidation in data centers," *2013 IEEE Global Communications Conference (GLOBECOM)*. Atlanta, GA, pp. 2872–2878, 2013.

[80] L. Yu, T. Jiang, Y. Cao, S. Yang, and Z. Wang, "Risk management in internet data center operations under smart grid environment," *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*. Tainan, Taiwan, pp. 384–388, 2012.

[81] L. Yu, T. Jiang, S. Member, Y. Cao, and S. Member, "Risk-constrained operation for internet data centers in deregulated electricity markets," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 5, 2014, pp. 1306–1316.

[82] United Nations, "Report of the World Commission on Environment and Development: Our Common Future," 1987.

[83] "Data Center Site Infrastructure Tier Standard: Operational Sustainability," Uptime Institute, 2014.

[84] "DCOS - Data Centre Operations Standard," EPI, 2016.

[85] M. Levy, "Análisis de Riesgos y Selección de Localidad: Dos Pilares Imprescindibles [Risk analysis and site selection: two essential pillars]," *Datacenter Dynamics Focus*, pp. 32–34, 2014.

[86] M. Levy and J. O. Hallstrom, "A New Approach to Data Center Infrastructure Monitoring and Management (DCIMM)," *IEEE CCWC 2017. The 7th IEEE Annual Computing and Communication Workshop and Conference*. Las Vegas, NV, 2017.

[87] M. Levy and J. O. Hallstrom, "A Reliable, Non-Invasive Approach to Data Center Monitoring and Management," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3. pp. 1577–1584, 2017.

[88] J. Koomey and J. Stanley, "The Science of Measurement: Improving Data Center Performance with Continuous Monitoring and Measurement of Site Infrastructure," 2009.

[89] J. L. Vincent and J. Kuzma, "Using Platform Level Telemetry to Reduce Power Consumption in a Datacenter," *2015 31st Thermal Measurement, Modeling & Management Symposium (SEMI-THERM)*. Folsom, CA, 2015.

Appendix A

The following table presents a summary of the reviewed existing data center metrics, classified by type with the main promoter. In some cases, the main promoter could not be identified. Must be noted that each metric uses its own definition for some terms (e.g., efficiency, productivity, performance, and risk), and for comparison purposes the same metric must be used.

Table A.1. Existing data center metrics

Metric	Type	Promoter
Power Usage Effectiveness (PUE) Data Center infrastructure Efficiency (DCiE)	Efficiency. Energy.	The Green Grid
Data Center Size	Efficiency. Space.	Data Center Institute Standards endorsed (AFCOM)
Rack Density	Efficiency. Rack. Space and energy.	Data Center Institute Standards endorsed (AFCOM)
Data Center Density (DCD)	Efficiency. Space and energy.	The Green Grid
Fixed and proportional overhead	Efficiency. Energy.	BCS Data Centre Specialist Group
HVAC effectiveness Airflow efficiency Cooling system efficiency Cooling System Sizing Factor Air Economizer Utilization Factor Water Economizer Utilization Factor Air temperature Relative humidity Rack Cooling Index (RCI) Return Temperature Index (RTI). Sensible Coefficient of Performance (SCOP)	Efficiency. Cooling system.	American Society of Heating, Refrigerating and Air-Conditioning (ASHRAE)
Coefficient of Performance of the Ensemble (COP)	Efficiency. Cooling system.	Hewlett-Packard
Cooling Capacity Factor (CCF)	Efficiency. Cooling system.	Upsite Technologies Inc.

Energy Efficiency Ratio (EER) Seasonal Energy Efficiency Ratio (EER)	Efficiency. Cooling system.	Air-Conditioning Heating and Refrigeration Institute
UPS load factor UPS system efficiency Lighting density.	Efficiency. Electrical equipment.	---
Compute Power Efficiency (CPE)	Efficiency. IT equipment (server).	The Green Grid
SPECpower_ssj2008	Efficiency. IT equipment (server).	Standard Performance Evaluation Corporation
SPECvirt_sc2013	Efficiency. IT equipment (server). Consolidation and virtualization.	Standard Performance Evaluation Corporation
SPEComp2012	Efficiency. IT equipment (server). Highly parallel complex computer calculations.	Standard Performance Evaluation Corporation
SPECweb2009	Efficiency. IT equipment (server). Web applications.	Standard Performance Evaluation Corporation
FLOPS per Watt	Efficiency. IT equipment (server)/	TheGreen500
TPC-Energy: TPC-C and TPC-E	Efficiency. IT equipment (server). Online transaction processing	Transaction Processing Performance Council
TPC-Energy: TPC-H and TPC-DS	Efficiency. IT equipment (server). Business intelligence or data warehouse applications	Transaction Processing Performance Council
TPC-Energy: TPC-VMS	Efficiency. IT equipment (server). Virtualized environment	Transaction Processing Performance Council
Vmmark	Efficiency. IT equipment (server). Virtualization platforms	VMWare
Storage performance	Efficiency. IT equipment (storage).	Storage Networking Industry Association, Emerald Program and the Storage Performance Council
Space Wattage and Performance (SWaP)	Efficiency. IT equipment (server).	Sun Microsystems
Idle-to-peak power ratio (IPR)	Efficiency. IT equipment (server). Idle consumption.	---
Linear deviation ratio (LDR)	Efficiency. IT equipment (server). Power linearity.	---
Green Energy Coefficient (GEC)	Sustainability. Alternative energy.	The Green Grid
Carbon Usage Effectiveness (CUE)	Sustainability. Carbon emissions.	The Green Grid
Water Usage Effectiveness (WUE)	Sustainability. Water usage.	The Green Grid
Energy Reuse Effectiveness (ERE) Energy Reuse Factor (ERF)	Sustainability. Energy reuse.	The Green Grid
Electronic Disposal Efficiency (EDE)	Sustainability. Decommissioned IT equipment.	The Green Grid
Site-Infrastructure Power Overhead Multiplier (SI-POM)	Sustainability. Site physical infrastructure overhead.	Uptime Institute
IT Hardware Power Overhead Multiplier (H-POM)	Sustainability. IT equipment efficiency.	Uptime Institute
Deployed Hardware Utilization Ratio (DH-UR)	Sustainability. IT equipment efficiency.	Uptime Institute
Deployed Hardware Utilization Efficiency (DH-UE)	Sustainability. IT equipment efficiency.	Uptime Institute
Free Cooling	Sustainability. Free cooling.	Uptime Institute
Energy Save	Sustainability. IT equipment hibernate.	Uptime Institute
Datacenter Performance per Energy (DPPE)	Performance. Sustainability. IT Equipment.	The Green IT Promotion Council (Japan)
Data Center Performance Efficiency (DCPE)	Performance. Productivity.	The Green Grid
Data Center energy Productivity (DCeP)	Performance. Productivity.	The Green Grid
Data Center compute Efficiency (DCcE)	Performance. Compute resources.	The Green Grid
Data Center Storage Productivity (DCsP)	Performance. Storage systems.	The Green Grid
Data Center Fixed to Variable Energy Ratio (DC-FVER)	Performance. Productivity. Wasted energy.	BCS Data Centre Specialist Group
Digital Service Efficiency (DSE)	Performance. Sustainability. Cost.	Ebay
Availability, Capacity and Efficiency (ACE)	Performance. Efficiency. Cooling system.	Future Facilities
Corporate Average Datacenter Efficiency (CADE)	Performance. Efficiency.	Uptime Institute
Data Center Energy Efficiency and Productivity (DC-EEP)	Performance. Efficiency. Productivity	Uptime Institute
Performance Indicator (PI)	Performance. Cooling system.	The Green Grid
Engineering Operational Ratio	Performance. Cooling system.	Infrastructure Masons

Data Center Performance Index	Performance. Availability, Efficiency. Environmental.	Infrastructure Masons
Class metric	Risk. Probability of failure.	MTechnology
Data Center Risk Index	Risk. Operations.	Cushman & Wakefield
Data Center Site Risk Metric	Risk. Site location. Operations.	M. Levy

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 252-256 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Frameworks for Performing on Cloud Automated Software Testing Using Swarm Intelligence Algorithm: Brief Survey

Mohammad Hossain*, Sameer Abufardeh, Sumeet Kumar

Department of Math Science and Technology, University of Minnesota Crookston, Crookston, MN 56716, USA.

ARTICLE INFO

Article history:
Received: 01 March, 2018
Accepted: 16 March, 2018
Online: 11 April, 2018

Keywords:
Swarm Intelligence
Ant Colony Algorithm
Bee Colony Optimization
Cloud Based Testing
Test Case Prioritization

ABSTRACT

This paper surveys on Cloud Based Automated Testing Software that is able to perform Black-box testing, White-box testing, as well as Unit and Integration Testing as a whole. In this paper, we discuss few of the available automated software testing frameworks on the cloud. These frameworks are found to be more efficient and cost effective because they execute test suites over a distributed cloud infrastructure. One of the framework effectiveness was attributed to having a module that accepts manual test cases from users and it prioritize them accordingly. Software testing, in general, accounts for as much as 50% of the total efforts of the software development project. To lessen the efforts, one the frameworks discussed in this paper used swarm intelligence algorithms. It uses the Ant Colony Algorithm for complete path coverage to minimize time and the Bee Colony Optimization (BCO) for regression testing to ensure backward compatibility.

1. Introduction

With the increase of code complexity in modern software, chances of errors are exponentially increasing. These errors can cause loss of money and innocent human lives [1]. For example, in April 24 1994, a China airline airbus A-300 crashed, due to a software bug, resulting the death of 264 innocent lives [2, 3]. Another famous incident was reported in April 1999, where a small software bug in a military satellite was behind a $1.2 billion loss: one of the costliest unmanned accidents in the history of Cape Canaveral launches [4, 5].

Therefore, to reduce the risks of errors, researchers have developed a variety of testing techniques to find and fix software bug early and before the deployment of the software. One of the most critical tests is unit testing, where each module of program is tested separately. Another critical test is integration testing that occurs after unit testing, where individual software modules are combined and tested as a group. Since unit testing requires access to the system code, it is done during the initial stages of a program, detecting an estimated 65% of the errors [6, 7, 8]. Other types of tests includes, system testing and acceptance testing. In system testing, the system is tested as a whole to verify that it meets the specified requirements. After that, acceptance testing is done to verify that, the system meets the client/user requirements.

Testing is generally a lengthy and costly process. Therefore, automated software testing generally intended to reduce the time and the cost of testing. It also can increase the depth and scope of tests to help improve software quality. However, most automated testing tools fails to provide efficient results, because they only focus on specific testing techniques [9, 10] and in sometimes they may be unsuitable for large-scale software.

This survey paper focuses on testing based on Cloud platforms tools in order to develop cost effective, efficient and time saving tools that follows the rules of research based techniques to produce software free of or with few errors or bugs.

2. Testing

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to identify and understand the risks of software implementation. Testing techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs (errors or other defects). It is the process of validating and verifying a software program, application or product [11]. Software testing is a huge domain, but it can be broadly categorized into two areas: manual testing and automated testing:

*Corresponding Author: Mohammad Hossain, Crookston, MN 56716, 218-281-8222, Email: hossain@crk.umn.edu

https://dx.doi.org/10.25046/aj030229

2.1. Manual Testing

An individual or a group of individuals performing all of the software quality assurance testing, checking for errors and defects, is knows as manual testing.

2.2. Automated Testing

The main purpose of this testing is to replace manual testing with automated cloud based testing without loss of efficiency, in such a way that it can not only save time but also produce high quality software.

3. Terminologies and Abbreviations

The following abbreviations and terminologies were used in this research paper.

3.1. Regression Testing

Software undergoes constant changes. Such changes are necessitated because of defects to be fixed, enhancements to be made to existing functionality, or new functionality to be added. Anytime such changes are made, it is important to ensure that, first changes or additions work as designed. Second changes or additions are something that is already working and should continue to work. Regression testing is carried out to ensure that any new feature introduced to the existing product does not adversely affect the current [11].

3.2. Traceability

Traceability is defined as the ability to describe and follow the life of a requirement, in both forward and backward direction, throughout the software life cycle. Traceability relations can assist with several activities of the software development process such as evolution of software systems, compliance verification of code, reuse of parts of the system, requirement validation, understanding of the rationale for certain design decisions, identification of common aspects of the system, and system change and impact [12,13].

3.3. White-Box Testing (WBT)

White-Box Testing, also known as clear box testing, gives verification engineers full access to the source code and the internal structure of the software. It is the detailed investigation of internal logic and structure of the code [14]. In WBT, it is necessary for a tester to have full knowledge of source code. Some important types of WBT techniques includes Statement Coverage (SC), where tester tests every single line of code, and Condition Coverage (CC) in which all the conditions of the code are checked by providing true and false values to the conditional statements in the code.

3.4. Black-Box Testing (BBT)

BBT treats the software as a “Black Box” without any knowledge of internal working of the system and it only examines the fundamental aspects of the system. In BBT the tester has knowledge of the system architecture but he/she does not have access to the source code [15].

4. Swarm Intelligence Algorithms to Optimize Regression Testing.

Regression Testing ensures that any changes or enhancement made to the system will not adversely affect the functionality of software. The execution of all test cases can be an costly and time consuming process. With this in hand, prioritization of test cases can help in reduction in cost of regression testing. Swarm intelligence is an emerging area in the field of optimization and researchers have developed various algorithms by modeling the behaviors of different swarm of animals and insects such as ants, termites, bees, birds, fishes [16]. These algorithms are being used to reduce the time and cost of testing in general and more specifically regression testing [11, 17]. Two such widely used algorithms are Ant Colony Algorithm and Bee Colony Optimization.

4.1. Ant Colony Algorithm (ACA)

Ant colony algorithms are based on the behavior of a colony of ants when looking for food. In their search they mark the trails they are using by laying a substance called pheromone. The amount of pheromone in a path tells other ants if it is a promising path or not. This observation inspired Colorni, Dorigo and Maniezo [18] for proposing a metaheuristics technique: ants are procedures that build solutions to an optimization problem. Based on how the solution space is being explored, some values are recorded in a similar way as pheromone acts, and objective values of solutions are associated with food sources. An important aspect of this algorithm is parallelism: several solutions are built at the same time and they interchange information during the procedure and use information of previous iterations [19]. In [20] Li and Lam proposed to use UML State chart diagrams and ACA for test data generation. The advantages of their proposed approach are that this approach directly uses the standard UML artifacts created in software design processes and it also automatically generated feasible test sequence, non-redundant, and it achieves the stated coverage criteria.

4.2. Bee Colony Optimization (BCO)

Bee swarm behavior in nature is characterized by autonomy and distributed functioning, and it is self-organizing. Recently, researchers started studying the behavior of social insects in an attempt to use the Swarm Intelligence concept in order to develop various Artificial Systems [21]. In software engineering, BCO can be used as a method of regression testing and traceability. Its purpose is to verify that the current version of the software is compatible with pervious test results and the data is comparable to previous versions of the software.

Karnavel and Santhoshkumar proposed a fault coverage regression system exploiting the BCO algorithm discussed in [11]. The idea is based on the natural bee colony with two types of worker bees that are responsible for the development and maintenance of the colony: scout bees and forager bee. The BCO algorithm developed for the fault coverage regression test suite is based on the behavior of these two bees. The algorithm has been formulated for fault coverage to attain maximum fault coverage in minimal units of execution time of each test case. Two examples were used whose results are comparable to the optimal solution [22, 23]. The system was divided into the following components: *Testing Phase, Traceability Phase, Exploration Test, Generating Reports, and Storing in Database* (figure 1):

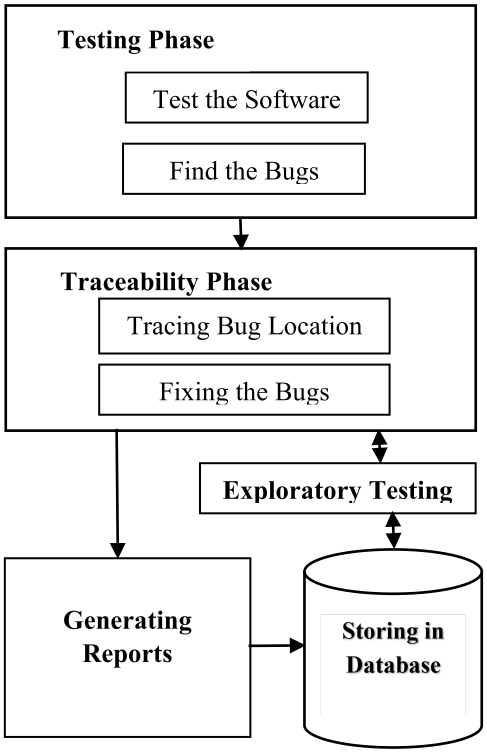


Figure 1: Architecture of BCO based fault coverage regression test.

5. Cloud Testing Framework Example

Large software requires a huge number of test cases. Often, these test cases require a lot of time and effort even with automated testing [24, 25, 26]. Because each test consumes execution time, the execution time required generally decreases when parallelization is used. Cloud Testing Frameworks presented in [24, 26] improves the execution process time by performing the parallel execution of test cases using current computational resources without source code modification [24, 26, 27].

Using a distributed and parallelized test can result in significant reduction of time required for test cases execution. It can also reduce the needed to identify and correct faults. Hence, reducing the total cost of development. Furthermore, the proposed framework in [24] increases the reliability of the test results by using heterogeneous environment that can result in the exposure of hidden failures ahead of the production phase [24].

One of the earliest Cloud Testing Framework "*CouldTesting*" presented by Oliveira and Duarte in 2013 is shown in figure 2. The framework distributes the unit tests using reflection on the local classes and then schedules the machine on cloud. It then the load is distributed over machines, using the round robin scheduling algorithm to ensure even distribution of the requests to the available machines in the test infrastructure [24].

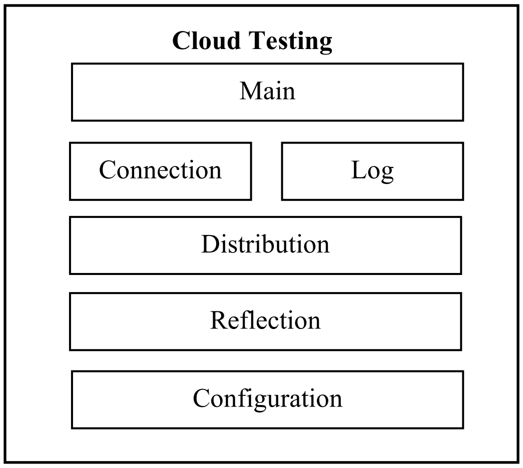


Figure 2: Cloud testing components from [24]

Figure 2 presents the architectural components of the framework. The main components of the framework are: Configuration, Reflection, Distribution, Connection, Log and Main.

The *configuration* component help in defining information for *paths, hosts,* and for *load balancing* [24]. This component deals with issues related to local storage space allocation for test results, selecting the libraries needed for proper test execution, and file access permission. It includes the list of machines and the parameters for the load balancer. The Reflection component extracts the tests cases [24].

Figure 3 depicts the distribution component being used to intermediate the execution of test suites over a parallel infrastructure. To work with a given IDE and parallel infrastructure the framework must be extended to include specific plugins. The *connection* component provides an interface on the client side to communicate with the cloud provider. In the cloud site this component provides a service that manages the execution of each test and it sends the test results back to the client in real-time. The log component records events generated in the process. The main component is a facade that encapsulates the components [24].

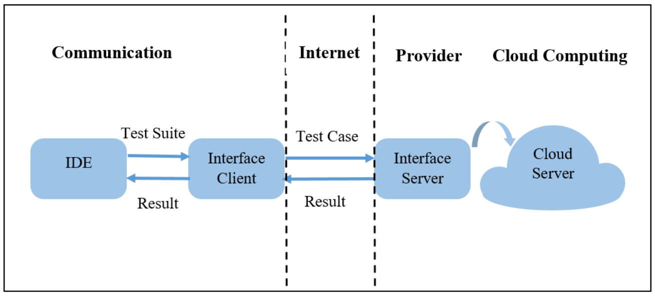


Figure 3: Distribution component being used to intermediate the execution of test suites over a parallel infrastructure from [24].

6. Test Case Prioritization Techniques

With the limited testing budget, it has been always a big challenge of software testing to optimize the order of test case execution in a test suite, so that they detect maximum number of errors. Three solutions to this problem were discussed in [28] such as test suite reduction, test case selection, and test case prioritization.

As the name suggests, the test suite reduction removes test cases from a test suite, which are redundant, and test case selection selects the most fault revealing tests based on a given heuristic. On the other hand, Test Case Prioritization (TCP) considers all the test cases without removing any of them. Instead it ranks all existing test cases thus prioritizes the test cases. While testing is performed testers executes the test cases with the higher priorities first as long as the testing budged supports [25]. Following techniques are discussed in [25, 28] to implement TCP.

6.1. Code/Topic Coverage Based

Most popular technique that has been reported in the literature is the Code coverage-based TCP that prioritize the test cases effectively [29]. It requires knowing the source code information of the software to measure the code coverage. This technique is not applicable in black- box systems tests because of lack of information about the code coverage. In [30] Thomas et. al proposed a modified a TCP technique, which they called topic coverage, provides an alternative concept to code coverage. The goal was to rank tests so that they cover more topics sooner [25, 28].

6.2. Text Diversity Based

Another common technique for TCP is to diversify the test cases. In [25] authors described a test case diversifying techniques that analyzes the test scripts directly hence this approach is called a text diversity-based TCP. The technique treats test cases as single, continuous strings of words. Then it applies different string distance metrics, such as the Hamming distance, on pairs of test cases and determines their dissimilarity. The idea is that the more two test cases are dissimilar textually the more they are likely to detect faults in different part of the source code [25, 28].

6.3. Risk Based Clustering

This TCP needs to access to execution results of the previous test cases, typically examining only the last execution of the test cases. It can be extended to as many previous executions as possible. This technique must ensure to run those test cases that failed in their previous execution provided that they are still relevant [31]. The technique might be combined with other TCP techniques, as well. For example, one can prioritize the previously failed test cases using a coverage-based approach to provide a full ordering of the test cases. In [28] authors modified this approach to have several clusters of test cases of different riskiness factor rather than having only two clusters of failed and non-failed test cases. In their approach, the highest risk is assigned to the tests that failed in the immediate version before the current version. The next riskiest cluster are tests that did not fail in the previous version but failed in the two versions before the current version, and so on.

7. Proposed Frameworks Discussed

As we discussed earlier, Swarm Intelligence algorithms as an emerging technique in the field of optimization are being integrated in many of the automation testing frameworks. These algorithms were instrumental in improving the performance and the efficiency of software testing on the cloud. In this section, we briefly highlight few of the frameworks that used Swarm Intelligence algorithms:

In [23] A. Kaur and S. Goyal proposed a Bee Colony Optimization algorithm for fault coverage-based regression test suite prioritization. The framework imitates the behavior of two types of worker bees found in nature. The behavior of the bees has been observed and mapped to prioritize software test suite.

A similar system has been presented by Karnavel and Santhoshkumar in [11] that used Bee Colony Optimization algorithm for test suite prioritization. The authors modified an existing framework to reduce the number of test cases from the retest test pool.

G. Oliveira and A. Duarte presented one of the aerialist frameworks called ‘CloudTesting’. The framework executed test cases in parallel over a distributed cloud infrastructure [24]. In the framework, cloud infrastructure is used as the runtime environment for automated software testing. Their experimental results indicate remarkable performance gain without significantly increasing the cost involved in facilitating the cloud infrastructure. The framework also simplified the execution of automatic tests in distributed system.

The framework presented by S. Faeghe and S. Emadi in [32] used the Ant Colony algorithm to automate software path test generation. The authors proposed a solution based on ant colony optimization algorithm and model-based testing for faster generation of test paths with maximum coverage and minimum time and cost. According to authors’ evaluation, the framework showed better performance over the existing methods in terms of cost, coverage and time.

A. Kaur and D. Bhatt in [29] proposed a regression testing based on Hybrid Particle Swarm Optimization (HPSO). The HPSO is an algorithm where a Genetic Algorithm (GA) is being introduced to the Particle Swarm Optimization (PSO) concept. The authors used the hybrid approach to prioritize tests for regression testing. The authors also mentioned that use of algorithm improved the effectiveness of their proposed framework.

8. Conclusion

Testing is one of the most complex and time-consuming activities. Automated testing on the cloud is one of the most popular solutions to reduce the time and the cost of software testing. In this paper, we discussed examples of automated frameworks proposed for testing software on the cloud [11, 23, 24, 29, 32]. Based on our review to such frameworks, it is evident that the use of the cloud as runtime environment for software testing are more efficient and effective solution when compared to traditional methods. Furthermore, on the cloud automated testing frameworks that used Swarm Intelligence Algorithms such ACA and BCO were able to produce significant reduction in the time required to execute large test sets and cover a diverse and heterogenic testing coverage. The use of such effective algorithms facilitates and enhances parallel execution and distribution of large testing loads on the cloud. In addition, cloud-testing frameworks generally simplifies the execution of automatic tests in distributed environments, hence gains in performance, reliability and simplicity of configuration.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Authors would like to thank the reviewers of this work for their help. This work was accomplished by the support of University of Minnesota Crookston.

References

[1] R. Hower, "What are some recent major computer system failures caused by software bugs," http://www.softwareqatest. com/qatfaq1.html,[February 2015], 2005.

[2] P. Ladkin, "The crash of flight ci676." [Online]. Available:http://www.rvs.uni-bielefeld.de/publications/Reports/taipei/taipei.html

[3] P. Krömer, J. Platoš, V. Snášel, "Nature-inspired meta-heuristics on modern GPUs: state of the art and brief survey of selected algorithms." International Journal of Parallel Programming 42.5 (2014): 681-709.

[4] B. Beizer, Software testing techniques. Dreamtech Press, 2003.

[5] M. Mavrovouniotis, C. Li, S. Yang, "A survey of swarm intelligence for dynamic optimization: algorithms and applications", Swarm and Evolutionary Computation 33 (2017): 1-17.

[6] N. Gupta, "Different approaches to white box testing to find bug" International Journal of Advanced Research in Computer Sciency and Technology (IJARCST 2014) 2.3 (2014): 46-9.

[7] M. Nouman, U. Pervez, O. Hasan, K. Saghar, "Software testing: A survey and tutorial on white and black-box testing of c/c++ programs," in Region 10 Symposium (TENSYMP), 2016 IEEE. IEEE, 2016, pp. 225–230.

[8] B. Balamurugan, J. Sridhar, D. Dhamodaran, P. Venkata Krishna. "Bio-inspired algorithms for cloud computing: a review." International Journal of Innovative Computing and Applications 6, no. 3-4 (2015): 181-202.

[9] R. Khalid, "Towards an automated tool for software testing and analysis" In Applied Sciences and Technology (IBCAST), 2017 14th International Bhurban Conference on, pp. 461-465. IEEE, 2017.

[10] E. Pacini, C. Mateos, C.G. Garino. "Distributed job scheduling based on Swarm Intelligence: A survey." Computers & Electrical Engineering 40.1 (2014): 252-269.

[11] K. Karnavel1, J. Santhoshkumar, "Automated Software Testing for Application Maintenance by using Bee Colony Optimization algorithms (BCO)", Information Communication and Embedded Systems (ICICES), 2013 International Conference on. IEEE, 2013.

[12] U. Salima and A. Askarunisha, "Enhancing the Efficiency of Regression Testing Through Intelligent Agents" International Conference on Computational Intelligence and Multimedia Applications 2007, pp. 103-108.

[13] S. Dick, A. Kandel. Computational intelligence in software quality assurance. Vol. 63. World Scientific, 2005.

[14] M. E. Khan, F. Khan. "A comparative study of white box, black box and grey box testing techniques." Int. J. Adv. Comput. Sci. Appl 3.6 (2012).

[15] M. Khan, "Different Approaches to Black Box Testing Technique for Finding Errors," IJSEA, Vol. 2, No. 4, pp 31-40, October 2011

[16] K. Dervis, B. Akay. "A survey: algorithms simulating bee swarm intelligence." Artificial intelligence review 31.1-4 (2009): 61.

[17] A. Kaur, D. Bhatt "Hybrid Particle Swarm Optimization for Regression Testing", International Journal on Computer Science and Engineering, Vol. 3 No. 5 (May-11), pp1815~1824, ISSN: 0975-3397

[18] A. Colorni, M. Dorigo, V. Maniezzo. "Distributed Optimization by Ant Colonies". First European Conference on Artificial Life, 134-142, 1991.

[19] S. Mazzeo, and I. Loiseau. "An ant colony algorithm for the capacitated vehicle routing." Electronic Notes in Discrete Mathematics 18 (2004): 181-186.

[20] H. Li, P.L. Chiou, "Software Test Data Generation using Ant Colony Optimization." International conference on computational intelligence. 2004.

[21] D. Teodorovic, M. Dell'Orco. "Bee colony optimization–a cooperative learning approach to complex transportation problems." Advanced OR and AI methods in transportation (2005): 51-60.

[22] O. Gotel, and A. finkelstein "An analysis of the Requirements Traceability Problem", International conference on Requirements Engineering, USA, 1994

[23] A. Kaur and S. Goyal, "A Bee Colony Optimization Algorithm for Fault Coverage Based Regression Test Suite Prioritization", International Journal of Advanced Science and Technology Vol. 29, April, 2011

[24] G. Oliveira, A. Duarte "A Framework for Automated Software Testing on the Cloud", 2013 International Conference on Parallel and Distributed Computing, Applications and Technologies

[25] Y. Ledru, A. Petrenko, S. Boroday, and N. Mandran, "Prioritizing test cases with string distances," Automated Software Engineering, vol. 19, no. 1, pp. 65–95, 2011.

[26] A. Duarte, W. Cirne, F. Brasileiro andP. Machado, "Gridunit: software testing on the grid." In Proceedings of the 28th international conference on Software engineering (pp. 779-782). ACM, May 2006

[27] T. Banzai, H. Koizumi, R. Kanbayashi, T. Imada, T. Hanawa, M. Sato. "D-cloud: Design of a software testing environment for reliable distributed systems using cloud computing technology." In Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on, pp. 631-636. IEEE, 2010.

[28] H. Hemmat, Z. Fang, M. V. Mantyla , "Prioritizing Manual Test Cases in Traditional and Rapid Release Environments", Software Testing, Verification and Validation (ICST), 2015 IEEE 8th International Conference on. IEEE, 2015.

[29] G. Rothermel, R. H. Untch, C. Chu, and M. J. Harrold, "Prioritizing test cases for regression testing," Software Engineering, IEEE Transactions on, vol. 27, no. 10, pp. 929–948, 2001.

[30] S. Thomas, H. Hemmati, A. Hassan, and D. Blostein, "Static test case prioritization using topic models," Empirical Software Engineering, vol. 19, no. 1, pp. 182–212, 2014

[31] A. Onoma, W. Tsai, M. Poonawala, H. Suganuma, "Regression testing in an industrial environment," Communications of the ACM, vol. 41, no. 5, pp. 81–86, 1998.

[32] S. Faeghe, S. Emadi. "Automated generation of software testing path based on ant colony." In Technology, Communication and Knowledge (ICTCK), 2015 International Congress on, pp. 435-440. IEEE, 2015.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 257-262 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

What Should Be Considered for Acceptance Mobile Payment: An Investigation of the Factors Affecting of the Intention to Use System Services T-Cash

Riyan Rizkyandy*, Djoko Budiyanto Setyohadi, Suyoto

Universitas Atma Jaya Yogyakarta, Yogyakarta, 55281, Indonesia

ARTICLE INFO

Article history:
Received: 01 March, 2018
Accepted: 28 March, 2018
Online: 25 April, 2018

Keywords:
Mobile Payment
E-money
T-cash
Technology Acceptance Model

ABSTRACT

E-Money mobile payments, also called digital money, are electronic payments, payment transactions using an Internet network integrated with NFC-enabled smartphones and prepaid cards. In Indonesia not only banks that issue e-money products, telecom operators from Telkomsel also issued an e-money product called T-cash. T-cash is a new innovation of electronic money presented by Telkomsel. The purpose of this study was to check the effect of responsiveness, smartness, perceived ease of use, perceived usefulness, social influence, and security against the intention to use T-cash. The data used in this study include primary and secondary data. Respondents in this study are users of T-cash products in Yogyakarta as many as 115 respondents. While the data were collected by using the questionnaire to then be analyzed using the amos analysis technique 22.0. The results of the analysis prove that two characteristics of technology, responsiveness and smartness have a significant effect on perceived usefulness. Ease of use has a significant effect on perceived usefulness. Ease of use, usefulness and security have a significant effect on intention to use. The higher the level of responsiveness, smartness, ease of use, perceived usefulness and security will also increase the use of T-cash social influence factors have no effect on intention to use.

1. Introduction

The current payment system has evolved and has evolved to abandon the old way and switch to mobile devices (m-device) now known as mobile payment (m-payment) [1]. The new payment system is the result of the development of information and communication technology in the field of economic transactions between companies and customers [2]. This payment system emerges as one way to solve certain problems related to the handling of the circulation of cash. This payment system ensures flexibility for small purchases and instant payments, enhances security and protection against fraud and other forms of crime, the emergence of e-commerce on the internet and online payments [3,4]. The current payment system should also take into account social influencing factors that will influence consumers to use mobile payment services [5,6].

E-money or electronic money is the amount of money a person keeps in an electronic medium that is officially accepted as a means of payment [7]. These payments are used for small-value transactions such as parking and public transport payments that are now beginning to use electronic cards containing e-money. And large-value transactions such as payments on the sale of goods in online shopping and transactions between other business actors [8].

In Indonesia one of the companies offering NFC based mobile payment is T-cash. T-cash is one of the services offered by Telkomsel. Telkomsel which started its business from cellular operator services then expanded its business network by offering T-cash mobile payment service. Telkomsel has obtained a license from Bank Indonesia as an electronic money service provider. T-cash can be used by all Telkomsel subscribers, either postpaid or prepaid.

Telkomsel introduced T-cash as a new generation of electronic money services as well as mobile payments. T-cash comes with innovations that will give customers and merchants an exciting new experience in making payments via mobile phones. T-cash is a product of e-money that facilitates transactions to various circles

*Corresponding Author: Riyan Rizkyandy, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia. Email: rizkyandyriyan@gmail.com

https://dx.doi.org/10.25046/aj030230

of society, including communities that have not been served by an authorized financial institution. T-cash was first introduced in 2015, the use of the T-cash service is easy, simply by enabling and sticking the NFC T-cash sticker to the phone [9].

With the number of e-money products in Indonesia emerging and competing, one of them is T-cash shows that technology is growing and innovation payment using e-money increasingly popular with the public. People's attention to e-money is increasingly high cause a lot of various e-money products issued by banks and other companies in Indonesia. People feel a lot of benefits and assume that using e-money products can improve or improve their social status. And some people think that new technology will only make them difficult because they are used to and comfortable with the payment of the old way. Therefore, the authors are interested to examine the issue with the title "What Should Be Considered for Acceptance Mobile Payment: An Investigation of the Factors Affecting of the Intent to Use System Services T-Cash".

Contribution in this research is social influence factor, security factor and two factor of technology characteristic that is responsiveness and smartness as an important variable that can influence the consumer to use t-cash. The contribution is important given the issue of security using electronic money greatly affects the intention to use mobile e-money payments. And in this research can also be explained that the factors responsiveness, smartness, perceived ease of use, perceived usefulness, social influence, and security are factors that may affect the intention to use e-money.

2. Hypothesis Development

2.1. Technology Characteristics

In human-computer interaction (HCI) responsiveness is an important factor because users want a responsive mobile payment system in accordance with what users expect. And using the payment system makes users look smart [10]. In [11], it is explained that two characteristics of responsiveness and smartness technologies have a significant impact on perceived usefulness.

H1: Responsiveness has a positive and significant impact on the perceived usefulness of mobile t-cash (e-money) payment system service in Yogyakarta.

H2: Smartness has a positive and significant impact on the perceived usefulness of mobile t-cash (e-money) payment system service in Yogyakarta.

2.2. TAM (Technology Acceptance Model)

In [12], it states that the Technology Acceptance Model (TAM) was developed by Davis in 1986 [13], which originated from the theory of Fishbein and Ajzen (1975) Theory of Reasoned. TAM provides a theoretical basis for knowing what factors influence the acceptance of a technology in an organization [14]. And in [15], This study examines user acceptance and use of mobile payments, focusing on mobile ticket technology applied in the context of public transport. The results show that the intention to use technology is influenced by perceived usefulness, Ease of Use and Security of that technology. In addition, perceived usefulness is simultaneously affected by ease of use.

H3: Perceived ease of use has a positive and significant impact on the perceived usefulness of mobile t-cash (e-money) payment system service in Yogyakarta.

H4: Perceived ease of use has a positive and significant impact on the intention to use mobile t-cash (e-money) mobile payment system service in Yogyakarta.

H5: Perceived usefulness has a positive and significant impact on the intention to use mobile t-cash (e-money) payment system service in Yogyakarta.

2.3. Social Influences

According to [16,17], social influencing factors are how a group or environmental factors can influence a person's behavior decisions. Social influence is a reflection of the results of communication and interaction with others so that with the occurrence of these influences can change a person's attitude or behavior. In [18], about the intent and behavior to use e-learning, those social factors are more important than usability perceptions and perceptions of ease of use, because social factors that influence students to use e-learning come from seniors as well as instructors, and lecturers are also a factor successful implementation of e-learning. In [19], explains Subjective Norm His influence can come from the views and roles of friends, peers, families, and superiors. Subjective norms have an important role in the study of the adoption of new technologies.

H6: Social influencing factors have a positive and significant impact on the intention to use mobile t-cash (e-money) payment system service in Yogyakarta.

2.4. Security

In [20], explains security is an important factor in shaping consumer confidence. The perception of security is the consumer's trust in the service user can control and maintain personal data as well as transaction data from abuse by an irresponsible person. In the study [21], found that security had a significant positive effect on online purchasing decisions. Security becomes one of the important factors because the payment transaction is done through internet network.

H7: Security has a positive and significant impact on the intention to use mobile t-cash (e-money) payment system service in Yogyakarta.

The framework underlying this research can be shown in Figure 1:

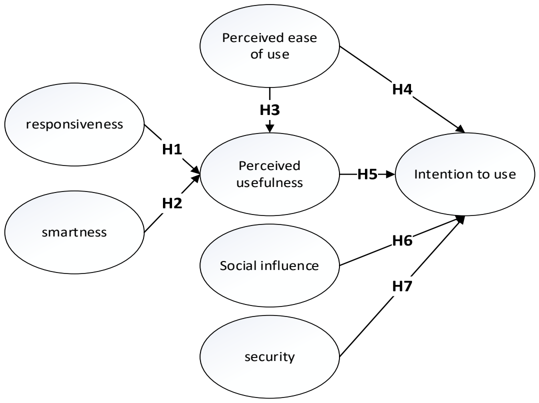


Fig 1: Research Model and Hypothesis

3. Method

The population in this research is Telkomsel subscribers who use T-cash service or who have never used T-cash respondents have unlimited nature because the number and characteristics of research respondents are not known for certain. The sampling technique used in this research is the nonprobability sampling technique. The respondents of this research are all Telkomsel subscribers who use T-cash who have access and have made T-cash mobile payment transaction in Yogyakarta.

The sample size plays an important role in the estimation and interpretation of results. In other structural methods, the sample size becomes the basis for the estimation of sampling error [22]. The sample size guidelines are 5-10 times the estimated number of parameters, the samples were taken in this study were 115 respondents (23 indicators x 5) [23,24].

In the questionnaire, there are two parts, the respondent's demographic information section, and 23 questions representing the factors in this study. Measurement of variables using Likert scale 1-5. List of questions can be seen in table 1.

Table 1: Questionnaire

Factor	Item	Question
Responsiveness (RES)	RES1	T-Cash mobile payment will offer prompt service to me.
	RES2	T-Cash mobile payment has willingness to help me.
	RES3	T-Cash mobile payment gives me individual attention.
Smartness (SMA)	SMA1	T-Cash mobile payment is intelligent.
	SMA2	T-Cash mobile payment is smart.
	SMA3	T-Cash mobile payment makes me look smart.
Perceive Ease of Use (PEOU)	PEOU1	Learning to operate an T-Cash mobile payment system would be easy for me.
	PEOU2	I would find an T-Cash mobile payment system easy to use.
	PEOU3	I think it is easy to use an T-Cash mobile payment system.
Perceive Usefulness (PU)	PU1	Using T-Cash mobile payment system in my finance related job would enable me to accomplish tasks more quickly.
	PU2	Using T-Cash mobile payment would enhance the effectiveness of my finance related job.
	PU3	I would find T-Cash mobile payment system useful in my finance job.
Social Influence (SI)	SI1	Friend's suggestions and recommendations will affect my decision to use T-Cash
	SI2	Family/relatives have influence on my decision to use T-Cash
	SI3	I will use T-Cash if my colleagues use it
	SI4	I will use T-Cash if the service is widely used by people in my community
	SI5	T-Cash will enable me to improve my social status
Security (SEC)	SEC1	I think that purchasing through a T-Cash mobile payment is secure.
	SEC2	I feel secure entering my card details for payment within the T-Cash.
	SEC3	I find T-Cash mobile payment services secure for conducting my payment transactions.
Intention to Use (ITU)	ITU1	I intend to use the T-Cash mobile payment system.
	ITU2	I predict I would continue using the T-Cash mobile payment system.
	ITU3	I plan to continue using the T-Cash mobile payment system.

Table 2 summarizes the demographic characteristics of respondents. Age is divided into 3 groups, and the most dominant age is the age group of 20-25 years of 67.0%. 51.3% were male and 48.7% were female. At the education level, there are 4 groups, namely SMP (junior high school), SMA (senior high school), S1 (undergraduate), S2 (postgraduate). which dominates the level of education S1 by 43.5%. 55.7% active transactions using T-cash and 44.3% inactive transactions using T-cash.

Table 2: Demographic of the Respondents

Demographic		Frequency	Percent%
Age	< 20 years	33	28.7
	20-25 years	75	67.0
	>25 years	5	4.3
Gender	Male	59	51.3
	Female	56	48.7
Level of Education	SMP	18	15.7
	SMA	44	38.3
	S1	50	43.5
	S2	3	2.6
Active Transaction T-cash	Yes	64	55.7
	No	51	44.3

4. Results and Discussion

4.1. Measurement Model Assessment

Assessing the goodness of fit is the primary goal in SEM to know to what extent the model is hypothesized "Fit" or matched the data sample [22,25]. The result of goodness of fit is shown in the data in table 3. Based on the Results in Table 3, it can be seen that the research model is approaching as a good fit model.

The result of CMIN / DF in this study 1,433 showed that the fit research model. The GFI value in this model is 0.823. The value close to the recommended level ≥ 0.90 shows the marginal fit research model. The RMSEA value of this study was 0.062 with the recommended value of ≤ 0.08 this shows the fit research model. The AGFI value in this model is 0.711. The value close to the recommended level ≥ 0.80 shows the marginal fit research model. TLI value in this study is 0.941 with the recommended value of ≥ 0.90 it shows fit research model. The NFI value in this study was 0.829 closer to the recommended level ≥ 0.90 showing the marginal fit research model. Based on the overall measurement of goodness of fit above indicates that the model proposed in this research is accepted.

Table 3: Assessing the Goodness of Fit

Goodness of fit index	Cut-off value	Reserach Model	Model
Significant probability	≥ 0.05	0.000	Marginal
RMSEA	≤ 0.08	0.062	Fit
GFI	≥ 0.90	0.823	Marginal
AGFI	≥ 0.90	0.711	Marginal
CMIN/DF	≤ 2.0	1.433	Fit
TLI	≥ 0.90	0.929	Fit
CFI	≥ 0.90	0.940	Fit

4.2. Structural Model Assessment and Hypotheses Testing

Instrument Validity is evaluated based on convergent validity and discriminant validity of the indicator calculated using Amos 22.0. Convergent validity is used to determine the validity of each relationship between indicators and their latent constructs (variables). Convergent validity parameters include 3 items, the value of item questionnaire loading (≥ 0.7), the value of Average Variance Extracted (AVE) with values (≥ 0.5) and communally (≥ 0.5). (Convergent validity) is said to be high if the loading value is above 0.7. Whereas discriminant validity is seen from the loading factor of each questionnaire item with the construct representing it [22,25].

Reliable instruments are not necessarily valid, while valid instruments are generally reliable. Thus, the reliability testing of the instrument must be done because it is a requirement for testing the validity. In this connection, this study measures the reliability of data with internal consistency reliability. Internal consistent reliability testing is done by testing the instrument once, then to test the reliability of the data used indicator based on the formula Variance Extracted (AVE) and Construct Reliability (CR). And the indicator of the variable is said to be reliable if the AVE value is ≥ 0.05 and CR ≥ 0.07 [22,25]. Table 4 shows the results of Variance Extracted (AVE) and Construct Reliability (CR).

The process of testing this statistic can be seen in table 5. From the data processing, it is known that the CR value is related to showing values above 1.96 and below 0.05 for the value of P [26]. H1, H2, H3, H4, H5, and H7 support or significant (S), while H6 does not support or not significant (NS).

The findings of this study have significant implications for the intention to use mobile t-cash (e-money) payment system service in Yogyakarta. In this research use TAM model. The result of two technological characteristics of responsiveness and smartness in T-Cash mobile payments has a significant impact on perceived usefulness. The results in this study support the results of previous studies conducted in [11], which ssuggest that responsiveness and smartness have a significant impact on perceived usefulness in mobile NFC payment services in Korea. The responsiveness and smartness presented in t-cash bring a fast service to the consumer in around and small-sized form of NFC sticker. Transactions can be quickly and easily done by simply bringing the sticker to the sensor, then simply enter the pin, and the transaction is done, and make the customer look smarter by using the new model payment system presented by the mobile operator Telkomsel. With the innovation of a fast and unique shaped payment model will affect consumer interest to use t-cash (e-money).

Table 4: Standardized Item Loading, AVE, CR

Variable	**Item**	**Factor Loading**	**CR**	**AVE**
ITU	ITU1	0.756	0.8334	0.6254
	ITU2	0.826		
	ITU3	0.789		
SEC	SEC1	0.729	0.8344	0.6276
	SEC2	0.821		
	SEC3	0.823		
PU	PU1	0.769	0.8567	0.6664
	PU2	0.868		
	PU3	0.809		
PEOU	PEOU1	0.778	0.8670	0.6863
	PEOU2	0.777		
	PEOU3	0.922		
SMA	SMA1	0.835	0.8584	0.6707
	SMA2	0.717		
	SMA3	0.895		
RES	RES1	0.878	0.8474	0.6502
	RES2	0.758		
	RES3	0.778		
SI	SI1	0.825	0.8737	0.5814
	SI2	0.718		
	SI3	0.736		
	SI4	0.712		
	SI5	0.814		

Table 5: Hypothesis Test Results

	Estimate	S.E.	C.R.	P	Label	Result
PU ← RES	.372	.113	3.283	.001	par_13	S
PU ← SMA	.192	.090	2.142	.032	par_14	S
PU ← PEOU	.277	.089	3.111	.002	par_23	S
ITU ← PEOU	.293	.118	2.477	.013	par_15	S
ITU ← PU	.359	.082	4.369	***	par_16	S
ITU ← SI	-.063	.078	-.801	.423	par_21	NS
ITU ← SEC	.382	.130	2.936	.003	par_22	S

Perceived ease of use has a significant effect on perceived usefulness, the results of which support the results of previous studies conducted in [15], which suggest that perceived ease of use has a significant impact on the perceived usefulness of mobile payments (IMMPA) for public transport. The perceived benefit of consumers towards the ease of using t-cash is to settle transactions in work and daily activities quickly. In [27], mentions if in a way perceived benefits by a consumer, the positive information from mouth to mouth can be spread widely through community forums and social media. So it is influential in increasing the intention of other consumers to use t-cash.

Perceived ease of use and perceived usefulness significantly influence intention to use. In This study support the results of previous studies conducted in [15], which in his research stated that perceived ease of use and perceived usefulness have a significant effect on intention to use mobile public transportation payment. In the use of t-cash is easy to learn in a short time to get used to the operation. Because the use of t-cash is very practical and efficient without the need to swipe like other e-money prepaid cards. Enough with tap sticker then the transaction is done. So in line with research [28], explaining that consumers will use e-commerce when a convenient technology is used easily and simply. So that easily t-cash to be used and perceived benefits in terms of effectiveness and quickly in the process of influential transactions in increasing the intention to use t-cash.

Security on T-Cash mobile payments has a significant impact on intention to use. in this study also supports the results of previous research, in [20], which in his research mentions that security has a significant effect on the intention to use mobile payment in the restaurant industry. In [29], it also reinforces that there is a positive relationship between security against intent to use mobile payment services. In transactions and their use, t-cash has a sufficiently secure and secure level of security. To activate t-cash requires confirmation from active mobile phone number to fill out the identity and give notification of each transaction history of t-cash and matching between usage of operator card of Telkomsel and t-cash card number. And t-cash will ask for a pin of every transaction made. And in [30] explaining, security is a belief in a person that the activities of transactions conducted have a high enough level of security and all about the personal information provided is guaranteed and safeguarded. With the increase of security services provided to consumers, it is influential in the intention to use t-cash.

In social influence on intention to use cannot be accepted with a p-value greater than at the 0.05 level of significance. This finding explains that social environment factors around respondents such as friends, relatives, family, and community do not support or influence the respondent to use T-cash and in the use of T-cash will not improve the user's social status. The results of this study are different from previous studies that has been done in [19] which in his research stated that the influence of social or social influence significantly influences the intention to use a mobile credit card. In [31], it also shows there is a positive relationship of social influence factors on the intention to use ICT in teaching. In [32], the results show that social influences have a significant positive impact on the intentions of use for mobile data services. In [33,34,35,36], also explains that social influences have a positive and significant influence on the use of SI. However, the results of this study are also supported by research [37], which finds that the influence of social factors on the interest of information systems utilization and social influencing factors on the use of information systems is rejected or has no effect. And the results of this study are also supported by interviews to some respondents who stated that using T-cash is encouraged by personal factors or self-interest that they feel the need to use T-cash, without any influence from friends or family.

In [38], explaining that a person's decision to choose a good and a service to be used will be influenced by several factors namely, environmental factors such as family, environment, culture, and personal factors. However, this research does not show any social influence which gives significant influence to the intention of using mobile t-cash (e-money) payment system service in Yogyakarta.

In [39], explaining factors that may affect purchasing behavior or intention to use are personal factors. Personal factors are psychologically someone who is not the same as others, therefore personal factors tend to have relatively consistent and long-lasting responses to environmental influences. In personal factors, there are characteristics: age and stage of the cycle, profession, economic situation, lifestyle, personality as well as self-concept [40]. Based on these characteristics that the differences in the results of this study can be caused by respondents who almost most of the profession are students and students. Which is the group of professions still rarely use t-cash facilities (e-money) because the profession does not have the intensity of payment transactions are too many, they choose to conduct transactions in cash or via atm. Personal factors are an important consideration in taking a decision other than environmental factors. So if the environment of the respondent has given the recommendation to use t-cash, the personal factor will become consumer consideration to determine consumer decision intention in using service of the mobile payment system of t-cash (e-money) in the future [41,42].

5. Conclusion

In this study found that, with respect to the intention to use mobile payment service system T-cash in Yogyakarta. Technological factors characteristic of responsiveness and smartness have a significant effect on perceived usefulness, perceived ease of use also shows significant relation to perceived usefulness. In perceived ease of use and perceived usefulness have a significant effect on the intention to use. The security factor also has a significant effect on the intention to use. With the conclusion of H1, H2, H3, H4, H5, and H7 are supported or accepted. Only one factor is social influence has no relationship to the intention to use. So, H6 is neither support nor rejected. That way social influencing factors will become less important than personal factors that ultimately are a consideration and consumer decisions to use them.

This study also shows that perceived usefulness, security and efficiency are important factors. The interview results concluded that for those who have never used e-money before will be a little confused how to use this service, and what benefits they get. This is due to a lack of knowledge about mobile payments [43]. So the challenge for the company is to develop and create new perceptions about consumer lifestyles in the use of T-cash, where T-cash can make the payment system easier, convenient to use and become the trend of payment, and better explain what happened and need to be done during the T-cash mobile payment process. by doing this may have an effect on the intention to use T-cash [29].

Based on the conclusions obtained from the results of the analysis in this study, researchers provide suggestions for further research, the suggestion is expected by researchers to review research at this time and previous research by adding or using other variables not contained in this study. So it can be further known what factors can affect a person to use new technologies such as e-money products in this study. subsequent research can also use a sample with a larger number, and also increase the number of indicators of the variables for the results of research can be obtained more optimal.

Acknowledgment

I would like to thank my supervisor for his guidance on this paper and to Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia.

References

[1] I. Mackensen, "Mobile Payments – A Strategic Forecasting Approach –," 2015.

[2] F. Liébana-Cabanillas, J. Sánchez-Fernández, and F. Muñoz-Leiva, "The moderating effect of experience in the adoption of mobile payment tools in Virtual Social Networks: The m-Payment Acceptance Model in Virtual Social Networks (MPAM-VSN)," *Int. J. Inf. Manage.*, 2014.

[3] F. José Liébana-Cabanillas Juan Sánchez-Fernández Francisco Muñoz-Leiva *et al.*, "Industrial Management & Data Systems Role of gender on acceptance of mobile payment," *Ind. Manag. Data Syst. Internet Res. Ind. Manag. &amp Data Syst. Iss Tao Zhou Ind. Manag. &amp Data Syst.*, vol. 114, no. 6, pp. 369–392, 2014.

[4] L.-C. Francisco, M.-L. Francisco, and S.-F. Juan, "Payment Systems in New Electronic Environments: Consumer Behavior in Payment Systems via SMS," 2015.

[5] K. Pousttchi and D. G. Wiedemann, "What_influences_consumers_intention_to_use_mobile," 2014.

[6] H. Khatimah and F. Halim, "The effect of attitude and its decomposed, perceived behavioral control and its decomposed and awareness on intention to use e-money mobile in Indonesia," *J. Sci. Res. Dev.*, 2016.

[7] T. J. Gerpott and K. Kornmeier, "Determinants of customer acceptance of mobile payment systems," *Int. J. Electron. Financ.*, 2009.

[8] D. Flood, T. West, and D. Wheadon, "Trends in Mobile Payments in Developing and Advanced Economies," 2013.

[9] M. S. Arifin *et al.*, "Smart Vending Machine Based on SMS Gateway For General Transactions," *IEEE*, 2017.

[10] S. Megadewandanu, Suyoto, and Pranowo, "Exploring mobile wallet adoption in Indonesia using UTAUT2: An approach from consumer perspective," in *Proceedings - 2016 2nd International Conference on Science and Technology-Computer, ICST 2016*, 2017.

[11] S. Shin, "The Effects Of Technology Readiness And Technology Acceptance On NFC Mobile Payment Services In Korea," *J. Appl. Bus. Res.*, vol. 30, no. 6.

[12] P. Moses, S. L. Wong, K. A. Bakar, and R. Mahmud, "Perceived Usefulness and Perceived Ease of Use: Antecedents of Attitude Towards Laptop Use Among Science and Mathematics Teachers in Malaysia," *Asia-Pacific Educ. Res.*, 2013.

[13] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *Source MIS Q.*, vol. 13, no. 3, pp. 319–340, 1989.

[14] R. Rauniar, G. Rawski, J. Yang, and B. Johnson, "Technology acceptance model (TAM) and social media usage: an empirical study on Facebook," *J. Enterp. Inf. Manag.*, 2014.

[15] L. Di Pietro, R. Guglielmetti Mugion, G. Mattia, M. Renzi, and M. Toni, "The Integrated Model on Mobile Payment Acceptance (IMMPA): An empirical application to public transport," *Transp. Res. Part C*, vol. 56, pp. 463–479, 2015.

[16] E. Shih, -Tse Wang, N. Pei, and -Yu Chou, "Consumer Characteristics, Social Influence, and System Factors on Online Group-Buying Repurchasing Intention," *J. Electron. Commer. Res.*, vol. 15, no. 2, 2014.

[17] G. Harvey Tanakinjal, R. MacDonell Andrias, S. Laison Sondoh Jr, and A. Asri Ag Ibrahim, "Relationship between Perceived Benefits and Social Influence towards Self- Disclosure and Behavioral Intention in Web 2.0," *Eur. J. Bus. Soc. Sci.*, vol. 1, no. 4, pp. 63–75, 2012.

[18] D. B. Setyohadi, M. Aristian, B. L. Sinaga, and N. A. A. Hamid, "Social critical factors affecting intentions and behaviours to use E-Learning: An empirical investigation using technology acceptance model," *Asian J. Sci. Res.*, 2017.

[19] G. W. H. Tan, K. B. Ooi, S. C. Chong, and T. S. Hew, "NFC mobile credit card: The next frontier of mobile payment?," *Telemat. Informatics*, 2014.

[20] C. Cobanoglu, W. Yang, A. Shatskikh, and A. Agarwal, "Hospitality Review Are Consumers Ready for Mobile Payment? An Examination of Consumer Acceptance of Mobile Payment Technology in Restaurant Industry "Are Consumers Ready for Mobile Payment? An," 2015.

[21] A. Raman and V. Annamalai, "Web Services and e-Shopping Decisions: A Study on Malaysian e-Consumer," *IJCA Spec. Issue "Wireless Inf. Networks Bus. Inf. Syst.*, 2011.

[22] T. Wijaya, *Analisis Structural Equation Modeling Menggunakan AMOS*, 5 ed. Yogyakarta: Universitas Atma Jaya Yogyakarta, 2009.

[23] J. W. Creswell, "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches," *Can. J. Univ. Contin. Educ.*, vol. 35, no. 2, 2009.

[24] A. Ferdinand, *Structural Equation Modelling dalam Penelitian Manajemen*. Semarang: FE UNDIP, 2002.

[25] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, 3rd ED. New York: Guilford Press, 2011.

[26] J. f Hair Jr, W. C. Black, B. J. Babin, and R. E. Anderson, "Multivariate Data Analysis, 7/e," *Pearson Prentice Hall*, 2010.

[27] J. C. Roca *et al.*, "Information Management & Computer Security The importance of perceived trust, security and privacy in online trading systems'The importance of perceived trust, security and privacy in online trading systems' The importance of perceived trust, security and privacy in online trading systems," *Inf. Manag. Comput. Secur. Ind. Manag. &amp; Data Syst. Online Inf. Rev. Iss Downloaded by MAHIDOL Univ.*, vol. 17, no. 22, pp. 96–113, 2015.

[28] M. Bonera, "The propensity of e-commerce usage: the influencing variables," *Manag. Res. Rev.*, 2011.

[29] J. Hauff, "Consumer Acceptance of Mobile Payment Services," 2011.

[30] C. L. Hsu, C. F. Wang, and J. C. C. Lin, "Investigating customer adoption behaviours in Mobile Financial Services," *Int. J. Mob. Commun.*, 2011.

[31] S. M. Salleh and K. Laxman, "Investigating the factors influencing teachers' use of ICT in teaching in Bruneian secondary schools," *Educ. Inf. Technol.*, 2013.

[32] T. Faziharudean and T. Li-Ly, "Consumers' behavioral intentions to use mobile data services in Malaysia," *African J. Bus. Manag.*, vol. 5, no. 5, pp. 1811–1821, 2011.

[33] R. L. Thompson, C. A. Higgins, and J. M. Howell, "Personal Computing: Toward a Conceptual Model of Utilization," *MIS Q.*, 1991.

[34] V. Venkatesh and F. D. Davis, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Manage. Sci.*, vol. 46, no. 2, pp. 186–204, 2000.

[35] G. C. Moore and I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," *Inf. Syst. Res.*, 1991.

[36] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View," vol. 27, no. 3, pp. 425–478, 2003.

[37] D. P. Triadmojo, "Analysis Of Factors Affecting The Interests Of The Use Of Information Systems And Use Of Information Systems (Empirical Study On Government Of Banyuwangi)," 2016.

[38] F. Liébana-Cabanillas, J. Sánchez-Fernández, and F. Muñoz-Leiva, "Antecedents of the adoption of the new mobile payment systems: The moderating effect of age," *Comput. Human Behav.*, 2014.

[39] Kotler and Philip, *Manajemen Pemasaran*, Edisi 11. Jakarta: PT Indeks Kelompok Gramedia, 2005.

[40] Kotler, Philip, and G. Amstrong, *Principles of Marketing*. New Jersey: Person Education, 2012.

[41] S. Y. Park, M.-W. Nam, and S.-B. Cha, "University students' behavioral intention to use mobile learning: Evaluating the technology acceptance model," *Br. J. Educ. Technol.*, 2012.

[42] W. Nasri and L. Charfeddine, "Factors affecting the adoption of Internet banking in Tunisia: An integration theory of acceptance model and theory of planned behavior," *J. High Technol. Manag. Res.*, 2012.

[43] A. S. Etim, "Mobile banking and mobile money adoption for financial inclusion," *Res. Bus. Econ. JournalBold Rotman*, vol. 9, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No.2, 263-270 (2018)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Three port converters used as interface in photovoltaic energy systems

Sarab Al-Chlaihawi[1,2,*], Aurelian Craciunescu[2]

[1]*Al-Furat Al-Awsat Technical University, Center Education Continuous, 54001, Iraq*
[2]*Electrical Engineering Faculty, University Politehnica of Bucharest, 060042, Romania*

ARTICLE INFO

Article history:
Received: 12 January, 2018
Accepted: 03 April, 2018
Online: 25 April, 2018

Keywords:
Three port converter
DC-DC converter
PV panel
MPPT
Full bridge converter

ABSTRACT

The aim of this paper is to derive and study a full-bridge three-port converter. Based on the standard design of full-bridge converter, we have modeled and derived a three port converter. The three port converter can be used in renewable energy scenarios, such as solar cells or wind turbines connected to the input port. The input can be taken from two-ports simultaneously or from one port at a time. In order to balance the power mismatch between the input port and load port, the batteries are attached to the third port, to ensure there are no discrepancies in the power generated at the input and power demand at the load. In order to ensure isolation and reduced voltage stress on the switches, a high frequency transformer is also used in the design. The overall design contains four switches, and four diodes. MOSFETs are the strongest candidate for the switches owing to their high switching speed, lower losses and high resistance to higher voltage. Moreover, a buck-boost structure is modeled in order to ensure that it can work for a wide variety of different applications by adjusting the duty cycle of the switches properly. To minimize the switching losses in the converter, Zero-Voltage Switching (ZVS) is also achievable in the modeled system.

1. Introduction

Multiport converters have the inherited advantage of being compact and handy, unlike the traditional converters. Instead of using multiple two-port converters for various applications involving power flow between different ports, three-port converters which feature single-stage conversion between different ports are more desirable; they provide higher efficiency, fewer components, compact-design and reduced switching losses. This is possible because switching and storage elements are shared between different ports and hence less component count yields better efficiency. Moreover, the unified power flow management is possible in multiport converters due to centralized control which boosts dynamic performance of the system. The centralized control enables the system to operate without any possibility of communication delay or errors, since it doesn't require any communication between control units of different converters. This paper is an extension of work originally presented in conference the 10th international symposium on advanced topics in electrical engineering [1]. Where we have analyzed the state equations, inductor volt-second balance for each state, analyzed the design parameters of the system and how different parameters depend on inductor currents, transformer turn ratio and output load current, analyzed the effect of different duty ratios on the battery and output current and we have analyzed the Zero-Voltage Switching mechanism in the modeled converter. Recently, many researches have been carried out on the topology of the multiport converters; for example, one research proposed to interface multiple converters to a common dc bus, each one having its own control unit [2]. But, it has been established that integrated topology carries definite advantages over multiple converters. Owing to these capabilities, three-port converters have gained popularity in recent years. Three-port converters can be constructed using different topologies such as series-resonant, full-bridge and half-bridge with the use of multi-winding transformers [3]. The power flow management and Zero Voltage Switching (ZVS) can be achieved in three-port converters using phase-shift control between different switching bridges [4]. However, an N-port converter contains N-1 control inputs; that is two control inputs for a three-port converter. As a consequence, modeling effort is increased, which results in complicated control circuit, which can offset the good efficiency of the system, if not properly designed. It also requires proper decoupling of different control loops since it contains integrated power stage [5].

The multiport topologies can be isolated or non-isolated, depending upon the application for which they are being used [6].

Isolated converters are preferred in applications when a huge step is desired between input and output voltages. Isolated converters based on bridge topology renders good efficiency in such cases,

[*] Sarab Al-Chlaihawi, Al-Furat Al-Awsat Technical University, Center Education Continuous, Al-Najaf, Iraq, +9647740675223, sarab.haedar@yahoo.com

https://dx.doi.org/10.25046/aj030231

when combined with soft switching methods [5]. Full-bridge configuration is used where converter power levels are beyond 750 W. It is not recommended for applications below this power level since it has a higher number of electronic components present—it has four transistor switches and associated controller circuit [7].

As the transformer's magnetizing current is bipolar, the full-bridge's utilization of transformer core is very efficient, since the entire cycle of B-H loop is used by the full-bridge. Usually the secondary winding is center-tapped and each half of the secondary winding forward the input power during alternate switching periods. This results in winding power loss due to secondary winding current. A schematic of general full-bridge converter with one secondary winding is shown in Figure 1. It contains four switches SA1, SA2, SB1, SB2, transformer with inductance L_m, four diodes and an inductor L_f to filter out the ac components in the output voltage level.

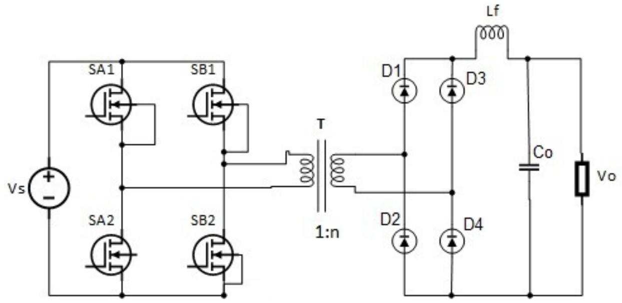
2. Derivation of Three Port Converter

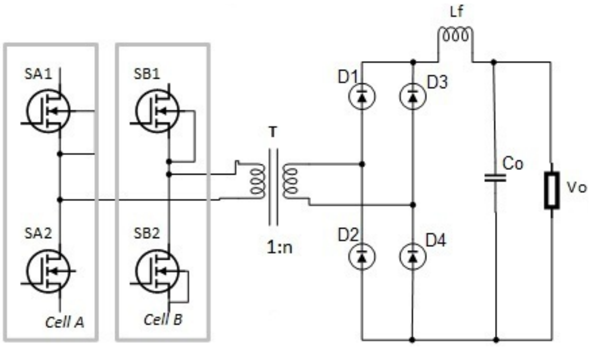
In order to derive the design of three port converter, again consider the Figure 1. It has two legs of switches, which are connected to the source Vs in parallel. The volt-second balance of the transformer inductance L_m represents the constraint condition of the full bridge converter. This implies that if inductance L_m satisfies the volt-second balance, then both the legs of the converter can be divided into two separate portions, known as cell A and cell B. This is shown in Figure 2.

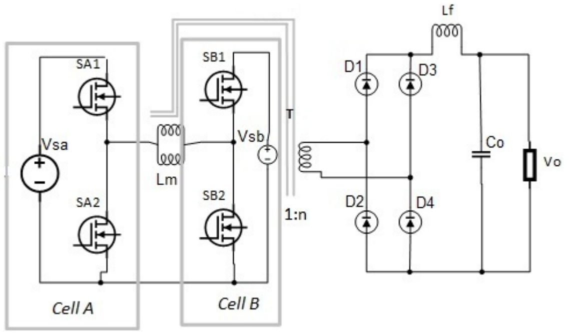
Furthermore, to introduce the battery in the system, we can connect cell A and cell B to different sources, such that cell A is connected to a source V_{sa} and cell B is connected to the battery source V_{sb}. If V_{sa} is equal to V_{sb}, the original full-bridge converter is derived. Hence, three port full-bridge converter is a general case in which V_{sa} and V_{sb} are assumed to be different sources with different voltage. When both the voltages are always equal, the cells are paralleled again. The converter assumes the shape as shown in Figure 3. It is obvious that the proposed three port full-bridge converter has a symmetric configuration in which both the sources can equally participate in supplying the power to the load. The primary winding inductance L_m can act as an inductor, which implies that primary side of the converter contains an inherent buck-boost converter with bidirectional characteristics.

This feature makes this design even more flexible in terms of power transfer between the two sources i.e. the power flow can be managed between the two sources as required, such as PV sources charging the batteries. Either of the voltage sources can be at higher or lower potential than the other.

The concept of primary side buck-boost converter is illustrated in Figure 4. The proposed converter with three ports is suitable for systems which combine renewable energy sources and energy storing devices, since the power flow is possible between the sources. The sources can be PV arrays, wind turbines etc.







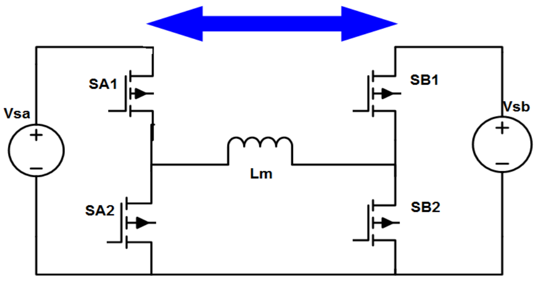


Figure 4-Primary Side of Buck-Boost

3. Full Bridge Three Port Converter

The three-port full-bridge converter with Photo Voltaic (PV) panel and battery bank is shown in Figure 5. As proposed, it has four switches which can transfer the power between PV array, the controller for Maximum Power Point Tracking (MPPT), batteries and one transformer with turn ratio 1:n, which steps up the voltage to drive the load. The Maximum Power Point Tracking is a mandatory circuit which is used with PV array to ensure that at the given sunlight intensity, maximum power is harnessed from the PV cells. Switches SA1 and SA2 provide an interface for the PV panel with a unidirectional flow from the PV panel, while switches SB1 and SB2 are connected to the battery port and provide bidirectional flow from and to the battery. This configuration is necessary since PV panels cannot intake power while batteries have to both charge and discharge over the course of the operation.

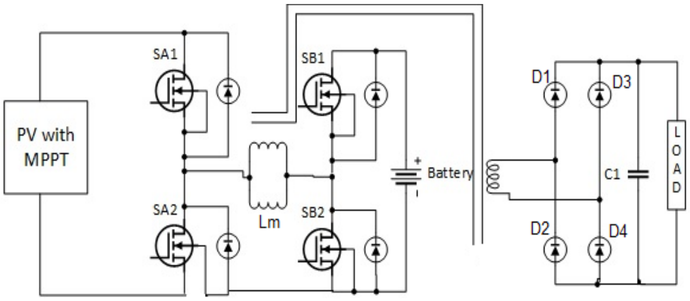


Figure 5- Three-port converter with PV panel and battery bank

A transformer is required to isolate the load output from the PV array and batteries. The transformer is also required to reduce the stress on the switches when the difference between input and output voltages is higher.

This helps in reducing the switch losses in the system. Since the transformer is introduced inside the buck-boost configuration of the converter, it has a higher frequency which depends on the switching frequency of the switches. This enables us to use much smaller transformers, even though the converter provides an output voltage within the frequency range of 50-60 Hz.

Table 1- Parameters of the full bridge three port converter

Component	Value
Transformer Frequency	10kHz
Transformer ratio	1:1
Output Capacitor (C1)	0.01F
Capacitor(Buck-boost)	4700μF
Inductor (Buck-boost)	47mH
Capacitor (PV panel)	100μF
Magnetizing Inductance (L_m)	50 H
Battery	12 V
Load	100 W
PV Irradiance	500 W/m^2
PV Maximum Power	150 W

3.1. Switching Modes of Three-Port Converter

In the ideal conditions, the power flow in the converter can be explained by the equation:

$$P_{pv} = P_b + P_o \tag{1}$$

In this equation, P_{pv} represents power from the PV panel, P_b is battery power and P_o is the output power. There are three possible modes in the converter:

- Single Input Single Output
- Single Input Dual Output
- Dual Input Single Output

The converter operates in Mode 1 when P_b= 0, that is, PV panel generates all power and meets the demand at the load side. In Mode 2, P_{pv}>P_o; the PV panels generate redundant power which goes into charging the batteries along with driving the load. In Mode 3, P_{pv}< P_o and PV array and batteries together supply the power to the load.

These modes have been made possible due to the flexible buck-boost structure of the converter. A summary of these modes is depicted in Table 2, while the functionality of each mode is explained in detail in the following section.

Table 2- Modes of the three port converter

Mode	PV Power Level	Battery Power Level
Single Input Single Output	P_{pv}= P_o	P_b=0
Single Input Dual Output	P_{pv}>P_o	P_b>0
Dual Input Single Output	P_{pv}< P_o	P_b<0

Mode 1

In mode 1, the configuration of the circuit is shown in Figure 6. Being a single input single output configuration, only PV panel feeds the power to the load. At the end of each cycle, the battery is also charged as evident from Figure 15, when SB1 turns ON and SB2 turns OFF. During the operation in Mode 1, the PV panel generates power to drive the load. However, the battery also gets charged slightly at the end of each switching period. The main flow of power is from PV panel to the output load via switches SA1 and SB2.

The voltage at the primary side of transformer is then boosted by a factor of 'n', which represents turns ratio at the secondary side. At the secondary side, diodes D1 and D4 are forward biased as the polarity of secondary voltage is positive at the top side.

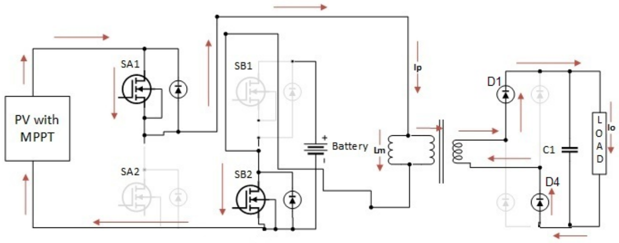


Figure 6- Mode 1 of the modeled converter

Mode 2

Mode 2 is the single input dual output mode, hence PV panels have enough power to feed the load side as well as charging the battery via the switch SB1 whose duty cycle is increased to the maximum value. The configuration for mode 2 is shown in Figure 7.

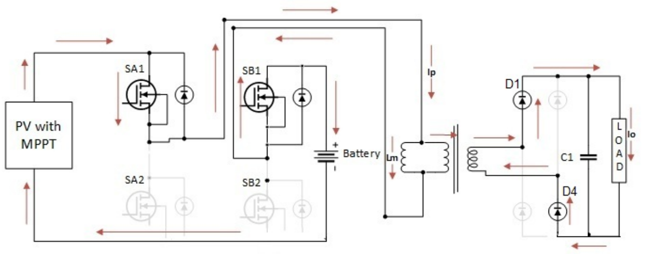


Figure 7- Mode 2 of the modeled converter

In Mode 2, PV panel alone feeds the load and charges the batteries. The PV panel charges the batteries through switch SB1 while catering to the demand of the load via switch SA1. The inherent buck-boost characteristics boost the PV voltage at the primary side of the transformer to a higher value. At the secondary side, diodes D1 and D4 are turned on and supply the positive voltage to the load.

Mode 3

Mode 3 represents the Dual input single output scenario, where the PV panel's generated power is not sufficient to drive the load; hence the batteries join in to cater to the needs of the load. However, when PV=0, for example during night, the battery has to provide the power to the load all alone, this is shown in Figure 8. Hence, the capacity of the batteries should be decided by considering the expected load. In mode 3, power flow is from batteries to the load via switch SB1 and the transformer.

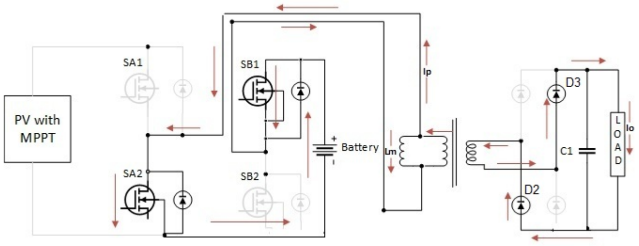


Figure 8-Mode 3 of the modeled converter

3.2. Switching States

In this section, we analyze the different switching states within the converter operations. Irrespective of the mode of the converter, the switching states are same. The only difference the modes make is in the value of inductor current i_{Lm}, which changes with the power generated at the PV panel P_{pv} and power at the load P_o. This current change in different modes in terms of value and direction. Below is the case study for the Dual Output mode of the converter. It is assumed that output capacitor C1 is large enough, three voltages V_{pv}, V_b and V_o are stable during the operation, once the steady state of the converter is reached and that $V_{pv} > V_b$.

State I [t0-t1]: At the time t0, switches SA1 and SB2 are ON, while SA2 and SB1 are OFF. The equations for inductor voltage V_{Lm} and capacitor current i_{c1} are as under.

$$V_p = V_{Lm} = V_{pv} \tag{2}$$

$$\frac{di_{Lm}}{dt} = \frac{V_{pv}}{L_m} \tag{3}$$

$$I_{c1} = I_s - I_o \tag{4}$$

Where V_p is the voltage at the primary side of the transformer while I_s is the current at the secondary turn.

State II [t1-t2]: During this time interval, switches SA1 and SB1 are ON, while SA2 and SB2 are OFF. The equations for inductor voltage V_{Lm} and capacitor current i_{c1} are as under.

$$V_p = V_{Lm} = V_{pv} - V_b \tag{5}$$

$$\frac{di_{Lm}}{dt} = \frac{V_{pv} - V_b}{L_m} \tag{6}$$

$$I_{c1} = I_s - I_o \tag{7}$$

State III [t2-t3]: During this time interval, switches SA2 and SB1 are ON, while SA1 and SB2 are OFF. During state III, battery supplies power to the load alone. If the power generated at the PV panels becomes is not utilized during this state and is isolated from the rest of the system.

In Figure 8, it is shown that the flow of power is from the battery to the load side via the switch SB1. On the secondary side, diodes D2 and D3 are turned on, due to negative voltage appearing at the secondary top side. The equations for inductor voltage V_{Lm} and capacitor current i_{c1} are as under.

$$V_p = V_{Lm} = -V_b \tag{8}$$

$$\frac{di_{Lm}}{dt} = \frac{-V_b}{L_m} \tag{9}$$

$$I_{c1} = I_s - I_o \tag{10}$$

State IV [t3-t0]: During this time interval, switches SA2 and SB2 are ON, while SA1 and SB1 are OFF. This configuration isolates the PV source and battery from each other and the output. Hence, during interval, inductor current i_{Lm} freewheels through the switches SA2 and SB2, while output voltage is maintained by the output capacitor C1.

The equations for inductor voltage V_{Lm} and capacitor current i_{c1} are as under.

$$V_p = V_{Lm} = 0 \tag{11}$$

$$\frac{di_{Lm}}{dx} = \frac{-V_b}{L_m} = 0 \tag{12}$$

$$I_{c1} = I_s - I_o \tag{13}$$

3.3. Volt-Second Balance

If we assume that small ripple approximation, writing the volt-second balance for the primary inductance L_m, the following equation can be obtained.

$$\langle v_{Lm} \rangle = \frac{1}{T}\int_0^T v_{LM}(t)dt \tag{14}$$

$$L_m \frac{di_{L_m}}{dt} = V_{Lm} = D_1(V_{pv}) + D_2(V_{pv} - V_b) + D_3(-V_b) + D_4(0) \tag{15}$$

3.4. ZVS Analysis

The operation of the proposed Full-Bridge Three Port Converter is almost similar to the operation of a conventional full-bridge converter, except for the difference that in our converter, the primary inductance of transformer is also used as an inductor. Zero Voltage Switching can be realized in this converter using inductors, leakage inductances and drain-to-source capacitances (CDS) of the switches, only if the currents i_{Lm} and i_p have the similar signs. This is explained in Figure 9, where we have explained SA2 as an example for the Zero-Voltage Switching.

If the leakage inductance of the transformer is L_K (as shown in Figure 9), then during state III, when SB1 is ON and SA1 OFF, the primary transformer current becomes $i_p= i_{Lm}+nI_s$ and leakage inductance L_K and transformer inductance L_m would release energy that would charge the CDS capacitances of SA1 and SA2.

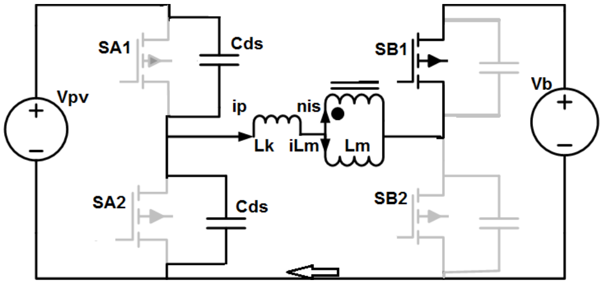


Figure 9-Zero-Voltage Switching in SA2

4. Controller Design

The controller for the converter is designed in Simulink which generates PWM signal for the four switches used in this converter. It is shown in Figure 10. It consists of a MATLAB function block with two inputs, which decide the mode of operation of the converter.

The input 1 initiates mode 1, while input 2 has two parameters, named PV power and Load demand, on the basis of which, mode 2 and 3 are decided. If power at PV is higher than load demand, mode 2 is activated and if load demand is higher than the PV power, mode 3 is activated. On the other side, there are 3 outputs of the MATLAB based controller.

The output SA represents the control signal for switches SA1 and SA2, SB represents the output for switches SB1 and SB2, while the third output C is a control output that controls the delay in the pulses SA and SB. It is to be noted that each pair of switches i.e. SA1/SA2 and SB1/SB2 operate in the complementary fashion, hence one input signal is enough for each pair of switches.

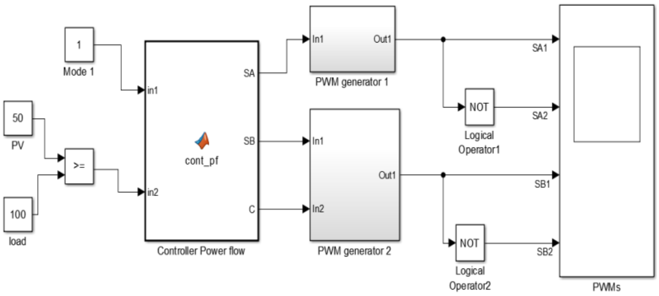


Figure 10- Design of controller for PWM Generation

5. Modeling of PV panel and controller

Photo Voltaic cells provide the cheapest and cleanest energy; however, their efficiency is lower than other modes. In order to utilize the maximum output of the PV arrays, different control mechanism are suggested which are given the name Maximum Power Point Trackers (MPPT). The solution involves the use of I-V characteristic curve of the solar panels and making the PV system operate at the point on the curve that gives maximum power known as Maximum Power Point. In the given I-V curve, as shown in Figure 11, the current value depends on the sunlight evident at the cells. The current can be calculated using the following equation:

$$I_{pv} = I_{ph} - I_0\left[\exp(\frac{V_{pv}+I_{pv}R_s}{V_{th}}) - 1\right] - \frac{V_{pv}+I_{pv}R_s}{R_p} \tag{16}$$

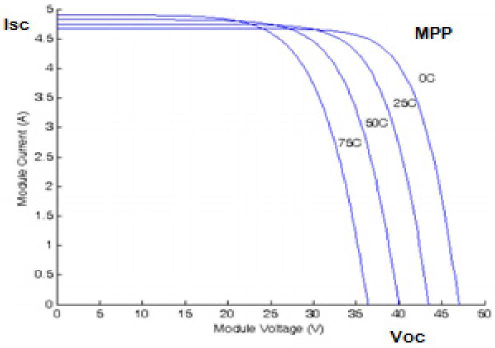


Figure 11-Typical I-V curve with dependence on incident sunlight

The V_{th} refers to the thermal voltage given by following equation:

$$V_{th} = \frac{nKT}{q} \tag{17}$$

As shown in Figure 11, when the voltage of the I-V curve approaches zero, at that time the current from the PV panel is maximum, which is called Short Circuit current (I_{sc}). It depends on the properties of the cells such as absorption and reflection constant and incident light etc. Similarly on the opposite extreme, the point at which the current approaches zero, that voltage as known as Open Circuit Voltage (V_{oc}), which is the maximum attainable voltage by the PV panel. It depends on the material used in the solar panel, e.g. silicon solar cells have V_{oc} at around 730 mV. It is also shown in the figure that Maximum Power Point occurs at the knee of the curve; this is the point where the product of current and voltage gives highest value and hence the best strategy is to keep the panel operating at this point for the given light intensity. The term ‘PV panel’ not only refers to the PV array for converting apparent sunlight into electronic current, but also to the converter (usually buck-boost or SEPIC) installed with the PV array in order to adjust the output voltage of the whole block, to get the maximum power. This converter also contains a separate controller which generates PWM signals for each of its switches. In order to maximize the power generation at the PV panels, different algorithms are used. A certain ‘perturb and observe’ method, as proposed by [9] for the Maximum Power Point tracking, is an efficient scheme for PV panels which ensures maximum output power. This method is used in the control for the PV panels in this converter scheme and can be visualized in Figure 12.

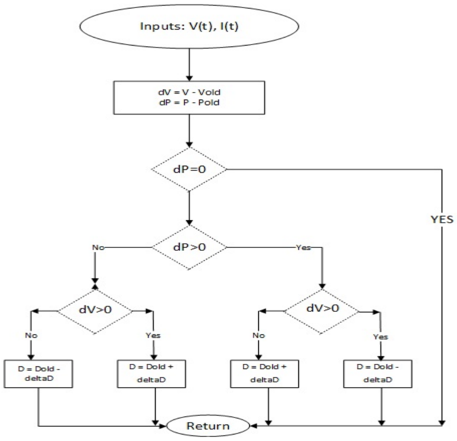


Figure 12-Perturb and Observe Algorithm for PV panels

6. Results and Discussion

6.1 Mode 1 operation

The waveforms for four switches, voltage at primary side v_p, the evolution of load voltage and power are shown below.

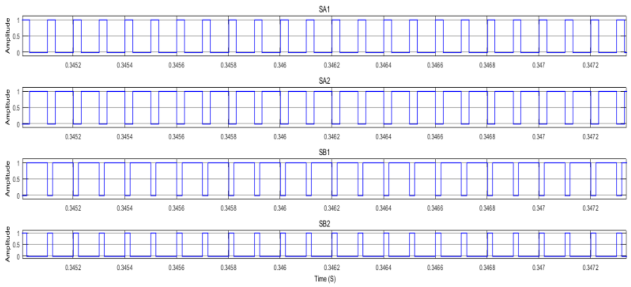


Figure 13- Waveforms for each switch in Mode 1.

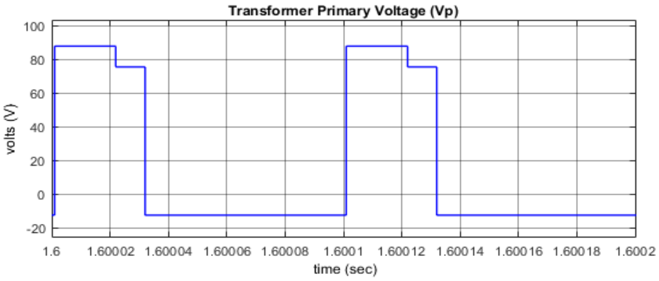


Figure 14-Primary Voltage in Mode 1

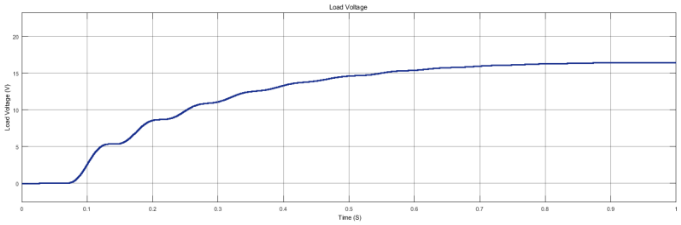


Figure 15- Load voltage curve for mode 1

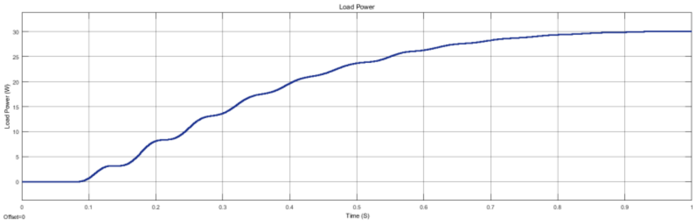


Figure 16-Load power curve for mode 1

As apparent from Figure 13, the switches SA1 and SA2 work on the complementary signal and likewise switches SB1 and SB2 operate on complementary drive signal. Similarly in Figure 14, primary voltage of the transformer is shown. After 1.6 second, the system response can be deemed as steady state response, in which during the first two states, primary voltage across the transformer is positive, while it is negative in the remaining two states, while the total area under the waveform is zero. This is expected, after all; the transformer can forward only alternating voltage across its winding. The three different levels of voltage V_p correspond to each state of the converter in each of which V_p depends on the topology of the converter.

6.2 Mode 2 operation

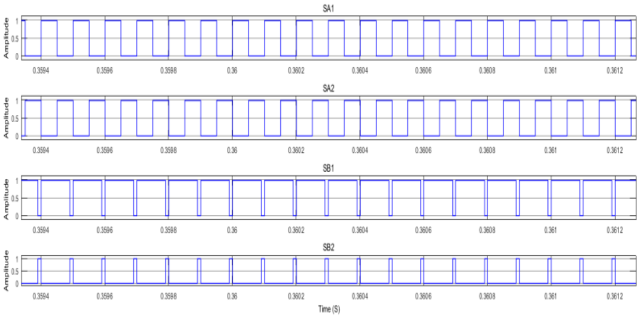


Figure 17-Waveforms for each switch in Mode 2

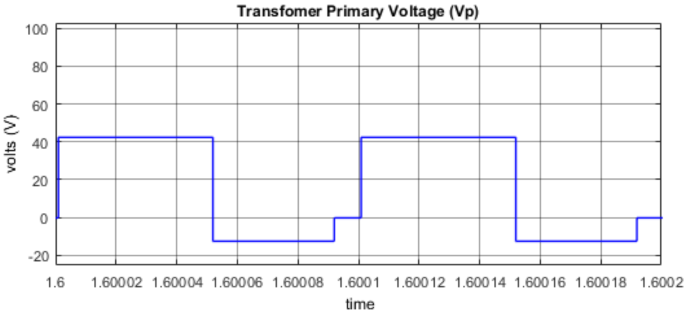


Figure 18- Primary Voltage in Mode 2

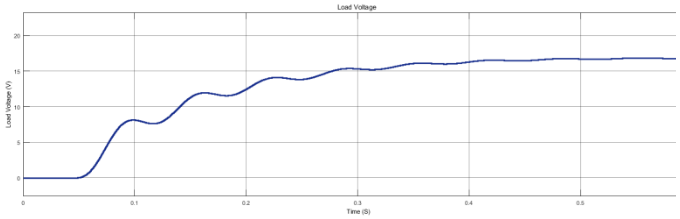


Figure 19- Load Voltage in Mode 2

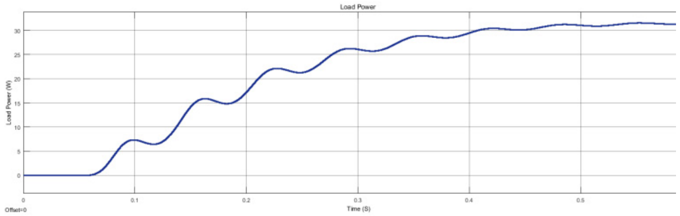


Figure 20- Load Power in Mode 2

The switch waveforms for operation in Mode 2 are shown in Figure 17. Similarly, in Figure 18, the transformer primary voltage V_p is shown. The three different steps in the V_p represent different stages of the transformer. Since in the beginning, V_p equal V_{pv}, hence the first step in the waveform is higher. In the second step, the voltage is below zero because the transformer must balance the voltages to keep its core from saturation. In the third step, the transformer voltage becomes zero, and it gets the time to reset itself. In case of load voltage and power, it can be seen that voltage rises from zero in transient response and stabilizes itself to a value of around 17 V.

6.3 Mode 3 Operation

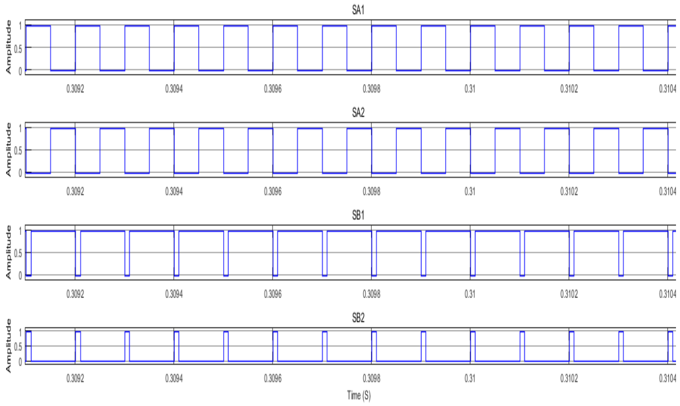


Figure 21- Waveforms for each switch in Mode 3

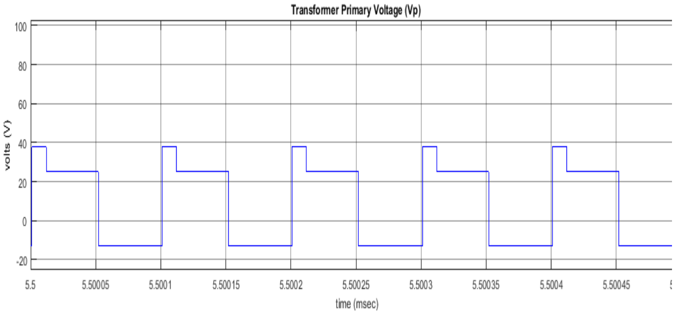


Figure 22- Primary Voltage in Mode 3

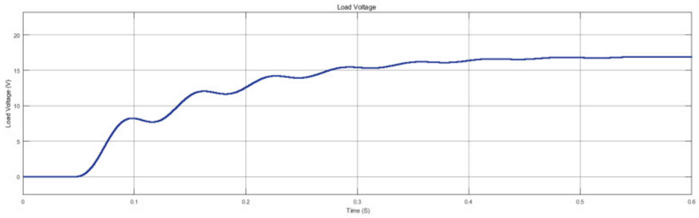


Figure 23-Load Voltage in Mode 3

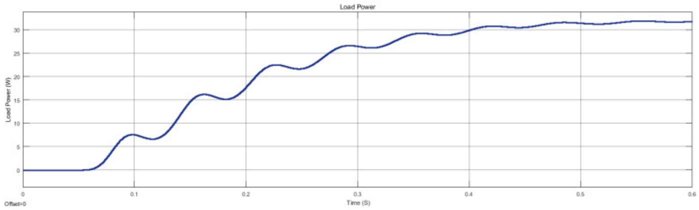


Figure 24- Load power in Mode 3

Similar to the previous two modes, in mode 3 operation, switch waveforms for each pair work in complementary mode, while Figure 22 corresponds to the transformer primary waveform. Each step represents a different mode of the converter, which were discussed previously. Most importantly, if the transformer voltage is positive for some time period, it should be negative for some other time period, in order to avoid saturation of the core.

7. Conclusion

In this paper, full bridge three port converter is discussed. Based on the standard design of full-bridge converter, we have modeled and derived a three port converter. The presented method splits the two switching legs of the full bridge converter into two switching cells with different sources. The presented converters offer the advantages of simple topologies and control, reduced number of devices, and a single-stage power conversion between any of the three ports. They are suitable for renewable power systems that are sourced by solar cells or wind turbines, etc.

The converter can operate in three different modes including Single Input Single Output, Single Input Dual Output and Dual Input Single Output. The operation in each of these modes depends on the conditions at hand, for example, the intensity of sunlight and load demand etc. We have analyzed the state equations, inductor volt-second balance and capacitor charge balance for each state.

Additionally, we have analyzed the Zero-Voltage Switching mechanism in the modeled converter. The transformer inductances such as leakage inductance and magnetizing inductance play crucial role in enabling ZVS mode in the presented converter. Furthermore, we have discussed the PV panel in detail which is the primary source of power in the converter. The PV panel incorporates standard PV cells made of silicon, along with Maximum Power Point Tracker (MPPT) which includes a buck-boost converter with a separate controller to keep the operation of the cells to produce maximum power at the output.

The input to the PV panel includes three parameter to select the mode of operation. We have also discussed the control algorithm called Perturb and Observe for keeping the maximum power. The results of the converters for each of the mode of operation are produced and discussed. These include the key waveforms of switches present in the converter, the voltage at the primary winding of the transformer, output voltage and power waveforms.

The variation of the primary voltage during each state is rationalized.

References

[1] S. J. AL-Chlaihawi, A. Craciunescu and A. G. Al-Gizi, "Power Flow Management in Three Port Converter Using PV Panel with Maximum Power Point Tracker," 2017 10th IEEE International Symposium on Advanced Topics in Electrical Engineering (ATEE 2017), Bucharest, Romania, March 23-25, 2017, pp.585-590, doi:10.1109/ATEE.2017.7905136.

[2] A. D. Napoli, F. Crescimbini, L. Solero, F. Caricchi and F. G. Capponi, "Multiple-input DC–DC power converter for power-flow management in Hyrbid Vehicles," Proceedings of IEEE Industrial Applications Conference, pp. 1578-1585, 2002.

[3] H. Tao, A. Kotsopoulos, J. Duarte and M. Hendrix, "Transformer-coupled multiport ZVS bidirectional dc-dc converter with wide input range," IEEE Transactions on Power Electronics, vol. 23, no. 2, pp. 771-781, 2008.

[4] D. Xu, C. Zhao and H. Fan, "A PWM plus phase-shift control bidirectional dc-dc converter," IEEE Transaction on Power Electronics , vol. 19, no. 5, pp. 666-671, 2004.

[5] Z. Qian, O. Abdel-Rahman, H. Al-Atrash and I. Batarseh, "Modeling and Control of Three-Port DC/DC Converter Interface for Satellite Applications," IEEE TRANSACTIONS ON POWER ELECTRONICS, vol. 25, no. 3, pp. 637-649, 2010.

[6] S. Al-Chlaihawi1, A. Al-Gizi. and A. Craciunescu, "The Analysis and Comparison of Multiport Converter used for Renewable Energy Sources," Advances in Science, Technology and Engineering Systems Journal (ASTESJ), vol. 2, no. 3, pp. 906-912, 2017.

[7] R. W. Erickson and D. Maksimovic, "Converter Circuits," in Fundamentals of Power Electronics, New York, KLUWER ACADEMIC PUBLISHERS, 2001.

[8] H. Wu, K. Sun, R. Chen and H. Hu, "Full-Bridge Three-Port Converters With Wide Input Voltage Range for Renewable Power Systems," IEEE TRANSACTIONS ON POWER ELECTRONICS, vol. 27, no. 9, 2012.

[9] M. L. FLOREA and A. BĂLTĂTANU, "Modeling Photovoltaic Arrays with MPPT Perturb & Observe Algorithm," in The 8th international symposium on advanced topics in electrical engineering (ATEE), Bucharest, Romania, 2013.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 271-282 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Steganography System with Application to Crypto-Currency Cold Storage and Secure Transfer

Michael J. Pelosi*[,1], Nimesh Poudel, Pratap Lamichhane[1], Danyal Badar Soomro[2]

[1] *Computer Science and Technology Faculty, East Central University, 74820, USA*

[2] *School of Software Engineering, Chongqing University, Chongqing 400044, China*

ARTICLE INFO

Article history:
Received: 15 November, 2017
Accepted: 05 February, 2018
Online: 25 April, 2018

Keywords:
Steganography
One-time pad
Steganalysis
Information hiding
Digital forensics

ABSTRACT

In this paper, we introduce and describe a novel approach to adaptive image steganography which is combined with One-Time Pad encryption and demonstrate the software which implements this methodology. Testing using the state-of-the-art steganalysis software tool StegExpose concludes the image hiding is reliably secure and undetectable using reasonably-sized message payloads (≤25% message bits per image pixel; bpp). Payload image file format outputs from the software include PNG, BMP, JP2, JXR, J2K, TIFF, and WEBP. A variety of file output formats is empirically important as most steganalysis programs will only accept PNG, BMP, and possibly JPG, as the file inputs. In this extended reprint, we introduce additional application and discussion regarding cold storage of crypto-currency account and password information, as well as applications for secure transfer in hostile or insecure network circumstances.

1. Introduction

In this paper, we introduce a comprehensive steganography software system and platform framework based on One-Time Pad (OTP) encryption and adaptive steganography technology. We provide usage recommendations and advice guidelines. The system is tested and shown to be resistant to many common steganalysis attacks. In the context of this paper we are assumed advocate of the steganography; someone who may be a political dissident in an oppressive regime, a religiously persecuted individual, a friendly agent engaging in covert communication, or a lawful individual desiring complete communication privacy, among other compelling examples.

2. What is Steganography?

Steganography, the art of invisible communication, is achieved by hiding secret data inside a carrier file such as an image. After hiding the secret data, the carrier file should appear unsuspicious so that the very existence of the embedded data is concealed. A major drawback to encryption is that the existence of the message data is not hidden. Data that has been encrypted, although unreadable, still exists as a suspicion arousing file transfer. If given enough time, once alerted, someone could potentially decrypt the data or derive other intelligence regarding either sender or receiver. A solution to this problem is steganography. This is the ancient art of hiding messages so that they are not detectable.

In steganography, the possible cover carriers are unsuspicious appearing files (images, audio, video, text, or some other digitally representative code) which will hold the hidden information. A message is the information hidden and may be plaintext, cipher text, images, or anything that can be embedded into a bit stream. Together the cover carrier and the embedded message create a stego-carrier. Hiding information may require a stego key which is additional secret information, such as a password or OTP key, required for embedding the information. For example, when a secret message is hidden within a cover image, the resulting product is a stego-image. A possible formula of the process may be represented as: cover medium + embedded message + stego key = stego-medium.

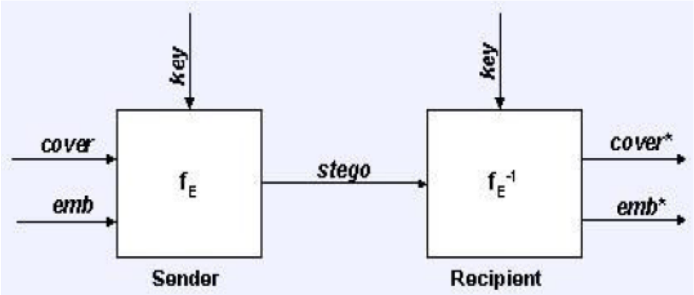


Fig:1 Graphical Version of Steganographic system

*Corresponding Author: Michael J. Pelosi, East Central University, Email: mpelosi@ecok.edu

https://dx.doi.org/10.25046/aj030232

f_E: steganographic function "embedding".
f_E^{-1}: steganographic function "extracting".
cover: cover data in which emb will be hidden.
emb: message file to be hidden.
stego: cover data w

The advantage of steganography is that it can be used to secretly transmit messages without the fact of the transmission being discovered. Often, using encryption might identify the sender or receiver as someone with something to hide. It is believed that steganography was first practiced during the Golden Age in ith the hidden message. Greece. An ancient Greek record describes the practice of melting wax off wax tablets used for writing messages and then inscribing a message in the underlying wood. The wax was then reapplied to the wood, giving the appearance of a new, unused tablet. The resulting tablets could be conveniently transported without anyone suspecting the presence of a message beneath the wax.

2.1. LSB Steganography

The simplest and popular image steganographic method is the least significant bit (LSB) substitution. It embeds messages into cover image by replacing the least significant bits directly. The hiding capacity can be increased by using up to 4 least significant bits (one each for Red, Green, Blue, and Alpha color channels, respectively) in each pixel. It has a common weak point i.e. the sample value changes asymmetrically. When the LSB of cover medium sample value is equal to the message bit, no change is made. Otherwise the value 2n is changed to 2n+1 or 2n+1 is changed to 2n. There are many improvements and modifications that have been proposed to strengthen this technique, such as adaptive techniques that alter payload distribution based on image characteristics. If the message is first encrypted and then embedded, the security is enhanced.

2.2. One-Time Pad

The "one-time pad" encryption algorithm was invented in the early 1900's, and has since been proven as unbreakable. The one-time pad algorithm is derived from a previous cipher called Vernam Cipher, named after Gilbert Vernam. The Vernam Cipher was a cipher that combined a message with a key read from a paper tape or pad. The Vernam Cipher was not unbreakable until Joseph Mauborgne recognized that if the key was completely random the cryptanalytic difficultly would be equal to attempting every possible key (Kahn 1996). Even when trying every possible key, one would still have to review each attempt at decipherment to see if the proper key was used. The unbreakable aspect of the one-time pad comes from two assumptions: the key used is completely random and the key cannot be used more than once. The security of the one-time pad relies on keeping the key secret and using each key only once.

The one-time pad is typically implemented by using exclusive-or (XOR) addition to combine plaintext elements with key elements. An example of this is shown in Figure 2. The key used for encryption is also used for decryption. Applying the same key to the cipher text results in the output of the original plaintext.

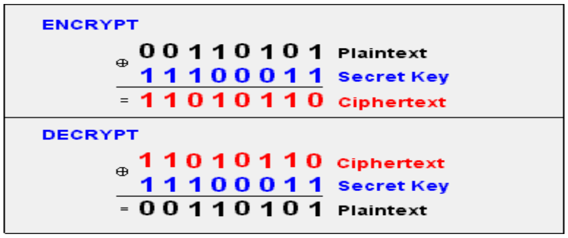

ENCRYPT
0 0 1 1 0 1 0 1 Plaintext
⊕ 1 1 1 0 0 0 1 1 Secret Key
= 1 1 0 1 0 1 1 0 Ciphertext
DECRYPT
1 1 0 1 0 1 1 0 Ciphertext
⊕ 1 1 1 0 0 0 1 1 Secret Key
= 0 0 1 1 0 1 0 1 Plaintext

Fig.2 Examples of a One-time Pad implementation using XOR addition.

OTP is immune even to unlimited resources brute-force attacks. Trying all keys simply yields all possible plaintexts, all equally likely to be the actual plaintext. Even with known plaintext, such as part of the message being known, brute force attacks cannot be used, since an attacker is unable to gain any information about the parts of the key needed to decrypt the rest of the message.

3. Methodology

The following describes the general method implemented in the software for key generation, encryption, embedding, message transfer, and decryption.

- Random image keys are generated using a key generator program. The key generator program generates one-time pad keys that consist of random colored pixels. Each random colored pixel consists of random values for red, green, and blue colors throughout the image. One image key is generated for every message that is intended to be sent.

- To encrypt a message, a cover image and random key image is selected. Each pixel in the cover image is XOR'ed with the key image X, Y coordinate pixel. Each pixel consists of a 32-bit long integer color value. One byte each corresponds to red, green, and blue components, respectively. The XOR'ed pixel values are then adjusted to hide the message. The bytes in the message are divided up into bits — one bit per pixel. The least significant bit (LSB) in the XOR'ed pixel colors are then adjusted to hide the message. Bit values that do not correspond are adjusted (in general 50% of the values will already be set correctly). LSB's for red, green, or blue are selected based on a local pixel variation score, contingent if the sum of the RGB LSB's are even or odd (even corresponds to a 0 bit, odd to a 1 bit).

- At this point, the newly derived color values are XOR'ed once again with the random image key to generate color values very close to the original image. These pixel color values will be used to construct the steganographic image that will be sent to the receiver.

- Ideally at this point, both the original cover image and the senders copy of the random image key can be destroyed (forensically wiped from the hard drive using a file erasure procedure). This is to prevent later detection and statistical comparisons.

- Upon receipt of the steganographic image, the receiver loads the intended image key and XOR's each pixel of the steganographic image with its respective corresponding X, Y pixel in the image key. This will derive a series of bit values that correspond to the plaintext message. The bits can be reassembled into bytes (and later 2-byte Unicode characters) that correspond to the plaintext message.

- The start and end of the message are delimited by randomly chosen 10-character delimiting strings that are embedded as EXIF comments into the random image key by the key generator program. Thus, random message padding is incorporated at the start and end of messages.
- The random image key also contains a random number seed, this is used for the random number generator algorithm in use and starts the generator at the proper sequence start value.

3.1. Random Number Generation

A cryptographically secure pseudo-random number generator (CSPRNG) or cryptographic pseudorandom number generator (CPRNG) is a pseudorandom number generator (PRNG) with properties that make it suitable for use in cryptography. Ideally, the generation of random numbers in CSPRNGs uses entropy obtained from a high-quality source, which might be a hardware random number generator or perhaps unpredictable system processes — though unexpected correlations have been found in several such ostensibly independent processes. Several robust CPRNGs are incorporated into the steganography software.

3.1.1 Mersenne Twister

The Mersenne Twister is a pseudorandom number generator (PRNG). It is by far the most widely used general-purpose PRNG. Its name derives from the fact that its period length is chosen to be a Mersenne prime. The Mersenne Twister was developed in 1997 by Makoto Matsumoto and Takuji Nishimura. It was designed specifically to rectify most of the flaws found in older PRNGs. It was the first PRNG to provide fast generation of high-quality pseudorandom integers. The most commonly used version of the Mersenne Twister algorithm is based on the Mersenne prime 219937−1. The standard library implementation of this, MT19937, uses a 32-bit word length. There is another implementation that uses a 64-bit word length, MT19937-64, that generates a different sequence. The software implements a cryptographically secure version of the Mersenne Twister provided by the algorithm authors Matsumoto and Nishimura.

3.1.2 Other Random Number Generators

Optional random number generator selections included in the OTP-Steg key generator program include the following (each of these can be optionally selected by the user):

- ISAAC— ISAAC (indirection, shift, accumulate, add, and count) is a cryptographically secure pseudorandom number generator and a stream cipher designed by Robert J. Jenkins, Jr. in 1996.
- CryptGenRandom-CryptGenRandom is a cryptographically secure pseudorandom number generator function that is included in Microsoft's Cryptographic Application Programming Interface. In Win32 programs, Microsoft recommends its use anywhere random number generation is needed.
- RtlGenRandom — On a default Windows XP and later install, CryptGenRandom calls into a function named ADVAPI32!RtlGenRandom, which does not require one to load all the CryptAPI classes for usage.
- Rnd () — Standard API random number generator (for research/testing purposes only – it is not cryptographically secure).

3.2 Key Delimiters

Upon key generation, a pair of key delimiters is also randomly chosen of 10 Unicode characters each for the start delimiter and end delimiter, respectively. These are used to indicate to the decryption program exactly where the message starts, and where it ends. Random padding is added to both ends of the message — the start and the end of the message embedded in the payload file. The key delimiters identify where to start the message text, and where to cut it short at the end of the message. These key delimiters are contained in the EXIF image comment data in the key file. No EXIF comment data whatsoever is contained in the payload file. Also, the key delimiter values are utilized for random number generation seed data used for encryption and decryption.

3.3 Expert System to Evaluate and Score Candidate Cover Images

It is well known from the literature that some cover images present much better candidates for steganographic security than others based on image characteristics. Typically, cover images with a high degree of pixel color variation, very few saturated white or black pixels, and few pixels next to each other of the same color, are excellent payload candidates. We implement an expert system to give the software user immediate knowledge of how good a candidate a potential color image is for detection security. We have incorporated a tentative scoring system that evaluates images based on several factors. The output score ranges from 0 to 100%, with greater than 90% score being a good candidate for a cover image. Scores of 80-90% are marginal, and less than 80% are considered not adequate. In the current preliminary version, a peak signal-to-noise ratio (PSNR) versus a solid color image is calculated. This rating is given a weighting of 25% in the overall score. Also, the number of same color pixels next to each other is given a weighting of 25% for up to 5% of the image pixels (in other words, a 5% of the image pixels are same color next to each other, this rating would be zero). Thirdly, a weighted rating of 25% is given to the number of white pixels, up to 5%. The same weighting is also calculated for black pixels. Each of the four factors is combined for the rating from 0% up to 100%. Ideally, a cover image will have zero white pixels, zero black pixels, very few colors next to each other that are the same, and a very high variation in color over comparison to a solid color image. Table 1 below lists the above and additional cover image scoring factors that could be evaluated in an expert system rating scheme.

3.4 Future Security = Small Payloads

To ensure robustness against potential future attacks we have limited payload relative sizes. The high limit for the bits-per-pixel pixel is approximately 25%. And since only half of pixels are typically altered based on the message, this corresponds to a practical limit of about 12.5%-pixel alteration. By limiting the pixel bit payload, it quite robustly limits detectability now and in the future.

Table 1. Potential Candidate Image Scoring Factors.

Factor	Description	Value
PSNR over solid color	Peak signal-to-noise ratio of image to solid color image.	Higher values are better.
Percentage of saturated colors	Portion of the image that is either all-white or all black.	Lower values are better.
Percentage of nearby same colors	Portion of image that has neighboring pixels of the same color.	Lower values are better.
Randomness of LSB's	Measures of randomness of the distribution of the significant bits.	Higher randomness is better.
Random RGB LSB distribution	Randomness of each color channel.	Higher values are better.
RS test on Cover Image	Clean RS test on cover image.	Lower values are better — indicates less probability of a threshold being reached after encoding.
Chi-squared test on Cover Image	Clean Chi-square test on cover image.	Lower values are better.
Pure Photograph	Photo has not previously been compression encoded using algorithm like JPEG.	Straight from a high-quality digital camera is best.
Original Photograph	No other copies of the photo exist in clean or altered state that can be used for comparison.	Known source and originality is best here.
Dimensions	It is well known that extremely large images have less pixel color variation and steganography here is more easily detected.	Approximate pixel dimensions of images frequently found on the Internet are best — about 1600?200 or less pixels.

Extremely advanced statistical detection techniques are being promulgated that are improving the odds of successfully detecting steganography efforts. There is no guarantee that these steganalysis efforts will not double or triple in effectiveness in the next few years. As a safety measure and margin of security, payload size is strictly limited by the software to an amount that should be reasonably safe for the foreseeable future. This equals future security for messages that may be encrypted today and subsequently intercepted and archived for several years for later decipherment.

3.5 Steganalysis

Steganalysis is "the process of detecting steganography by looking at variances between bit patterns and statistical norms". It is the art of discovering and revealing covert messages. The goal of steganalysis is to identify suspected information streams, determine whether or not they have hidden messages encoded into them, and, if possible, recover the hidden information. Unlike cryptanalysis, where it is evident that intercepted encrypted data contains a message, steganalysis generally starts with several suspect information streams but uncertainty whether any of these contain hidden message. The steganalyst starts by reducing the set of suspect information streams to a subset of most likely altered information streams. This is usually done with statistical analysis using advanced statistics techniques.

Analyzing repetitive patterns may reveal the identification of a steganography tool or hidden information. To inspect these patterns an approach is to compare the original cover image with the stego image and note visible differences. This is called a known-carrier attack. By comparing numerous images, it is possible that patterns emerge as signatures to a steganography tool. Another visual clue to the presence of hidden information is padding or cropping of an image. With some steganographic tools if an image does not fit into a fixed size it is cropped or padded with black spaces. There may also be a difference in the file size between the stego-image and the cover image. Another indicator is a large increase or decrease in the number of unique colors, or colors in a palette which increase incrementally rather than randomly. These are just examples among the many published and effective approaches.

StegExpose is a steganalysis tool specialized in detecting LSB (least significant bit) steganography in lossless images such as PNG and BMP. It has a command line interface and is designed to analyze images in bulk while providing reporting capabilities and customization which is comprehensible for non-forensic experts. The StegExpose rating algorithm is derived from an intelligent and thoroughly tested combination of pre-existing pixel based steganalysis methods including Sample Pairs by Dumitrescu (2003), RS Analysis by Fridrich (2001), Chi-Square Attack by Westfeld (2000) and Primary Sets by Dumitrescu (2002). In addition to detecting the presence of steganography, StegExpose also features the quantitative steganalysis (determining the length of the hidden message). We utilize StegExpose for steganalysis to test the software reliability in hiding messages effectively from steganalysis.

3.6 *Performance Speed and Robust Steganography*

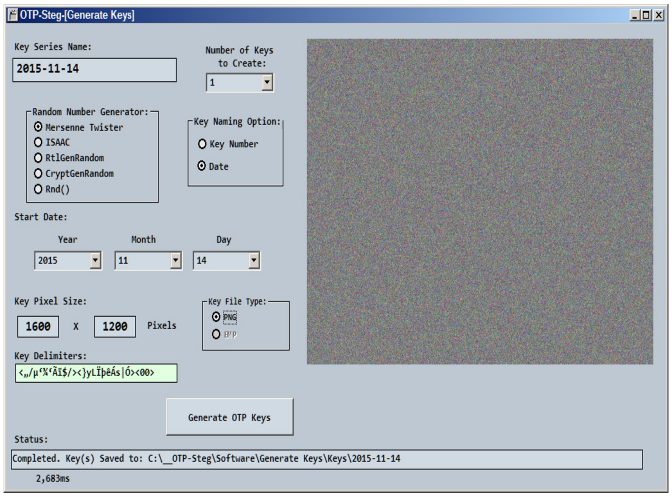
The straightforwardness of the embedding algorithm has also resulted in the good embedding speed. Most of the files worked with using the software take less than 60 to 90 seconds for embedding. Typically, about 30 seconds is required for decryption. Since the bit per pixel payload is less than 25%, the random number generator does not have to repeatedly struggle to find empty pixels that have not been previously encoded.

4. Software Implementation

The software implementation consists of three executable files: a key generator program, an encryption program, and a decryption program. The encryption program has image analysis functions and windows built-in to aid in cover image categorization.

4.1. *Key Generation*

A screenshot of the key generation program shown below. The key generation program constructs image keys of random colored pixels according to the user preference for size and file naming. Up to five previously discussed random number generators can be chosen from to generate the random colored pixels.


OTP-Steg-[Generate Keys]

Key Series Name: 2015-11-14

Number of Keys to Create: 1

Random Number Generator:
● Mersenne Twister
○ ISAAC
○ RtlGenRandom
○ CryptGenRandom
○ Rnd()

Key Naming Option:
○ Key Number
● Date

Start Date:
Year 2015 Month 11 Day 14

Key Pixel Size: 1600 X 1200 Pixels

Key File Type:
● PNG
○ BMP

Key Delimiters: <„/µ‘¾‘Ãï$/><}yLÏþêÁs|Ó><00>

Status:
Completed. Key(s) Saved to: C:\__OTP-Steg\Software\Generate Keys\Keys\2015-11-14
2,683ms

Figure 4. Key Generator executable program

4.2. *Encryption and Embedding*

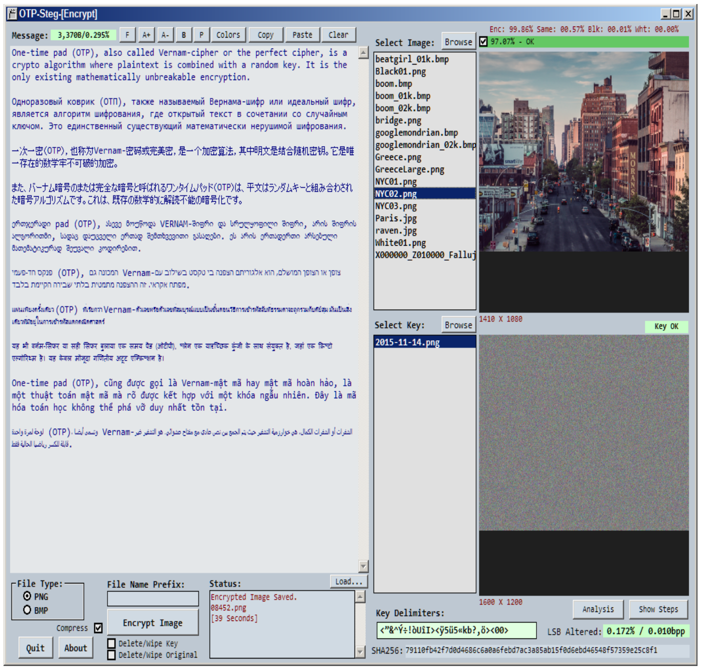
The encryption program has by wide margin the most features and functionality built-in. Also included are functions for deleting and forensically wiping the key file used for encryption, as well as the original cover image. By comparing the encrypted payload file to an original cover image, steganography could easily be detected as the difference between the two images. It is an extremely important security measure to eliminate the original cover image and key as soon as possible after encryption takes place.

4.3. *Message Hash Value*

SHA-256 values of the message are calculated in both the encryption and decryption steps. In the encryption step, the hashed value is incorporated into the end of the message string. Upon decryption, the transmitted hash value is compared to the hashed value of the decrypted message, and displayed in the decryption program graphical user interface. If the values match the user is informed that the message has not been altered in any way since it was encrypted by the sender. This is also a double check that successful decryption has taken place and the message is authentic.

4.4. *Text Compression:*

Compression prior to embedding the message generally reduces message size by 50 to 80%. The zLib compression library DLL is utilized and called as a function within both the encrypt and decrypt programs. The result makes encryption and decryption quicker and also has the benefit of reducing the bit per pixel payload size in the cover image, increasing security against detection.


OTP-Steg-[Encrypt]

Message: 3,370B/0.295%

One-time pad (OTP), also called Vernam-cipher or the perfect cipher, is a crypto algorithm where plaintext is combined with a random key. It is the only existing mathematically unbreakable encryption.

Одноразовый коврик (ОТП), также называемый Вернама-шифр или идеальный шифр, является алгоритм шифрования, где открытый текст в сочетании со случайным ключом. Это единственный существующий математически нерушимой шифрования.

一次一密(OTP),也称为Vernam-密码或完美密,是一个加密算法,其中明文是结合随机密钥。它是唯一存在的数学牢不可破的加密。

また、バーナム暗号のまたは完全な暗号と呼ばれるワンタイムパッド(OTP)は、平文はランダムキーと組み合わされた暗号アルゴリズムです。これは、既存の数学的に解読不能の暗号化です。

ერთჯერადი pad (OTP), ასევე მოუწოდა VERNAM-შიფრი და სრულყოფილი შიფრი, არის შიფრის ალგორითმი, სადაც დაუცველი ერთად შემთხვევითი გასაღები. ეს არის ერთადერთი არსებული მათემატიკურად შეუვალი კოდირებით.

פנקס חד-פעמי (OTP), המכונה גם Vernam-צופן או הצופן המושלם, הוא אלגוריתם הצפנה בי טקסט בשילוב עם
מפתח אקראי. זה ההצפנה מתמטית בלתי שבירה הקיימת בלבד.

แผ่นเพียงครั้งเดียว (OTP) ที่เรียกว่า Vernam-ตัวเลขหรือตัวเลขที่สมบูรณ์แบบเป็นขั้นตอนวิธีการเข้ารหัสลับที่ธรรมดาจะถูกรวมกับคีย์สุ่ม มันเป็นสิ่งเดียวที่มีอยู่ในการเข้ารหัสแตกคณิตศาสตร์

यह भी वर्नम-सिफर या सही सिफर बुलाया एक समय पैड (ओटीपी), प्लेन एक यादृच्छिक कुंजी के साथ संयुक्त है, जहां एक क्रिप्टो एल्गोरिथ्म है। यह केवल मौजूदा गणितीय अटूट एन्क्रिप्शन है।

One-time pad (OTP), cũng được gọi là Vernam-mật mã hay mật mã hoàn hảo, là một thuật toán mật mã mà rõ được kết hợp với một khóa ngẫu nhiên. Đây là mã hóa toán học không thể phá vỡ duy nhất tồn tại.

لوحة لمرة واحدة (OTP)، وتسمى أيضا Vernam-الشفرات أو الشفرات الكمال، هي خوارزمية التشفير حيث يتم الجمع بين نص عادي مع مفتاح عشوائي. هو التشفير
غير قابلة للكسر رياضيا الحالية فقط.

Select Image:
beatgirl_01k.bmp
Black01.png
boom.bmp
boom_01k.bmp
boom_02k.bmp
bridge.png
googlemondrian.bmp
googlemondrian_02k.bmp
Greece.png
GreeceLarge.png
NYC01.png
NYC02.png
NYC03.png
Paris.jpg
raven.jpg
White01.png
X000000_Z010000_Falluj

Enc: 99.86% Same: 00.57% Blk: 00.01% Wht: 00.00%
☑ 97.07% - OK
1410 X 1080

Select Key:
2015-11-14.png

Key OK
1600 X 1200

File Type:
● PNG
○ BMP

Compress ☑
File Name Prefix:
☐ Delete/Wipe Key
☐ Delete/Wipe Original

Status:
Encrypted Image Saved.
08452.png
[39 Seconds]

Key Delimiters: <"&^Ý÷!òUîI><ÿ5ü5«kb?,ö><00>
LSB Altered: 0.172% / 0.010bpp
SHA256: 79110fb42f7d0d4686c6a0a6febd7ac3a85ab15f0d6ebd46548f57359e25c8f1

Figure 5. Encrypt/Embed executable program

4.5. *Cover and key File Deletion and forensic Data Wiping*

File wiping utilities are used to delete individual files from an operating system mounted drive. The advantage of file wiping utilities is that they can accomplish their task in a relatively short amount of time as opposed to disk cleaning utilities which take much longer and must be run separately.

4.6. *Built-In Cover and Stego Image Analysis Tools*

Several image statistical analysis features are built into the encryption program. Peak signal-to-noise ratio (PSNR), RS, Chi-Squared, LSB visual analysis, color changes, and color variations. The distortion in the stego-image can be measured by parameters such as mean square error (MSE) and peak signal-to-noise ratio (PSNR) (see Equation 1 and 2 below), and correlation. The lesser distortion means, the lesser MSE, but higher PSNR. If p is a M 謀 grayscale image and q is its stego-image, then the MSE and PSNR values are computed using (1) and (2). For color images a pixel comprises 3 or 4 bytes. Each byte can be treated as a pixel and the same equations can be used to calculate the MSE and PSNR.

$$\mathrm{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(p_{ij} - q_{ij}\right)^2 \tag{1}$$

$$\mathrm{PSNR} = 10 \times \log_{10} \frac{C_{max}^2}{\mathrm{MSE}} \tag{2}$$

The software image analysis window in the encryption program is shown below. Using this window, several operations can be performed to estimate effectiveness of message embedding. Least significant bit (LSB) color values can be investigated visually. Shown on the right of the analysis window are the least significant bit values of the photo on the left. If the least significant bit for red, green, and/or blue is set, this color is added at full intensity to the respective pixel in the image on the right. In Figure 7, individual least significant bit color values can be investigated as well in the red, green, and blue channels. In this image, it is obvious there is a problem with the blue channel — the sky has full intensity for all values. Encoding message data here would be risky, as the pixel variation is nonexistent. Steganography would be very easily detected by any encoding in this area. As a result, the software spreads out the message embedding adaptively, and ignores the blue channel in the area of the sky. Since the red channel has the most random variation throughout the image, it carries the largest brunt of the payload, leveraging its random character throughout the image.

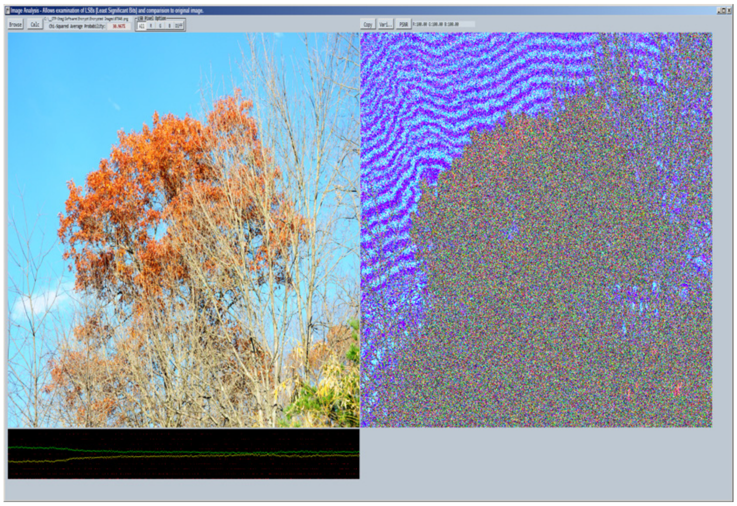


Image Analysis - Allows examination of LSBs (Least Significant Bits) and comparision to original image.

Chi-Squared Average Probability: 30.967%

R:100.00 G:100.00 B:100.00

Figure 6. Image Analysis window, cover image on left, LSB analysis on the right

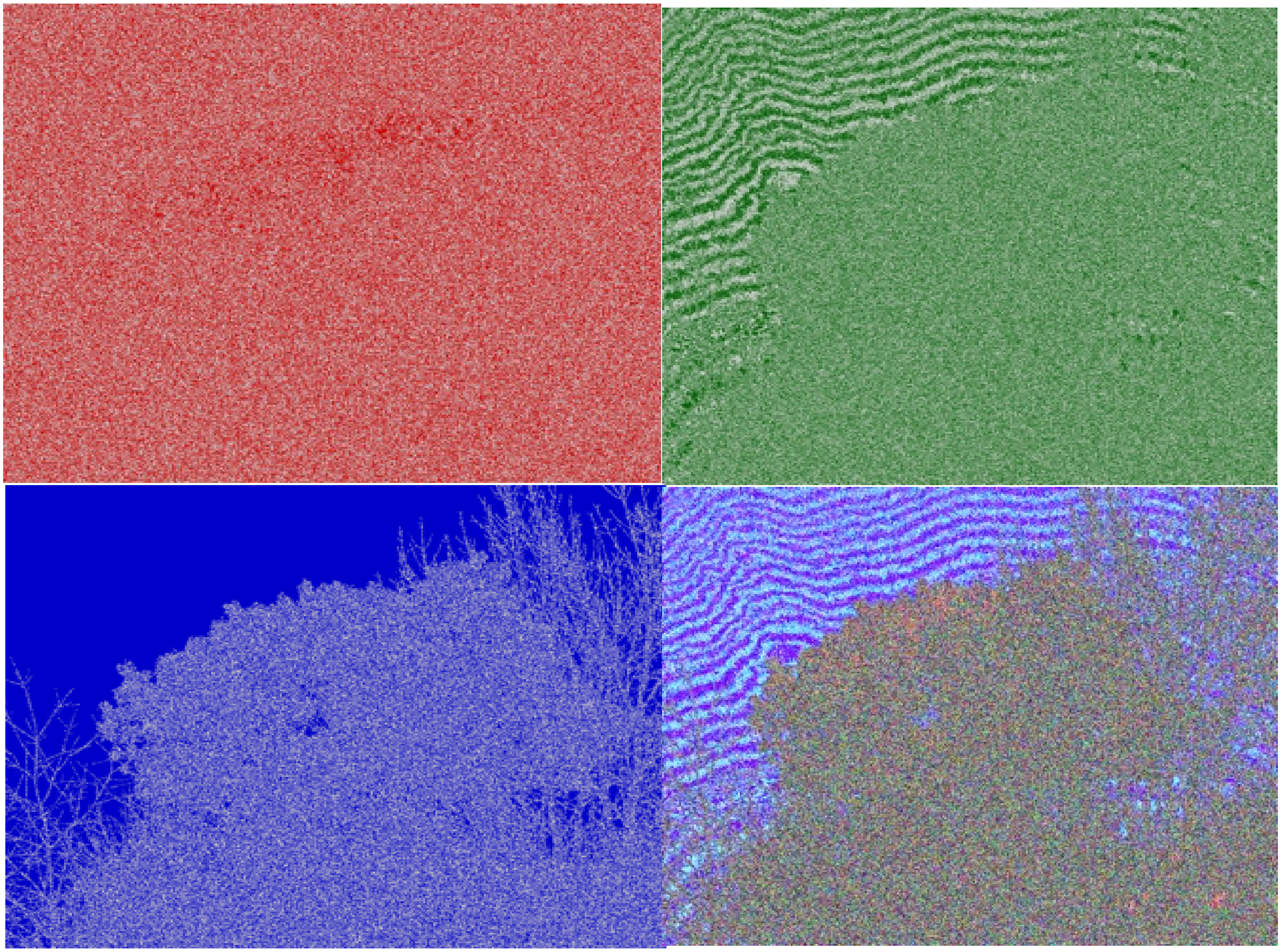


Figure 7. Analysis LSB color analysis graphics (Red, Green, Blue, All Colors).

Shown below are the variations in the red channel, the blue channel in Figure 9 shows the lack of variation in the sky area.

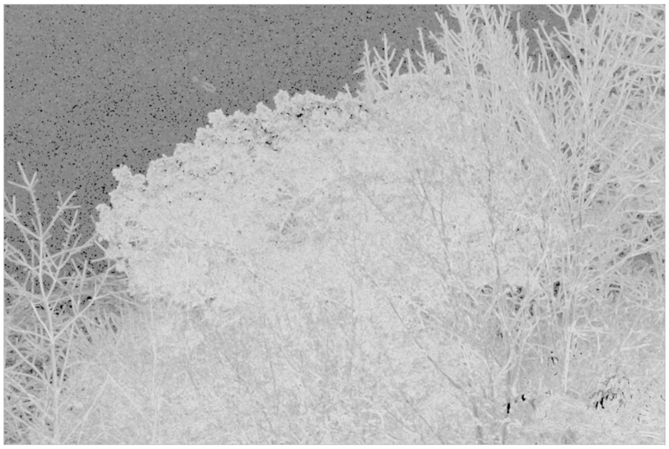
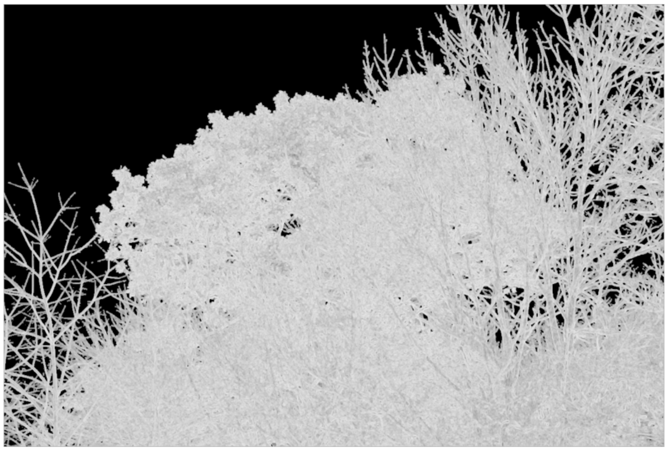


Figure 8. Red channel variation score (normalized to 0-255).



In Figure 10 below, the pixel least significant bit encodings are shown. Notice that in the area of the sky, only red pixels are encoded in the least significant bit. Other areas of the image vary between green and blue embedding depending on which color has the most variation in that pixel general area. Figure 11 shows a blowup of the pixel least significant coding in the area of the transition between the trees and the sky. Notice that the pixel encodings shift from primarily blue-green to the red color at this transition.

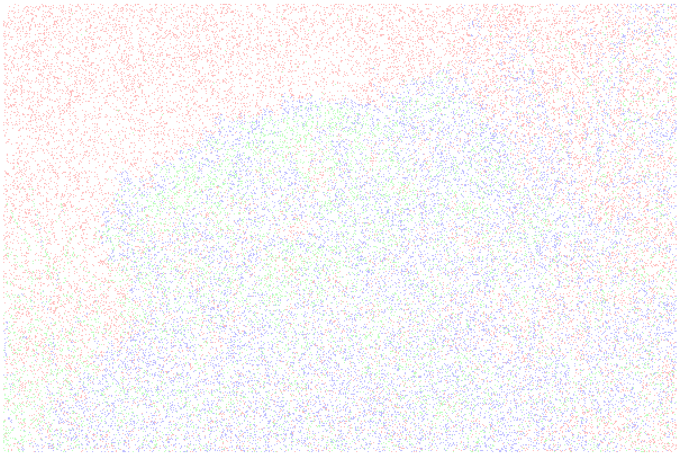
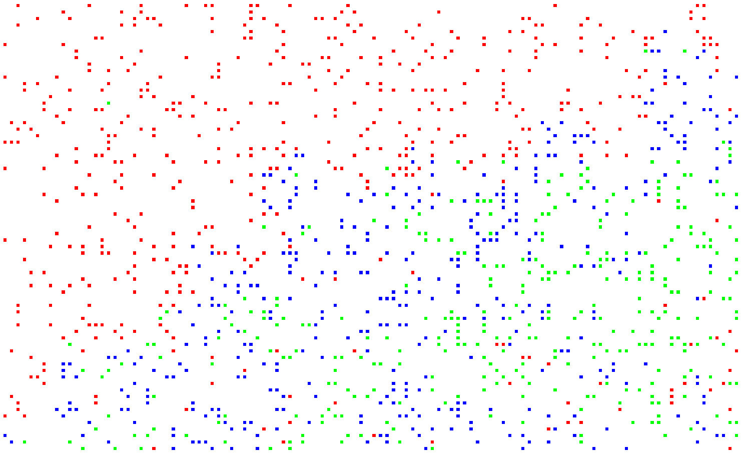
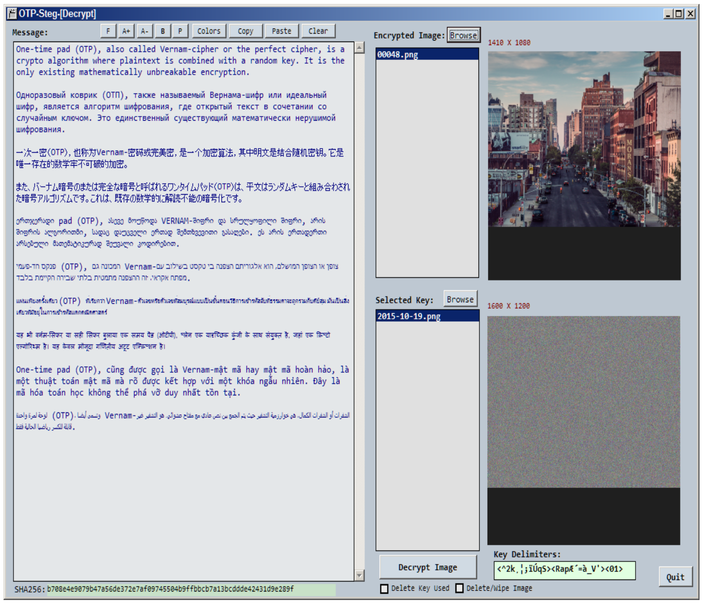


Figure 10. Pixel LSB modification encodings (Red, Green, or Blue).



4.7. Decryption

The decryption process largely reverses the encryption process using the decryption program. A SHA-256 hash value is computed from the decrypted message and compared to the hashed value contained within the payload file. If the two values matches, the hashed value is presented with a green colored background. If not, the background is reddish. A green value indicates to the receiver that the message has not been altered in any way since it was written.



One-time pad (OTP), also called Vernam-cipher or the perfect cipher, is a crypto algorithm where plaintext is combined with a random key. It is the only existing mathematically unbreakable encryption.

Одноразовый коврик (ОТП), также называемый Вернама-шифр или идеальный шифр, является алгоритм шифрования, где открытый текст в сочетании со случайным ключом. Это единственный существующий математически нерушимой шифрования.

一次一密（OTP），也称为Vernam-密码或完美密，是一个加密算法，其中明文是结合随机密钥。它是唯一存在的数学牢不可破的加密。

また、バーナム暗号のまたは完全な暗号と呼ばれるワンタイムパッド（OTP）は、平文はランダムキーと組み合わされた暗号アルゴリズムです。これは、既存の数学的に解読不能の暗号化です。

ერთჯერადი pad (OTP), ასევე მოუწოდა VERNAM-შიფრი და სრულყოფილი შიფრი, არის შიფრის ალგორითმი, სადაც დაუცველი ერთად შემთხვევითი გასაღები. ეს არის ერთადერთი არსებული მათემატიკურად შეუვალი კოდირებით.

פנקס חד-פעמי (OTP), המכונה גם Vernam-צופן או הצופן המושלם, הוא אלגוריתם הצפנה בי טקסט בשילוב עם מפתח אקראי. זה ההצפנה מתמטית בלתי שבירה הקיימת בלבד.

แผ่นเดียวครั้งเดียว (OTP) ที่เรียกว่า Vernam-ตัวเลขหรือตัวเลขที่สมบูรณ์แบบเป็นขั้นตอนวิธีการเข้ารหัสลับที่ธรรมดาจะถูกรวมกับคีย์สุ่ม มันเป็นสิ่งเดียวที่มีอยู่ในการเข้ารหัสแตกคณิตศาสตร์

यह भी वर्नम-सिफर या सही सिफर बुलाया एक समय पैड (ओटीपी), प्लेन एक यादृच्छिक कुंजी के साथ संयुक्त है, जहां एक क्रिप्टो एल्गोरिथ्म है। यह केवल मौजूदा गणितीय अटूट एन्क्रिप्शन है।

One-time pad (OTP), cũng được gọi là Vernam-mật mã hay mật mã hoàn hảo, là một thuật toán mật mã mà rõ được kết hợp với một khóa ngẫu nhiên. Đây là mã hóa toán học không thể phá vỡ duy nhất tồn tại.

لوحة لمرة واحدة (OTP)، وتسمى أيضا Vernam-الشفرات أو الشفرات الكمال، هي خوارزمية التشفير حيث يتم الجمع بين نص عادي مع مفتاح عشوائي. هو التشفير غير قابلة للكسر رياضيا الحالية فقط.

SHA256: b708e4e9079b47a56de372e7af09745504b9ffbbcb7a13bcddde42431d9e289f

Encrypted Image: 00048.png

Selected Key: 2015-10-19.png

Key Delimiters: <^2k.¦;ïÚqS><RapÆ´=à_V'><01>

□ Delete Key Used □ Delete/Wipe Image

Figure 12. Decrypt/Extract executable program

5.1. Photo Selection

There are several general guidelines for photo image selection to increase security. Original photos taken with the user's own camera should be selected as cover images. This is to ensure that the duplicate of the original photo does not exist somewhere on the Internet for comparison. The photo should never have previously been encoded to JPEG to ensure full CMOS pixel sensor color variations throughout the image. As mentioned previously, once these criteria are satisfied, the user can evaluate an encodability score that is calculated by the encryption program that ranges from 0 to 100%.

5.2. Encodability Score

The user should choose in general images that score above 90% for encodability to enhance steganalysis security. The following is the weighting breakdown for the encodability score:

25% Overall PSNR (dB) variation score (0-100%) (more color variation = higher score)

25% Same colors next to each other (0-5%) (less same colors = higher score)

25% Black pixels (0-5%) (less black = higher score)

25% White pixels (0-5%) (less white = higher score)

100% - (100-90% = OK, 90-80% = Marginal, <80% = Unacceptable).

5.3. Recommended Steganographic Practices

Table 2. Recommended Steganographic Practices.

No.	Practice	Description
1	Software Operation	Steganography software should be operated on a computer that is not connected to any network or the Internet. Files should be transferred using write only media such as DVD or CD, or less securely by USB drive.
2	Original Photos	Only original photos taken by the users high-quality camera should be considered as cover images. This is to ensure that the image does not have duplicates available on the Internet. Use RAW (original camera file format) images where possible. The software directly accepts all RAW image file types including Nikon, Canon, Sony, etc.
3	Software USB Loaded	The software should be run off of a USB drive plugged into the isolated computer. Further, USB drive containing software, keys, and cover images, should be located separately from the isolated computer in a safe and secure location.
4	Isolated Computers	The isolated computer used to run the software should be well secured and not networked in any way. The operating system should be directly installed from DVD, and antivirus and checks for malware should be regularly run to ensure there is no keystroke loggers, rootkits, or other security compromises installed.
5	"To" and "From" Keys	Both sender and receiver should have their own set of unique keys. Sender A to B, and B to A, each use their own one-way key series. This is to prevent key reuse. Each key must be used only one time, and one time only. Security using OTP depends on this precept.
6	Exchanging Keys	Key exchange should take place upon *physical meeting* using write only media such as DVD or CD. Key exchange must not take place over a network. Keys should be securely generated on isolated computers. Keys must be stored on removable USB drives separate from the isolated computer.
7	Deleting Files	All files including cover image files and key files should be forensically deleted and wiped once used. Forensic wiping utilities in the encryption and decryption programs can be used for this purpose. Wiping consists of randomly overwriting the previous file seven times with random data.
8	Sending Encrypted Files	Encrypted files should be sent as anonymously as possible. Direct email exchange should be avoided. A preferable alternative is to upload files periodically to gallery websites which have potentially thousands of viewers and downloaders daily. Identifying the specific receiver will be difficult in this situation. Each sender should upload to a different anonymous gallery.
9	Monitoring Windows Vulnerabilities	It should be known that just the act of plugging in a USB drive into a Windows computer creates a digital trail throughout the system registry. Installing software using a setup program also creates numerous records within the operating system registry. As a result, the software should be run off of a USB drive without running a separate install/setup utility. Windows must be isolated off of any network to ensure malware is not installed.

10	Malware	Malware can cause a compromise in the steganography system at any time. A keystroke logger that is uploading typed messages is an instant fail. Users must be extremely cautious and knowledgeable about potential malware threats before using the software. In particular, any networked computer presents a point of vulnerability — the software and keys must never be used here. Only transfer of files previously encrypted on an isolated computer can be conducted over a network.
11	Usage Limitations	The biggest limitation is the human factor. Operational security must be observed that all times in addition to technical security. This means aggressive securing of the USB drive use for the software and keys, as well as limited knowledge by parties involved. People should be informed on a need to know basis only.
13	Encrypted Keys	For further security, keys can be encrypted for storage. As a result they will not be able to be used unless the user has knowledge of the encryption key.
14	Wipe Original Photos	Original photos must be deleted and erased from the camera, storage medium, and USB drive as soon as possible after they are used.
15	Wipe Used Keys	Keys must be deleted and erased as soon as they are used.
16	Internet Computer "Clean"	The computer connected to the Internet must be clean of viruses and malware or keystroke loggers. Special care must be taken in this area.
17	Camera Secured	The photo CMOS sensor output profile can be mapped to a particular individual camera. Photo sent on the Internet can be matched up to the users camera. As a result an effort should be made to keep the camera secure.
18	File Upload Galleries	File upload galleries should be selected for anonymity and high traffic volume.
19	Carefully Selected Cover Images	Cover images should be conforming to high encodability statistics and originality. Also they should be of subject matter that will not raise any suspicions.
20	Image File Format	Various image file formats can be chosen, leveraging the fact that steganalysis software will not run on many different types of image file types. Take advantage of other lossless file formats besides PNG and BMP such as TIF, J2K, EXR, WEBP, and JXR.
21	Must-Dos	1: Keep software and keys in secure locations on USB drives. 2: Use software on isolated computer not connected to Internet. 3: Use keys and photos only once and be sure to erase files as soon as possible, especially original cover images and keys.

5.4. *Steganographic Communication Security State Level Estimation*

We envision certain levels of steganographic communication security levels that correspondents can use for planning, analysis, and security estimations. Thresholds can be established for protective measures using these security level guidelines.

Table 3. Notional Steganographic Security Levels.

Security Level	**Name**	**Description**	**Impact**
10	Secure	Communication commencing securely. Operational security and human threat and insider threat must be strongly monitored and evaluated here.	None-success
9	Communication Suspected	Authorities suspect communication without knowledge of sender and receiver.	Low
8	Steg Statistically Detected	Positive steganography screening results indicating further investigation.	Moderate

7	Internet Computer Searched	If proper security measures recommended previously are followed, nothing should be derived. Duress codeword should be immediately used and communication ceased.	Moderate
6	Transmitted Files Discovered	If proper procedures are used, locating these files should not present much evidence.	Moderate
5	Software Computer Discovered	Traces of software use should be detectable in Windows registry.	High
4	Steg Known	Investigators conclude illicit communication has taken place, without acquiring USB drive(s).	High
	USB Drive discovered	User should make efforts to inform receiver communication is compromised.	Severe
2	Software Discovered/ Acquired	Knowledge of message text should be assumed at this point.	Severe
1	Key(s) acquired	Complete security compromise.	Severe
0	Suspect Detained	All communicating parties should make efforts to destroy any remaining evidence.	Failure

Communicators should have a procedure in place to indicate ceasing of messages and also to destroy related software and keys systematically.

Steganographers should consider incorporating a duress codeword into their communication security protocol. The duress code word should be a predetermined word or phrase that indicates to the receiving party that communication security has been compromised. For example, capture by authorities may have created a situation where one party is succumbing to efforts to be "turned". The duress code word indicates such a situation and should be carefully chosen to arouse no suspicion should authorities have knowledge of its inclusion in a "trap" message.

5.5. *Software Availability*

The software is available as a free educational and research download to be used for digital forensics education and related projects. Please feel free to use the software for your own educational and research purposes. The software can be acquired here: http://199.175.52.196/OTP-Steg/.

6. Steganalysis Results

StegExpose is a Java based steganalysis tool heavily geared towards bulk analysis of lossless images. It is a steganalysis tool specialized in detecting LSB (least significant bit) steganography in lossless images such as PNG and BMP. It has a command line interface and is designed to analyze images in bulk while providing reporting capabilities and customization which is comprehensible for non-forensic experts. The *StegExpose* rating algorithm is derived from an intelligent and thoroughly tested combination of pre-existing pixel based steganalysis methods Two new fusion detectors, standard and fast fusion were derived from four well known steganalysis methods and successfully implemented in the tool. Standard fusion is more accurate than any of the component detectors it is derived from.

The following LSB steganalysis methods have been incorporated in *StegExpose*. RS analysis (Fridrich, Goljan, and Du 2001) detects randomly scattered LSB embedding in grayscale and color images by inspecting the differences in the number of regular and singular groups for the LSB and "shifted" LSB plane. Sample pair analysis (Dumitrescu, Wu, and Wang 2003) is based on a finite state machine whose states are selected multisets of sample pairs called trace multisets (Dumitrescu, Wu, and Wang 2003). The chi-square attack (Westfeld and Piltzmann 2000) is a statistical analysis of pairs of values (PoV's) exchanged during LSB embedding. PoV's are groups of binary values within a object's LSB's. Primary sets (Dumitrescu, Wu, and Memon 2002) is based on a statistical identity related to certain sets of pixels in an image. The difference histogram analysis (Zhang and Ping 2003) is a statistical attack on an image's histogram, measuring the correlation between the least significant and all other bit planes. Two new fusion detectors, standard and fast fusion, were derived from four well known steganalysis methods and successfully implemented in the tool. The standard fusion test is more accurate than any of the component detectors it is derived from.

StegExpose (the free open source download), was run on a batch of 27 image files that were encoded using the *OTP-Steg* software. Test specifications and results are shown below. None of the embedded files were detectable above the preset default threshold. Standard fusion was the test run which consists of all of the specific steganalysis tests mentioned above.

Table 4. StegExpose Steganalysis Test Specifications.

Test Spec	**Description**
Embedded Text File:	U.S. Constitution; 52,782 Bytes Unicode (422,256 bits)
Images:	27 Various landscape PNG photos, 1200?97 pixels (956,400 pixels) Nikon D90.

Uncompressed:	Approximate Bits per Pixel (bpp) 0.442 bpp
Compressed (zLib):	Approximate Bits per Pixel (bpp) 0.086 bpp
Alterations:	1.445% LSBs altered, 4.335% of pixels
File Archive:	http://199.175.52.196/OTP-Steg/USConstitution/

Table 5. StegExpose Steganalysis Test Results using "Standard Fusion" test.

File name	Above Stego Threshold?	Primary Sets	Chi Square	Sample Pairs	RS analysis	Fusion (mean)
00247.png	***FALSE***	0.023408176	0.00353364	null	0.02018579	0.01570926
02155.png	***FALSE***	0.068625394	0.01936033	null	0.04494632	0.04431068
02664.png	***FALSE***	NaN	5.03E-13	null	0.08658666	0.04329333
03090.png	***FALSE***	NaN	0.0037011	null	0.23737088	0.12053603
03164.png	***FALSE***	0.136200359	0	null	0.02282364	0.05300800
03504.png	***FALSE***	NaN	0.00363950	null	0.19724031	0.10043991
03509.png	***FALSE***	0.120022314	0.00140079	null	0.03595793	0.05246034
04031.png	***FALSE***	0.004125309	3.57E-04	null	0.04380402	0.01609549
04095.png	***FALSE***	NaN	0.00743453	null	0.09919615	0.05331534
04164.png	***FALSE***	NaN	0.01840689	null	0.07674373	0.04757531
04378.png	***FALSE***	NaN	4.83E-04	null	0.17958705	0.09003513
04479.png	***FALSE***	0.047114348	0.00183215	null	0.06152043	0.03682231
04637.png	***FALSE***	NaN	3.57E-04	null	0.09375770	0.04705742
05169.png	***FALSE***	0.030743209	3.57E-04	null	0.03714153	0.02274730
05255.png	***FALSE***	NaN	3.57E-04	null	0.11245120	0.05640417
05262.png	***FALSE***	0.018022058	0.00206287	null	0.01099853	0.01036115
05777.png	***FALSE***	0.017279631	6.59E-13	null	0.0070650	0.00811488
06202.png	***FALSE***	NaN	4.25E-04	null	0.09364117	0.04703301
06672.png	***FALSE***	0.06420808	0	null	0.06458346	0.04293051
07134.png	***FALSE***	0.03542274	3.57E-04	null	0.017337435	0.01770577
07140.png	***FALSE***	NaN	0.00142365	null	0.165817881	0.08362076
07946.png	***FALSE***	NaN	1.02E-11	null	0.072127587	0.03606379
08145.png	***FALSE***	0.033316358	2.77E-04	null	0.023061286	0.01888482
09061.png	***FALSE***	0.014700003	0.00485040	null	0.025382546	0.01497765
09252.png	***FALSE***	0.074362745	7.14E-04	null	0.01319539	0.02942414
09431.png	***FALSE***	0.040878552	0.00328135	null	0.031193448	0.02511778
09988.png	***FALSE***	0.062680713	3.54E-04	null	0.054694774	0.03924330

Test Results Summary: Zero (0%) steganalysis detections using the "Standard Fusion" detection algorithm in *StegExpose* software. *StegExpose* can be downloaded here: https://github.com/b3dk7/StegExpose.

7. Applications Regarding Crypto-Currency Cold-Storage and Transfer

Recently the global phenomenon of crypto-currencies has made headlines and changed the face of financial technology worldwide. The seminal block-chain algorithm has been applied to at least 1200 new crypto-currencies in addition to original, the Bitcoin. However, the future of this revolution is uncertain as security vulnerabilities and regulatory restrictions may make the future of ownership and transfer questionable.

To promote safe usage and encourage investment and speculation for the future, steganography applications such as the one discussed in this paper can be taken advantage of with robustly secure results. For example, Bitcoin public address and private key values can be securely encrypted and hidden in images uploaded to the cloud, transferred through e-mail, or stored locally. This allows applications in so-called cold storage scenarios as well as covert transfer of crypto-currency amounts when necessary.

Transfers using the standard Bitcoin network protocols are easily detectable, and blocked by firewalls and deep TCP/IP packet inspection. As a result, any activity in the clear over the network or protocol whatsoever can be used to flag the IP address and responsible parties--severely compromising one of the precepts of the crypto-currencies which is both anonymity and privacy. VPN traffic has been and is virtually entirely shut down in several entire nations, making this obvious alternative for network usage infeasible.

In the case of Bitcoin, TCP/IP ports 8332 and 8333 are two default ports used for Bitcoin peer-to-peer transfers and can be easily detected, inspected, and blocked, and the information derived used for further investigation into the origin and nature of the network traffic. As a result, it is highly desirable to have secondary means for engaging in transfers other than the standard protocol in situations where deep packet inspection or firewalls may be interfering with the transaction process. This is the case not only for Bitcoin but all of the other crypto currencies as well, now and most likely in the future.

One-Time-Pad Steganography offers this potential in a highly secure manner. Further, in the instance of cold storage with retrieval backup, several individuals can be entrusted with encrypted images and keys, and some of the information can be doubly or triply encrypted to ensure the agreement of several individuals to the release of crypto-currency funds. Encrypted images can be stored on the cloud in multiple locations, as well as locally on permanent storage media. Keys, however, must be stored with the person of the control authority. Additional customized protocols and policies should be designed in the future to address crypto-currency authentication and management; however, the steganography technology offers an ideal starting point for both hidden and theoretically unbreakable storage.

8. Conclusion

In this paper, we have presented a complete One-Time Pad encryption and steganography system including all software necessary to complete practical communication. We have compiled recommended best practices and identified potential security levels. Finally, we have tested the software using robust state-of-the-art steganalysis techniques and found the low payload threshold maintained in the software produces a high margin of communication security safety. No payload files were detected (0% detections), despite each file containing the entire content of the U.S. Constitution as embedded text.

Conflict of Interest

The authors declare no conflict of interest regarding the publication of this paper.

References

[1] R. J. Anderson, and F. A. P. Petitcolas, In the limits of steganography? IEEE Journal of Selected Areas in Communications, vol.16, no.4, pp.474-481, 1998.

[2] M. Bashardoust, G. B. Sulong, and P. Gerami, Enhanced LSB image steganography method by using knight tour algorithm, vignere encryption and LZW compression? International Journal of Computer Science Issues, vol.10, no.2, pp.221-227, 2013.

[3] Bhatacharya, I. Banerjee, and G. Sanyal, A survey of steganography and steganalysis techniques in image, text, audio and video cover carrier? Journal of Global Research in Computer Science, vol.2, no.4, pp.1-16, 2011.

[4] K. Chan, and L. M. Chang, Hiding data in images by simple LSB substitution? Pattern Recognition, vol.37, pp.469-474, 2004.

[5] Cheddad, J. Condell, K. Curran, and P.M. Kevitt, Digital image steganography: survey and analysis of current methods? Signal Processing, vol. 90, pp.727-752, 2010.

[6] S. M. Douiri, M. B. O. Medeni, S. Elbernoussi, and E. M. Souidi, A new steganographic method for gray scale image using graph coloring problem? Applied Mathematics & Information Sciences, vol.7, no.2, pp.521-527, 2013.

[7] Gangwar, and V. Srivastava, Improved RGB-LSB steganography using secret key? International Journal of Computer Trends and Technology, vol.4, no.2, pp.85-89, 2013.

[8] R. S. Gutta, Y. D. Chincholkar, and P. U. Lahane, Steganography for two and three LSBs using extended substitution algorithm? ICTAT Journal on Communication Technology, vol.4, no.1, pp.685-690, 2013.

[9] Gutub, M. Ankeer, M. Abu-Ghalioun, A. Shaheen, and A. Alvi, Pixel indicator high capacity technique for RGB image based steganography? in Proceedings of Fifth IEEE International Workshop on Signal Processing and its Applications, 2008, University of Sharjah, U.A.E.

[10] N. Hamid, A. Yahya, R.B. Ahmad, D. Nejim, and L. Kannon, Steganography in image files: a survey? Australian Journal of Basic and Applied Sciences, vol.7, no.1, pp.35-55, 2013.

[11] J. He, S. Tang, and T. Wu, In adaptive steganography based on depth-varying embedding? in Proceedings of 2008 Congress on Image and Signal Processing, 2008, pp.660-663.

[12] M. Hussain, and M. Hussain, A survey of image steganography techniques? International Journal of Advanced Science and Technology, vol. 54, pp.113-123, 2013.

[13] Y. K. Jain, and R. R. Ahirwal, A novel image steganography method with adaptive number of least significant bits modification based on private stego-keys? International Journal of Computer Science and Security, vol.4, no.1, pp.40-49, 2010.

[14] M. Juneja, and P.S. Sandhu, Designing of robust image steganography technique based on LSB insertion and encryption? in Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, 2009, pp.302-305.

[15] Kamaldeep, Image steganography techniques in spatial domain, their parameters and analytical techniques: a review article?, IJAIR, vol.2, no.5, pp.85-92, 2013.

[16] H. B. Kekre, A. A. Athawale, and P. N. Halarnkar, Increased capacity of information hiding in LSB's method for text in image? International Journal of Electrical, Computer and System Engineering, vol.2, no.4, pp.246-249, 2008.

[17] Y. K. Lee, G. Bell, S.Y. Huang, R.Z. Wang, and S.J. Shyu, In advanced least-significant-bit embedding scheme for steganographic encoding? LNCS, vol.5414, 2009, pp.349-360.

[18] Li, J. He, J. Huang, and Y.Q. Shi, A survey on image steganography and steganalysis? Journal of Information Hiding and Multimedia Signal processing, vol.2, no.2, pp.142-172, 2011.

[19] C. Lou, and C. H. Hu, LSB steganographic method based on reversible histogram transformation function for resisting statistical steganalysis? Information Sciences, vol.188, pp.346-358, 2012. Application of a large key cipher in image steganography by exploring the darkest and brightest pixels? International Journal of Computer Science and Communication, vol. 3, no.1, pp.49-53, 2012.

[20] R. S. Marcal, and P.R. Pereira, A steganographic method for digital images robust to RS steganalysis? LNCS, vol.3656, 2005, pp.1192-1199.

[21] Martin, G. Sapiro, and G. Seroussi, Is image steganography natural? IEEE Transactions on Image Processing, vol.14, no.12, pp.2040-2050, 2005.

[22] M. K. Meena, S. Kumar, and N. Gupta, Image steganography tool using adaptive encoding approach to maximize image hiding capacity? International Journal of Soft Computing and Engineering, vol.1, no.2, pp.7-11, 2011.

[23] Mishra, A. Gupta, and D. K. Vishwakarma, Proposal of a new steganography approach? in Proceedings of International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009, pp.175-178.

[24] H. Mathkour, G. M. R. Assassa, A. A. Muharib, and I. Kiady, A novel approach for hiding messages in images? in Proceedings of International Conference on Signal Acquisition and Processing, 2009, pp.89-93.

[25] H. Motameni, M. Norouzi, and A. Hatami, Labeling method in steganography? World Academy of Science, Engineering and Technology, vol. 24, pp.349-354, 2007. vol. 270, part II, 2012, pp.479-488.

[26] M. T. Parvez, and A. A. Gutub, RGB intensity based variable-bits image steganography? in Proceedings of IEEE Asia-pacific Services Computing Conference, 2008, pp.1322-1327. Gandharba Swain et al. / International Journal of Computer Science & Engineering Technology (IJCSET)

[27] P. S. Pharwaha, Secure data communication using moderate bit substitution for data hiding with three-layer security? IE(I) Journal-ET, vol.91, pp.45-50, 2010., International Journal of Security and Its Applications, vol.6, no.2, pp.1-12, 2012.

[28] G. Swain, and S. K. Lenka, RSB array based image steganography technique by exploring the four least significant bits? CCIS

[29] G. Swain, D. R. Kumar, A. Pradhan, and S. K. Lenka, A technique for secure communication using message dependent steganography? International Journal of Computer and Communication Technology, vol.2, no. 2- 4, pp.177-181, 2010.

[30] G. Swain, and S. K. Lenka, Steganography using the twelve-square substitution cipher and an index variable? in Proceedings of ICECT, 2011, vol.3, pp.84-88.

[31] G. Swain, and S. K. Lenka, A robust image steganography technique using dynamic embedding with two least significant bits? Advanced Materials Research, vols. 403-408, pp.835-841, 2012.

[32] G. Swain, and S. K. Lenka, A dynamic approach to image steganography using the three least significant bits and extended hill cipher? Advanced Materials Research, vols. 403-408 pp.842-849, 2012.

[33] G. Swain, and S. K. Lenka, A technique for secret communication by using a new block cipher with dynamic steganography"

[34] G. Swain, and S. K. Lenka, A hybrid approach to steganography- embedding at darkest and brightest pixels?, in Proceedings of International Conference on Communication and Computational Intelligence, 2010, pp.529-534.

[35] M. A. B. Younes, and A. Jantan, A new steganography approach for image encryption exchange by using least significant bit insertion, International Journal of Computer Science and Network Security, vol.8, no.6, pp.247-254, 2008.

[36] H. J. Zhang, and H. J. Tang, "A novel image steganography algorithm against statistical analysis", in Proceedings of Sixth International Conference on Machine Learning and Cybernetics, 2007, pp.3884-3888.

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 283-308 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Design and Implementation of Closed-loop PI Control Strategies in Real-time MATLAB Simulation Environment for Nonlinear and Linear ARMAX Models of HVAC Centrifugal Chiller Control Systems

Nicolae Tudoroiu*,[1], Mohammed Zaheeruddin[1], Songchun Li[1], Elena-Roxana Tudoroiu[2]

[1]*Faculty of Engineering and Computer Science, Building Civil and Environmental Engineering, Concordia University, H3G 1M8, Canada*

[2]*Faculty of Sciences, Mathematics and Informatics, University of Petrosani, 332006, Romania*

ARTICLE INFO

Article history:
Received: 22 February, 2018
Accepted: 04 April, 2018
Online: 25 April, 2018

Keywords:
Centrifugal chiller
ARMAX and ARX models
PI control
PI-plus control
Parameters Estimation

ABSTRACT

The objective of this paper is to investigate three different approaches of modeling, design and discrete-time implementation of PI closed-loop control strategies in SIMULINK simulation environment, applied to a centrifugal chiller system. Centrifugal chillers are widely used in large building HVAC systems. The system consists of an evaporator, a condenser, a centrifugal compressor and an expansion valve. The overall system is an interconnection of two main control loops, namely the chilled water temperature inside the evaporator, and the refrigerant liquid level control in condenser. The centrifugal chiller dynamics model in a discrete-time state-space representation is of high complexity in terms of dimension and encountered nonlinearities. For simulation purpose the centrifugal chiller model is simplified by using different approaches, especially the development of linear polynomials ARMAX and ARX models. The aim to build linear ARMAX models for centrifugal chiller is to simplify considerable the control design strategies that are investigated in this research paper. The novelty of this research is a new controller design approach, more precisely an improved version of proportional - integral control, the so called Proportional-Integral-Plus control for systems with time delay, based on linear ARMAX models. It is conceived within the context of non-minimum state space control system that "seems to be the natural description of a discrete-time transfer function, since its dimension is dictated by the complete structure of the model". The effectiveness of this new controller design, its implementation simplicity, convergence speed and robustness are proved in the last section of the paper.

1. Introduction

Nowadays significant amount of energy is consumed in commercial or residential buildings by the heating, ventilation and air conditioning (HVAC) systems to provide thermally comfortable indoor environment. Consequently, improving their efficiency becomes a critical issue for energy and environmental sustainability. In most commercial buildings, chilled-water from centrifugal chillers is supplied to air-handling units to meet the cooling needs of the building. The overall system as shown in Figure 1 consists of two water circuits: in one circuit water is circulated through the evaporator to produce chilled water and the second water circuit is used to reject heat to outdoors through a cooling tower [1-7]. During the last two decades, for a wide variety of HVAC control systems applications, the centrifugal chillers have become the most widely used due to their high capacity, high reliability, and low maintenance [7]. Moreover, among the major devices in chilled-water systems, the centrifugal chiller is the most energy-consuming device, and its efficiency can be improved by implementing advanced model-based controller design strategies and fault detection and isolation (FDI) methodologies.

Consequently, the development of high-fidelity dynamic centrifugal chiller models has become a priority task of our research. A brief literature review in this field reveals that a significant amount of work has been done on transient and steady state modeling for centrifugal chillers [1]-[4], [7-11].

On the other hand, several dynamic models for vapor compression cycle also have been extensively studied, such as in [6] where is modeled a reciprocating compressor with polytrophic efficiency, that includes also a compressor housing and refrigerant mixing with the oil in the sump. The heat exchanger is modeled

*Corresponding Author: Nicolae Tudoroiu, John Abbott College, 2127 Lakeshore Road, Sainte-Anne-de-Bellevue, QC. ntudoroiu@gmail.com

https://dx.doi.org/10.25046/aj030233

based on moving boundary method that assumes an average state for each phase region according to well mixed modeling assumption. The expansion valve is modeled based on isenthalpic process, and in addition the dynamics of the sensing bulb is also considered. The simulation models are tested with shut-down and start-up and different operations. An interesting approach we find in [9] where a centrifugal compressor model is developed based on control volume methodology that designs the impeller and diffuser separately as control volumes; the impeller is modeled in detail based on the Euler equations, and the diffuser is modeled as ideal without any losses. The mass flow rate and the exit state enthalpy specific for low pressure refrigerant R234a are predicted by two empirical relations. A substantial improvement of this model we find in [10], where a liquid chiller lumped - parameter model of the heat exchangers is developed. Input step changes in the condenser-side water flow rates are studied in order to simulate the system performance under disturbances. Furthermore, in [8] is developed a mechanistic, single stage and two-stage centrifugal chiller models where the centrifugal compressor is modeled based on the Euler turbo-machinery theory, the energy balance and the impeller velocity equations. The coefficient of performance (COP) of the chiller is simulated by considering the compressor polytropic efficiency, hydrodynamic, mechanical and electrical losses. On the other hand, the condenser and evaporator are modeled based on lumped parameter approach and the heat transfer is calculated based on effectiveness model. The chiller model is validated with a water-to-water chiller test facility. The simulations reveal that the chiller dynamic model predicts higher start-up condenser and evaporator pressures than the measurements, more precisely the modeled dynamics seems to appear faster than the experiment results for convergence to the steady-state. Also, the evaporator-side model performs better than the condenser-side in terms of predicting the transient responses. In an another study [2] a dynamic centrifugal liquid chiller model with flooded-type shell-and-tube heat exchangers is developed, where the compressor model is based on a constant speed and its capacity control is achieved by variable inlet guide vane (IGV). An interesting regression model was applied in [8] to determine the maximum capacity condition, i.e., with wide-open IGV, and the actual mass flow rate was then computed based on a linear relationship of the maximum mass flow rate and the IGV position.

During the transient periods, in a numerical simulation environment was found the optimal values of the initial and the overall system charges. Moreover in [8] is mentioned that for simulations purpose, due to complex nature of the refrigeration cycle, it seems that all the existing simulation models developed for centrifugal chillers include some simplifications for some components. For the centrifugal compressor, the dynamic performance was often approximated based on the compressor characteristic map rather than parameterized dynamic models as is mentioned in [1, 4, 6]. Such approximation is sufficient for simulating the steady-state performance of the chiller, whereas may lead to difficulty in simulating transient processes such as startup, load change and shutdown. In [11] is developed a dynamic model for the centrifugal compressor where the model of the heat exchangers is based on lumped parameters approach.

The objective of our research is to develop a comprehensive dynamic model for the centrifugal chiller suitable for control analysis and design. In particular we aim to achieve quality simulation performance in MATLAB and SIMULINK modeling, and simulations environment for all components of centrifugal chiller. The significance of the proposed research includes the following:

- The dynamic characteristics of the centrifugal chiller system was modeled based on detailed mass, momentum and energy balance equations
- The shell-and-tube heat exchangers were modeled based on lumped parameters approach.
- The developed control strategies based on the centrifugal chiller model cover different operating conditions in order to be robust to several changes of the parameters control structure or modeling errors, and to achieve the best transient and steady-state performance.

In this paper we are focused to get accurate models for centrifugal chiller and based on these models to develop a few PI closed-loop control strategies to find the most suitable control strategy for these kinds of applications that performs better in terms of accuracy, robustness, convergence speed, overshoot, and disturbance rejection. The performance comparison of the proposed control strategies is very useful to choose the most suitable control strategy that performs better in terms of meeting the control design requirements and process control constraints, rejection of the effect of possible disturbances that act on the controlled process, as well as a tracking error accuracy.

2. The Centrifugal Chiller Nonlinear Model

2.1. *Modeling and description of components*

In Figure 1 is shown a schematic of the water-cooled centrifugal chiller system which consists of four major components: a centrifugal compressor, an expansion valve, an evaporator, and a condenser [8]. In general, the overall chilled-water system consists of one circuiting refrigerant loop and two water loops. The first water loop is circuited between the condenser and the cooling tower, and the second water loop is circuited between the evaporator and the air handling unit (AHU). In our research, a detailed nonlinear model of a centrifugal chiller system is developed. The modeling methodology and the model equations are described in Annex-1 appended at the end of this paper. The nonlinear model consists of four major sub-component models, namely, a flooded evaporator, a flooded water-cooled condenser, a centrifugal compressor and an electronic expansion valve. The simulation of transient and steady-state behavior of the water-cooled centrifugal chiller is performed in a well-known MATLAB / SIMULINK environment. As described in Annex-1, the dynamic model of the overall centrifugal chiller system is of high complexity in terms of state dimension and nonlinearity. The nonlinear model equations were then reformulated into a state space model by a set of 40 first order differential equations (ODE), as a mathematical form to represent the physical chiller system.

The significant aspects of the nonlinear model are presented briefly in this section.

First, we focus our attention to centrifugal compressor that is the much faster responding device in the chiller plant. Similar as is shown in [8] the centrifugal compressor capacity balance is achieved with variable rotor speed control. In our research, the capacity control is manipulated by adjusting the IGV position and input torque, as shown in Figure 2. The model takes into account also, the effects of incidence and fluid friction losses, mentioned in [8]. Figure 2 schematically depicts the centrifugal model with boundary conditions and capacity control. It is important to note that the working medium in the centrifugal chillers is refrigerant that works under multi-phase conditions, which makes it more complex and difficult for the underlying modeling framework and consistent numerical initialization [8].

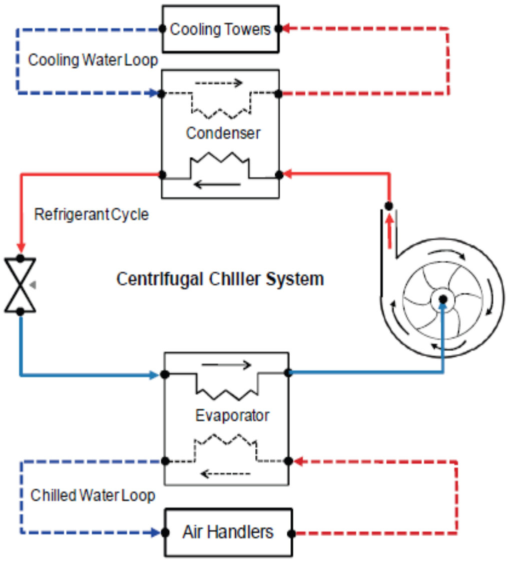


Figure 1. Schematic of water-cooled centrifugal chiller system (see [8])

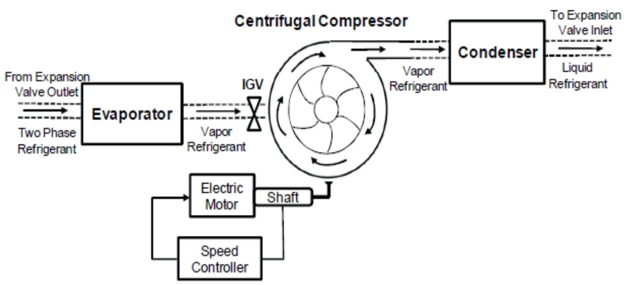


Figure 2. The schematic of the centrifugal compression system in chiller (see [8])

The evaporator and condenser modeling requires quality heat exchanger models. The flooded evaporator was divided into a two-phase section and a superheat section as detailed in Annex-1. Likewise, the flooded condenser was divided into two superheat sections, a two-phase section and a sub-cool section.

The electronic expansion valve (EXV) was modeled to regulate the flow rate of refrigerant in the system and in this study the EXV was used to maintain the liquid level of the refrigerant in the condenser. The modeling details are described in Annex-1.

The dynamic model for centrifugal chiller system presented in Annex-1 is reformulated in state-space representation, by a linear set of ordinary differential equations (ODE), as a natural representation of a physical system, and easily to be solved using a suitable MATLAB solver.

3. MATLAB Open-Loop Simulation results

Simulations were performed with the nonlinear centrifugal chiller model to investigate its open-loop dynamic behavior and steady-state performance using R134a as refrigerant, in a MATLAB/SIMULINK simulation environment. In open-loop the centrifugal chiller is a multi-input, multi-output (MIMO) control system that has two inputs, namely the compressor motor drive speed (Ucom), and the expansion valve opening (U_EXV), two outputs corresponding to chilled water temperature (Tchwsp), and the liquid level in condenser (Level)). Using a consistent set of initial conditions, simulation runs were conducted under constant design load conditions. The open-loop simulation results are shown in Figure 3. The temperature and pressure responses are smooth and reach their respective design operating conditions as the system reaches steady state. The responses in Figure 3 depict the condensing pressure Pc in 3(a), the evaporation pressure Pev in 3(b), chilled water temperature (Tchw,sp) in 3(c), water temperature in the condenser 3(d), the heat exchange rate in 3(e) and the coefficient of performance in 3(f). From these results it can be noted that the system responses reach steady state in about 30 minutes when subjected to constant load. At steady state the coefficient of performance of the system is 5.5 and the heat rejection rate is about 20% greater than the cooling capacity as a result of heat generated by the compressor motor.

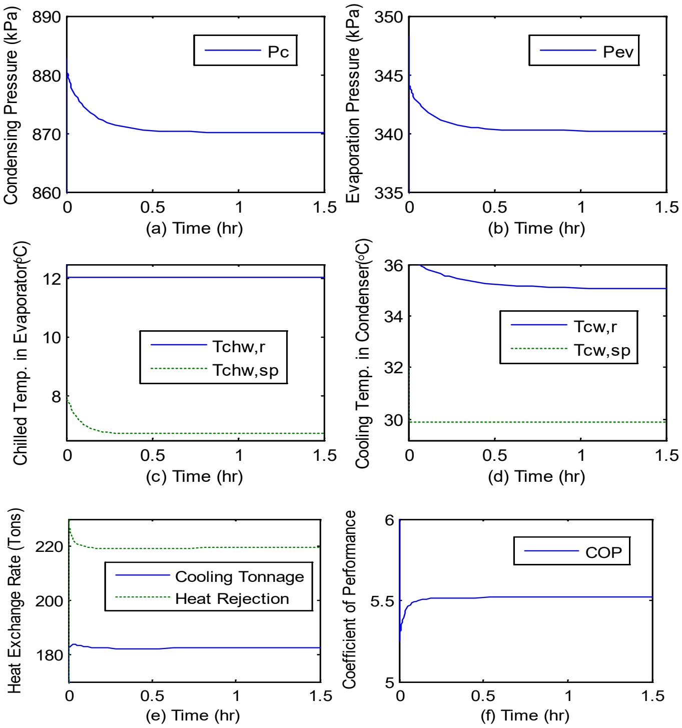


Figure 3. Open-loop responses of the chiller system

Legend: (a) Compression pressure (Pc); (b) Evaporation pressure (Pev); (c)Chilled water temperature Tchw, sp; (d) Water temperature in condenser Tcw; (e) Cooling tonnage and heat rejection; (f) Coefficient of performance (COP);

The SIMULINK model of centrifugal chiller system in open-loop and its details are shown in Figures 4 and 5. These figures with higher resolution are also depicted in Annex-2.

The MATLAB Function1 in the SIMULINK model from Figure 5 has the following code lines:

```
function [U1, U2, Y1, Y2] = fcn(x)
Tchw_sp = x(18);
Level = x(35);
Y1 = Tchw_sp;
Y2 = Level;
U1 = x(33);
U2 = x(37);
end
```

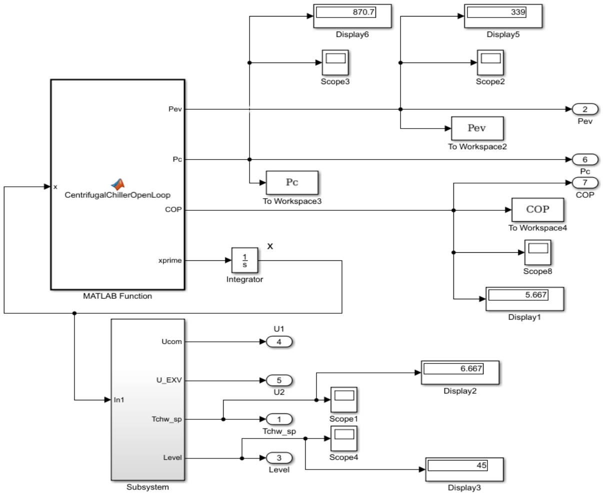


Figure 4: The SIMULINK model of centrifugal chiller in open-loop

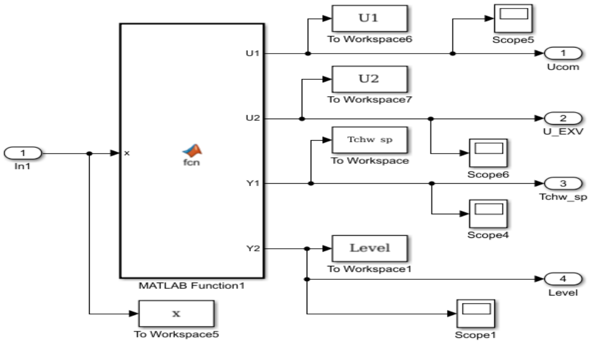


Figure 5. The SIMULINK model of the bottom Subsystem Block

The Centrifugal Chiller open-loop model is very useful to build the PI control strategy as described in the next section for the overall system seen as centralized system, since in "real-life" there exist some interferences between the loops. Furthermore, through open-loop simulations we generate sets of input-output data required to build the ARMAX models for Centrifugal chiller. This is the reason to start with a good open-loop model capable to capture entire dynamics of the overall chiller system under various operating conditions. This give us more flexibility for control design.

4. Closed – Loop Control Strategies Design

In this section we develop and implement in a MATLAB/SIMULINK environment three closed-loop control strategies, namely an overall Proportional-Integral (PI) controller based on the chiller system model in open-loop developed and implemented in MATLAB simulation environment in previous section 3. By extensive simulations we found that the interference between the both loops, temperature and liquid Refrigerant level, respectively, is very week, and thus we can simplify the overall chiller system model by considering that the loops are decoupled.

This is an essential modeling aspect that is investigated in this research paper, by building for each loop separately a PI controller, as is developed in subsection 4.1. The SIMULINK model of PI controllers for each separate loop is shown in Annex-2.

The open-loop model is useful also to generate the set of input-output data measurements necessary to build the linear ARMAX SISO models. The input-output data set measurements file was loaded in a new MATLAB code file that estimates the ARMAX SISO models.

However, from accuracy considerations, in the future research work we will investigate also the extension of ARMAX multi-input single output (MISO) models for building predictive control strategies. In subsection 4.2 the ARMAX SISO decoupled models will be used to build one linear PI controller for each loop separately. The third control strategy is an improved PI control version that is conceived within the context of non-minimum state space control system. The non-minimal state space representation in contrast to minimal state space realizations "seems to be the natural description of a discrete-time transfer function, since its dimension is dictated by the complete structure of the model", as is mentioned in [12-16]. More precisely, we talk about on improved version of Proportional - integral control, so called Proportional-integral-plus (PIP) control for systems with time delay [12-16], that is developed in subsection 4.3. Based on ARMAX models is proved also the superiority of new PPI controller compared with the first two PI versions in terms of efficiency, implementation simplicity, convergence speed and robustness to the load changes disturbance and to the changes in the noise level of measurement sensors. It is worth to mention again the particular structure for this kind of applications of the MIMO centrifugal chiller control system to be split into two single input-single output (SISO) independent closed - loops control subsystems (i.e. the Evaporator chilled water temperature loop to control the Tchwsp, and the Condenser liquid Refrigerant level loop to control the Refrigerant liquid Level. The preset values of the first closed-loop is Tchw_set = 6.67 [°C], and for the second closed-loop is Level set = 45(%) of the maximum level of the liquid Refrigerant inside the Condenser. Also, a PIP closed-loop control is developed in detail in subsection 4.3. Concluding, at the end of section 4 we can decide on a potential controller choice amongst these three control strategies to find the one that performs better in terms of the accuracy, robustness to the changes in the load magnitude, considered as a disturbance, and to the level of measurement noises injected additionally in each forward loop to Tchwsp, and to liquid Level respectively.

4.1. Closed-loop PI Control Strategy of MIMO
4.2. Centrifugal Chiller System

In this subsection we design for each open-loop a PI control strategy given by [12-16]:

- closed-loop PI control strategy of chilled water temperature Tchwsp of Evaporator:

$$U_{com}(t) = k_{p_1}\varepsilon_1(t) + k_{i_1}\int_0^t \varepsilon_1(\tau)d\tau \tag{1}$$

$$\varepsilon_1(t) = y_{1sp} - y_{1mes}(t) = Tchw_{set} - Tchw_{sp}(t) \tag{2}$$

where $\varepsilon_1(t)$ is the error between the set point value of the closed-loop output temperature $Tchw_{set}(t)$ value and its measured value $Tchw_{sp}(t)$; k_{p_1} , k_{i_1} are the tuning proportional, integral coefficients of the first PI controller.

- closed-loop PI control strategy of liquid level inside the condenser:

$$U_{EXV}(t) = k_{p_2}\varepsilon_2(t) + k_{i_2}\int_0^t \varepsilon_2(\tau)d\tau + k_{d_2}\frac{d\varepsilon_2}{dt} \tag{3}$$

$$\varepsilon_2(t) = y_{2sp} - y_{2mes}(t) = Level_{set} - Level(t) \tag{4}$$

where $\varepsilon_2(t)$ is the error between the set point value of the closed-loop output liquid level $Level_{set}(t)$ value and its measured value $Level(t)$; k_{p_1} , k_{i_1} are the tuning proportional, integral coefficients of the second PI controller.

An extensive number of simulations for adjusting the coefficients of both PI controllers in order to perform well in terms of overshoot, settling time and steady-state error led to the following values:

- For closed-loop Evaporator subsystem:

$k_{p_1} = 0.1; k_{i_1} = 0.0001; Tchwsp\_set = 6.67[°C]$,

i.e. the Temperature set-point value.

- For closed-loop Condenser subsystem

$k_{p_2} = 0.05; k_{i_2} = 0.0001; Level\_set = 45[\%]$, i.e. the Liquid level setting point value.

The closed-loop simulation results related to PI controller performance for Evaporator subsystem are shown in Figure 6 for Condenser subsystem, and in Figure 7 they are related to the overall performance in terms of coefficient of performance COP of entire MIMO control system. In Figure 6 are shown the chilled water Evaporator temperatures (Tchwsp, Tchwsp_set) as in 6(a), the evaporation pressure Pev in 6(b), and the Compressor motor speed actuator efforts in 6(c).

In Figure 7 is showing the PI controller closed-loop performance for Condenser subsystem related to the Refrigerant liquid level, such in 7(a), condensing pressure Pc as is revealed in 7(b), and expansion valve actuator effort presented in 7(c).

The overall performance for entire MIMO control system in the particular case of nonlinear plant dynamics in terms of coefficient of performance is shown in Figure 8.

Concluding, in this section we can see that the closed-loop PI controller of Condenser subsystem perform better compared to the closed-loop PI controller of Evaporator subsystem in terms of accuracy, fast transient, and disturbance rejection, as is shown in Figure 9 for the Evaporator chilled water temperature, and in Figure 10 for Refrigerant liquid level in the Condenser.

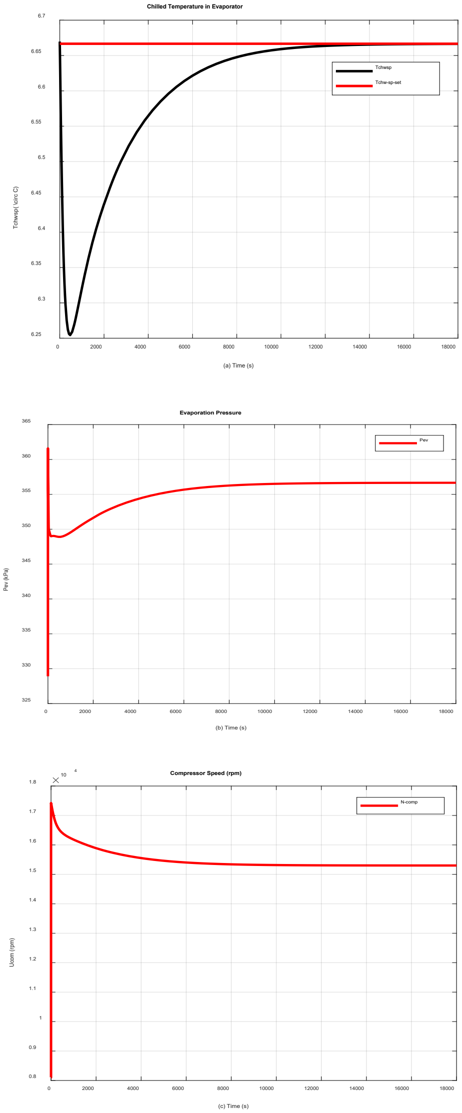


Figure 6. PI control strategy for closed-loop simulations for chiller plant model- Evaporator Subsystem

Legend: (a) Temperature water supply pressure Tchwsp; (b) Evaporation pressure Pev; (c) Compressor motor drive actuator speed

As disturbance, is considered a sharp variation of the load, i.e. the temperature of chilled water Tchw_rr that increases from 48[°C] to 52[°C]), injected inside the MIMO control system by addition in the forward path of the chilled water temperature of Evaporator subsystem before the measured chilled water temperature output, after 9000 seconds (2 hours and 30 minutes), persisting further 2 hours and 30 minutes. This time assures for MIMO control system enough time to reach a new steady-state.

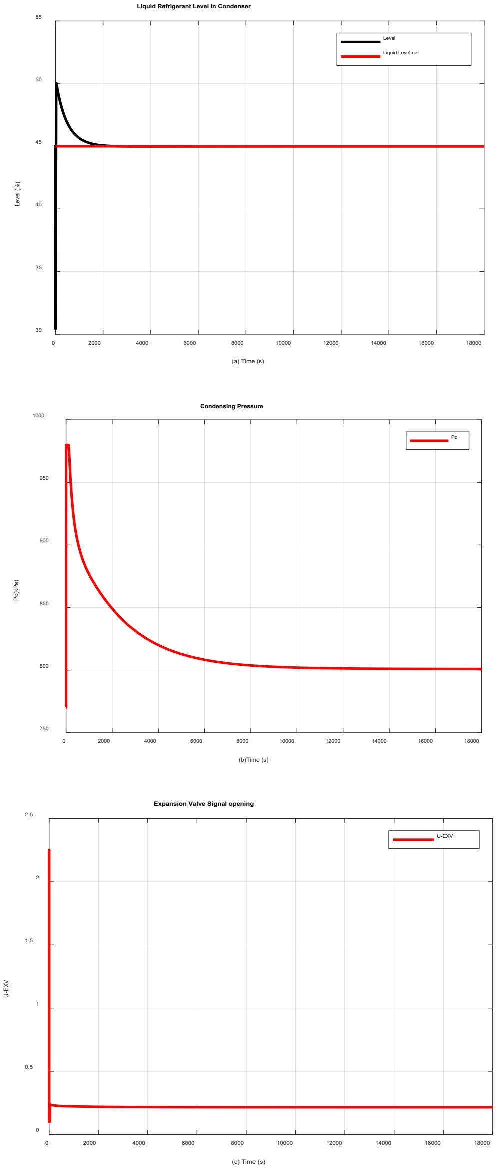


Figure 7. PI control strategy for closed-loop simulations for chiller plant model – Condenser Subsystem

Legend: (a) Refrigerant liquid level; (b) Condensing pressure Pc; (c) Expansion valve actuator opening

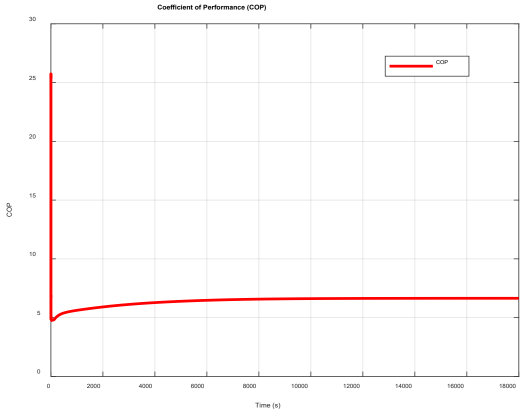


Figure 8 Overall coefficient of performance COP

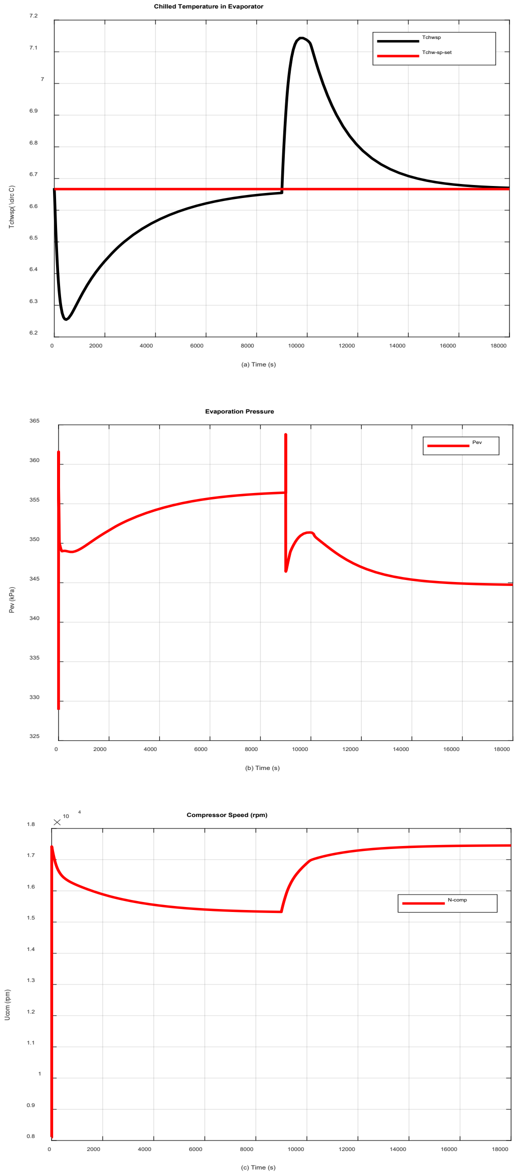


Figure 9. PI control strategy for closed-loop simulations for chiller plant model - Evaporator subsystem, disturbance rejection

Legend: (a) Temperature water supply pressure Tchwsp; (b) Evaporation pressure Pev; (c) Compressor motor drive actuator speed

The both PI controllers perform very well and prove a good robustness to the changes in the control system load, with a significant amount effort of both actuators, especially by the compressor motor drive. The SIMULINK models of both PI control strategies are attached for clearness in the Annex-2. The simulation results reveal also that the PI controller of the Refrigerant liquid level in Condenser subsystem compared to PI controller of the Evaporator subsystem performs faster, during a very short transient time to reject the disturbance impact, therefore is more robust. Despite the changes in the load is worth to mention that the coefficient of performance for entire MIMO control system, as is shown in Figure 11, changes smoothly and regains in short time the optimal value reached in steady-state before the disturbance injection.

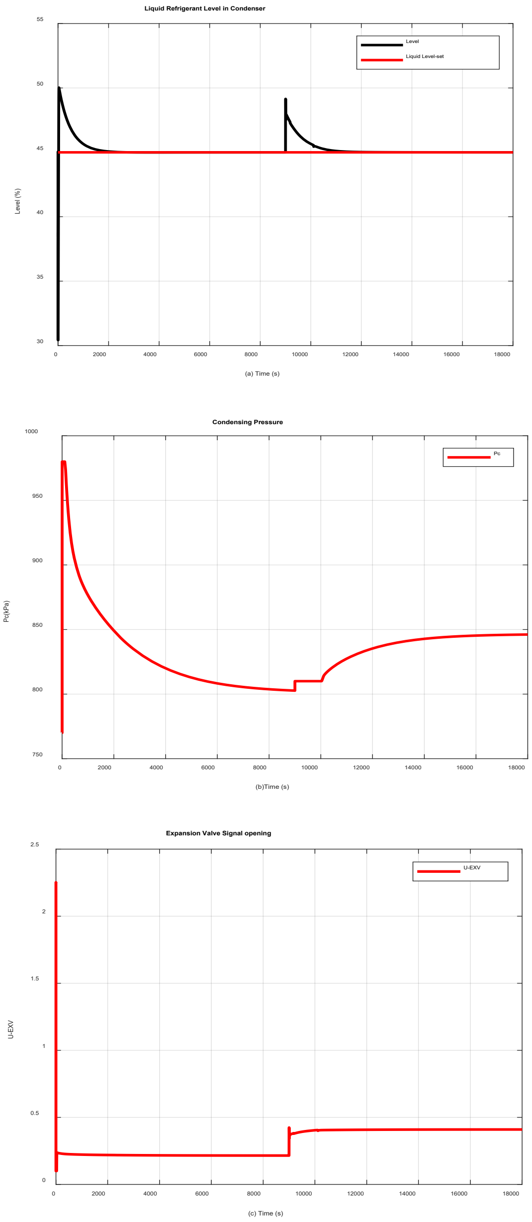


Figure 10. PI control strategy for closed-loop simulations for chiller plant model – Condenser subsystem, disturbance rejection
Legend: (a) Refrigerant liquid level; (b) Condensing pressure Pc; (c) Expansion valve actuator opening U_EXV

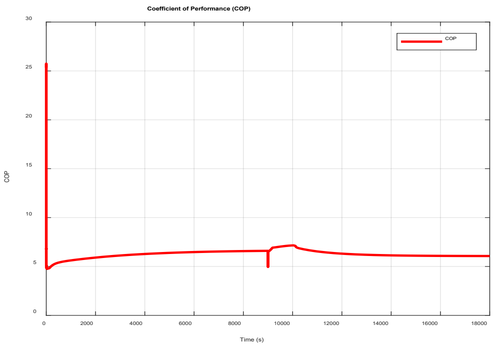


Figure 11. Robustness of the overall coefficient of performance COP to a sharp disturbance variation in the load

Additionally, a normally distributed Gaussian random signal is injected in the output chilled water temperature loop and Refrigerant liquid level loop in order to simulate the impact of the measurement noise level. For a variance in measurement noise level of 0.0001 in chilled water temperature, and 0.1 in the Refrigerant liquid level, their impact on the both PI controllers performance are shown in Figure 12 and Figure 13. The simulation results shown in Figure 13 reveal a good behavior of the closed-loop PI controller of the Refrigerant liquid level compared to closed-loop PI controller of chilled water temperature.

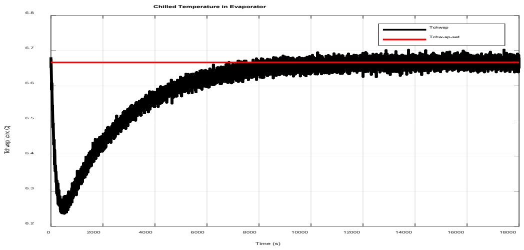


Figure 12. Tchwsp - Measurement noise impact on the MIMO control system response (noise level variance 0.001)

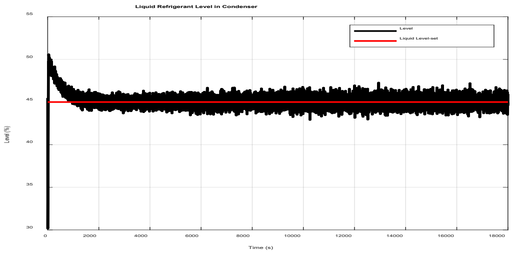


Figure 13. Refrigerant liquid level - Measurement noise on the MIMO control system response (noise level variance 0.1)

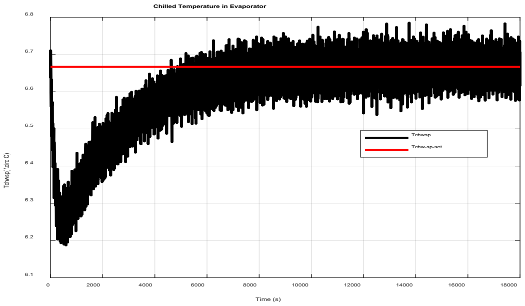


Figure 14 Tchwsp - Measurement noise impact on the MIMO control system response (noise level variance increased to 0.001)

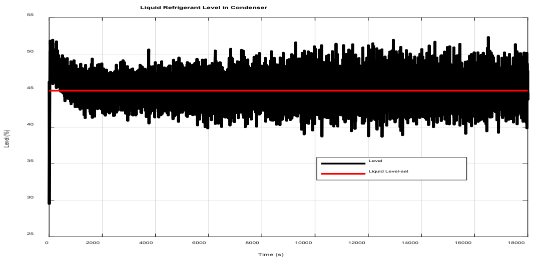


Figure 15. Refrigerant liquid level - Measurement noise impact on the MIMO control system response (noise level variance increased to 1)

The impact due to ten times rise in the measurement noise levels becomes significant, as is shown in Figure 14, and Figure 15. The simulation results reveal a fragile robustness of the both PI controllers to changes in noise variance level.

The noise variance in the measurements can be attenuated by using two Moving Average filters coded in MATLAB, with a windows length of 10, and 100 samples respectively. The filtered responses of MIMO control system are shown in Figure 16 and Figure 17.

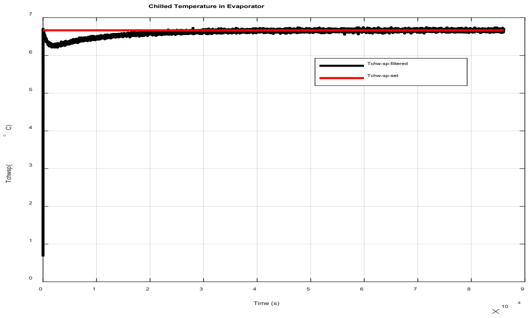


Figure 16. Tchwsp – Moving Average filtered measurement noise impact on the MIMO control system response (noise level variance increased to 0.001, windows length = 10)

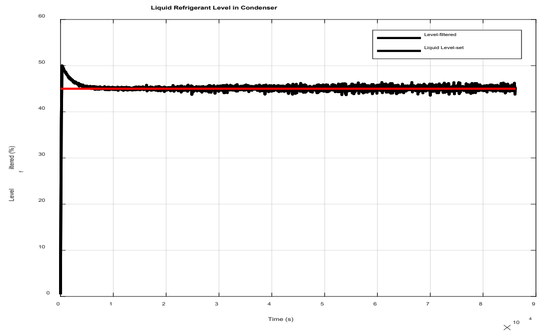


Figure 17. Refrigerant liquid level - Moving Average filtered measurement noise impact on the MIMO control system response (noise level variance increased to 1, window length =100 samples)

4.3. *Closed-loop PI Control Strategies using Linear ARMAX SISO Models for Centrifugal Chiller System*

In this section we develop two linear Auto Regressive Moving Average (ARMAX) SISO models for the both subsystems (Evaporator and Condenser) of the centrifugal chiller system (CCS). These two linear ARMA models are useful in this section to build and implement two PI closed-loop control strategies, and furthermore an interesting approach of a new control design approach, namely a PIP control strategy developed in next subsection 4.3. To estimate the models' parameters are used two particular MTLAB functions from Identification Toolbox, more precisely armax and arx. First one, armax MATLAB function estimates the parameters of ARMAX MIMO or SISO polynomial discrete-time models using time-domain data [17]. The data set of input-output chiller system measurements are generated by open-loop simulations performed on the SIMULINK model of chiller system shown in Figure 4, Figure 5.

The second one, *arx* MATLAB function estimates the parameters of two polynomials discrete-time models known as Autoregressive with an exogenous input (ARX) and a more simple polynomial Autoregressive without exogenous input (AR) that estimates the parameters of the scalar time series, by means of the well-known least squares method the most used in control systems identification and parameters estimation. The both MATLAB functions use a prediction-error method and specified polynomial orders. For each of these models pure transport delays of the signals flow in the feedback path from the measurement sensors to the controllers for each input/output pair are also specified in their ARAMAX and ARX structures [17]. In addition, in each ARMAX and ARX model structures noise channels are incorporated, hence the most general polynomial form of these discrete-time models is given by [18 - 21]:

$$
\begin{aligned}
&A(q)y(t) = B(q)u(t-n_k) + C(q)e(t) \\
&A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_{n_a} q^{-n_a} = 1 + \sum_{i=1}^{n_a} a_i q^{-i} \\
&\qquad = 1 + q^{-1} \sum_{i=1}^{n_a} a_i q^{-i+1} = 1 + q^{-1} A^*(q^{-1}), \\
&B(q) = b_1 + b_2 q^{-1} + b_2 q^{-2} + \ldots + b_{n_b} q^{-n_b+1} = \sum_{i=1}^{n_b-1} b_i q^{-i} \\
&\qquad = q^{-1} \sum_{i=1}^{n_b-1} b_i q^{-i+1} = q^{-1} B^*(q^{-1}), \\
&C(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \ldots + c_{n_c} q^{-n_c} \\
&\qquad = 1 + \sum_{i=1}^{n_c} c_i q^{-i} = 1 + q^{-1} \sum_{i=1}^{n_c} c_i q^{-i+1} = 1 + q^{-1} C^*(q^{-1})
\end{aligned} \tag{5}
$$

where

- $y(t)$ is the output at discrete-time t, n_a, n_b, n_c represent the degrees of the polynomials $A(q), B(q)$, and $C(q)$ respectively,
- $u(t)$ is the control system input at the discrete instant t,
- n_k is an integer number of sampling periods, the so-called *dead time* and most used in the control systems (pure transport delay of the signal flow between the measurement sensors and controllers in the feedback path,
- a_i, b_i, c_i represent the coefficients (parameters) of the polynomials $A(q), B(q)$, and $C(q)$ of ARMAX models respectively,
- q @ forward shift operator, i.e. $q(y(t)) = y(t+1)$, and q^{-1} @ backward shift operator, i.e. $q^{-1}(y(t)) = y(t-1)$,
- t @ $\frac{t}{T_s}$, denotes the normalized sampling time, only symbolic that replaces the most used typical notation $t = kT_s, k \in \mathbf{Z}^+$ used to designate the description of dynamical discrete-time system, T_s representing the sampling period ,

- $e(t)$ denotes the white noise disturbance value at the discrete instant t

With these notations the general discrete-time representation of a dynamic system as ARMAX model can be put in the following form, as is given in [20]:

$$y(t+1) = -A^{*}(q^{-1})y(t) + q^{-n_k}B^{*}(q^{-1})u(t) + (1+q^{-1}C^{*}(q^{-1}))e(t) \quad (6)$$

If the data are as time series with no input channels and with only one output channel, then armax calculates an ARMA model for the time series, given by [1, 20]:

$$y(t+1) = -A^{*}(q^{-1})y(t) + (1+q^{-1}C^{*}(q^{-1}))e(t) \quad (7)$$

Furthermore, if in the noise source channel there is an integrator ARMAX model becomes an ARIMAX model, described by [19]:

$$y(t+1) = -A^{*}(q^{-1})y(t) + q^{-n_k}B^{*}(q^{-1})u(t) + \frac{(1+q^{-1}C^{*}(q^{-1}))}{1-q^{-1}}e(t) \quad (8)$$

If there are no inputs, the ARIMAX structure becomes an ARIMA model of the following form:

$$y(t+1) = -A^{*}(q^{-1})y(t) + \frac{(1+q^{-1}C^{*}(q^{-1}))}{1-q^{-1}}e(t) \quad (9)$$

It is worth also to mention that the discrete-time representation in (6) can be put in a so called "regression form" useful to make the link with the least square estimation method of the ARMAX model coefficients and parameters based on the input-output data set measurements, as is mentioned in [19]:

$$y(t+1) = \theta^{T}\varphi(t) \quad (10)$$

where the vectors θ, φ represent the vector of parameters:

$$\theta = [a_1, a_2, ..., a_{n_a}, b_1, b_2, ..., b_{n_b}]^{T} \quad (11)$$

and the so called the "*regressor vector*" of measured values of the current outputs and the inputs and theirs the past values, respectively:

$$\varphi(t) = [-y(t), -y(t-1), ..., -y(t-n_a+1), u(t-n_k), u(t-n_k-1), ..., u(t-n_k-n_b+1)] \quad (12)$$

By means of some symbolic algebraic manipulations performed in the first equation (5) we get an interesting result that makes the link with the well-known discrete time "*transfer operator*" for open or closed loop control systems:

$$\begin{aligned} y(t) &= q^{-n_k}\frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \\ &= G_u(q^{-1})u(t) + G_e(q^{-1})e(t) \end{aligned} \quad (13)$$

$$G_u(q^{-1}) = q^{-n_k}\frac{B(q^{-1})}{A(q^{-1})}, G_e(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})} \quad (14)$$

where $G_u(q^{-1})$ and $G_e(q^{-1})$ represents the discrete time transfer operators for the channels input-output control system in open or closed-loop, $u(t) \rightarrow y(t)$ and "*white*" or "*colored*" noise channel – output control system, $e(t) \rightarrow y(t)$ respectively.

Moreover, if we replace formally in the equations (13) and (14) the shift (backward or forward) operator q by a complex variable $z \in C, z = e^{st}, s = \sigma + j\omega, \sigma = \mathrm{Re}(s), \omega = \mathrm{Im}(s)$ representing the real and imaginary parts of the complex variable s respectively, the most used by Laplace transform to compute the transfer functions of the control systems in the complex domain s, a new representation in z complex domain is found, as follows:

$$\begin{aligned} Y(z) &= z^{-n_k}\frac{B(z^{-1})}{A(z^{-1})}U(z) + \frac{C(z^{-1})}{A(z^{-1})}E(z) \\ &= G_u(z^{-1})U(z) + G_e(z^{-1})E(z) \end{aligned} \quad (15)$$

$$G_u(z^{-1}) = z^{-n_k}\frac{B(z^{-1})}{A(z^{-1})}, G_e(z^{-1}) = \frac{C(z^{-1})}{A(z^{-1})} \quad (16)$$

This description is typically used for the control systems represented in discrete-time, where $U(z) @Z\{u(t)\}, Y(z) @Z\{y(t)\}$, and $E(z) = Z\{e(t)\}$ are the z-images of the input $u(t)$, white noise $e(t)$, and the output $y(t)$, obtained by applying the "transform Z" operator to all three discrete - time variables $u(t)$, $e(t)$, and $y(t)$ respectively [19]. The transfer operators $G_u(z^{-1})$, and $G_e(z^{-1})$ defined in equation (16) are so-called z-transforms functions of the discrete-time control system.

In our case study, the discrete-time representations of the both SISO ARMAX models assigned to the centrifugal chiller control system are obtained in MATLAB R2017b by some manipulations of typical MATLAB functions in MATLAB code, given by [18-21]:

A. Chilled water temperature closed-loop SISO ARMAX discrete-time model:

$$\begin{aligned} y_1(t) &= q^{-2}\frac{B(q^{-1})}{A(q^{-1})}u_1(t) + \frac{C(q^{-1})}{A(q^{-1})}e_1(t) \\ &= G_{u_1}(q^{-1})u(t) + G_{e_1}(q^{-1})e_1(t) \end{aligned} \quad (17)$$

$$G_{u_1}(q^{-1}) = q^{-2}\frac{B(q^{-1})}{A(q^{-1})}, G_{e_1}(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})} \tag{18}$$

and the polynomials $A(q^{-1}), B(q^{-1})$ are given by:

$$A(q^{-1}) = 1 - 2.909q^{-1} + 2.822q^{-2} - 0.9131q^{-3} \tag{19}$$

$$B(q^{-1}) = 0.01843 - 0.005826q^{-1} - 0.001592q^{-2} \tag{20}$$

where the polynomials $A(q^{-1}), B(q^{-1})$, and $C(q^{-1})$ have the orders 3, 3, and 6 respectively. The pure transport delay is $n_k = 2$ samples, so the polynomials coefficients are given by:

$$\begin{aligned} a_{1T} &= -2.909, a_{2T} = 2.822, a_{3T} = -0.9131 \\ b_{1T} &= 0.01843, b_{2T} = -0.005826, b_{3T} = -0.001592 \end{aligned} \tag{21}$$

B. Refrigerant liquid level closed-loop SISO ARMAX discrete-time model:

$$\begin{aligned} y_2(t) &= q^{-2}\frac{B(q^{-1})}{A(q^{-1})}u_2(t) + \frac{C(q^{-1})}{A(q^{-1})}e_2(t) \\ &= G_{u_2}(q^{-1})u_2(t) + G_{e_2}(q^{-1})e_2(t) \end{aligned} \tag{22}$$

$$G_{u_2}(q^{-1}) = q^{-2}\frac{B(q^{-1})}{A(q^{-1})}, G_{e_2}(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})} \tag{23}$$

$$A(q^{-1}) = 1 - 0.9938q^{-1} \tag{24}$$

$$B(q^{-1}) = 1.129q^{-1} + 0.152\,q^{-2} \tag{25}$$

$$C(q^{-1}) = 1 + 0.7744q^{-1} \tag{26}$$

$$\begin{aligned} a_{1L} &= -0.9938 \\ b_{1L} &= 1.129, b_{2L} = 0.152 \\ c_{1L} &= 0.7744 \end{aligned}$$

In Figure 18 we gather a valuable information about the validation of the both ARMAX models for entire range of the input-output measurements data set. To validate these two models in MATLAB can be used the typical MATLAB function called *compare* found in Identification MATLAB Toolbox. The input-output data set measurement of length 3006 samples are generated in open-loop simulations based on SIMULINK model of chiller system developed in section 3. The first segment of 1800 samples from input-output data set is used in prediction phase, and the second segment containing the samples between 1800 and 3006 is used in validation phase. The simulation results shown in Figure 18 reveal a very good estimation accuracy of the both ARMAX SISO models.

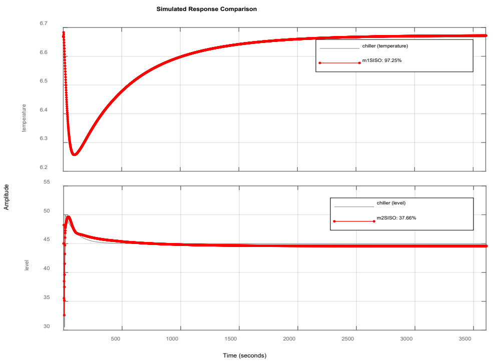


Figure 18. Estimated ARMA SISO model validation for the booth closed-loops, chilled water temperature and Refrigerant liquid level

The closed-loop PI control laws for the both SISO ARMAX models are similar to those developed in the equations (1) - (4), but their design is more suitable in discrete-time to match the both SISO ARMAX models. The both PI closed loops control strategies are built in SIMULINK and the simulations results are performed partially in SIMULINK and finalized in MATLAB, as is shown in Figure 19, and Figure 20. In Figure 19 are shown the simulations results for chilled water temperature in Evaporator closed-loop control subsystem, and in Figure 20 can be seen the simulation results of closed-loop PI controller performance for Refrigerant liquid level in Condenser subsystem. The both PI control loops are completely independent, without any interference between them. They are based on the decoupling loops assumption, proved in open-loop MATLAB simulation environment. The simulation results reveal good accuracy and fast transient, especially for Refrigerant liquid level PI control shown in Figure 20. For the first approximate 500 samples the chilled water temperature in Evaporator reaches low values compared to measured temperature from the input-output measurements data set required for the estimation, prediction and validation of both ARMAX models. Fortunately, after this starting transient period the controlled chilled water temperature reaches the output measurements data set values with high accuracy. This is an interesting modelling key issue generated by the linearization of the plant dynamics, leading for a short period of time during the transient to a considerable degradation in PI controller performance. In addition the constraints on the operating range of the actuators could be also a big issue for PI controller design. In contrast, the simulation results from Figure 20 reveal a very good accuracy, and also a fast transient.

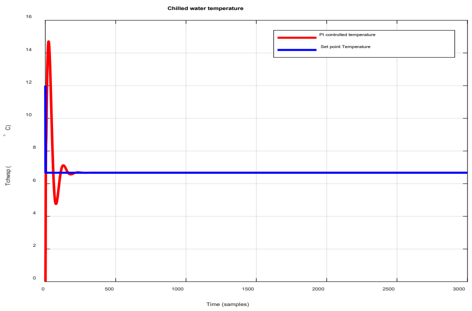


Figure 19. PI controlled chilling water Evaporator temperature closed-loop ARMA SISO model compared to the set point temperature value

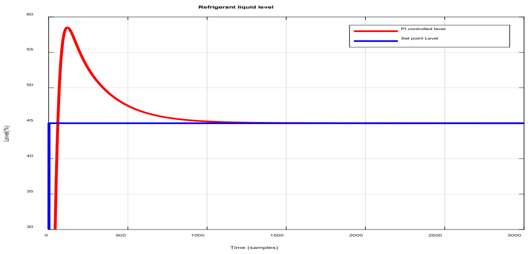


Figure 20. PI controlled liquid level closed-loop ARMA SISO model compared

Set point level value

The Simulink models for the both PI controllers based on ARMAX SISO models are shown in Figure 21 and Figure 22, and also for more clearness they are attached in Annex-2. The rejection disturbance performance of both PI controllers is shown in Figure 23 and Figure 24.

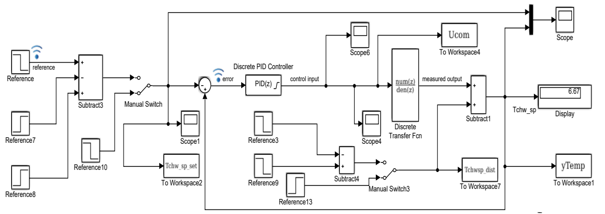


Figure 21. The SIMULINK model of chilled water temperature PI closed-loop Evaporator control subsystem

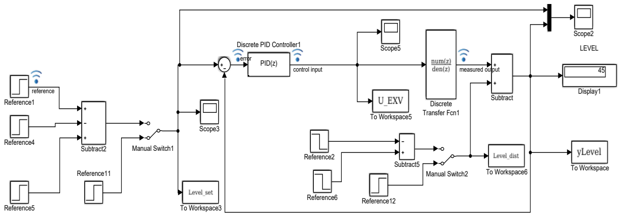


Figure 22. The SIMULINK model of the Refrigerant liquid level PI closed-loop Condenser control subsystem

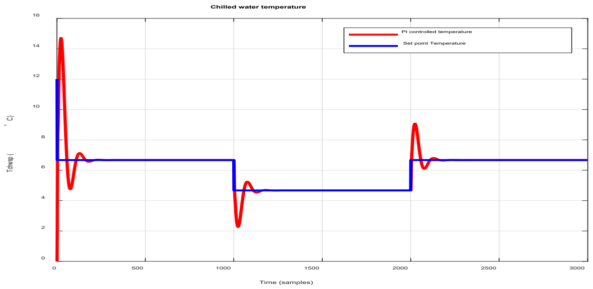


Figure 23. Robustness of PI controlled chilling water Evaporator temperature closed-loop ARMA SISO model to the changes in set point temperature values

The simulations results in Figure 23 and Figure 24 show a great load disturbance rejection performance for both PI controllers. In fact, these results prove also the robustness of the both PI controllers for changes in the chiller system load.

Concluding, compared to the both PI control strategies developed in subsection 4.1, the proposed PI controllers in this subsection perform successful when using linear ARMAX SISO models for both control subsystems Evaporator and Condenser.

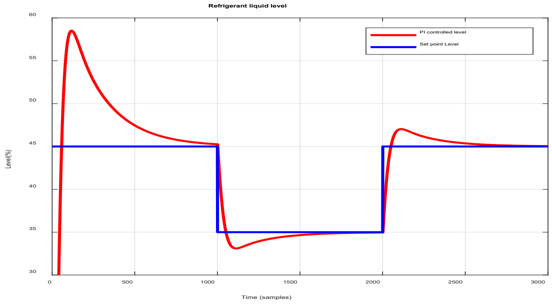


Figure 24 Robustness of PI controlled refrigerant liquid level in Condenser closed-loop ARMA SISO model to the changes in set point level values

4.4. PPI Closed-Loop Controller Design using Linear ARMAX Models for Centrifugal Chiller Control System with Time Delay

The novelty of this subsection is a new design approach of the standard PI control strategy as a predictive discrete-time PI-plus (PIP) controller within the context of non-minimum state space (NMSS) control system, as is formulated in [13]. Furthermore, the non-minimal state space (NMSS) representation in contrast to minimal state space realizations "*seems to be the natural description of a discrete-time transfer function, since its dimension is dictated by the complete structure of the model*" as is stated in [13]. The minimal state space descriptions account only for the order of the denominator of the transfer function, and also the state variables assigned to each description not always have physical meaning, usually representing combinations of input and output signals [13]. Therefore, the resulting control algorithm in the new PIP control design approach within the NMSS context "*can be interpreted as a logical extension of the conventional PI controller, facilitating its straightforward implementation using a standard hardware-software arrangement*", as is stated also in [13]. Further, the controller design methodology is the same as for an equivalent discrete-time Smith predictor (SP) controller for time delay systems, under certain non-restrictive pole assignment conditions, as is mentioned in [13].

The proposed PPI controller design follows the same methodology described in [13], encouraged by its great results obtained when applied in a MIMO ALSTOM nonlinear gasifier control plant [15]. Compared to the well-known Smith Predictor controller, in the new design approach the predictive PI-plus controller (PIPC) has more design flexibility and robustness [13].

Furthermore, the tuning parameters optimal selection, such as the weighting matrices required in the optimal performance criterion formulation can achieve multiple objectives. In addition, this new approach can be easy extended to MIMO PIPCs based on ARMAX models, such in [15]. In our case study the centrifugal chiller control system is represented in discrete-time by two SISO loops ARMAX models, given by the equations (17) - (21), and also in z-domain described by the equations (15) –(16). To simplify further the description in z-domain we neglect the white noise term, and the equations (15) and16) can be written in the following compact form, similar to those introduced in [13]:

$$y(k)=\frac{(b_{n_k}z^{-1}+b_{n_{k+1}}z^{-2}+b_{n_{k+2}}z^{-3}+...+b_{n_{k+m-1}}z^{-m})z^{-n_k+1}}{1+a_1z^{-1}+a_2z^{-2}+...+a_nz^{-n}}u(k) \qquad (27)$$

$$=\frac{B(z^{-1})z^{-n_k+1}}{A(z^{-1})}=H_y^u(z)$$

In (27) $H_y^u(z)$ represents the transfer function in z-domain of the output y of centrifugal chiller control SISO loop system with respect to corresponding control input u. Following the design control methodology described in [14], the model of the control system defined by the discrete transfer function given in (27) can be represented by the following NMSS equations:

$$x(k)=Fx(k-1)+gu(k-1)+dy_{sp}(k) \qquad (28)$$

$$y(k)=hx(k) \qquad (29)$$

where the $n+m+n_k-1$ dimensional NMSS state vector $x(k)$ is defined as follows:

$$x(k)=[y(k)\ y(k-1)...y(k-n-1)\ u(k-1)...u(k-m-n_k+2)\ z(k)]^T \qquad (30)$$

The new variable $z(k)$ introduced in (30) represents the discrete-time integral of error between the set point input (reference) $y_{sp}(k)$ and the control system output $y(k)$:

$$z(k)=z(k-1)+(y_{sp}(k)-y(k)) \qquad (31)$$

The matrix F, and the vectors g,d,h are calculated based on the methodology described in [13-17]. Due to the editing constraints we give below only the matrix F and the vectors g,d,h for a delay $n_k=2$ since matches the SISO temperature loop, and if is the case, for $n_k\geq 2$ you can see the general description used in [13] for SISO systems, that can be also extended for MIMO systems. Similar, in [15] is investigated the case $n_k=1$ for a MIMO system that can be tailored on the Refrigerant liquid level SISO loop in our case study. For SISO systems with a delay $n_k=2$ that matches also the temperature SISO loop, the matrix F and the vectors g,d,h can be written as following [13]:

$$F=\begin{bmatrix} -a_1 & -a_2 & ... & -a_{n-1} & -a_n & b_{n_k} & ... & b_{n_k+m-2} & b_{n_k+m-1} & 0 \\ 1 & 0 & ... & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & ... & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & ... & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ . & . & ... & . & . & . & ... & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ a_1 & a_2 & ... & a_{n-1} & a_n & -b_{n_k} & ... & -b_{n_k+m-2} & -b_{n_k+m-1} & 1 \end{bmatrix} \qquad (32)$$

$$\begin{aligned} g&=[0\ 0\ ...0\ 1\ 0\ 0.....0\ 0\ 0]^T \\ &=[0\ 0\ ...0\ g_i{=}1\ 0\ 0.....0\ 0\ 0]^T \\ d&=[0\ 0\ ...0\ 0\ 0\ 0...0\ 0\ 1]^T \\ h&=[1\ 0\ ...0\ 1\ 0\ 0.....0\ 0\ 0] \end{aligned} \qquad (33)$$

The element $g_i=1$ of the vector g occupies the same position as the component $u(k-1)$ occupies in the state vector $x(k)$ defined by (30). The components of the vectors d,h are identified straightforward due to their simple structures. A MIMO control system characterized by a transport delay $n_k=1$ is described in [15] by the following expressions:

$$\begin{aligned} G&=\begin{bmatrix} B_1 & Z_p & Z_p & ... & Z_p & I_p & Z_p & ... & Z_p & Z_p & -B_1 \end{bmatrix}^T \\ D&=\begin{bmatrix} Z_p & Z_p & Z_p & ... & Z_p & Z_p & Z_p & ... & Z_p & Z_p & I_p \end{bmatrix}^T \\ H&=\begin{bmatrix} I_p & Z_p & Z_p & ... & Z_p & Z_p & Z_p & ... & Z_p & Z_p & Z_p \end{bmatrix}^T \end{aligned} \qquad (34)$$

$$I_p=\begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & . & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{bmatrix}_{p\times p} \quad , \quad Z_p=O_{p\times p}$$

$$F=\begin{bmatrix} -A_1 & -A_2 & ... & -A_{n-1} & -A_n & B_2 & B_3 & ... & B_{m-1} & B_m & 0 \\ I_p & Z_p & ... & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p \\ Z_p & I_p & ... & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p \\ . & . & . & . & . & . & . & . & . & . & . \\ Z_p & Z_p & ... & I_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p \\ Z_p & Z_p & ... & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p & Z_p \\ Z_p & Z_p & ... & Z_p & Z_p & I_p & Z_p & Z_p & Z_p & Z_p & Z_p \\ Z_p & Z_p & ... & Z_p & Z_p & Z_p & I_p & Z_p & Z_p & Z_p & Z_p \\ . & . & ... & . & . & . & . & . & . & . & . \\ I_p & Z_p & ... & Z_p & Z_p & Z_p & Z_p & Z_p & I_p & Z_p & Z_p \\ A_1 & A_2 & ... & A_{n-1} & A_n & B_2 & B_3 & ... & B_{m-1} & B_m & I_p \end{bmatrix} \qquad (35)$$

The control law $u(k)$ related to the NMSS realization (28-30) takes the following state variable feedback (SVF) form [14-16]:

$$u(k)=-k_Fx(k) \qquad (36)$$

$$k_F=[f_0\ f_1\ f_2...f_{n-1}\ g_1\ g_2......g_{m+n_k-2}\ -k_I]^T \qquad (37)$$

where the last component k_I of k_F is the integrator gain of the first block located in the forward path of the PIP control diagram shown in Figure 25.

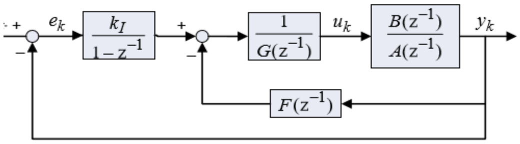


Figure 25. PIP control block diagram (see [16])

It is worth to mention that the linear SVF control law (33) is easy to be implemented in practice, due to the facility storage of the variables in the MATLAB workspace. Furthermore, the inherent SVF formulation allows to investigate any SVF design methodology, such as:

- Poles assignment
- Linear quadratic (LQ) optimization
- Linear Quadratic Gaussian optimization
- H_∞ optimization
- Linear exponential of quadratic Gaussian (LEQG) optimization

The poles assignment PIP controller design methodology in many cases is much simpler and intuitive, since the algebraic results are more perceptibly. In this paper we are focused only on the second approach, linear quadratic (LQ) optimization, but for the interested readers we try to provide only the steps of the poles assignment PIP controller design methodology, as follows:

Step1. Define a characteristic polynomial for the desired behavior of closed-loop PIP control system, i.e. choose the roots of the desired characteristic polynomial inside of unit circle $|z|<1$ to assure the stability of the PIP control system in closed-loop, and to reach the desired performance (poles assignment).

The desired characteristic polynomial that meet the requirements from first step can be written as follows:

$$L(z^{-1}) = 1 + l_1 z^{-1} + l_2 z^{-2} + \ldots + l_{n+m+n_k-1} z^{-n-m-n_k+1} \qquad (38)$$

Step2. Compute the discrete time transfer function in closed-loop for the PIP control system that has the block diagram presented in Figure 24, as is given in [13]:

$$H_y^{y_{sp}}(z) = \frac{k_I B(z^{-1}) z^{-n_k+1}}{(1-z^{-1})[G(z^{-1})A(z^{-1}) + F(z^{-1})B(z^{-1})z^{-n_k+1}] + k_I B(z^{-1}) z^{-n_k+1}}$$

$$H_y^{y_{sp}}(z) = \frac{y(k)}{y_{sp}(k)} = \frac{B_{cl}(z^{-1})}{A_{cl}(z^{-1})}, B_{cl}(z^{-1}) = k_I B(z^{-1}) z^{-n_k+1}, \qquad (39)$$

$$A_{cl}(z^{-1}) = (1-z^{-1})[G(z^{-1})A(z^{-1}) + F(z^{-1})B(z^{-1})z^{-n_k+1}] + k_I B(z^{-1}) z^{-n_k+1}$$

Step 3. Identify the polynomials coefficients $f_i|_{i=\overline{0,n-1}}$, $g_i|_{i=\overline{1,m+n_k-2}}$ of $F(z^{-1})$, and $G(z^{-1})$ respectively, as well as the integrator gain k_I:

$$F(z^{-1}) = f_0 + f_1 z^{-1} + \ldots + f_{n-1} z^{-n+1} \qquad (40)$$

$$G(z^{-1}) = g_0 + g_1 z^{-1} + \ldots + g_{m+n_k-2} z^{-m-n_k+2} \qquad (41)$$

Identification of the polynomials coefficients is based on the following poles assignment equation:

$$A_{cl}(z^{-1}) = L(z^{-1}) \qquad (42)$$

assuming that the closed-loop desired poles are located inside the unit circle to assure the stability, a good transient, as well as a good set point tracking.

The LQ optimization design approach is applied to the both closed-loops SISO ARMAX models (19-20), (25-26), considered as a starting point in order to design two PIP control laws.

Typically, the optimization criterion of LQ optimization design is defined in the same manner as for any quadratic optimization form written for a SISO control system:

$$J = \sum_{k=0}^{\infty} (x(k)^T Q x(k) + r u(k)^2) \qquad (43)$$

where Q is a diagonal weighting matrix of the form:

$$Q = diag([q_1 \quad q_2 \ldots\ldots \quad q_i \quad \ldots\ldots \quad q_{n-1} \quad q_n]), q_{i|_{i=\overline{1,n}}} > 0 \qquad (44)$$

and $r > 0$ is a positive scalar weight on the scalar input u.

The resulting SVF gains are then obtained recursively from the steady state solution of the Algebraic Riccati Equation (ARE), derived from the standard LQ cost function (14) as follows [16]:

$$k_F(k) = [g^T P^{(k+1)} g + r]^{-1} g^T P^{(k+1)} F \qquad (45)$$

$$P^{(k)} = F^T P^{(k+1)} [F - g k_F] + Q \qquad (46)$$

where the matrix P is a symmetrical positive definite matrix with its initial value $P^{(0)} = Q$, and k_F is the control gain vector for SVF.

A. PIP control law for chilled water temperature closed-loop SISO ARMAX discrete-time Evaporator model described by:

$$A(z^{-1}) = 1 - 2.909z^{-1} + 2.822z^{-2} - 0.9131z^{-3}$$

$$B(z^{-1}) = 0.01843z^{-1} - 0.01436z^{-2} - 0.004013z^{-3}$$

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} \quad ,$$

$$a_1 = -2.909, a_2 = 2.822, a_3 = -0.9131$$

$$B(z^{-1}) = b_2 z^{-1} + b_3 z^{-2} + b_4 z^{-3}, b_2 = 0.01843,$$

$$b_3 = -0.01436, b_4 = -0.004013$$

$$n = 3, m = 3, n_k = 2$$

$$F(z^{-1}) = f_0 + f_1 z^{-1} + f_2 z^{-2}$$

$$G(z^{-1}) = 1 + g_1 z^{-1} + g_2 z^{-2} + g_3 z^{-3}$$

The state associated to the PPI controller is given by:

$$x(k) = [y(k) \quad y(k-1) \quad y(k-2) \quad u(k-1) \quad u(k-2) \quad u(k-3) \quad z(k)]^T$$

and the PIP controller parameters computed according to (32) and (33) are given by:

$$F = \begin{bmatrix} \text{-a}_{1T} & \text{-a}_{2T} & \text{-a}_{3T} & \text{b}_{2T} & \text{b}_{3T} & \text{b}_{4T} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \text{a}_{1T} & \text{a}_{2T} & \text{a}_{3T} & \text{-b}_{2T} & \text{-b}_{3T} & \text{-b}_{4T} & 1 \end{bmatrix}$$

$$g=\begin{bmatrix}0 & 0 & 0 & 1 & 0 & 0 & 0\end{bmatrix}^T$$
$$k_{sp}=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}^T$$
$$h=\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

The tuning parameters are set to:

$$Q=\begin{bmatrix}150 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix},\ \text{and } r=1.$$

The simulation results are shown in Figure 26 that reveal a very good tracking performance, fast transient, no oscillations, and a very small overshoot, perhaps the best performance obtained until now compared to previous temperature controllers.

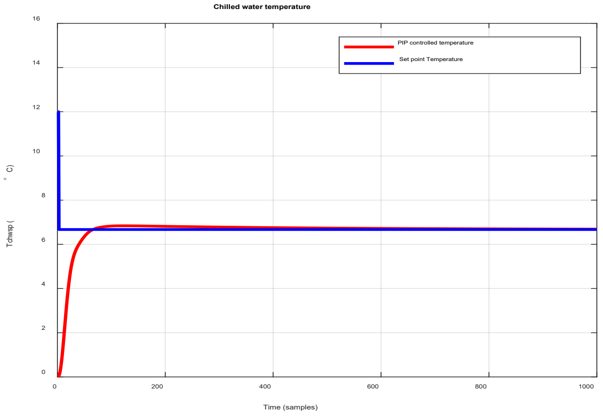


Figure 26. PIP controlled chilling water Evaporator temperature closed-loop ARMA SISO model compared to the set point temperature value

B. PIP control law for closed-loop Refrigerant liquid level SISO ARMAX discrete-time Condenser model described by:

$$\begin{aligned}&A(q^{-1})=1-0.9938q^{-1}\\&B(q^{-1})=1.129q^{-1}+0.152\,q^{-2}\\&C(q^{-1})=1+0.7744q^{-1}\\&a_{1L}=-0.9938\\&b_{1L}=1.129, b_{2L}=0.152\\&c_{1L}=0.7744\end{aligned}$$

$$Z_2=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}=\text{zeros}(2), I_2=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\text{eye}(2)$$

The polynomials degree is smaller compare to temperature loop description, i.e. $n=2, m=2, n_k=1$, thus the state vector will be of small dimension that simplify very much the PIP controller design, such as:

$$x(k)=\begin{bmatrix}y(k) & u(k-1) & z(k)\end{bmatrix}^T$$

and the model parameters are calculated according to (34) replacing $I_p=I_1=1, Z_p=Z_1=0$ to match to an ARMAX SISO closed-loop control, as follows:

$$F=\begin{bmatrix}\text{-a}_{1L} & \text{b}_{2L} & 0\\ 1 & 0 & 0\\ \text{a}_{1L} & \text{-b}_{2L} & 1\end{bmatrix}, g=\begin{bmatrix}\text{b}_{1L} & 1 & \text{-b}_{1L}\end{bmatrix}^T$$

$$\text{ksp}=\begin{bmatrix}0 & 0 & 1\end{bmatrix}^T, \text{h}=\begin{bmatrix}1 & 0 & 0\end{bmatrix}$$

and the tuning parameters are set to

$$Q=\begin{bmatrix}0.1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0.1\end{bmatrix}, r=1$$

The Simulink models of PIP controllers are shown in Figure 27 for chilled water temperature control in Evaporator subsystem, and Figure 28 for liquid Refrigerant level control in Condenser subsystem, respectively.

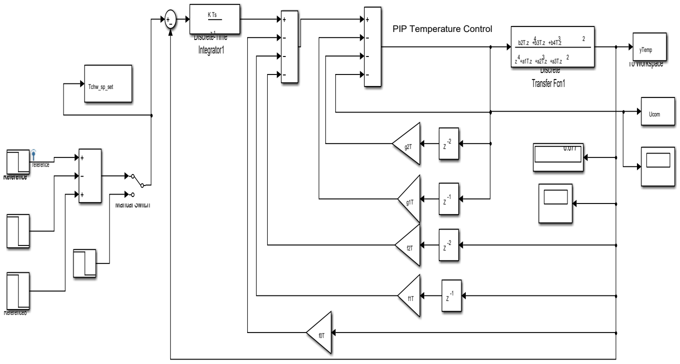


Figure 27. The PIP controller of chilled water temperature in Evaporator Subsystem

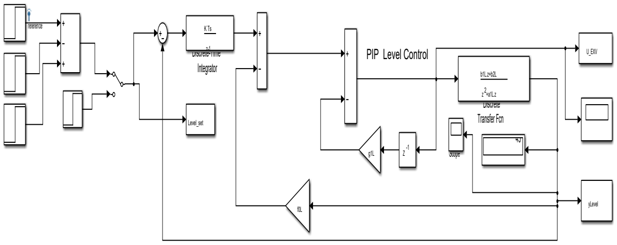


Figure 28. The PIP controller of the liquid Refrigerant level in Condenser Subsystem

The simulation results are shown in Figure 29 that reveal a very fast transient and no oscillations, thus no overshoot and the overall behavior of PIP controller is as an aperiodic element. The same conclusion we can formulate that is the best performance obtained until now compared to previous Refrigerant liquid level controllers. Furthermore, we test the both PIP controllers for robustness to changes in set points, and the results reveal an excellent tracking performance for both, as is shown in Figure 30, and Figure 31 respectively.

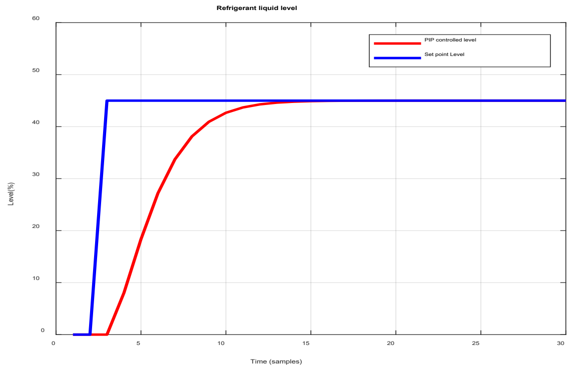


Figure 29. PIP controlled Refrigerant liquid level closed-loop ARMA SISO model compared to the set point level value

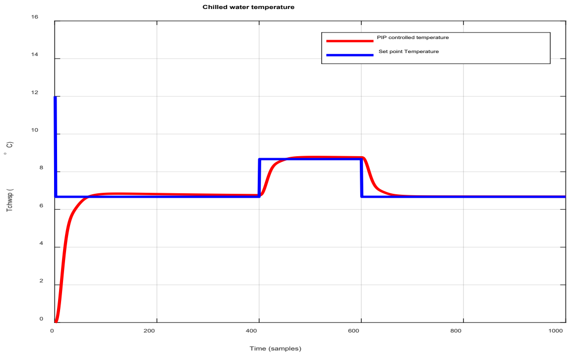


Figure 30.The robustness to changes in input set point for closed-loop PIP controlled chilled water temperature in Evaporator subsystem described by an ARMA SISO model

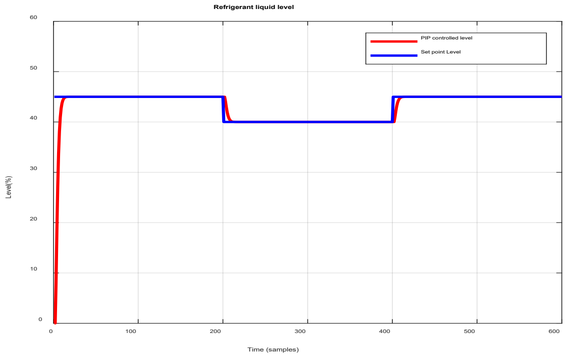


Figure 31. The robustness to changes in input point set for closed-loop PIP controlled Refrigerant liquid level in Condenser subsystem described by an ARMA SISO model

Conflict of Interest

The authors declare no conflict of interest.

Conclusions

In our research work we introduce a new approach controller design, the so called in the literature as PI controller plus [13-16], definitely a new version of PI controller, but much simpler for design, and no more tuning parameters are required compared to standard version of PI controller, i.e. the coefficients k_p, k_I. The third control strategy is an improved PI control version conceived within the context of non-minimum state space control systems with time delay.

The non-minimal state space representation in contrast to minimal state space realizations “seems to be the natural description of a discrete-time transfer function, since its dimension is dictated by the complete structure of the model”, as is stated in [13]. Also, overall the new PIP controller developed in subsection 4.3 outperforms the first two PI controller’s versions proposed in the subsection 4.1 and 4.2 in terms of tracking performance, robustness, convergence speed and overshoot. Furthermore, the PIP controller can be easily extended to control MIMO systems with delay, as is done in [15], thus remains an open research topic for further investigations in the future work.

References

[1] S. Bendapudi, J. E. Braun, et al., “Dynamic Model of a Centrifugal Chiller System - Model Development, Numerical Study, and Validation”, *ASHRAE Trans.*, vol. 111, pp.132-148, 2005.

[2] A. Beyene, H. Guven, et al., (1994),” Conventional Chiller Performances Simulation and Field Data”, *International Journal of Energy Research*, vol.18, pp.391-399, 1994.

[3] J. E. Braun, J. W. Mitchell, et al., “Models for Variable-Speed Centrifugal Chillers”, *ASHRAE Trans.* , New York, NY, USA, 1987.

[4] M. W. Browne, P. K. Bansal, “Steady-State Model of Centrifugal Liquid Chillers", *International Journal of Refrigeration*, vol. 21, no.5, pp. 343-358, 1998.

[5] B.E. Deepa Th. Mannath, “Study and Simulation of the Predictive Proportional Integral Controller in comparison with Proportional Integral Derivative controllers for a two-zone heater system”, Master Thesis, Texas Tech University, 2002.

[6] M. Dhar, W. Soedel, “Transient Analysis of a Vapor Compression Refrigeration System”, in *The XV International Congress of Refrigeration*, Venice, 1979.

[7] J.M. Gordon, K. C. Ng, H.T. Chua, "Centrifugal chillers: thermodynamic modeling and a diagnostic case study", *International Journal of Refrigeration*, vol.18, no. 4, pp.253-257, 1995.

[8] Li Pengfei, Li Yaoyu, J. E. Seem, “Modelica Based Dynamic Modeling of Water-Cooled Centrifugal Chillers”, *International Refrigeration and Air Conditioning Conference*, Purdue University, pp.1-8, 2010. *http://docs.lib.purdue.edu/iracc/1091*.

[9] P. Popovic, H. N. Shapiro, “Modeling Study of a Centrifugal Compressor”, *ASHRAE Trans.,* Toronto, 1998.

[10] M. C. Svensson, “Non-Steady-State Modeling of a Water-to-Water Heat Pump Unit”, in *Proceedings of 20th International Congress of Refrigeration*, Sydney, 1999.

[11] H. Wang, S. Wang, “A Mechanistic Model of a Centrifugal Chiller to Study HVAC Dynamics”, in *Building Services Engineering Research and Technology*, vol. 21, no.2, pp. 73-83, 2000.

[12] A. A, Nada, E. M. Shaban, “The Development of Proportional-Integral-Plus Control Using Field Programmable Gate Array Technology Applied to Mechatronics System”, American Journal of Research Communication, vol. 2, no.4, pp.14-27, 2014.

[13] C J Taylor, A Chotai and P C Young, “Proportional-integral-plus (PIP) control of time delay systems”, in *Proceedings of the Institute Mechanical of Engineers*, vol. 212, no.1, pp.37-47, 1998.

[14] P. C. Young, A. P. McCabe, A. Chotai “State-dependent parameter nonlinear systems: Identification, Estimation and Control”, in *IFAC, 15th Triennial World Congress*, Barcelona, Spain, 2002.

[15] C. J. Taylor and E.M. Shaban, “Multivariable Proportional-Integral-Plus (PIP) control of the ALSTOM nonlinear gasifier simulation”, in *IEE Proceedings Control Theory and Applications*, vol. 153, no.3, pp.277–285, 2006.

[16] E. M. Shaban, A. A. Nada, “Proportional Integral Derivative versus Proportional Integral plus Control Applied to Mobile Robotic System”, *Journal of American Science*, vol.9, no.12, pp.583-591, 2013, doi:10.7537/marsjas091213.76.

[17] M. Zaheeruddin, N. Tudoroiu, “Neuro - PID tracking control of a discharge air temperature system”, Elsevier, *Energy Conversion and Management*, vol. 45, pp.2405–2415, 2004.

[18] MATLAB R2017b Documentation, *https://www.mathworks.com/help/ident/ref/armax.html.* Accessed at February 3rd, 2018.

[19] I.D. Landau, R. Lozano, M. M'SAAD, A. Karimi, *Adaptive Control – Chapter 2: Discrete-Time System Models for Contro*l, Springer Link, pp. 35-53, 2011, DOI: 10.1007/978-0-85729-664-1_2.

[20] L. Ang, "Comparison between Model Predictive Control and PID Control for water-level maintenance in a two-tank system", Master Thesis, University of Pittsburgh, 2010.

[21] Carnegie Mellon Lab, University of Michigan, "Control Tutorials for MATLAB", http://ctms.engin.umich.edu.

[22] M. Ning, "Neural Network Based Optimal Control of HVAC&R Systems", PHD Thesis, Department of Building, Civil and Environmental Engineering. Montreal, Quebec: Concordia University, 2008.

[23] ASHRAE Handbook-HVAC Systems and Equipment, 2008.

[24] S.A. Korpela, Principles of turbomachinery. A John Wiley & Sons, Inc., Publication. 2011.

[25] S.P.W. Wong and S.K. Wang," System Simulation of the performance of a Centrifugal chiller Using a Shell-and-Tube-Type Water-Cooled Condenser and R-11 as Refrigerant", in ASHRAE Trans, pp.445-454, 1989.

[26] F.W. Yu and K. T. Chan, "Improved Condenser Design and Condenser-Fan Operation for Air-Cooled Chillers", Applied Energy vol. 83, pp. 628-648, 2006.

Annex-1. Nonlinear Model for Centrifugal Chiller System

1.1. Development of the centrifugal chiller model

A schematic diagram of a centrifugal chiller system is shown in Figure 1. The main components of the system include: a flooded evaporator; a flooded water-cooled condenser; a centrifugal compressor and an electronic expansion valve. These components and flow direction of the fluids, including the water and the refrigerant loops are depicted in Figure 1.

In order to simplify the model, the following assumptions are made [22-26]:

1) The heat exchange between the chiller and the ambient air is neglected;
2) 15% of the total refrigerant mass flow is used for cooling the hermetic compressor motor;
3) The refrigerant distributes in the shell side of the evaporator and condenser evenly in superheated state, or saturated mixture state of liquid and vapour, or sub-cooled state.
4) The pressure loss due to refrigerant flow in the tubes, discharge line and suction elbow due to friction are neglected, which means that the evaporation pressure is considered as the suction pressure of the compressor, and condensing pressure is regarded as the discharge pressure of the compressor.
5) Hot gas bypass is closed during the operation and simulation.

By applying the fundamental principles of mass, momentum and energy balances, sets of equations for the component models and the overall system model were developed. These are described in the following sections.

1.1.1. Flooded evaporator model

A flooded evaporator is commonly used in centrifugal chiller systems. In the evaporator, the refrigerant R134a flows in the shell side and the chilled water flows inside the tubes in a two-pass circuit. According to the state of the refrigerant, a moving boundary exists, which divides the evaporator into two sections, a two-phase section (TP) where the refrigerant is a mixture of liquid and vapor, and a superheat section (SH) where the refrigerant is superheated. Based on the energy and mass conservation principle, the heat exchange between the chilled water and refrigerant are expressed by the following equations:

$$C_{chw,tp,1}\frac{dT_{chw,1}}{dt}=\dot{m}_{chw,tp,1}\,c_w(T_{chw,r}-T_{chw,1})-N_{tb,ev,tp,1}U_{ev,tp,}LMTD_{ev,tp,1} \tag{1}$$

$$C_{chw,tp,2}\frac{dT_{chw,sp,1}}{dt}=\dot{m}_{chw,tp,2}\,c_w(T_{chw,1}-T_{chw,sp,1})-N_{tb,ev,tp,2}U_{ev,tp}LMTD_{ev,tp,2} \tag{2}$$

$$C_{chw,sh}\frac{dT_{chw,sp,2}}{dt}=\dot{m}_{chw,sh}\,c_w(T_{chw,1}-T_{chw,sp,2})-N_{tb,ev,sh}U_{ev,sh}LMTD_{ev,sh} \tag{3}$$

$$T_{chw,sp}=\frac{N_{tb,ev,tp,2}T_{chw,sp,1}+N_{tb,ev,tp,1}T_{chw,sp,2}}{N_{tb,ev,tp,2}+N_{tb,ev,tp,1}} \tag{4}$$

$$C_{re,ev,sh}\frac{dT_{re,ev,out}}{dt}=N_{tb,ev,sh}U_{ev,sh}LMTD_{ev,sh}-\dot{m}_{re,ev}\,c_{p,re,v}(T_{re,ev,out}-T_{re,tp,ev}) \tag{5}$$

where $\dot{m}_{re,ev}$ is the refrigerant vapour generation rate in the evaporator, kg/s, which is calculated from (6).

$$\dot{m}_{re,ev}=\frac{N_{tb,ev,tp,1}U_{ev,tp,}LMTD_{ev,tp,1}+N_{tb,ev,tp,2}U_{ev,tp,}LMTD_{ev,tp,2}}{h_{gh,Pev}}+x_v\,(\dot{m}_{re,EXV}+\dot{m}_{re,motor}) \tag{6}$$

In addition, from the mass conservation principle, the following equation was derived for the refrigerant flowing in the shell [22]:

$$\dot{m}_{re,EXV}+\dot{m}_{re,motor}-\dot{m}_{re,com}=\int_0^{Vtp,ev}\frac{\partial\rho_{re}}{\partial t}dV+\int_{Vtp,ev}^{Vev}\frac{\partial\rho_{re}}{\partial t}dV \tag{7}$$

$$=\int_0^{Vtp,ev}\frac{\partial(\overline{\gamma_{v,ev}}\rho_{re,satv,Pe}+(1-\overline{\gamma_{v,ev}})\rho_{re,satl,Pe})}{\partial t}dV+\int_{Vtp,ev}^{Vev}\frac{\partial\rho_{re,sh,Pe}}{\partial t}dV$$

According to the Leibniz integral rule,

$$\frac{d}{dz}\left(\int_{x_1(z)}^{x_2(z)} f(x,z)dx\right)=\int_{x_1(z)}^{x_2(z)}\frac{\partial f}{\partial z}dx+f(x_2(z),z)\frac{dx_2}{dz}-f(x_1(z),z)\frac{dx_1}{dz} \tag{8}$$

and by assuming that the density of the refrigerant vapor in the SH section is equal to the saturated vapor density at evaporator pressure, equation(7) can be rewritten as:

$$\overset{\bullet}{m}_{re,EXV}+\overset{\bullet}{m}_{re,motor}-\overset{\bullet}{m}_{re,com}=$$

$$\left(\frac{d}{dt}\left(\int_0^{Vtp,ev}[\overline{\gamma_{v,ev}}\rho_{re,satv,Pev}+(1-\overline{\gamma_{v,ev}})\rho_{re,satl,Pev}]dV\right)-\rho_{re,satv,Pev}\frac{dV_{tp,ev}}{dt}\right)$$

$$+\left(\frac{d}{dt}\left(\int_{Vtp,ev}^{Vtp}\rho_{re,v}dV\right)+\rho_{re,satv,Pev}\frac{dV_{tp,ev}}{dt}\right)$$

$$=(V_{ev}-V_{tp,ev}+\overline{\gamma_{v,ev}}V_{tp,ev})\frac{d\rho_{re,satv,Pev}}{dP_{ev}}\frac{dP_{ev}}{dt}+(1-\overline{\gamma_{v,ev}})V_{tp,ev}\frac{d\rho_{re,satl,Pev}}{dP_{ev}}\frac{dP_{ev}}{dt} \tag{9}$$

From equation (9), we can get the relationship between the evaporation pressure and the mass of the refrigerant inside the evaporator.

The total number of tubes $N_{tb,ev,tp,1}$, $N_{tb,ev,tp,2}$ and $N_{tb,ev,tp,sh}$ in the two-phase section 1, two-phase section 2 and superheat section respectively are determined by the tubes distribution in the evaporator vessel and the refrigerant mass in the evaporator. By neglecting the refrigerant mass stored in the circulating pipes and compressors, and by assuming that the tubes are distributed evenly in the lower half of evaporator vessel, $N_{tb,ev,tp,1}$, $N_{tb,ev,tp,2}$ and $N_{tb,ev,tp,sh}$ can be determined as follows:

$$V_{net,ev}=0.125\pi D_{ev}^2L_{ev}-0.25\pi N_{tb,ev,total}D_{tb}^2L_{ev} \tag{10}$$

$$V_{tp,ev}=\frac{m_{re,charge}-m_{re,c}}{\overline{\gamma_{v,ev}}\rho_{re,satv,Pev}+(1-\overline{\gamma_{v,ev}})\rho_{re,satl,Pev}} \tag{11}$$

$$N_{tb,ev}=\frac{V_{re,ev}}{V_{net,ev}}N_{tb,ev,total} \tag{12}$$

if $N_{tb,ev}>N_{tb,ev,total}$, $N_{tb,ev,tp,1}=N_{tb,ev,tp,2}=0.5N_{tb,ev,total}$, $N_{tb,ev,sh}=0$;

else if $0.5N_{tb,ev,total}<N_{tb,ev}<N_{tb,ev,total}$, $N_{tb,ev,tp,1}=0.5N_{tb,ev,total}$, $N_{tb,ev,tp,2}=N_{tb,ev}-N_{tb,ev,tp,1}$, $N_{tb,ev,sh}=N_{tb,ev,total}-N_{tb,ev}$;

else if $N_{tb,ev}<0.5N_{tb,ev,total}$, $N_{tb,ev,tp,1}=N_{tb,ev}$, $N_{tb,ev,tp,2}=0$, $N_{tb,ev,sh}=N_{tb,ev,total}-N_{tb,ev}$.

1.1.2. Flooded condenser model

Based on the mass and energy conservation principle, the heat transfer between the refrigerant and the cooling water inside the condenser was described by the following dynamic equations:

$$C_{cw}\frac{dT_{cw,r}}{dt}=N_{tb,1,c}U_{tb,sh,c}LMTD_{sh,1,c}-\frac{2N_{tb,1,c}}{N_{total,c}}\overset{\bullet}{m}_{cw}c_w(T_{cw,r}-T_{cw,1}) \tag{13}$$

$$C_{cw}\frac{dT_{cw,sh}}{dt}=N_{tb,2,c}U_{tb,sh,c}LMTD_{sh,2,c}-\frac{2N_{tb,2,c}}{N_{total,c}}\dot{m}_{cw}c_w(T_{cw,sh}-T_{cw,sp}) \tag{14}$$

$$C_{cw}\frac{dT_{cw,tp}}{dt}=N_{tb,3,c}U_{tb,tp,c}LMTD_{tp,c}-\frac{2N_{tb,3,c}}{N_{total,c}}\dot{m}_{cw}c_w(T_{cw,tp}-T_{cw,sp}) \tag{15}$$

$$C_{cw}\frac{dT_{cw,sc}}{dt}=N_{tb,4,c}U_{tb,sc,c}LMTD_{sc,c}-\frac{2N_{tb,4,c}}{N_{total,c}}\dot{m}_{cw}c_w(T_{cw,sc}-T_{cw,sp}) \tag{16}$$

$$\begin{aligned}C_{cw}\frac{dT_{cw,1}}{dt}&=N_{tb,2,c}U_{tb,sh,c}LMTD_{sh,2,c}+N_{tb,3,c}U_{tb,tp,c}LMTD_{tp,c}+N_{tb,4,c}U_{tb,sc,c}LMTD_{sc,c}-\\&\frac{2(N_{tb,2,c}+N_{tb,3,c}+N_{tb,4,c})}{N_{total,c}}\dot{m}_{cw}c_w(T_{cw,1}-T_{cw,sp})\end{aligned} \tag{17}$$

$$C_{re,sc}\frac{dT_{re,sc,c}}{dt}=(\dot{m}_{re,EXV}+\dot{m}_{re,motor})c_{p,\mathrm{Re},l,c}(T_{re,tp,c}-T_{re,sc,c})-N_{tb,4,c}U_{tb,sc,c}LMTD_{sc,c} \tag{18}$$

Note that the saturated refrigerant temperature $T_{re,tp,c}$ is determined by the condensing pressure P_c which will be described in the next section. Furthermore, the number of tubes $N_{tb,1,c}$, $N_{tb,2,c}$, $N_{tb,3,c}$ and $N_{tb,4,c}$ in the1st superheat section, the 2nd superheat section, the two-phase section and the sub-cool section, respectively, are determined by the mass of the refrigerant and the tube distribution in the condenser, which will be described later.

1.1.3. Centrifugal compressor model

The percent volumetric flow Θ and the percent head of the refrigerant through the compressor are modeled as in reference [23]:

$$\Omega=\mu(V_p^2/a^2) \tag{19}$$

$$\Theta=\frac{V_p}{a\pi}\frac{\dot{Q}}{U_{com}N_{comd}D^3} \tag{20}$$

where U_{com} is the input signal of the compressor rotational speed, which is used to control the chilled water supply temperature. Friction losses in the pipes and in the evaporator were neglected. In other words, the evaporation pressure is considered as equal to the suction pressure of the compressor by neglecting the pressure loss across the inlet guide vane (IGV). Similarly, the condensing pressure is assumed to be equal to the discharge pressure.

In addition, the polytropic efficiency [24], the polytropic compression work [1] per unit mass of refrigerant, and the motor efficiency [25] were modeled using the approach cited in the above references:

$$\eta_p=(\frac{n-1}{n})(\frac{\gamma}{\gamma-1}) \tag{21}$$

$$W_P=(\frac{P_{disc}}{\rho_{re,disc}}-\frac{P_{suc}}{\rho_{re,suc}})\frac{\ln(P_{disc}/P_{suc})}{\ln[(P_{disc}\rho_{re,suc})/(P_{suc}\rho_{re,disc})]} \tag{22}$$

$$\eta_{motor} = \alpha_1 + \alpha_2(PLR) + \alpha_3(PLR)^2 \tag{23}$$

From the above equations, the total electrical power input to the compressor is computed from (24).

$$P_{elec} = \frac{\dot{m}_{re,com} W_p}{\eta_P \eta_{motor}} \tag{24}$$

1.1.4. Electronic expansion valve (EXV) model

EXV mainly regulates the flow rate of the liquid refrigerant entering the evaporator to maintain the refrigerant mass balance in the flooded evaporator and flooded condenser. In practice, EXV is used to maintain the liquid level of the refrigerant in the condenser. When the liquid level is higher than the set-point, the EXV opening will increase; and *vice versa*. In the absence of EXV physical characteristics and performance data, the mass flow rate was modeled as a function of opening area as described in reference [26].

$$\dot{m}_{re,EXV} = U_{EXV} C_d A_{EXV,\max} \sqrt{2(P_c - P_{ev})\rho_{re,c,out}} \tag{25}$$

By using the mass balance relationship between the total refrigerant mass inside the condenser, including the liquid refrigerant and the refrigerant mixture, and the liquid level L_{evel}, the total number of tubes in each section of the condenser is calculated as follows:

$$\begin{gathered} m_{re,c} = m_{0,c} + \int_0^t \dot{m}_{re,com} dt - \int_0^t \dot{m}_{re,EXV}\, dt - \int_0^t \dot{m}_{re.motor} dt \\ = [\overline{\gamma_{v,c}} \rho_{re,satv,Pc} + (1 - \overline{\gamma_{v.,c}})\rho_{re,satl,Pc}] \frac{N_{tb,3,c}}{N_{tb,total,c}} V_{net,c} + \rho_{re,satl,Pc} \frac{N_{tb,4,c}}{N_{tb,total,c}} V_{net} \end{gathered} \tag{26}$$

$$V_{L,c} = \beta_1 L_{evel}{}^3 + \beta_2 L_{evel}{}^2 + \beta_3 L_{evel} + \beta_4 \tag{27}$$

$$N_{tb,3,c} + N_{tb,4,c} = N_{tb,total,c} - N_{tb,1,c} - N_{tb,2,c} = \frac{V_{L,c}}{V_{net,c}} N_{tb,total,c} \tag{28}$$

$N_{tb,1,c} = 0.5 N_{tb,total,c}$, while $L_{evel} \le 50\%$

$$N_{tb,3,c} = \frac{\dot{m}_{re,com} h_{re,disc} - N_{tb,1,c} U_{tb,sh,c} LMTD_{sh,1,c} - N_{tb,2,c} U_{tb,sh,c} LMTD_{sh,2,c}}{U_{tb,tp,c} LMTD_{tp,c}} \tag{29}$$

$$\begin{gathered} N_{tb,2,c} = N_{tb,total,c} - \{(m_{0,c} + \int_0^t \dot{m}_{re,com} dt - \int_0^t \dot{m}_{re,EXV}\, dt - \int_0^t \dot{m}_{re.motor} dt) N_{tb,total,c} / V_{net,c} - \\ N_{tb,3,c}[\overline{\gamma_{v,c}} \rho_{re,satv,Pc} + (1 - \overline{\gamma_{v.,c}})\rho_{re,satl,Pc}]\} / \rho_{re,satl,Pc} - N_{tb,3,c} - N_{tb,1,c} \end{gathered} \tag{30}$$

In addition, by assuming an isenthalpic process across the valve, and by applying the energy and mass balance principles, the enthalpy of the refrigerant entering the evaporator is computed from:

$$h_{re,ev,in} = \frac{\dot{Q}_{motor}}{0.15\, \dot{m}_{re,com}} + h_{re,c,out} \tag{31}$$

The model equations were solved using a consistent set of initial conditions. A single compressor chiller system with a cooling capacity of 190-ton was simulated. The open-loop responses of the chiller system at the design operating conditions are presented in Figure 3.

Nomenclature

$A_{EXV,\max}$	maximum opening area of the EXV, m^2
C	thermal capacity, J/℃
C_d	discharge coefficient of the expansion valve
c_w	specific heat of the water, J/(kg*℃)
D	diameter, m
L	length of the evaporator vessel or condenser vessel, m
g	gravitational acceleration, m/s^2
h	rise in pressure head of the refrigerant across the compressor, m
h_{fg}	enthalpy change per unit mass of refrigerant due to phase change, J/kg
h_{re}	enthalpy of the refrigerant, J/kg
L_{evel}	refrigerant liquid level in the condenser
$LMTD$	log-mean temperature difference between the water and the refrigerant, ℃
$\dot{m}$	mass flow rate of the fluid, kg/s
$m_{0,c}$	initial mass of the refrigerant in the condenser, kg
$m_{re,c}$	total mass of the refrigerant in the condenser, kg
$m_{re,charge}$	total mass of the refrigerant charged in the chiller, kg
$\dot{m}_{re,com}$	refrigerant mass flow rate compressed by the compressor, kg/s
$\dot{m}_{re,ev}$	refrigerant vapour generation rate in the evaporator, kg/s
$\dot{m}_{re,EXV}$	refrigerant mass flow rate through the electronic expansion valve, kg/s
$\dot{m}_{re,motor}$	refrigerant mass flow rate used to cool the compressor motor, kg/s
n	polytropic index
$N_{com,\max}$	maximum rotational speed of the compressor, rpm
N_{tb}	number of cropper tubes
P_c	condensing pressure, kPa
P_{elec}	electrical power consumption of the compressor, kW

P_{ev}	evaporation pressure, kPa
PLR	partial load ratio
ΔP_{re}	pressure rise of the refrigerant across the compressor, m
$\dot{Q}_{motor}$	heat generation rate of the compressor, W
t	time, s
T	temperature, ℃
T_{re}	refrigerant temperature, ℃
U	overall heat transfer coefficient between the water and the refrigerant per unit length of copper tube, W/℃
U_{IGV}	Inlet Guide Vane opening position
U_{EXV}	Expansion valve opening position
U_{com}	compressor rotational speed signal
v_{disc}	specific volume of the refrigerant vapor at the compressor discharge, kg/m^3
V_{ev}	total volume of the shell side in the evaporator, m^3
$V_{L,c}$	total volume of the liquid refrigerant and saturated mixture, m^3
V_p	tip speed of the compressor impeller, m/s
$\dot{V}_{re}$	volume flow rate of the refrigerant, m^3/s
$V_{tp,ev}$	total refrigerant volume of the two-phase section, m^3
W_P	polytropic work input per unit mass of refrigerant, kW/kg
x_v	ratio of the vapor refrigerant mass to the total refrigerant mass in the mixture before entering the evaporator

Greek letters

$\bar{\gamma}_v$	average volume fraction of the vapour refrigerant in the TP section
ρ_{re}	density of the refrigerant, kg/m^3
β	angel of the impeller, rad
γ	specific heat ratio of the refrigerant
η_P	polytropic efficiency
η_{motor}	compressor motor efficiency

Subscripts

1	1st section	2	2nd section
c	condenser	cw	cooling water
chw	chilled water	ev	evaporator

imp	impeller	out	outlet
Pc	condensing pressure	Pev	evaporation pressure
r	return water	re	refrigerant
$satv$	saturated refrigerant vapour	$satl$	saturated refrigerant liquid
sc	subcool section	sh	superheat section
sp	supply water	suc	suction of the compressor
$total$	the total number of the tubes	tp	two phase section

Annex-2. SIMULINK models of open-loop and PI and PIP closed-loop control of chiller system

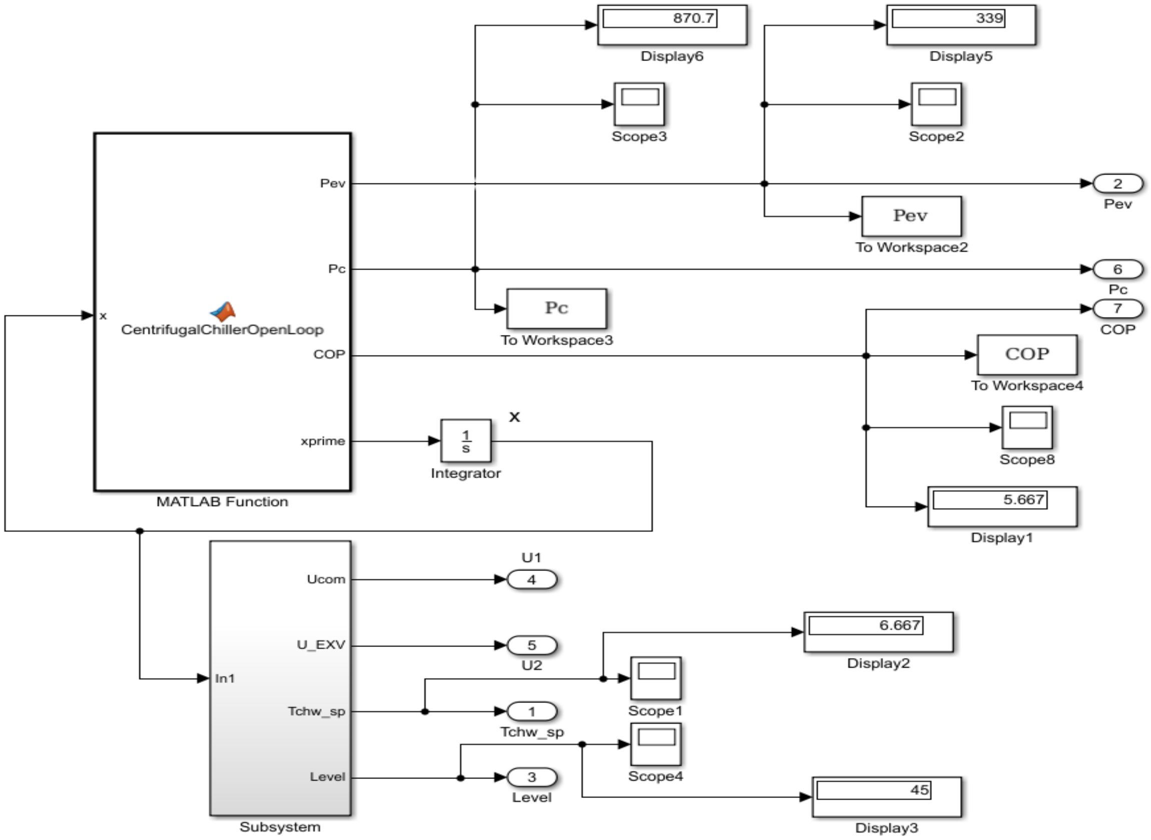


Figure 32. The SIMULINK model of chiller system in open-loop

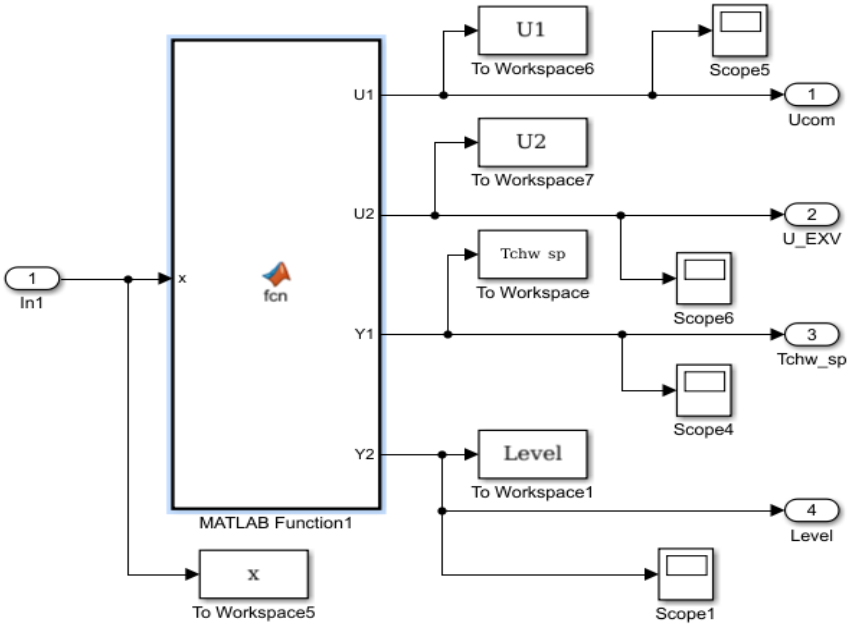


Figure 33. The SIMULINK block selector of chiller system inputs and outputs

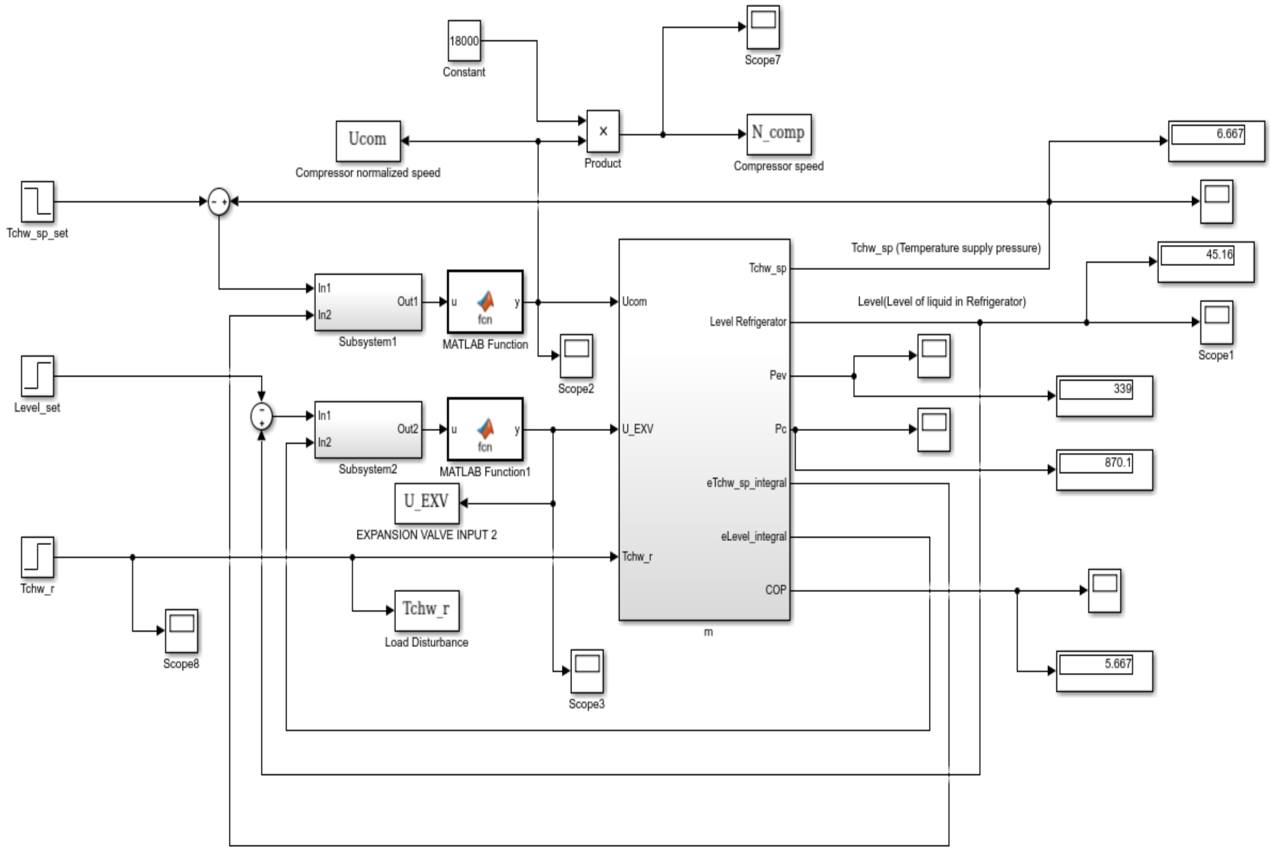


Figure 34. The closed-loop overall PI control of chiller system

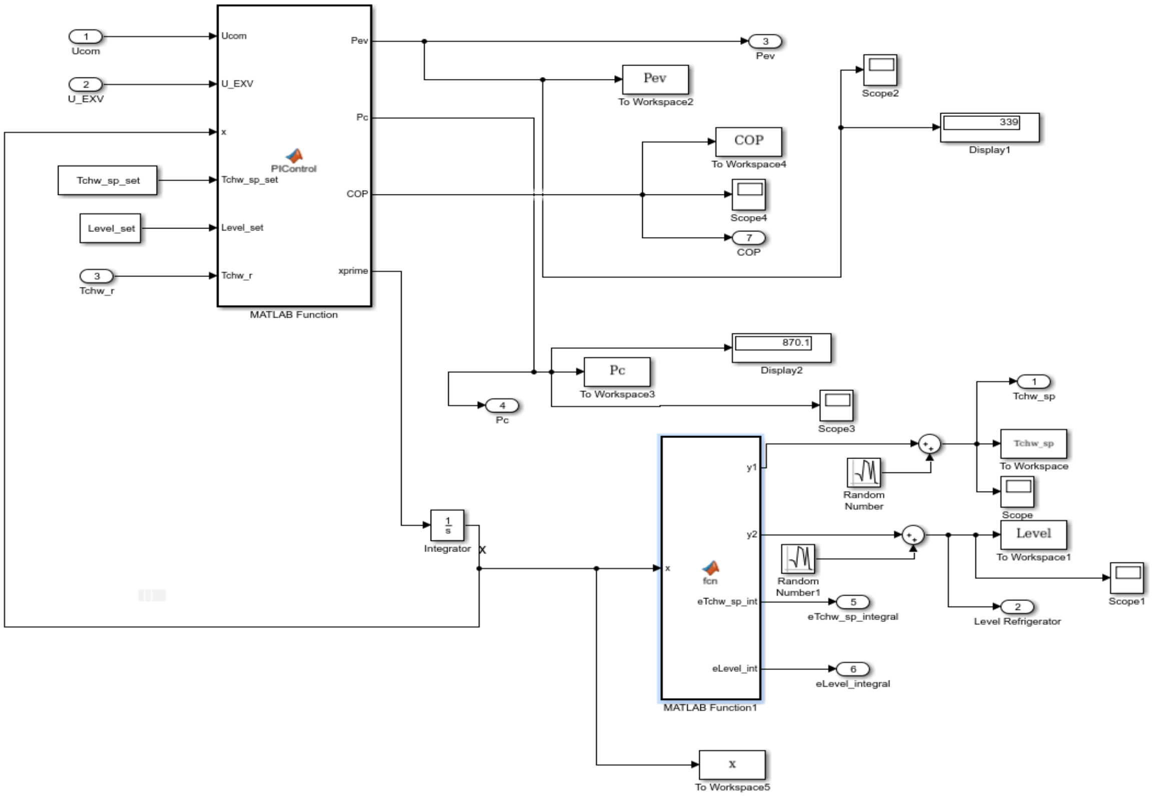


Figure 35. The SIMULINK block models of the state vector integrator and output selector

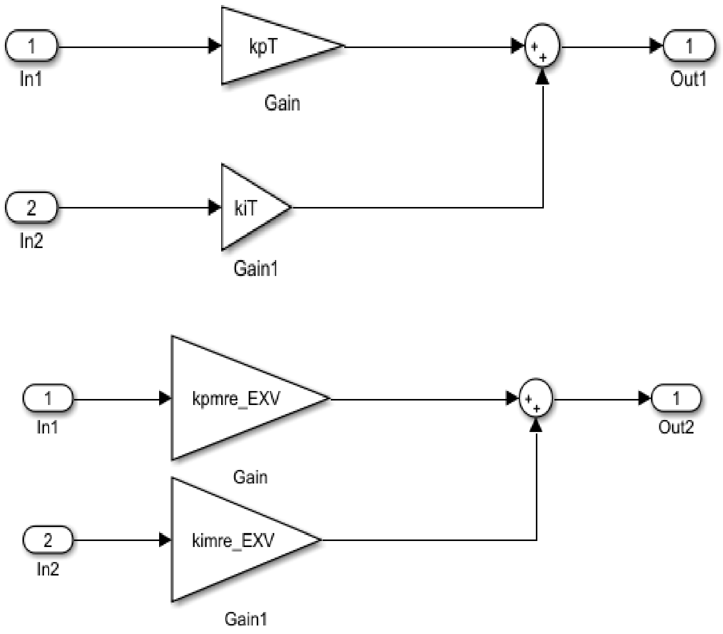


Figure 36 SIMULINK block of PI controllers for chilled water temperature (left) and liquid Refrigerant level (right)

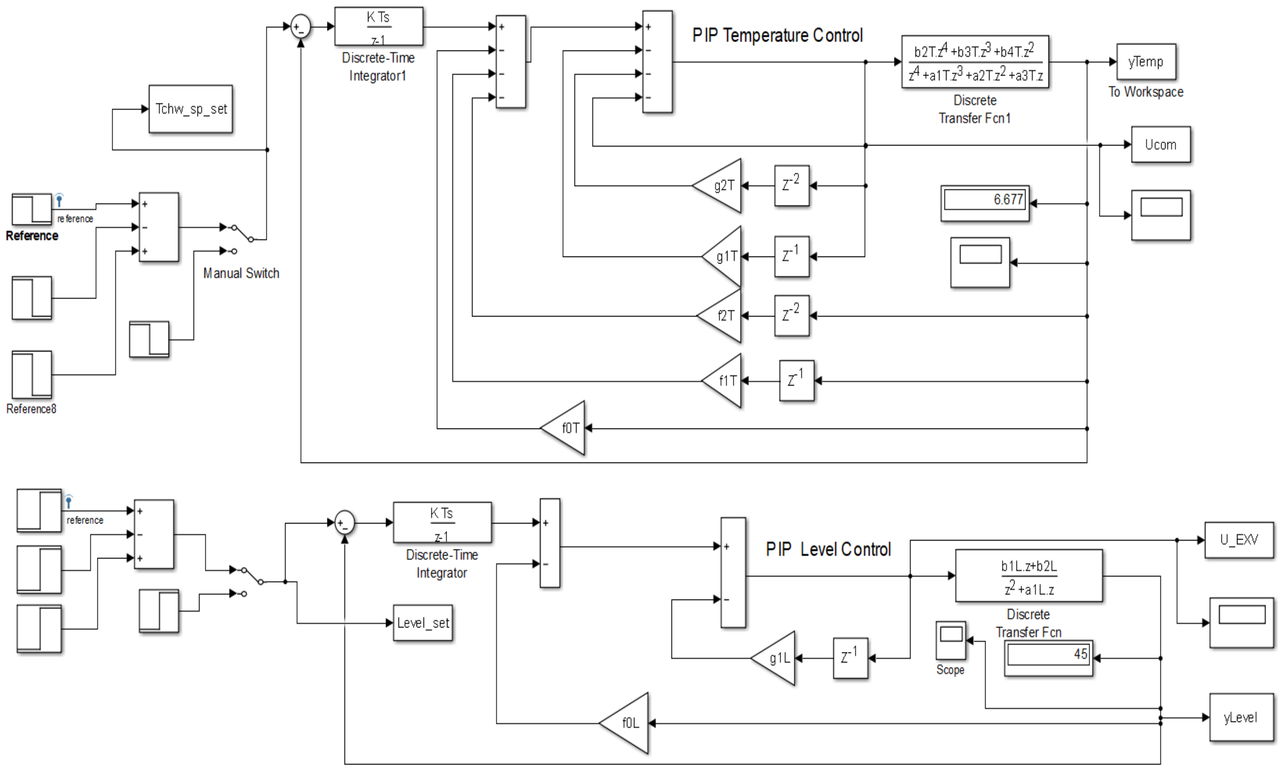


Figure 37. The PIP controller of chilled water temperature in Evaporator (top) and PIP controller of Liquid Refrigerant level in Condenser (right)

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2,309-318 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Which User of technology? Perspectivising the UTAUT model by application of the SFL language Pronoun System towards a systems perspective of technology acceptance and use

Cheryl Marie Cordeiro[*]

Centre for International Business Studies (CIBS), School of Business, Economics and Law, University of Gothenburg, 405 30, Sweden

ARTICLE INFO

Article history:
Received: 24 February, 2018
Accepted: 09 April, 2018
Online: 25 April, 2018

Keywords:
Research methodology
Unified theory of acceptance and use of Technology (UTAUT)
Systemic functional linguistics (SFL)
International business
Technology management
Remote service

ABSTRACT

This study applies systemic functional linguistics (SFL) as complementary framework of analysis of technology acceptance models (TAMs). The purpose is to bridge research methodology language in international business (IB) studies and engineering management science. Currently TAMs and its consolidated version, the Unified Theory of Acceptance and Use of Technology (UTAUT) provides for a typology of one user in one context scenario. The need for the UTAUT model to account for multiple users in multiple work contexts in a single framework of analysis was foregrounded in the study of the workflow processes of a remote services business model of a European founded multinational business enterprise (MBE) with regards to its (i) intra-firm improvements in managing remote services cases, and its (ii) extra-firm selling of life cycle management remote services contracts. The Enterprise has global operations in over 100 countries, of which this study focused on its European operations of improving the quality of remote services for the marine industry. Through an application of SFL unto UTAUT, this study illustrates how multiple users in multiple contexts can be analysed simultaneously, and whose behaviours can be accounted for in a single framework of analysis. The combined SFL UTAUT model addresses the initial statisticity of the UTAUT model, whilst at the same time, expands upon current theoretical perspectives of technology use and acceptance that can be applied in practice.

1. Introduction

Technology acceptance models (TAMs) [1-3] and its consolidated version, the Unified Theory of Acceptance and Use of Technology (UTAUT) [4,5] whilst widely applied to various industry context [6-8], currently provides for a typology of a single user to a single context of use of a specific technology [9-12]. In an era of converging technologies and digital platforms across multiple work spaces, a less static, more sophisticated approach towards the study of technology use and acceptance is needed.

The purpose of this study is to bridge research methodology in language in international business (IB) and engineering management science. It applies the pronoun system found in systemic functional linguistics (SFL) [13,14] from language science as complementary framework of analysis to the UTAUT model. Most technology acceptance studies are quantitative oriented studies. This study contributes to the existing literature by extending the UTAUT model applications by use of SFL, a primarily qualitative analysis approach. The effect of applying the pronoun system in SFL unto UTAUT enables for a simultaneous analysis of multiple users in multiple context of use for a single technology. The need for a multiple user, multiple context perspective in the study of technology use and acceptance arose in the context of studying the workflow processes of a remote services business model of a European founded multinational business enterprise (MBE) with regards to its (i) intra-firm improvements in managing remote services cases and its (ii) extra-firm securing of remote services life cycle management contracts. The Enterprise of study has global operations in over 100 countries, of which this study focuses mostly on its European operations of improving the quality of remote services for the marine industry. The simultaneous understanding of intra- and extra- firm UTAUT processes was necessary for a systems

[*]Corresponding Author: Cheryl Marie Cordeiro, University of Gothenburg, Sweden. Email: cheryl.cordeiro@handels.gu.se

https://dx.doi.org/10.25046/aj030234

integral view of the overall Enterprise remote services business model.

A systems perspective [15,16] of the remote services business model for the Enterprise was needed, because if there was a bottleneck of product-services efficiency, the Enterprise should be able to identify the area of inefficiency quickly. In order to do this, a *systemic* systems overview of workflow operations is necessary by top management of the Enterprise. This called for a research practice method that could give a systems integral perspective of UTAUT. But the simultaneous application of the UTAUT model posed a challenge in different contexts in its current form due to the linear staticity of perspective i.e. one UTAUT model for one context, targeted to understand one type of pre-defined User. The aim of this research methods study is to illustrate how this linear staticity can be circumvented by the application of the pronoun system in systemic functional linguistics (SFL) as theory and framework of data analysis. This is done by collecting and analysing data from both intra-firm and extra-firm perspectives given by Enterprise and Enterprise-related respondents for remote services technology use and acceptance.

This paper begins with a literature review of UTAUT and SFL as fields of research, focusing on UTAUT and SFL as methods for research. This study contains two methods sections. The first pertains to the Enterprise remote services marine sector as case example in illustration of how the SFL pronoun system can be applied to the UTAUT model to unfold various User perspectives. The second methods section illustrates how SFL can be combined in complementarity to the UTAUT model, to form an integrated SFL UTAUT model in studying technology acceptance and use. This is followed by a findings and discussion section, where the empirical findings are presented incorporating a discussion on the SFL UTAUT combined model. This paper ends with an outline of study limitations, future research directions and conclusions.

2. Literature Review

2.1. Technology Acceptance Studies

The past forty years have seen scholars design theories and models in understanding of the influencing elements of acceptance and use of technologies. During the 1970s, the Theory of Reasoned Action (TRA) was put forth by Fishbein and Ajzen [1], which explained a person's behavioural tendencies with the aim of predicting changes and interpreting particular personal behaviour. TRA was formulated based on the assumption that behaviour is shaped by intentions that in turn depend on personal attitudes and subjective norms. A decade later, the Theory of Planned Behaviour (TPB), was proposed as an extension of TRA, working on the assumption that behaviours could be controlled by certain parameters in context [2]. In similar timeframe, Davis, Bagozzi and Warshaw [3] proposed the Technology Acceptance Model (TAM) to explain the causal relationships between internal psychological variables such as beliefs, attitudes and behavioural intention and actual information technology (IT) system. The widely studied and considered valid TAM model was based on the two constructs of the User, which was Perceived Usefulness (U) and Perceived Ease of Use (E). These constructs were considered effective when applied to the understanding (even predicting) of individual acceptance behaviour across various information technologies and their users. Subsequent models developed include the combined TAM and TPB that focused on the impact of experience of the use of technology [4], the TAM2/TAM 3 models as a theoretical extension of the TAM that included the perspectives of subjective norms and job relevance when accepting the use of new technologies [5,6], and the Unified Theory of Acceptance and Use of Technology, or UTAUT [13].

UTAUT [6] arose from a comprehensive conceptual review and empirical study of eight technology acceptance models. The unified model consists of six broad constructs deemed to be significant direct determinants of technology acceptance and use that include:

(i) performance expectancy (PE) – the degree to which an individual believes that using the system will help them improve on job performance
(ii) effort expectancy (EE) – the degree of ease of use by the individual of the system
(iii) social influence (SI) – the degree to which the individual perceives it important that others perceive them to use the new system
(iv) behavioural intention (BI) – the degree to which the individual intends to use the system
(v) use behaviour (UB) – degree of affect on the part of the individual when using the system
(vi) facilitating conditions (FC) – the degree to which the individual feels they have the resources and support (technical / organizational) to use the system

An additional four constructs that moderate technology acceptance and use are gender, age, experience and voluntariness of use. Figure 1 illustrates UTAUT as it appears in Venkatesh *et al.* [13:447]. UTAUT continues to be widely used across various technology management studies even if other models of technology acceptance such as the Model of Acceptance with peer support [14] and the Content Acceptance model [15] have been proposed.

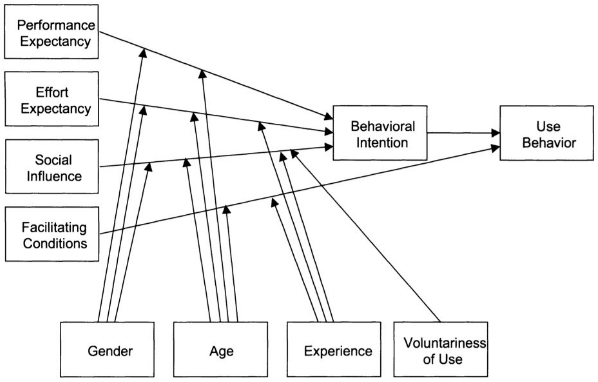


Figure 1. Model of Unified Theory of Acceptance and Use of Technology (UTAUT) [13: 447].

The UTAUT model generally focuses on the causal (cause and effect) relationship between individual attitudes towards using a technology, personal tendencies towards using a technology, actual use of a technology and identifying performance expectancy of a technology. In this model, FCs are taken as the main determinant factor in the use of a technology or system [6].

But while the model of technology acceptance predominantly explain a User's behavioural *expectation* [16] and *intention to use*

[13], *perceived ease of use* [17,18] and *actual use* [6,7] in context (*facilitating conditions*), most studies have implicitly defined the User in a linear model of product-service to the User. In this construct, the User is often defined as the business enterprise end-customer/user that includes a broad range of social actors that include customers [8,19,20], teachers / students [21-24], academics, physicians and nurses [25-27], civilians and military personnel [28,29].

The linear product-service workflow from enterprise to customer as (end-)User disregards within enterprises users who range from service technicians to engineers and product lifecycle managers. This internally defined enterprise users of technology are those who work in support of or are even the creators of the technology for enterprise end-customer/user. The enterprise internal users need to use the same technologies or technological platform to support the product-service workflow from enterprise to end-customer/user. The linear UTAUT concept flow from Engineer to Customer use poses a challenge to the model's inability to overlay contexts of use from intra- to extra- enterprise environments, leaving the perspective of Engineer to Engineer as a field of knowledge under researched.

2.2. Systemic Functional Linguistics

Language is an open adaptive system of choice that humans use language to describe and circumscribe reality [30-32]. We use language in its functional purpose both as process of acting/transacting through time, and as product that helps us situate and identify ourselves in relation to others [33-35]. "Functional linguistic theories are predicated on the claim that language is first and foremost a means for communication between human beings, and that this fact has a deep and all-pervasive influence on the forms that languages take." [36:619]. Language is a resource for making meaning and meaning resides in systemic patterns of options so that language presents itself as a system network potential for meaning [9,10,40].

The usefulness of the UTAUT model when used in research design and method is that it condenses the semiotics (meaning) attributed by Users to technology used in context. Applying UTAUT helps researchers find answers to *why* and *how* Users accept and use technology. User experiences and interactions with technology are expressed through the language system. Because the language system is open and adaptive, the factors abstracted and condensed into UTAUT more often condenses and delimits the full experiential context expressed by Users. In a complementarity, systemic functional linguistics (SFL) is a theory and model of language as a social semiotic system. SFL as theory allows us to account for how language enables us to communicate with each other and with our surrounding in the manner that we do [36]. In that sense, language is its own meta-language, because the language system can be used to study language typology, as well as subjects of other disciplines i.e. other systems. In order to simultaneously apply UTAUT in different contexts of use, this study turns to the underlying unifying theory and framework of language, reflected in the SFL theory and model of language.

The architecture of language reflected in SFL comprises sets of systems and options. When people talk, what is said is usually derived from systemic choice, and SFL gives an open adaptive framework to a general theory of meaning. Language in use is a "system network [that] can represent any domain of activity where choosing can be analysed into small closed sets of options. This does not imply that all such sets of options will be independent of each other – in language they never are." [30:20]. To that end, it would be the consistent dialogic of text and talk carried out by individuals, between individuals and groups of individuals that would also have the power to influence, circumscribe and shape the intersectional relations between producers and users of technology product-services.

Digitalization and the context of Industry 4.0 creates an environment of converging technologies that increases interconnectivity. This increased connectivity between humans and between humans and technology, is also reflected in parallel with the meaning creating system in language when systems simultaneous (rather than dependent) functioning increases its semiotic potential [37]. As such, in the case of the application of UTAUT in understanding technology acceptance, what is needed is a more holistic perspective of not just how a set of defined Users use and accept technology, but how the technology in itself is shaped by feedback from Users in a process of co-evolution of technology product-services. SFL is used in this study to show how it can provide a meta-language framework that allows for the investigation of these dialogic processes between Producers and Users of technology in a context that form today's system of modern ecological habitus. As Bourdieu's [38 81ff] notion of *habitus* has it, "logogenesis provides the material (i.e. semiotic goods) for ontogenesis, which in turn provides the material for phylogenesis; in other words, texts provide the means through which individuals interact to learn the system" and it is through this individual heteroglossic aggregation that a social system evolves. What needs to be investigated in such an interconnected habitus is simultaneity in UTAUT application. This is the ability and capacity for research design and methodology to perspectivise systematically, all actors in all contexts, so that one part of the system can be studied in relation to other parts of the system, larger or smaller. This study focuses in particular, the pronoun system of language that perspectivises deictic points of view such as *I (You), We (They), It* and *Its* applied to the simultaneous application of UTAUT in different contexts. The research questions addressed are:

> RQ1: How can SFL be applied to the UTAUT model towards simultaneous use to reflect intra- and extra- firm perspectives of User acceptance and use of technologies?
>
> RQ2: What contributing factors can be established by using SFL as language theory and framework to broaden the applicability of UTAUT into various contexts of use?

3. Method: Case Example

3.1. Remote Services in the Marine Sector as Enterprise Case Example

The business model of the Enterprise of study since its founding in the late 1800s is sales of products from manufacturing in a business to business context. In the past twenty years, its business model has needed to shift from manufactured products alone to digitally connected product-services. The operations of the Enterprise spans over 100 countries with about 135 000 employees working in global teams that speak different languages across different business sectors. In the era of digitalisation and the Industry Internet of Things (IIoT), the Enterprise began to face differing national and regional data sharing policies as potential barriers to efficiently providing customers with advanced product-service business solutions. Empirical data for this study comes in the form of semi-structured interviews and shadowing of engineers from field studies doing remote services and product maintenance in the marine sector. The field studies were conducted in the Enterprise's office locations for the marine sector. These offices were located in 5 European countries that include Finland, Netherlands, Norway, Sweden and Switzerland (regional remote service centre headquarters location).

Remote services in the Enterprise case example context is defined as a type of system maintenance that allows for a round-the-clock remote monitoring and observation of Enterprise manufactured products. It is part of the Enterprise's advanced services portfolio for the marine sector. Built on big data analytics, such systems of maintenance potentially allow for longer upkeep times of automation processes, sending out alarm signals to Enterprise system engineers, reporting potential system failures and predicting hardware life spans so that components can be changed prior to breakdown. In its efforts towards providing advanced services to its customers, the Enterprise is currently in the phase of building shared digital platforms, both intra- and extra- Enterprise.

The concurrent building of software platforms internal and external to the Enterprise in address of Enterprise efficiency, gave rise to a synchronous duo-User of technology scenario with an overlapping work process timeframe. The first scenario (ScenA) is of a need for a shared internal remote services platform for improved work efficiency for Enterprise employees. ScenA will help employees better coordinate their work efforts across departments, across local business units (LBUs) and its European regional remote services centre located in Switzerland. The second scenario (ScenB), is of a shared external platform for customers that interfaces with ScenA. ScenB is so that the Enterprise remote services team can have continuous contact with its global clients when the vessels are out at sea. ScenA and ScenB Users comprise this duo-User scenario.

3.2. Respondent Profile

In terms of the UTAUT model, what was needed was thus a congruent, synchronous understanding of User acceptance and use of technologies for both ScenA and ScenB, for which the User is differently defined. ScenA would have User defined as Enterprise employees. ScenB would have User defined as Enterprise end-customer, purchasers of the Enterprise remote service systems. Both sets of defined Users for ScenA and ScenB are not homogenous groups of individuals because ScenA for example consists of different employee profiles from Enterprise top managers, system engineers to customer service personnel. ScenA User profiles might also be regionally dispersed, working in different Enterprise business units from headquarters to local business units (LBUs). With most sea faring vessels regionally unconfined, ScenB User profile is also variegated to customers who are located in different parts of the world and who speak different languages. It is not unusual for ScenB Users to encounter different vessel product system use and regional policies. ScenB respondent feedback for this study comes the Enterprise's Product Life Cycle Managers who act as points of contact between the Enterprise and its end-customers. Enterprise end-customers can be located as far as Singapore in Southeast-Asia and are referred to the closest remote services centre in times of distress signals or need of maintenance.

Although the Enterprise has different departments managing remote services that could in total include more than 400 employees in different job functions, there was a challenge in setting up field studies that coincided with respondent time availability for interviews and shadow operations. Enterprise Engineers for example, had to be onsite for both interviews and shadowing of operations to be conducted. As such the total number of respondents available for representation in this study was limited in relation to the given research project timeframe. Interviews and focus group discussions were held with the following Enterprise individuals that categorised as ScenA Users:

- 9 Remote Services Engineer (Headquarters and Local Business Units). These individuals are located at global Enterprise headquarters (Switzerland) that also serves as regional remote service centre for the marine sector. Engineers have different types of specialist knowledge and expertise levels, ranging from Engineering Level 1 (for expert knowledge) and Engineering Level 2 and 3 for lower expertise levels and experience. Cases that cannot be solved for the end-customers are escalated in accordance to specialist knowledge from LBUs to the global centre. Engineers are on call for end-customers 24/7, with no allowance for more than 2 hours downtime for the end-customer. Some engineers are also building the remote diagnostics and maintenance platform whilst the Enterprise is providing this as product-service to the end-customers, so that the engineers in this aspect are both Producer and User of the (same) technology platform.
- 3 Field Engineers. These individuals are located both at global headquarters and at LBUs. Depending on area of expertise and knowledge, the field engineer closest to the vessel is deployed. They are on call 24/7, and can be flown (helicopter or private plane) to vessel sight with immediate notification. Some field engineers in the Enterprise have moved on to the role of Product Life Cycle Managers.
- 2 Marine Remote Service Customer Service (RSCS) Personnel. These individuals are usually the first point of contact between Enterprise end-customers the remote services engineers. They might also receive and manage calls for Enterprise related questions, not pertaining to remote services. Most calls from Enterprise end-customers for remote services in the marine sector are time critical. The remote services team have a response time of maximum two hours to get the parts/components onsite to the vessels. The challenge for these respondents is to get the correct connection between Enterprise expert engineer for its end-customer. They are Users of the Enterprise internal general computer services platform. Calls are almost always initiated by the Enterprise end-customers.
- 2 Enterprise top managers (Headquarters). Top management of the Enterprise consist of a team of individuals in leadership position, although all technological decisions have to be

approved by the Enterprise Chief Technology Officer (CTO) regardless of Enterprise division. The respondents here are important as decision makers and in their capacity to steer technology strategies for the Enterprise for long-term sustainable business. Their main interest is to develop a future internal standardised Enterprise interface that connects the departments of various divisions (even beyond the marine sector). This internal interface should have a corresponding standardized external interface for when end-customers login to their accounts. As such, some aspects of remote services need to have a consistent Enterprise branding and theme that is both employee and end-customer user friendly. A consistent brand and theme for digital interfaces for the Enterprise has been challenging to achieve due to the relative autonomity of LBU workings, and different workgroups within the Enterprise.

Enterprise respondents categorised for ScenB Users include:

- 6 Enterprise Product Life Cycle Managers (LCM). These individuals often have advanced engineering degrees and some have worked as field engineers prior to becoming LCMs. Their role is different from the RSCS personnel in the sense that the LCMs take care of end-customer needs from product purchase to product life-end or recycle of components. Part of their job is to sell diagnostics and service maintenance contracts. Acting as key account managers, they are in long-term direct feedback contact with Enterprise marine sector end-customers (EeC). Their feedback for this study is assumed reliable with regards to feedback from end-customers for the following reasons – (i) LCMs are in constant contact with EeCs due to an earnest effort in improving product-services portfolio for future technology developments, and towards future Enterprise sustainable business competitiveness and (ii) LCMs are held responsible by EeCs for product-service downtime, a poor follow-up of which might affect future service contract agreements.

Respondents were targeted specifically for product-service expert knowledge and because of experience and exposure to both intra- and extra- systems use that was to be improved upon by the Enterprise. All respondents are involved in providing advanced services in the Enterprise in different capacities and areas of expertise, the number of years spent with the Enterprise ranged from 2 to 24 years. Those who have spent 2 years at the Enterprise had mostly joined as master thesis students, who had then gained expert knowledge in a specific field or product before being fully employed by the equal opportunities employer Enterprise.

4. Method: SFL Pronoun System Combined With Constructs of UTAUT

The pronoun system of language in use, which is *I (You), We (They), It* and *Its,* serves as a reference point system to indicate point of view in answer to the questions of *who* is acting / saying and *what* about, under *which* circumstance. It is a language referencing perspectivity system that can be mapped in a four quadrant model (Figure 2). All pronoun perspectives in their various forms, singular/plural, subjective/intersubjective, objective/interobjective, can encompass an "inner" and "outer" experience and can be expressed as such. This model, reflecting of the evolving nature of language, is inherently adaptive and relative in perspective, depending upon researcher definition of the unit of analysis of what is singular/plural, subjective/objective etc. Consistent inquiries from each perspective will render a specific type of knowledge that can be classified under the eight primordial methodological perspectives such as phenomenology (singular subjective), hermeneutics, ethnomethodology, autopoiesis, empiricism and systems theory (plural interobjective). As such, the four quadrants can be said to reflect a type of knowledge related to that perspective of inquiry, reflecting both inner and outer worldviews.

Figure 2 maps the different perspectives. Within the SFL system, the semiosis and expression of inner and outer human experiences are reflected through transitivity processes in the metafunction of language, amongst them are acts of doing (material processes), acts of saying (verbal processes), or thinking (mental processes), indicated through use of verbs. Each transitivity processes encompasses agents who act within an activity context and circumstance.

In the Upper Left (UL) quadrant is the singular subjective perspective of 'I'. When applying UTAUT to technology, the assumed User would have this Agency / Actorship of 'I' reflected in the UL quadrant, such as "I use this technology for *x* purposes" or "Using this technology helps me accomplish <task *x*>". Collected knowledge of UTAUT constructs related to User's expectancy of use can be reflected in the UL quadrant. But the perspective is relative, depending on who is the defined 'I'. In ScenA, the 'I' User would be the Enterprise engineers and remote customer services personnel. In ScenB, it would be the Enterprise's end-customer who uses the remote services produced by the Enterprise. The Lower Left (LL) quadrant reflects the plural intersubjective perspective of "We". This collective perspective tends to reflect an ideology of shared perspectives attributed to the proximity of working with others in the same environment, given allowance for slight variation of these shared experiences. Knowledge collected on UTAUT constructs pertaining to social influence (corporate culture) or voluntariness of use (intra- inter-group workings and social norms) are reflected in the LL quadrant.

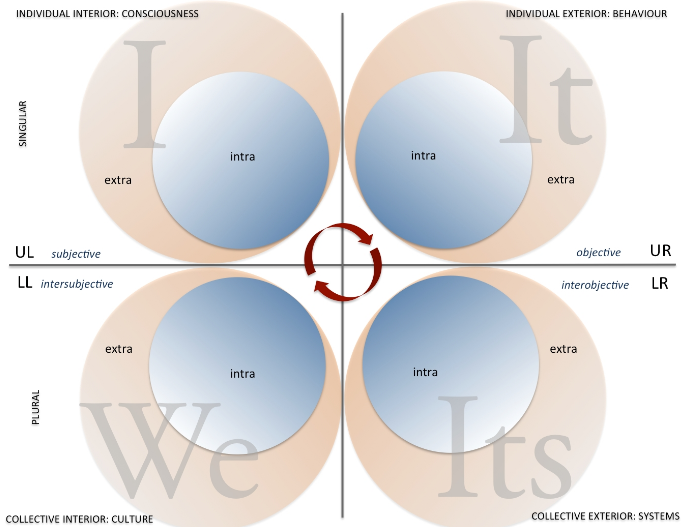

INDIVIDUAL INTERIOR: CONSCIOUSNESS
INDIVIDUAL EXTERIOR: BEHAVIOUR
I
It
SINGULAR
intra
extra
intra
extra
UL
subjective
objective
UR
LL
intersubjective
interobjective
LR
extra
intra
extra
intra
PLURAL
We
Its
COLLECTIVE INTERIOR: CULTURE
COLLECTIVE EXTERIOR: SYSTEMS

Figure 2. The pronoun system of language in use, mapped in four quadrants (adapted from Wilber [39]).

The Upper Right (UR) and Lower Right (LR) quadrants reflect singular objective and plural interobjective perspectives. Knowledge collected on UTUAT constructs pertaining to a particular technology 'It' and the technology system, 'Its', to which a particular technology belongs are reflected in these quadrants. In this study, it is noted that ScenA presents a perspective from the Enterprise engineers who design and implement the technology and technology system, to which they are Users of these systems themselves. ScenB presents a perspective of the Enterprise end-customer as User who purchases the technology and technology system from the Enterprise. Their technology acceptance experience is different, but both types of Users will need to feedback their experiences of use of technology into the Enterprise system in order for improvement in efficiency of Enterprise remote services delivery.

The perspectives, beginning in the UL quadrant working anti-clockwise presents perspectives that increase in complexity. The perspectives move from *Individual* to *Group* to *Object/s* in a system to a *Systems* view reflected in the LR quadrant. The LR quadrant that reflects a systems view of the subject of inquiry presents the broadest perspective, one that encompasses all other perspectives. All perspectives are in dialogic relation with each other, where a change in one perspective could well influence the context of the other perspectives. The broadest perspective reflected in the LR quadrant encompasses all other perspectives.

The SFL pronoun system model is useful for the simultaneous application of UTAUT in different contexts of use because the four quadrants unfold in a systemic manner, not just the Agency of a material/verbal action but it uncovers (human) Agency in relation to the object of use (UTAUT's type of technology) and circumstance of use (UTAUT's facilitating conditions).

5. Findings and Discussion

Figure 3 shows constructs of UTAUT for ScenA mapped into the SFL pronoun system quadrant, where Enterprise engineers as defined Users. Figure 4 shows constructs of UTAUT for ScenB mapped into the SFL pronoun system quadrant, where Enterprise end-customers are defined Users. The insights for ScenB Users come from Enterprise Product Life Cycle Managers who are in direct contact with the customers. All elements found in the quadrant perspectives are interrelated and in dialogic relationship to each other. The dialogic relations between the Actors are signalled by arrows drawn in circularity. All actions take place through the context of spacetime. Time is indicated as a fundamental background feature in which Enterprise activities and processes take place.

5.1. ScenA UTAUT SFL Pronoun System Findings

Due to that this study is specific in its Users for ScenA being Enterprise top managers, engineers and remote services customer service personnel, the normative definition of User often studied by application of the UTAUT model is slightly modified to reflect the different roles as Users of a single platform system. The single platform system is one that encompasses the marine sector remote diagnostics and product maintenance system both built and used by Enterprise engineers and LCMs. What is different for the different Users of this system is the Time perspective. Enterprise top managers have a longer timeline perspective, in view that one Enterprise technology strategy is to build a standard computer software platform for internal and external use. This standard platform of computer supported services in remote services necessarily interfaces with those used by the Enterprise end-customers, for them to independently login and check on product diagnostics. The Enterprise engineers have different roles, some have project timelines as short as 3 months to come up with computer support. A binding factor for ScenA Users is shared organizational values in technological innovation and service excellence towards end-customers.

ScenA Users hold various expert capacities who worked with the building of the Enterprise products and services that were to be sold to ScenB Users. As such, UTAUT moderating constructs [13] such as 'social influence', 'gender', 'age' and 'voluntariness of use' could have potentially played secondary influencing roles in this study to other factors such as 'experience' that ranked high on the Users list. Most Users for ScenA were expert knowledge workers and the 'social influence' could be redefined as the Enterprise culture of technological innovation, and pioneering remote services work not just in the marine sector but in other heavy industries in which the Enterprise has business operations. UTAUT models also predict User behaviour based on cognitive state of the user, their expectation and intention. These factors for ScenA users were subsumed under a broader corporate culture context of quality excellent in the product-services produced towards end-customers. This then needed a shift in perspective for the application of the UTAUT model towards a broader unit of analysis than the individual as User but towards the Enterprise top management defined as 'User' in this context.

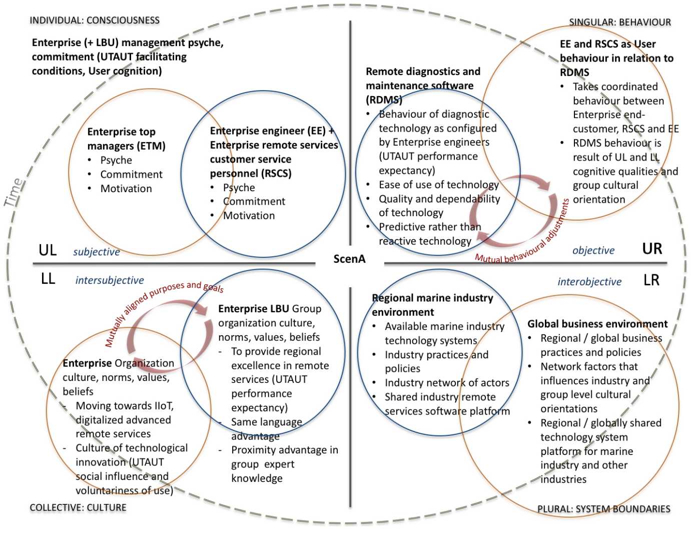

INDIVIDUAL: CONSCIOUSNESS
Enterprise (+ LBU) management psyche, commitment (UTAUT facilitating conditions, User cognition)
Enterprise top managers (ETM)
• Psyche
• Commitment
• Motivation
Enterprise engineer (EE) + Enterprise remote services customer service personnel (RSCS)
• Psyche
• Commitment
• Motivation
SINGULAR: BEHAVIOUR
Remote diagnostics and maintenance software (RDMS)
• Behaviour of diagnostic technology as configured by Enterprise engineers (UTAUT performance expectancy)
• Ease of use of technology
• Quality and dependability of technology
• Predictive rather than reactive technology
EE and RSCS as User behaviour in relation to RDMS
• Takes coordinated behaviour between Enterprise end-customer, RSCS and EE
• RDMS behaviour is result of UL and LL cognitive qualities and group cultural orientation
Mutual behavioural adjustments
Time
UL
subjective
ScenA
objective
UR
LL
intersubjective
interobjective
LR
Mutually aligned purposes and goals
Enterprise LBU Group organization culture, norms, values, beliefs
- To provide regional excellence in remote services (UTAUT performance expectancy)
- Same language advantage
- Proximity advantage in group expert knowledge
Enterprise Organization culture, norms, values, beliefs
- Moving towards IIoT, digitalized advanced remote services
- Culture of technological innovation (UTAUT social influence and voluntariness of use)
COLLECTIVE: CULTURE
Regional marine industry environment
• Available marine industry technology systems
• Industry practices and policies
• Industry network of actors
• Shared industry remote services software platform
Global business environment
• Regional / global business practices and policies
• Network factors that influences industry and group level cultural orientations
• Regional / globally shared technology system platform for marine industry and other industries
PLURAL: SYSTEM BOUNDARIES

Figure 3. SFL pronoun system perspective of UTAUT Users for ScenA: Users from the Enterprise

The Enterprise top management defined as 'User' towards new technologies comes in the form of Enterprise psyche, commitment and motivation (Figure 3, UL quadrant) of first producing the technology and then using it themselves, as platform towards an integrated advanced services portfolio in an era of digitalisation. To some extent Users of technology will need to be convinced of their intention to use that technology even prior to first testing. In the case of ScenA Users, the respondent feedback from field studies and interviews indicated that the Enterprise engineers would not (and could not) produce the remote diagnostics and maintenance platform if top management at the Enterprise were not (financially) commitment and believed that this technology would contribute to their competitiveness and to their end-customer's industry competitiveness. It is here that the SFL pronoun system is effective in helping researchers perspectivise and define 'which User' of technology, giving the research model more fluidity and adaptiveness for research purposes, so that the psyche and commitment of the top management of the Enterprise can be accounted for in understanding technology acceptance and use from the Enterprise perspective.

The Enterprise top management commitment reflected in the UL quadrant in Figure 3 not only bolsters both employee motivation and commitment at individual level but it encourages the building of a corporate culture towards constant technology innovation, reflected in the LL quadrant in Figure 3. Enterprise top management discourse would filter through the organization through management hierarchies targeted at group level meetings where Enterprise engineers and remote service customer service personnel would be motivated to provide excellent remote services for their end-customers. This organizational culture of technological innovation creates an understanding towards the building of an Enterprise internal common remote services platform in order to connect LBU employees with its global and regional centres for remote services. The main challenge experienced here, reflected in the LL quadrant in Figure 3, is the communication between the LBUs located in different European countries, and the Enterprise remote services headquarters in Switzerland. Two main challenges arose on why it was difficult building a common internal remote services platform that include (i) preference for local language and (ii) existing remote services technology platform that did not correlate with the architecture of the updated platform as suggested by headquarters.

5.2. ScenB UTAUT SFL Pronoun System Findings

Feedback from Enterprise end-customers (EeC Users, Figure 4) come from Enterprise LCMs who have the role of key account managers, who secure and follow-up on remote services life cycle management contracts with end-customers. LCM interview responses are assumed earnest due to that the proper managing of accurate end-customer feedback is crucial in building future remote services support for both Enterprise and end-customer use. The long-term Enterprise technology strategy is to have a shared platform of remote services between Enterprise and its end-customers. A challenging task much due to that currently, most end-customers are geographic proximity bound by registration of country of ownership of vessel, even if the vessels are globally seafaring. Different types of software systems are used, how much data information the systems can share between themselves are regulated by industry, national and regional policies. The challenge for seafaring vessels and remote services is that vessels lack consistent internet connections due to inconsistent satellite connections. As such, for varying reasons, Enterprise end-customers tend to be contextually (geographic proximity) bound in their immediate business networks when it comes to remote diagnostics and maintenance. Should a sea vessel come into crisis, it is referred to the nearest regional remote services centre or an LBU. Enterprise LBUs, although belonging to the Enterprise, work sometimes in competition with other to secure end-customers. LBUs have the advantage of local language, proximity to end-customers and shared remote services platform systems with their end-customers, built from more than twenty years ago. The strength of the LBU's system is also the weakness of the Enterprise's system due to that LBUs tend to have separate tools and systems to track cases that are not shared with the Enterprise global headquarters. This contributes to the challenges faced by ScenA Users whose current technology strategy is to create a common working platform for Enterprise and its EeCs. While there are Enterprise European initiatives to coordinate between departments and LBUs located in Europe to create a single platform for services, at the time of study, this has yet to be achieved. As such, Enterprise end-customer feedback is crucial and timely information is needed by the LCM in order to work towards a common shared platform for remote services.

The LCM/EeC as User perspective is reflected in Figure 4. The interdependent relations between Enterprise and EeCs are reflected in the UL and LL quadrants in Figure 4, where both Users of remote services need to share a sense of commitment to using the system bolstered by a general movement towards digitalisation and the Industry Internet of Things (IIoT). EeC corporate culture will also need to be one that is progressive and welcoming to technology innovation and change. Some motivating indicators as to why EeC would want to use Enterprise provided remote diagnostic and maintenance service and why they would want a signed contract would be SFL UTAUT constructs pertaining to the UR quadrant of singular objective 'It'. Reflected also in the summary findings in Table 1, what EeC Users find critically important is that the remote services provided is quick (time critical), reliable, user friendly systems that require little effort in learning from the EeC User. This is not to say that EeCs do not wish to put in learning time for acquired Enterprise systems. Rather, that the learning time should be short. EeCs also have as corporate goal, to be independent users of acquired systems, so that they can fix problems themselves without needing to go back to the Enterprise LCMs or call Enterprise crisis lines.

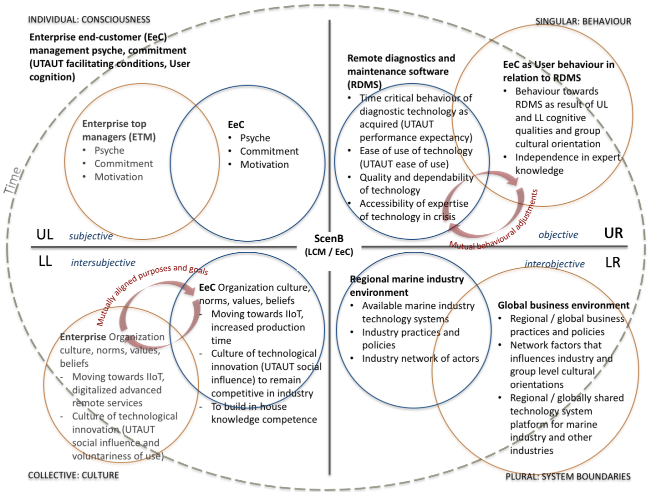

INDIVIDUAL: CONSCIOUSNESS
Enterprise end-customer (EeC) management psyche, commitment (UTAUT facilitating conditions, User cognition)
Enterprise top managers (ETM)
• Psyche
• Commitment
• Motivation
EeC
• Psyche
• Commitment
• Motivation
Time
UL
subjective
LL
intersubjective
Mutually aligned purposes and goals
Enterprise Organization culture, norms, values, beliefs
- Moving towards IIoT, digitalized advanced remote services
- Culture of technological innovation (UTAUT social influence and voluntariness of use)
EeC Organization culture, norms, values, beliefs
- Moving towards IIoT, increased production time
- Culture of technological innovation (UTAUT social influence) to remain competitive in industry
- To build in-house knowledge competence
COLLECTIVE: CULTURE
ScenB (LCM / EeC)
SINGULAR: BEHAVIOUR
Remote diagnostics and maintenance software (RDMS)
• Time critical behaviour of diagnostic technology as acquired (UTAUT performance expectancy)
• Ease of use of technology (UTAUT ease of use)
• Quality and dependability of technology
• Accessibility of expertise of technology in crisis
EeC as User behaviour in relation to RDMS
• Behaviour towards RDMS as result of UL and LL cognitive qualities and group cultural orientation
• Independence in expert knowledge
Mutual behavioural adjustments
objective
UR
interobjective
LR
Regional marine industry environment
• Available marine industry technology systems
• Industry practices and policies
• Industry network of actors
Global business environment
• Regional / global business practices and policies
• Network factors that influences industry and group level cultural orientations
• Regional / globally shared technology system platform for marine industry and other industries
PLURAL: SYSTEM BOUNDARIES

Figure 4. SFL pronoun system perspective of UTAUT Users for ScenB: Enterprise end-customer

The desire to be independent is due to EeC corporate sensitive and private data/information where it is not appreciated if Enterprise engineers can login remotely on unsecured internet connections to access EeC corporate systems data. EeCs are not welcoming of what they consider to be an intrusion of privacy. How can EeC trust that it is an Enterprise engineer and not a competitor who has logged in was one such question raised to an Enterprise LCM, even if confidentiality agreements have been signed. EeCs also tend to want to keep operating costs down, so that having a remote diagnostics and maintenance contract is deemed useful when the system is predictive rather than reactive in feedback. This allows for action to be taken prior to failure. A movement towards reuse and recycling of parts also lowers costs for EeC. But even as EeCs might be aware of what more advanced remote services might offer, the incompatibility with older, existing systems takes time to disinvest or incorporate with newer systems and services.

Table 1. SFL UTAUT constructs for technology acceptance for combined Users: Enterprise, Enterprise engineers and Enterprise end-customers

SFL defined Users ->	**ETM (ScenA)**	**EEHQ (ScenA)**	**EELBU (ScenA)**	**RSCS (ScenA)**	**LCM/EeC (ScenB)**
SFL UTAUT constructs					
Performance expectancy (singular subjective, ‘I’)	-	c√	c√	√	-
Performance expectancy (plural intersubjective, ‘We’)	√	√	√	√	-
Performance expectancy (singular objective, ‘It’)	√	c√	c√	c√	c√
Performance expectancy (plural interobjective, ‘Its’)	f√	f√	-	-	f√
Effort expectancy (singular subjective, ‘I’)	-	√	√	√	√
Effort expectancy (plural intersubjective, ‘We’)	√	√	√	√	√
Effort expectancy (singular objective, ‘It’)	c√	c√	c√	c√	c√
Effort expectancy (plural interobjective, ‘Its’)	f√	f√	-	-	f√
Social influence (singular subjective, ‘I’)	-	-	-	-	-
Social influence (plural intersubjective, ‘We’)	√	√	√	√	√
Social influence (singular objective, ‘It’)	c√	c√	√	√	c√
Social influence (plural interobjective, ‘Its’)	f√	f√	√	√	f√
Behavioural intention (singular subjective, ‘I’)	√	√	√	√	√
Behavioural intention (plural intersubjective, ‘We’)	c√	c√	√	√	c√
Behavioural intention (singular objective, ‘It’)	c√	c√	c√	c√	c√
Behavioural intention (plural interobjective, ‘Its’)	f√	f√	√	√	f√
Use behaviour (singular subjective, ‘I’)	-	-	-	-	-
Use behaviour (plural intersubjective, ‘We’)	-	-	-	-	-
Use behaviour (singular objective, ‘It’)	c√	c√	c√	c√	c√
Use behaviour (plural interobjective, ‘Its’)	f√	f√	f√	f√	f√
Facilitating conditions (singular subjective, ‘I’)	c√	c√	c√	c√	c√
Facilitating conditions (plural intersubjective, ‘We’)	c√	c√	c√	c√	c√
Facilitating conditions (singular objective, ‘It’)	c√	c√	c√	c√	c√
Facilitating conditions (plural interobjective, ‘Its’)	f√	f√	f√	f√	f√

Users Key
ETM = Enterprise top management, EEHQ = Enterprise engineers, Headquarters
EELBU = Enterprise engineers, Local Business Unit,
RSCS = Remote services customer service personnel
LCM/EeC = Product Life Cycle Managers (reflecting Enterprise end-customers)

5.3. Combined SFL UTAUT Users Findings Summary

Table 1 summarises the findings of the SFL pronoun system defined Users, in relation to the six broad UTAUT category constructs for technology acceptance when placed into the SFL pronoun system quadrants.

Factors such as ‘gender’, ‘age’, ‘experience’ and ‘voluntariness of use’ are incorporated into the more specific pronoun system quadrants. The factors of ‘gender’, ‘age’ and ‘experience’ were accounted for in respondent profile through the primary requirement that all respondents were knowledge experts in their field regardless of gender or age. The Enterprise emphasises it is an equal opportunity employer. The combined Users include perspectives from Enterprise top managers, Enterprise engineers, remote services customer service personnel and Enterprise end-customers who feedback directly to Enterprise Product LCMs. Table 1 indicates which technology acceptance constructs are relevant (by ‘√’), which are critically relevant (by ‘c√’) and those that will be relevant in future (by ‘f√’). UTAUT constructs with an indeterminate answer or that are non-relevant are indicated by ‘-‘.

In overview, Table 1 indicates that what was defined as ‘relevant’, ‘critically relevant’ and ‘future relevant’ as UTAUT construct, depended upon the User perspective, as helped defined by the SFL pronoun system. In keeping with organization area of expertise, that the Enterprise defines itself as world leader in innovative technology, what was deemed as critically relevant for all Users was SFL UTAUT construct Effort expectancy (singular objective, ‘It’), which meant that the remote services technology was expected to work for both the Enterprise and its end-customer. The technology was produced to be end-customer user friendly and accessible, so that learning input from the end-customer, whilst necessary, would be kept at minimal level. Critically relevant was UTAUT facilitating conditions for all SFL pronoun system perspectives as Users. This could be due to the fact that without adequate motivation from Individual level to conducive environment at Systems level, the technology would not be able to survive or evolve in the human technological ecosystem. Facilitating conditions also had to be right if a future (thus future relevant, ‘f√’) remote services standard platform for regional and global use is to be built / implemented. Its development will depend upon Enterprise commitment as well as customer demand and supporting industry policies and practices.

6. Study Limitations and Future Research

The strengths of this study are also its limits. This includes (i) using the Enterprise as case example unit of analysis and Enterprise only respondents, and (ii) using the SFL pronoun system in offering of a plurality of perspectives to UTAUT User constructs. Pertaining to strength and limitation (i), Enterprise as case example provides this study with a cohesive organization environment in which research method theory and analysis can be done. What has been assumed are shared corporate cultures and shared corporate strategies. The single organization as case example might in that sense crave comparative studies to be made, perhaps by definition of industry sector, even across industry sectors, comparing remote services technologies. Still, while single organization shared values has shown to be mostly true, it was not true that the Enterprise and its LBUs were homogeneous in their workflows. This might have proved to be a challenge with regards to placing all Enterprise respondents into the majority ScenA Users. This is because it reflects that ScenA Users share the same User experience, when in effect, there are variations of context of use for ScenA system Users, which are not fully accounted for in this study due to that the focus of this study was to illustrate how SFL could be used to unfold different versions of Users usually statically defined when applying UTUAT models. The capturing of User nuances in remote services could in this case, be a subject area for future research, in view that remote services would be a field that grows in application with increased digitalisation of industry processes, beyond the marine sector into robotics and process automation. Since language is still its own metalanguage, SFL could be a means of a cohesive theory and framework towards a general systems theory perspective of studies on technology innovations and combined product-services. There were practical challenges to obtaining data for this study that included time coordination for respondents to be present during the field study and for shadowing purposes.

Pertaining to strength and limitation (ii), the use and application of the SFL pronoun system, while offering an unfolding of perspectives on any subject of study, is also reliant on clear definition of research purpose and design on the part of the researcher. Relativity of perspective *x,* will need to be defined, in relation to point of interest *y,* in context *[n+...]* in order to be effective in use. While the SFL pronoun system framework opened up various User definitions for UTAUT and made troubleshooting easier, in the sense that the SFL UTAUT defined User could be exactly pinpointed for unease of use of technology for example, what still remained were the working processes of communication between Enterprise business units and their end-customer on how to improve remote services efficiency. It could also be said that the SFL pronoun quadrant approach might not be too appreciated with a management or practitioner audience because the fluidity of perspectives as exacting as it might be, is also confusing, depending on audience. The remote services customer service personnel for example, who were the individuals meeting crisis calls from end-customers had little concern for how much commitment top management had towards building a standardised internal remote services platform for the future. They were rather more concerned that they could pass on the crisis call to an immediate available remote services engineer.

7. Conclusion

The purpose of this study is to bridge research methodology across different disciplines. In particular, it is to apply a language science model of analysis the SFL, unto a model of technology use and acceptance, UTAUT. This was done because a practical need arose to address multiple users of technologies in multiple work contexts in a single framework of analysis. The need for a plurality of perspectives of defined User (i.e. *which* user of technology?) arose in a case example of an Enterprise needing to develop a shared remote services platform for employees (mostly engineers) as well as end-customers. In such a context, both Enterprise engineers and Enterprise end-customer were Users of this technology system. UTAUT in its original application, is a model of technology use and acceptance that renders a linear statically defined User to a single context of use. Because language is inherently human and we use language to express and encode many internal and external experiences and world views, this study turned to SFL as functional language theory and framework of data analysis with the questions of (RQ1) *how* SFL can be applied to the UTAUT model so that the UTAUT model can be simultaneously applied for differently defined Users in different contexts and (RQ2) *what* contributing factors can be established by use of SFL as language theory and framework that broadens the applicability of UTAUT into various contexts? To that end, in answer to RQ1, this study has illustrated how the architecture of language contains in itself a myriad of language systems, of which the pronoun system is one. The SFL pronoun system unfolds primordial perspectives of Agency and Actorship set in both group and environmental context. In answer to RQ2, this unfolding of perspectives by applying the SFL pronoun system is what allows for the plurality of views of the defined Users of UTAUT constructs. The plurality of User views in different contexts when combined with UTAUT constructs allows for the immediate identification of disjunctive views occurring in the business workflows, when an Enterprise engineer believes the system to be user friendly but when the Enterprise end-customer does not. Or when two Enterprise engineers have different User experiences of the same system. This SFL UTAUT model allows for such gap in knowledge identification, and it gives those who work in the context a chance to reconcile these differences in opinion and work towards closing the knowledge gaps. The fluidity of the SFL pronoun system that also maps different types of knowledge zones also means that researchers can use the pronoun quadrant model to visualise research design perspective and address gaps in knowledge.

The application of the SFL pronoun system in complement to the UTAUT model is novel in research methodology. The resulting combined SFL UTAUT theoretical construct and the subject of study of remote services could be better developed by similar type multi-enterprise studies or comparative multi-enterprise type studies. And perhaps what the SFL UTAUT construct does not and cannot do, is to address the actual communication patterns between Actors and their surrounding context. Miscommunication is identifiable by the SFL UTAUT construct, but the act of improving on communication across different Enterprise business units and between their end-customers remains very much a human cognitive process.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

This study is funded by Stiftelsen Riksbankens Jubileumsfond of Sweden, as part of its Flexit (2015-18) Program, RMP15-0954:2. The purpose of the main project is to investigate linguistic methodological applications of analysis to user acceptance and use of technology.

References

[1] Fishbein, M., Ajzen, I., Belief, attitude, intention and behavior: An introduction to theory and research, Reading, MA: Addison-Wesley Publishing Company, 1975

[2] Ajzen, I., “The theory of planned behaviour”, Organ Behav Hum Dec, 50(2), 179- 211, 1991. http://dx.doi.org/10.1016/0749-5978(91)90020-T

[3] Davis, F. D., Bagozzi, R. P., Warshaw, P. R. "User acceptance of computer technology: A comparison of two theoretical models", Manage Sci, 35, 982–1003, 1989. http://www.jstor.org.ezproxy.ub.gu.se/stable/2632151

[4] Venkatesh, V., Bala, H., "Technology acceptance model 3 and a research agenda on interventions", Decision Sci 39(2), 425-478, 2008. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[5] Taylor, S., Todd, P.A. "Assessing IT usage: The role of prior experience", MIS Quart 19(2), 561-570, 1995. http://www.jstor.org/stable/249633

[6] Blackwell, C., Lauricella, A., Wartella, E., Robb, M., Schomburg, R., "Adoption and use of technology in early education: The interplay of extrinsic barriers and teacher attitudes: The interplay of extrinsic barriers and teacher attitudes", *Comput Educ,* 69, 310-319, 2013. https://doi.org/10.1016/j.compedu.2013.07.024

[7] Haas, S., Senjo, S., "Perceptions of effectiveness and the actual use of technology-based methods of instruction: A study of California criminal justice and crime-related faculty", J Crim Just Educ, 15(2), 263-285, 2004. https://doi-org.ezproxy.ub.gu.se/10.1080/10511250400085981

[8] Thakur, R., "Customer adoption of mobile payment services by professionals across two cities in India: An empirical study using modified technology acceptance model", Bus Perspect Res 1(2), 17-30, 2013. https://doi-org.ezproxy.ub.gu.se/10.1177/2278533720130203

[9] Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., "User acceptance of information technology: Toward a unified view", MIS Quart 27(3), 425-478, 2003. http://www.jstor.org.ezproxy.ub.gu.se/stable/30036540

[10] Barelka, A., Jeyaraj, A., Walinski, R., "Content acceptance model and new media technologies", J Comput Inform Syst 53(3), 56-64, 2013. https://doi-org.ezproxy.ub.gu.se/10.1080/08874417.2013.11645632

[11] Sykes, T.A., Venkatesh, V., Gosain, S., "Model of acceptance with peer support: A social network perspective to understand employees' system use", MIS Quart 33(2), 371-393, 2009. http://www.jstor.org.ezproxy.ub.gu.se/stable/20650296

[12] Maruping, L., Bala, H., Venkatesh, V., Brown, S., "Going beyond intention: Integrating behavioral expectation into the unified theory of acceptance and use of technology", J Assoc Inf Sci Technol 68(3), 623-637, 2017. http://dx.doi.org/10.1002/asi.23699

[13] Halliday, M.A.K., Matthiessen, C., Halliday's Introduction to functional grammar, 4th ed. Abingdon: Routledge, 2014

[14] Halliday, M.A.K., Meaning as choice. In, L. Fontaine, T. Bartlett and G. O'Grady (eds.), Systemic functional linguistics: Exploring choice, Cambridge University Press, p. 15-26, 2013.

[15] Capra, F., & Luisi, P. (2014). *The Systems View of Life: A Unifying Vision.*

[16] Capra, F. (2005). Complexity and Life. *Theory, Culture & Society, 22*(5), 33-44

[17] Martins, C., Oliveira, T., Popovič, A., "Understanding the Internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application", Int J Inform Manage, 34(1), 1-13, 2014. https://doi.org/10.1016/j.ijinfomgt.2013.06.002

[18] Saadé, R. G., Kira, G., "Mediating the impact of technology usage on perceived ease of use by anxiety", Comput Educ, 49(4), 1189-1204, 2007. https://doi.org/10.1016/j.compedu.2006.01.009

[19] Shin, D., "Determinants of customer acceptance of multi-service network: An implication for IP-based technologies", Inform Manage 46(1), 16-22, 2009. https://doi.org/10.1016/j.im.2008.05.004

[20] Dalcher, I., Shine, J., "Extending the new Technology Acceptance Model to measure the end user information systems satisfaction in a mandatory environment: A bank's treasury", Technol Anal Strateg, 15(4), 441-455, 2003. https://doi-org.ezproxy.ub.gu.se/10.1080/095373203000136033

[21] Krishnaraju, V., Mathew, S., Sugumaran, K., "Web personalization for user acceptance of technology: An empirical investigation of E-government services", Inform Syst Front, 18(3), 579-595, 2016. https://doi-org.ezproxy.ub.gu.se/10.1007/s10796-015-9550-9

[22] Lee, J-H., Song, C-H., "Effects of trust and perceived risk on user acceptance of a new technology service", Soc Behav Personal, 41(4), 587-598, 2013. http://dx.doi.org/10.2224/sbp.2013.41.4.587

[23] Teo, T., "Modelling technology acceptance in education: A study of pre-service teachers", Comput Educ, 52(2), 302-312, 2009. https://doi.org/10.1016/j.compedu.2008.08.006

[24] Šumak, B., Heričko, M., Pušnik, M., "A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types", Comput Hum Behav, 27(6), 2067-2077, 2011. https://doi.org/10.1016/j.chb.2011.08.005

[25] Strudwick, G., McGillis Hall, L., "Nurse acceptance of electronic health record technology: A literature review", J Res Nurs, 20(7), 596-607, 2015. https://doi-org.ezproxy.ub.gu.se/10.1177/1744987115615658

[26] Pai, F-Y., Huang, K-I., "Applying the Technology Acceptance Model to the introduction of healthcare information systems", Technol Forecast Soc, 78(4), 650-660, 2011. https://doi.org/10.1016/j.techfore.2010.11.007

[27] Aggelidis, V. P., Chatzoglou, P. D.,"Using a modified technology acceptance model in hospitals", Int J Med Inform, 78(2), 115-126, 2009. https://doi.org/10.1016/j.ijmedinf.2008.06.006

[28] De Graaf, M., Ben Allouch, S., Van Dijk, J., "A phased framework for long-term user acceptance of interactive technology in domestic environments", New Media Soc, 1-22, 2017. https://doi-org.ezproxy.ub.gu.se/10.1177/1461444817727264

[29] Wagner, G. D., Flannery, D. D., "A quantitative study of factors affecting learner acceptance of a computer-based training support tool", J Eur Ind Training, 28(5), 383-399, 2004. https://doi-org.ezproxy.ub.gu.se/10.1108/03090590410533071

[30] Hasan, R. (2012). *Systemic Functional Linguistics: Exploring Choice* (Vol. 9781107036963). Cambridge University Press.

[31] Popova, O., Popov, B., Karandey, V. & Evseeva, M. (2015). Intelligence amplification via language of choice description as a mathematical object (Binary Tree of Question-answer System). *Procedia - Social and Behavioral Sciences, 214*, 897-905.

[32] Hasan, R., & Martin, J. (1989). *Language development: Learning language, learning culture* (Advances in discourse processes, 27). Norwood, N.J.: Ablex.

[33] Martin, J. R., "Meaning matters: A short history of systemic functional linguistics", WORD, 62:1, 35-58, 2016. https://doi-org.ezproxy.ub.gu.se/10.1080/00437956.2016.1141939

[34] Halliday, M. A. K. Language as social semiotic: The social interpretation of language and meaning, London: Edward Arnold, 1978.

[35] Whorf, B. L. Language, thought and reality: Selected writings, Cambridge, MA: MIT Press, 1956.

[36] Butler, C., "An ontological approach to the representational lexicon in Functional Discourse Grammar", Lang Sci, *34*(5), 619-634, 2012. https://doi.org/10.1016/j.langsci.2012.02.004

[37] Bartlett, T. "I'll manage the context": context, environment and the potential for institutional change. In, L. Fontaine, T. Bartlett and G. O'Grady (eds.), Systemic functional linguistics: Exploring choice, Cambridge University Press, p. 342-364, 2013.

[38] Bourdieu, P., Language and symbolic power, Cambridge: Polity Press, 1991.

[39] Wilber, K., A theory of everything, Boston: Shambhala, 2000.

[40] Butler, C., "Criteria of adequacy in functional linguistics", Folia Linguist, 43(1), 1-66, 2009. https://doi-org.ezproxy.ub.gu.se/10.1515/FLIN.2009.001

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 319-328(2018)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Ontology Modeling of Social Roles of Users in Mobile Computing Environments

Daniel Ekpenyong Asuquo, Patience Usoro Usip*

University of Uyo, Computer Science Department, Faculty of Science, 520003, Uyo, Nigeria

ARTICLE INFO

Article history:
Received: 30 January, 2018
Accepted: 14 March, 2018
Online: 25 April, 2018

Keywords:
Context Representation
Quality of Service
Semantic Approach
Mobile Communication Network

ABSTRACT

Today, computing devices of various types with wireless interconnections are used for diverse tasks and increasingly in ad hoc manners. It is not always obvious which devices are present, reachable, and connected when users and their devices are mobile. In such mobile computing environment, the number of registered lines on the network via network operators cannot qualify a user to carry out any service due to the unpredictable service quality (SQ), dynamic user context and the device in use. To properly manage the SQ, there is need to specify the roles applicable to mobile devices to effectively utilize the constrained and shared resources for the feature-rich applications. The user's context may change and adaptation to changing behavior, resource usage, and security settings also pose problems. This paper presents the use of semantic web approach in modeling ontology for a richer knowledge representation of users' activities and social roles on mobile devices. The developed ontology for users' social roles was implemented in Protégé to determine whether a mobile user with its interaction medium has a functional capability for specific social role or not. The importance on the use of context in interactive applications is shown and a proposed framework for development of context-aware applications is developed. Results revealed that the approach can effectively enhance partnership between mobile operators and content providers of next generation wireless networks for the provision of value-added mobile web services.

1. Introduction

Increased advances in mobile communications networks and its integration with web services for different application domains has resulted in proliferation of mobile web services. Mobile web services are defined as web services that are deployed on mobile devices and are published over the Internet, wireless network or within the operators' network [1]. Their goal is to offer new personalized services to users on their mobile devices such as mobile phones, personal digital assistants (PDAs) as well as laptop computers. In such environment, the context of a user (e.g. location, time, system resources, network state, user's activity, battery power level, etc.) is highly dynamic, and granting a user access without considering his current context can result in challenging situations. This is because individuals change in behaviour in different settings. However, there is relatively less developed research on factors affecting device usage based on the user's context and individual characteristics that arise from social environment [2, 3, 4]. According to [5], p*ersona* oriented research is a human computer interaction technique which offers a research paradigm that provides a ground on modeling individuals' behaviour and usage of the mobile phone. Since mobile phones are social devices, context specific research has the potential of significantly providing a more rational understanding of the factors influencing device usage and adoption of associated services at home, school, market, worship or work place, etc. Furthermore, the ability of mobile devices to handle multiple tasks and media types can facilitate mobile learning for information seeking subscribers and content delivery by content providers and mobile operators. Therefore, the developed content-based services will depend on the interaction capabilities of the device, the usage habits of the subscriber and the considered contextual information.

In [6], the authors classified personal digital assistants (PDAs), mobile phones, and Personal Media Players (PMPs) as mobile devices excluding tablets and laptop computers because tablets and laptops give relatively the same functionality as

*Patience Usoro Usip, Computer Science Department, University of Uyo, Uyo, +2348060546140, patiencebassey@uniuyo.edu.ng, patceeng@gmail.com

https://dx.doi.org/10.25046/aj030235

desktop computers and are more portable than mobile. Another reason is that they can easily be brought anywhere but not accessed anytime because of their bulky nature and relative start up time. In a typical mobile computing model, computation does not occur at a single location in a single context, as in desktop computing, but rather spans a multitude of situations and locations covering the office, meeting room, home, airport, hotel, classroom, market, bus, etc. Users might access their computing resources from wireless portable machines and also through stationary devices and computers connected to local area networks. This collection of mobile and stationary computing devices that are communicating and cooperating on the user's behalf is called a mobile distributed computing system. This form of computing is broader than mobile computing because it concerns mobile people not just mobile computers. These systems aim to provide ubiquitous access to information, communication, and computation. One significant aspect of this emerging mode of computing is the constantly changing execution environment. The processors available for a task, the devices accessible for user input and display, the network capacity, connectivity, and costs may all change over time and place. In short, the hardware configuration is continually changing. Similarly, the computer user may move from one location to another, joining and leaving groups of people, and frequently interacting with computers while in changing social situations.

Existing research involving adaptive user modeling for personalized services in the emerging areas of mobile and ubiquitous computing is limited. In [7], the author considered using social paradigms in smart cities for mobile context-aware computing but his approach was not ontology based. However, in [4], the authors used ontology to model user mobility and not social roles. This paper presents an ontology-based context-awareness model for context representation of mobile service contents that supports ubiquitous learning. This is expected to strengthen partnership between mobile operators and content providers.

The rest of the paper is structured as follows. Section 2 reviews literature on context and context-awareness models with emphasis on Ontology models. Section 3 presents the ontology modeling approach for mobile users' social role as well as a framework for provisioning of context-aware services to mobile users. In section 4, the developed ontology model is implemented in protégé-OWL and results discussed while section 5 summarizes and concludes the work in this paper.

2. Context and context-awareness models

In [8], the authors proposed the concept of context-aware ubiquitous learning to emphasize the characteristics of learning the 'right content' at the 'right time' and 'right place', and also to facilitate a seamless ubiquitous learning environment that supports learning without constraints of time or place. This concept requires the detection of learner changing context information to provide different learning content via mobile devices [9]. In [10], the authors showed that mobile context-aware computing is an essential component of the 'smart cities' infrastructure, where a collection of smart computing infrastructure including the new generation of integrated hardware, software and network technologies that provide a real-time awareness of the real-world, and actions that optimize business processes are required [11]. Such smart cities rely on the vision of ubiquitous computing whereby devices can communicate with each other to provide services and information to end users [12]. Therefore, the interaction capability that the current encoding demands on mobile platforms when navigating and getting information is a challenge that must be addressed. Improved representation of content for mobile services is required with new service provision models for more personalized and value added services. It is observed that semantic web technology can handle this challenge by creating a machine process-able web capable of providing a unique value proposition for the advancement of mobile services provision models [13, 14].

Context has been described differently by different authors [15,16,17] as no single definition of context exists. Nevertheless, In [18], the author defined context as any information that can be used to characterize the situation of an entity that is considered relevant to the interaction between a user and an application, including the user and the application themselves. An entity could be a person, a place or an object while information refers to any particular element or detailed piece of data that allows for the description of any condition or state of the participating entities. Context-awareness is a mobile computing paradigm in which applications can discover and take advantage of contextual information such as user location, time of day, neighbouring users and devices, and user activity [19]. In [20], the authors identified location, identity, time, and activity as primary context types for characterizing the situation of a particular entity. They showed that the primary pieces of context for one entity can be used as indices to find secondary context (e.g. phone number, email addresses, birthday, list of friends, relationship to other people, etc.) for that same entity as well as primary context for other related entities in the same location. With these conceptual definitions, an application developer can enumerate the context for a given application.

A system that can extract, interpret, use context information and adapt its functionalities to the current context is said to be context-aware. The challenge is to create a system that will adapt to the set of constraints imposed by the corresponding context of use. These constraints are set by various internal and external factors or dimensions of context [21] as shown on Table 1. Table 1 gives four primary context types - identity, location, time, and activity, for characterizing the situation of a particular entity. It includes specific context information of a user such as in a social environment, where a user may be in a public or private place or in a group setting. The idea is that people's actions are often predicted by their situation and contextual information and commands aim to exploit this fact. Different results from queries on contextual information can be produced based on the context they are issued. Also, whatever activity is performed will depend on its goals and tasks as well as the tools and resources needed to perform them. The infrastructure deployed will be affected by available communication bandwidth, cost and network connectivity while the product design will greatly be influenced by the physical features of the device and the digital capabilities embedded for its functionality. The location of the user specifically include temporal (date, time, etc) and spatial (position, orientation, and movement) context. Context-aware computing has been applied to different application domains and scenarios. For example, in [22], two applications, AwarePhone and AwareMedia were developed from AWARE architecture to support context sensing and management regarding the working context of users. AwarePhone provides context information about colleagues and workplaces while the AwareMedia acts as a whiteboard system for cooperation in workplaces. Similarly, in [23], the authors described a prototyped context-aware messenger called ConaMSN, which uses context information (emotion, stress, activity) obtained from wearable sensors (using probabilistic methods) and shares it among the application users.

Table 1 Dimensions of context

Dimension	Description
Internal	**Users profile:** beliefs, previous experience, physical and emotional state **Social environment:** work context, business processes (i.e. private, public, group) **Activity:** goals, tasks, tools, resources
External	***Physical environment:*** noise and illumination level, weather conditions, proximity to other objects ***Infrastructure**:* bandwidth, network connectivity, cost **Location:** spatial (orientation, position, velocity) and temporal (time, date, season) **Device technology:** physical and digital properties

The lack of conceptual models and tools to support the rapid development of rich context-aware applications for empirical investigation of interaction design and the social implications of context-aware computing is a challenge. Essentially, a context-awareness model is required to define and store context information in a machine-readable form. In [24], the authors summarized the most relevant context-modeling approaches based on data structures used for representing and exchanging contextual information in their respective systems. These include key-value models, mark-up models, graphical models, object-oriented models, logic-based models, and ontology models, etc. The key-value models are the simplest data structure for context modeling and are frequently used in various service frameworks where key-value pairs are used to describe the capabilities of a service. Service discovery is then applied with matching algorithms which use these key-value pairs. Key-value modeling does not support knowledge sharing across different systems. The mark-up models use a hierarchical data structure comprising mark-up tags, attributes and content to create profiles. Mark-up modeling approach is difficult and non-intuitive to capture complex contextual relationships and constraints. The graphical models make use of unified modeling language (UML) in modeling contextual aspects. The object-oriented models use various objects to represent different context information (e.g. location, time, temperature, etc.), and encapsulate details of context processing and representation. Modeling context using object-oriented techniques offers the full power of object orientation which includes encapsulation, reusability and inheritance while providing access to the context and context-processing logic using well-defined interfaces. In logic-based models, facts, expressions and rules are often used to define a context model with a high degree of formality. The logic-based system can then be used to manage the listed terms and allows for modification or updating of facts. The reasoning or inference process is used to derive new facts based on existing rules in the system. Contextual information is then represented in a formal way as facts. Although object-oriented modeling, graphical modeling and logic-based modeling approaches support formality and some of them capture context information, they do not address knowledge sharing and context reasoning issues. Ontology models are very promising for modeling contextual information due to their high and formal expressiveness (that can be attached to the user behaviour or the service) and possibilities for applying ontology reasoning techniques at the negotiation layer. Ontology in this work represents a description of concepts and their relationships. The main advantages for ontology modeling in context-aware computing systems are formal knowledge representation, logic reasoning, knowledge sharing and reuse. The paper therefore models ontology for knowledge representation of users' activities on mobile phones for a typical application or service.

3. Service domain ontology modeling approach

Context information needs to be expressed with richer definitions such as those deployed by ontology and rich vocabularies. The richer the definition expressed the more adaptive and value added the service provided through the interaction scenario becomes. Semantic web provides different forms of representation regarding the same information resource due to its interoperability property. It is aimed at facilitating the higher automation of service discovery, composition, invocation, and monitoring in an open environment. Ontology is therefore the semantic interoperability and knowledge sharing foundation for semantic web services matching and context reasoning. Figure 1 shows the ontology-based context-awareness model for web services data management [25], where a web application processes mobile client requests from multiple web browsers. The mobile client interacts with the server using Hypertext Transmission Protocol (HTTP) request to a Java Servlet. This servlet reads the ontology from the server using an application programming interface (API), a Java framework for building semantic web applications. The API provides a programmable environment for resource description framework (RDF), RDF-S (Schema) and the web ontology language (OWL). Protégé, an open source ontology editor and knowledge base framework can be used for developing the domain ontology while Pellet, an OWL reasoner which is a core component of ontology-based data management applications, can also be used in the architecture for extracting inferred knowledge from the ontology [26]. Adaptive web pages are returned to the client, which adapts Java Server Pages to the browser being used by the client device.

To successfully apply semantic web services into telecommunication systems, one must abstract the sharing domain concepts and reasonably organize them [27]. We adopt an efficient ontology hierarchy modeling approach and consider reusability and extensibility as two important ontology modeling factors in our layered ontology modeling method shown in figure 2. For instance, time, space, and people ontologies under common ontology, can be shared in the different domains, like Telecommunications, Medical, Transportation, Business domain or any other domain.

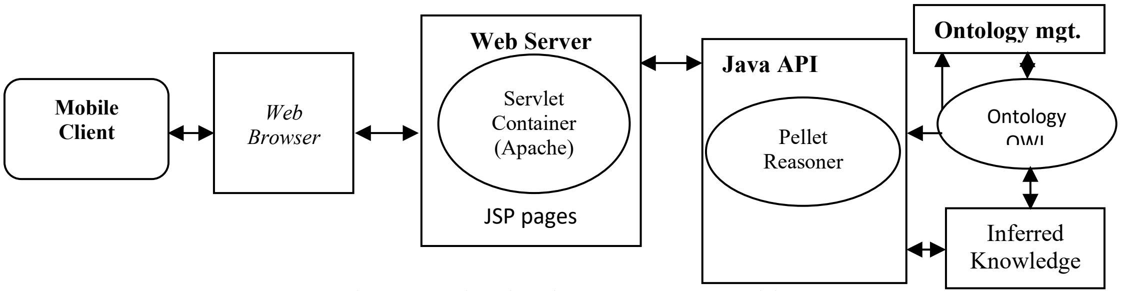


Figure 1 Ontology-based context-awareness model

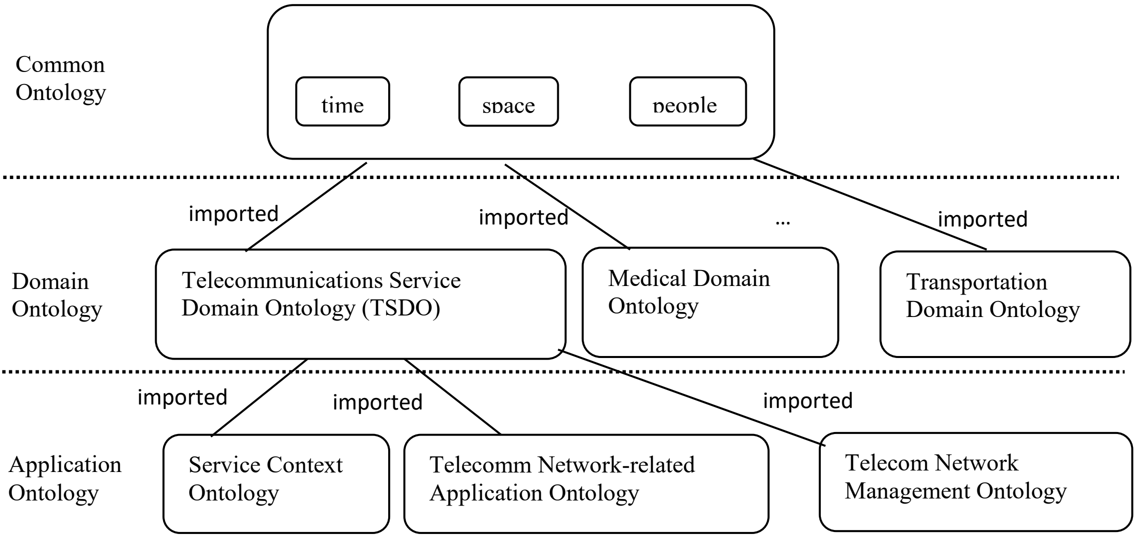


Figure 2 Layered ontology modeling approach

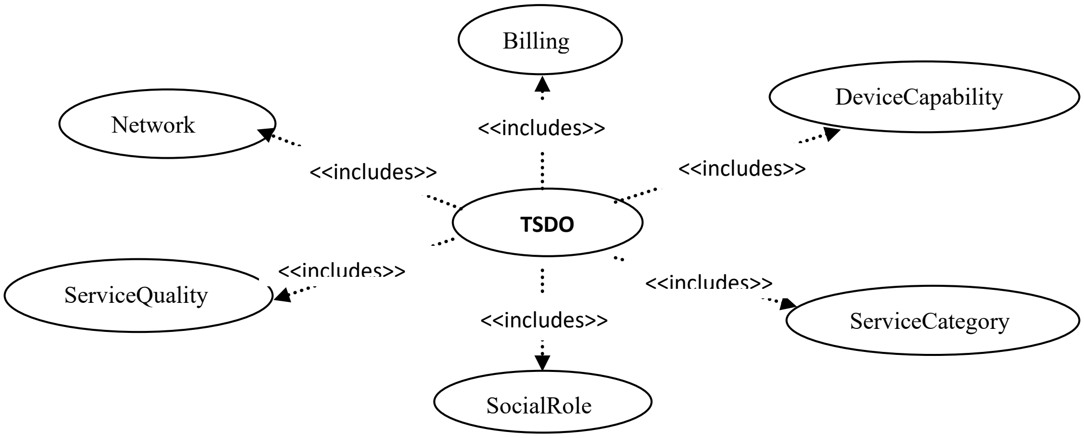


Figure 3 A framework for telecommunication service domain ontology

The concrete domain ontology can be shared by the different domain–related application ontologies. For example, telecommunication service domain ontology can be used to create the service context ontology, network management ontology, etc. Furthermore, a modular modeling approach is considered in the construction of telecommunication service domain ontology to ensure stronger cohesion and loose coupling in the correlation of concepts. The framework for the service domain ontology shown in figure 3 using UML use-case diagram comprises six sub-ontologies, including Device Capability Ontology, Network Ontology, Service Role Ontology, Billing Ontology, Service Quality Ontology, and Service Category Ontology. Features of the sub-ontologies are presented on Table 2. The various associations between the different levels of context information can be described and represented through semantic web languages and appropriate vocabularies.

Table 2 Features of the telecommunication service domain sub-ontologies

Sub-ontology	**Feature**
Network	Specifies network concepts, network categories, features and relationships between various networks e.g. mobile network, fixed network, wireless network, wired network, GSM, CDMA, WCDMA and WLAN, etc.
Device Capability	Defines concepts about device software, hardware, browser, and network characteristics supported by device.
Service Quality	Defines end-to-end QoS guarantee based on technical characteristics, e.g. access network QoS, core network QoS, and user's quality of experience such as call setup delay, throughput, etc.
Service Category	Defines relationship between various service classifications, like voice service, messaging service, data service, download service, browsing service, value-added service, etc.
Service Role	Describes the stakeholder's concept of the service supply chain, e.g. service provider, content provider, network operator, service user, etc.
Billing	Defines payment methods, service charge, and account balance concepts e.g. prepaid or postpaid payment; time-based, volume-based, event-based, or content-based service charge; bonus, etc.

3.1. System architecture for social roles ontology

Ideally, context-aware systems make use of context models expressed as ontologies to formalize and limit the notion of context. Since what is considered relevant information differs from one domain to another, effective use of this information is required. OWL was used to explicitly formalize the properties and structure of TSDO contextual information to guarantee common semantic understanding among different architectural components. OWL has well-defined syntax, formal semantics, reasoning support, and enhances information retrieval and interoperability. The developed mobile users' social roles ontology is shown in figure 4. Each layer indicates a representation of a state supporting the provision of the mobile service. The seven layers are represented by classes in the ontology developed with domain experts (except mobile device manufacturers), which include mobile operators, content providers and described using OWL. This enhances a wider level of interoperability and provides a more generic vocabulary on describing properties and classes with richer semantics than widely deployed RDF.

Table 3 Social roles and interaction characteristics

User social role	Service category	Interaction medium
Speaking	voice services only	Voice interface (mouthpiece, earpiece), wireless Bluetooth
Composing	messaging service in addition to voice	Keyboard, stylus, lightpen
Photography	snapshot service in addition to voice and messaging	Camera, colour display
Surfing	data and browsing service in addition to snapshot, messaging and voice services	GPRS, bluetooth, bandwidth, browser
Audio recording/ listening	audio services in addition to surfing	Media player, speaker, bluetooth
Video recording/watching	video services in addition to audio services	Media player, video player, memory card

This paper assumes that an individual using a mobile device is attached to a role that might be changed due to context whether social or locational. This role depends on the task at hand, the interaction capability of the device as well as the usage habits and interaction styles of the user. To resolve the critical design challenge of integrating the characteristics of a social setting with the services characteristics and the interaction styles available, we developed a model of service ontology that describes the social role with prescribed characteristics as well as specific interaction characteristics in order to address a value added mobile content provision. The model can support knowledge representation and communication interoperability with regards to different aspects of context. Third Generation Partnership Project (3GPP) categorization of the classes of service provided on a mobile communications system is considered. This includes *conversational* (voice), *interactive* (web browsing, interactive chats), *background* (SMS, FTP, Email), and *streaming* (audio/video conferencing) services. However, for clarity of ontology development, Table 3 describes the four service classes based on the proposed social roles and the related characteristics. The service category ontology defines the relationship between various telecommunications services, like voice service, messaging service, data service, download service, browsing service, video conference service, and value-added service.

Figure 4 presents the ontology of users' social roles, sub-ontology of the telecommunication service domain ontology. From the ontology representation, answers to basic competency questions are obtained. The competency questions include:

Who uses the ontology?
Who performs social role?
How is social role carried out?
What is a social role?
What determines the social role to be performed?

Deductions can reasonably be made from the following rules. The capability of the device in use defined by its interaction medium or service categories determines its social roles and its adaptation in a mobile computing environment. Thus, the capability of the device is defined by rules 1 to 6 as part of the ontology.

Rule 1: IF Mobile_device has Voice-interface THEN social-role is Speaking.
Rule 2: IF Mobile_device has Message-interface THEN social-role is Messaging.
Rule 3: IF Mobile_device has Photo-interface THEN social-role is Photography.
Rule 4: IF Mobile_device has Surfing-interface THEN social-role is Surfing.
Rule 5: IF Mobile_device has Audio-interface THEN social-role is Audio-recording-and-listening.
Rule 6: IF Mobile_device has Video-interface THEN social-role is Video-recording-and-watching.

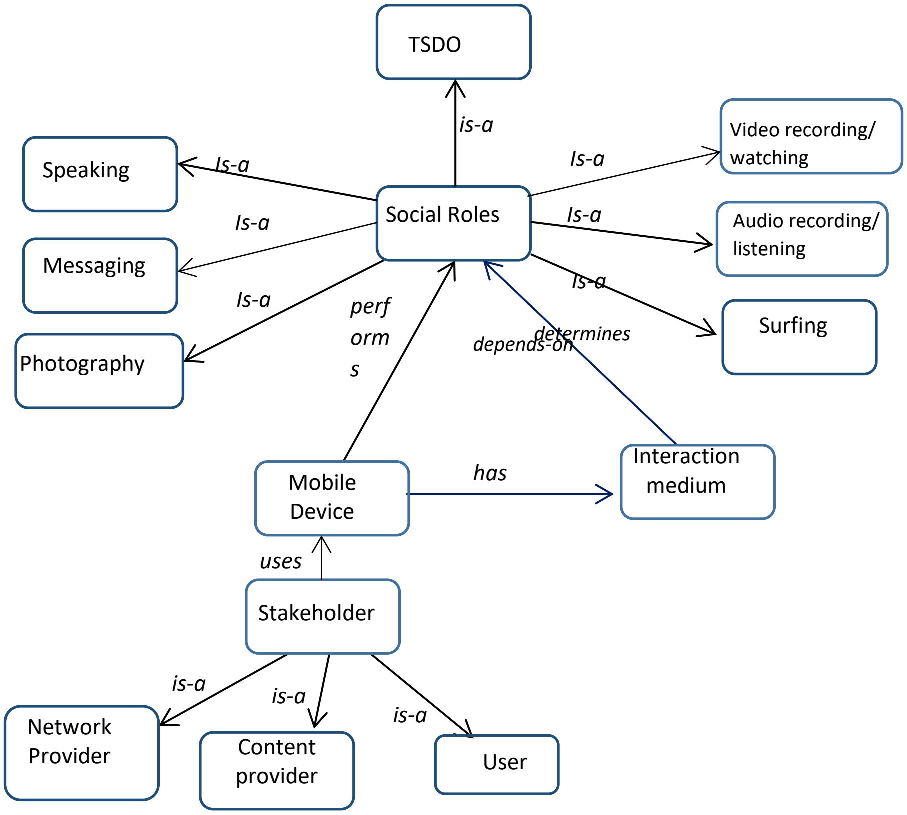


Figure 4 Mobile users' social role ontology

Devices with combined service categories can assume more than one service role. Hence, the number of service roles a device can perform determines its adaptation in mobile environments which, in this regard, is the interaction medium. Rules 7 to 12 in the ontology give the combined service categories as the antecedent and the resulting social roles as the conclusion.

Rule 7: IF InteractionMediumCategory include Voice-interface only THEN Speaking
Rule 8: IF InteractionMediumCategory include Message-interface AND Voice-interface THEN Messaging
Rule 9: IF InteractionMediumCategory include Photo-interface AND Message-interface AND Voice-interface THEN Photography

Rule 10: IF InteractionMediumCategory include Surfing-interface AND Photo-interface AND Message-interface AND Voice-interface THEN Surfing

Rule 11: IF InteractionMediumCategory include Audio-interface AND Surfing-interface AND Photo-interface AND Message-interface AND Voice-interface THEN Audio-recording-and-listening.

Rule 12: IF InteractionMediumCategory include Video-interface AND Audio-interface AND Surfing-interface AND Photo-interface AND Message-interface AND Voice-interface THEN Video-recording-and-watching.

Rules 7 to 12 depict the relationships among the various social roles as follows:

Video-recording-and-watching ⊃ Audio-recording-and-listening ⊃ Surfing ⊃ Photography ⊃ Messaging ⊃ Speaking.

This means that:

Speaking ⊂ Messaging ⊂ Photography ⊂ Surfing ⊂ Audio-recording-and-listening ⊂ Video-recording-and-watching.

This logical conclusion shows the requirement for social role of a device which determines its level of adaptation in a mobile environment.

The proposed framework for providing context-aware personalized web services is shown in figure 5. It indicates that the *User Device* is a smart device which is equipped with the context-aware sensors that gather entered user's request and acquire the necessary contextual information. *Context Acquisition and Assimilation* module is to capture contextual information, inserted, captured or stored, and manages their heterogeneity while *Context Processing and Reasoning* module is responsible for reasoning and deducing new situations from OWL representations of context and inference rules. *Service Adaptation* module allows content providers to adapt services to context at runtime based on services from situations derived by context processing and reasoning. In order to infer high level context from low level one, the context ontology is used as input to context processing and reasoning module. It is characterized by a two-level hierarchy: the general, domain independent and the domain dependent, application specific levels. The first level may specify the user's profile and preferences, service requested, activity, device and environment related properties.

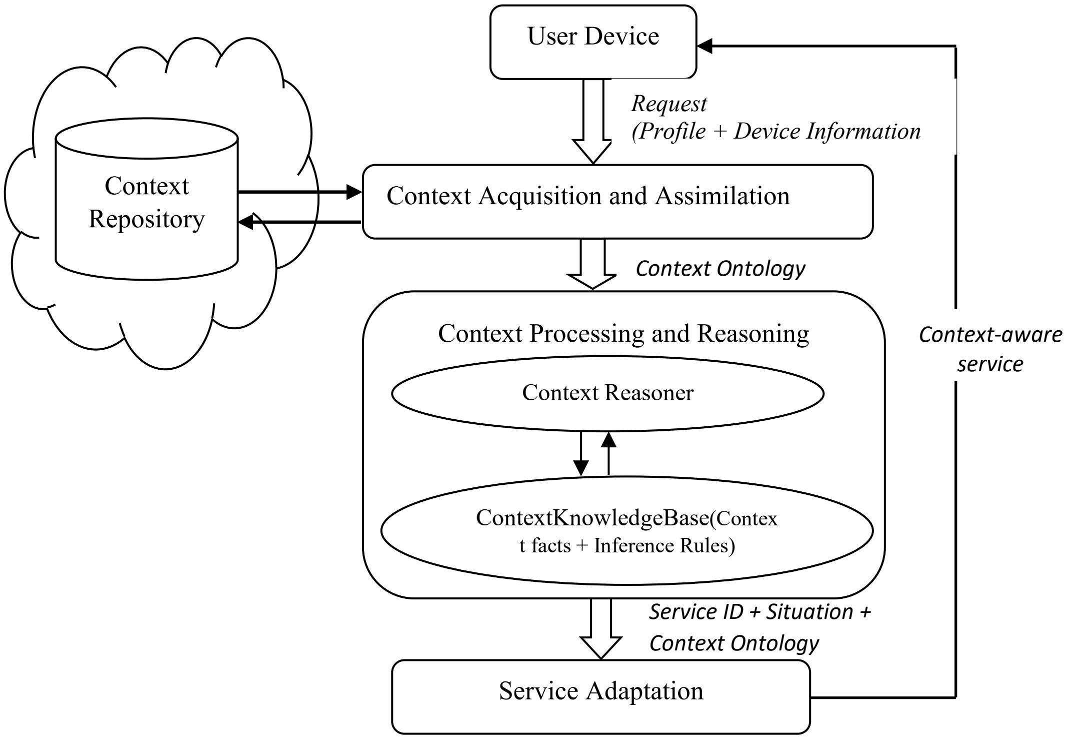

User Device
Request
(Profile + Device Information
Context Repository
Context Acquisition and Assimilation
Context Ontology
Context Processing and Reasoning
Context Reasoner
ContextKnowledgeBase(Contex t facts + Inference Rules)
Context-aware service
Service ID + Situation +
Context Ontology
Service Adaptation

Figure 5 Framework for context-aware web services provision

4. Model implementation

Next, we implement the ontology model for social roles that will enhance the provision of value added mobile web services. Implementing the social role ontology in Protégé allows us to determine whether a mobile device with its interaction medium is fit for a social role or not. The following axioms define the rules required to reason with facts in the ontology. Rules 1 to 12 of the previous section are represented in *Ax.1* to *Ax.12* using first order logic as the language of representation with predicates reified as entities or objects. The notations md, sr and im mean mobile device, social role and interaction medium respectively with *has* and *is* as predicates.

Ax.1: $\forall md, sr \exists im.\ has(md, im(Voice)) \Rightarrow is(sr, Speaking)$

Ax.2: $\forall md, sr \exists im.\ has(md, im(Message)) \Rightarrow is(sr, Messaging)$

Ax.3: $\forall md, sr \exists im.\ has(md, im(Photo)) \Rightarrow is(sr, Photography)$

Ax.4: $\forall md, sr \exists im.\ has(md, im(Surfing)) \Rightarrow is(sr, Surfing)$

Ax.5: $\forall md, sr \exists im.\ has(md, im(Audio)) \Rightarrow is(sr, Audio\text{-}recording\text{-}and\text{-}listening)$

Ax.6: $\forall md, sr \exists im.\ has(md, im(Video)) \Rightarrow is(sr, Video\text{-}recording\text{-}and\text{-}watching)$

Ax.7: $\forall md, im, sr.\ has(md, im(Voice)) \Rightarrow is(sr, Speaking)$

Ax.8: $\forall md, im, sr.\ has(md, im(Voice \wedge Message)) \Rightarrow is(sr, Messaging)$

Ax.9: $\forall md, im, sr.\ has(md, im(Voice \wedge Message \wedge Photo)) \Rightarrow is(sr, Photography)$

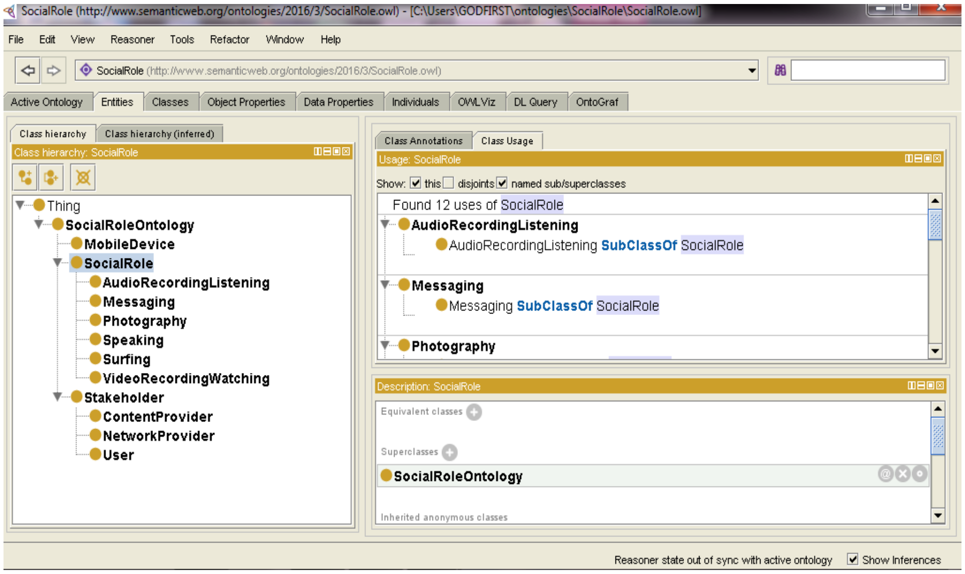
Ax.10: $\forall md, im, sr.\ has(md, im(Voice \wedge Message \wedge Photo \wedge Surfing)) \Rightarrow is(sr, Surfing)$

Ax.11: $\forall md, im, sr.\ has(md, im(Voice \wedge Message \wedge Photo \wedge Surfing \wedge Audio)) \Rightarrow is(sr, Audio\text{-}recording\text{-}and\text{-}listening)$

Ax.12: $\forall md, im, sr.\ has(md, im(Voice \wedge Message \wedge Photo \wedge Surfing \wedge Audio \wedge Video)) \Rightarrow is(sr, Video\text{-}recording\text{-}and\text{-}watching)$

Upon addition of these axioms to the ontology, some of the competency questions earlier stated will be answered.

The implementation of the service role ontology using Protégé results in the class hierarchy given in figure 6. From the class hierarchy, three main classes are defined. They are*SocialRole*, *Stakeholder* and*MobileDevice.*The usage of the *SocialRole* class is described. The *SocialRole* class has six sub-classes which include *Speaking*, *Messaging*, *Photography,Surfing,AudioRecordingListening,* and *VideoRecordingWatching*. The *Stakeholders* class has *ContentProvider, NetworkProvider* and the *User* as its sub-classes. The *MobileDevice* class is defined by the *BrandName*and InteractionMediumas its properties. The *InteractionMedium* obtained from Table 3 is classified according to the interface it provides and the classes are *VoiceInterface,MessageInterface, PhotoInterface, SurfingInterface, AudioInterface* and *VideoInterface*. Sample device types, brand names interaction medium are specified in table 4. The classes are linked together with the relations shown in the ontology in figure 4 and defined as the object properties and shown in the implementation interface in figure 7. These object properties include *has, is-a, uses, performs, depends-on* and *determines.*



Found 12 uses of SocialRole

AudioRecordingListening
AudioRecordingListening SubClassOf SocialRole

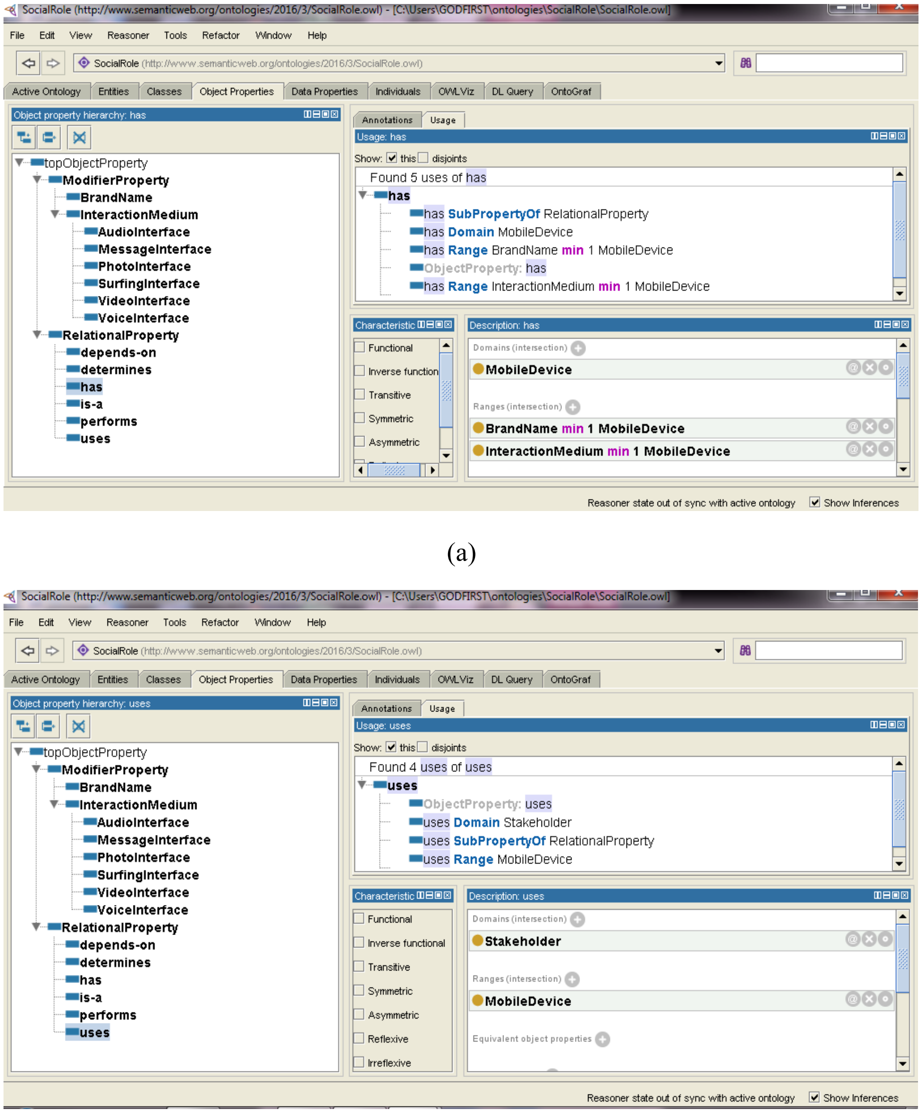
Messaging
Messaging SubClassOf SocialRole

Photography

Figure 6 Class hierarchy of SocialRole ontology

Table 4. Sample Brand Name of Phones and Interaction Medium

Brand Name	Phone Type	Voice Interface	Message Interface	Photo Interface	Surfing Interface	Audio Interface	Video Interface
Nokia	3310 (Dual SIM)	Yes	Yes	Yes	No	Yes	No
	2626	Yes	Yes	No	Yes	Yes	No
Tecno	Spark Plus K9	Yes	Yes	Yes	Yes	Yes	Yes
	T401 (Dual SIM)	Yes	Yes	Yes	Yes	Yes	Yes
	T401 (Triple SIM)	Yes	Yes	Yes	Yes	Yes	Yes
	T349	Yes	Yes	Yes	No	No	No
iMose	M3i (Dual SIM)	Yes	Yes	Yes	No	Yes	Yes
AIEK	E1	Yes	Yes	No	No	Yes	No
Fero	K24101	Yes	Yes	Yes	Yes	Yes	Yes
Bontel	3310	Yes	Yes	Yes	No	No	No
Generic	3D Animal Print	Yes	Yes	Yes	Yes	Yes	Yes


SocialRole (http://www.semanticweb.org/ontologies/2016/3/SocialRole.owl) - [C:\Users\GODFIRST\ontologies\SocialRole\SocialRole.owl]
File Edit View Reasoner Tools Refactor Window Help
SocialRole (http://www.semanticweb.org/ontologies/2016/3/SocialRole.owl)
Active Ontology Entities Classes Object Properties Data Properties Individuals OWLViz DL Query OntoGraf
Annotations Usage
Characteristic
Reasoner state out of sync with active ontology
Show Inferences

Object property hierarchy: has

- topObjectProperty
  - ModifierProperty
    - BrandName
    - InteractionMedium
      - AudioInterface
      - MessageInterface
      - PhotoInterface
      - SurfingInterface
      - VideoInterface
      - VoiceInterface
  - RelationalProperty
    - depends-on
    - determines
    - has
    - is-a
    - performs
    - uses

Usage: has

Show: ☑ this ☐ disjoints

Found 5 uses of has

- has
  - has SubPropertyOf RelationalProperty
  - has Domain MobileDevice
  - has Range BrandName min 1 MobileDevice
  - ObjectProperty: has
  - has Range InteractionMedium min 1 MobileDevice

☐ Functional
☐ Inverse function
☐ Transitive
☐ Symmetric
☐ Asymmetric

Description: has

Domains (intersection)

MobileDevice

Ranges (intersection)

BrandName min 1 MobileDevice

InteractionMedium min 1 MobileDevice

(a)

Object property hierarchy: uses

- topObjectProperty
  - ModifierProperty
    - BrandName
    - InteractionMedium
      - AudioInterface
      - MessageInterface
      - PhotoInterface
      - SurfingInterface
      - VideoInterface
      - VoiceInterface
  - RelationalProperty
    - depends-on
    - determines
    - has
    - is-a
    - performs
    - uses

Usage: uses

Show: ☑ this ☐ disjoints

Found 4 uses of uses

- uses
  - ObjectProperty: uses
  - uses Domain Stakeholder
  - uses SubPropertyOf RelationalProperty
  - uses Range MobileDevice

☐ Functional
☐ Inverse functional
☐ Transitive
☐ Symmetric
☐ Asymmetric
☐ Reflexive
☐ Irreflexive

Description: uses

Domains (intersection)

Stakeholder

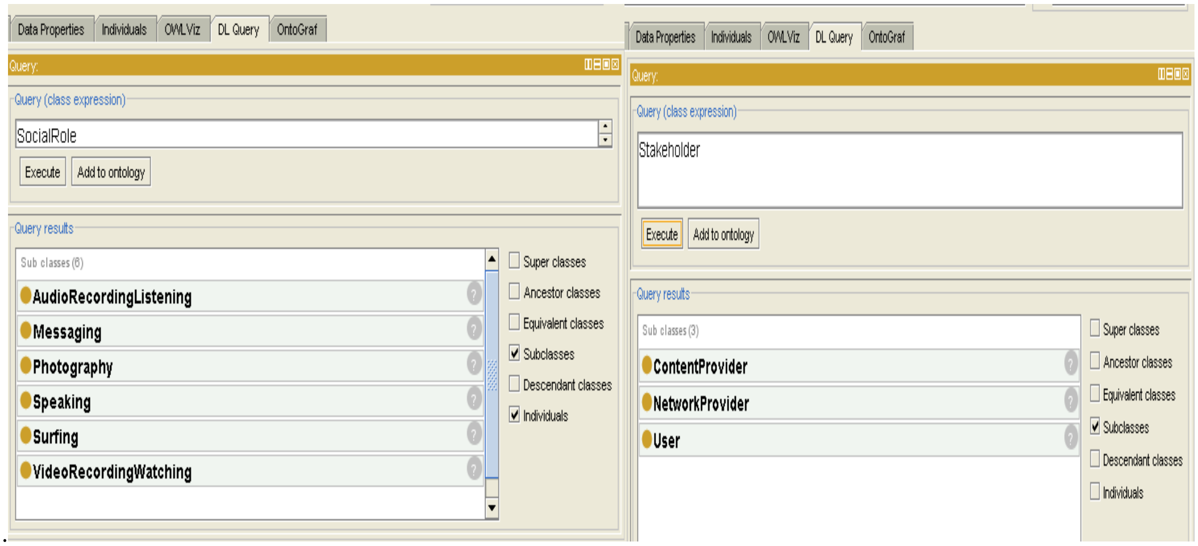
Ranges (intersection)

MobileDevice

Equivalent object properties

(b)

Figure 7 Object properties of *SocialRole* Ontology with usage description for '*has*' and '*uses*' relations in (a) and (b) respectively


Data Properties Individuals OWLViz DL Query OntoGraf
Query:
Execute
Add to ontology
Data Properties Individuals OWLViz DL Query OntoGraf
Query:
Execute
Add to ontology

Query (class expression)

SocialRole

Query results

Sub classes (6)

- AudioRecordingListening
- Messaging
- Photography
- Speaking
- Surfing
- VideoRecordingWatching

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses
☐ Descendant classes
☑ Individuals

Query (class expression)

Stakeholder

Query results

Sub classes (3)

- ContentProvider
- NetworkProvider
- User

☐ Super classes
☐ Ancestor classes
☐ Equivalent classes
☑ Subclasses
☐ Descendant classes
☐ Individuals

(a) (b)

Figure 8 Query results from the *SocialRole* ontology

A sample query on the *SocialRole* ontology gives the results from the knowledge base as shown in figure 8 (a) social roles and (b) stakeholder

5. Conclusion

The advancement with ubiquitous computing technologies significantly enhance people's ability to communicate, learn, relax, and be entertained in a safe and comfortable environment such as home, office, campus, library, classroom, laboratory, market, or church. The proposed framework for mobile phone services provisioning supporting context-aware system performs reasoning over contexts to support dynamic adaptation to changing environment. This guarantees the quality of the context, deduce implicit information and pass decisions about the actions to be triggered. The paper shows that ontology-based context-aware model can provide situation and social-aware services in much pervasive way for different domains. This provides useful tool for development and deployment of needed applications by content developers and service providers to mobile users. The end effect will be provision of satisfied mobile users' quality of service and quality of experience requirements anytime, anywhere.

References

[1] Farley, P. and Capp, M. (2005) 'Mobile web services', *BT Technology Journal*, 23(2), 202-213.

[2] Palen, L., Salzman, M. and Youngs, E. (2000) 'Going wireless: behaviour and practice of new mobile phone users', *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, ACM Press, New York, USA, 201-210.

[3] Tamminen, S., Oulasvirta, A, Toiskallio, K. and Kankainen, A. (2004) 'Understanding mobile contexts', *Personal and Ubiquitous Computing*, 8(2), 135-143.

[4] Skillen KL., Chen L., Nugent C.D., Donnelly M.P., Burns W., Solheim I. (2012) Ontological User Profile Modeling for Context-Aware Application Personalization. In: Bravo J., López-de-Ipiña D., Moya F. (eds) Ubiquitous Computing and Ambient Intelligence. UCAmI 2012. Lecture Notes in Computer Science, vol 7656. Springer, Berlin, Heidelberg.

[5] Pruitt, J. and Adlin, T. (2006) *The Persona Lifestyle: Keeping People in mind throughout Product Design,* 1st ed., Morgan Kaufmann, San Francisco.

[6] Anderson, P. and Blackwood, A. (2004) *Mobile and PDA technologies and their future use in education*, *JISC Technology and Standards Watch*, Bristol, UK.

[7] Kamberov, R. (2016). Using social paradigms in smart cities mobile context-aware computing. *11th IEEE Iberian Conference onInformation Systems and Technologies (CISTI),* (1-5).

[8] Ogata, H. and Yano, Y. (2004) 'Context-aware support for computer-supported ubiquitous learning', *Proceedings of IEEE WMTE Conference*, Taoyuan, Taiwan, Computer Society Press, 27-34.

[9] Rogers, Y., Price, S., Randell, C., Fraser, D. S., Weal, M. and Fitzpatrick, G. (2005) 'Ubi-learning integrating indoor and outdoor learning experiences', *Communications of the ACM*, 48(1), 55-59.

[10] Meier, R and Lee, D. (2011) 'Context-aware pervasive services for smart cities', Ubiquitous Developments in Ambient Computing and Intelligence: Human-Centered Applications, IGI Global, 1, 1-6.

[11] Khan, Z., Kiani, S. L. and Soomro, K. (2014), 'A framework for cloud-based context-aware information services for citizens in smart cities', Journal of Cloud Computing: Advances, Systems and Applications, 3(14), 1-17.

[12] Schmidt, C. (2011) '*Context-aware computing*', Berlin Inst. Technology Tech, 1-9, retrieved from http://diuf.unifr.ch/pai/education/2002_2003/seminar/winter/ubicomp/02_Pervasive.pdf/

[13] Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The Semantic Web', *Scientific Americana, 284(5), 34-43*.

[14] Lassila, O. and Adler, M. (2003) 'Semantic gadgets: Ubiquitous computing meets the semantic web', In Dieter Fensel *et al.,* (Eds.), Spinning the Semantic Web, MIT Press, 363-376.

[15] Brown, G., Bull, J. and Pendlebury, M. (1997) London.

[16] Chen, G. and Kotz, D. (2000) *A survey of context-aware mobile computing research*, Tech. Rep. TR2000-381, Darthmouth.

[17] Schilit, B. and Theimer, M. (1994) 'Disseminating active map information to mobile hosts', *IEEE Network*, 8(5), 22-32.

[18] Dey, A. K. (2001) 'Understanding and using context', *Personal and Ubiquitous Computing*, 5(1), 4-7.

[19] Musumba, G. W. and Nyongesa, H. O. (2013) 'Context awareness in mobile computing: A review', *International Journal of Machine Learning Applications*, 2(1), 1-10.

[20] Dey, A. K. and Abowd, G. D. (2000), 'Towards a better understanding of context and context-awareness', *Proceedings of the 2000 Conference on Human Factors in Computing Systems (CHI 2000) in the workshop of the What, Who, Where, When and How of Context-Awareness,* The Hague, The Netherlands, 371.

[21] Prekop, P. and Burnett, M. (2003) 'Activities, context and ubiquitous computing', *Computer Communications*, 26(11), 1168-1176.

[22] Bardram, J. E. and Hansen, T. R. (2010) 'Context-based workplace awareness', *Computer Supported Cooperative Work (CSCW)*, 19, 105-138.

[23] Hong, J. H., Yang, S. I. and Cho, S. B. (2010) 'ConaMSN: A context-aware messenger using dynamic Bayesian networks with wearable sensors', *Expert Systems with Applications*, 37(6), 4680-4686.

[24] Strang, T. and Linnhoff-Popien, C. (2004), 'A context modeling survey', Proceedings of the *First International Workshop on Advanced Context Modeling, Reasoning and Management at Ubicomp 2004*, Nottingham, UK, 34-41.

[25] [25] Ahmed, S. and Parsons, D. (2011), 'ThinknLearn: An ontology-driven mobile web application for science enquiry based learning', *Proceedings ofthe 7th International Conference on Information Technology and Applications (ICITA 2011), IEEE Computer Chapter NSW, Sydney, Australia,* 255-260.

[26] Protégé-OWL Editor and Application Programming Interface. [online] http://protege.stanford.edu/ (Accessed 12 February, 2016).

[27] Veijalainen, J. (2008) 'Mobile ontologies: Concepts, development, usage, and business potential, *International Journal on Semantic Web and Information Systems,* 4(1), 20-34.

Advances in Science, Technology and Engineering Systems Journal
Vol. 3, No. 2, 329-345 (2018)
www.astesj.com
Special Issue on Advancement in Engineering Technology

ASTES Journal
ISSN: 2415-6698

Efficient Alignment of Very Long Sequences

Chunchun Zhao*, Sartaj Sahni

Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA

ARTICLEINFO

Article history:
Received: 20 March, 2018
Accepted: 16 April, 2018
Online: 30 April, 2018

Keywords:
Sequence alignment
Energy efficient
Multi core

ABSTRACT

We consider the problem of aligning two very long biological sequences. The score for the best alignment may be found using the Smith-Waterman scoring algorithm while the best alignment itself may be determined using Myers and Miller's alignment algorithm. Neither of these algorithms takes advantage of computer caches to obtain high efficiency. We propose cache-efficient algorithms to determine the score of the best alignment as well as the best alignment itself. All algorithms were implemented using C and OpenMP, and benchmarked using real data sets from the National Center for Biotechnology Information (NCBI) database. The test computational platforms were Xeon E5 2603, I7-x980 and Xeon E5 2695. Our best single-core cache-efficient scoring algorithm reduces the running time by as much as 19.7% relative to the Smith-Waterman scoring algorithm and our best cache-efficient alignment algorithm reduces the running time by as much as 17.1% relative to the Myers and Miller alignment algorithm. Multicore versions of our cache-efficient algorithms scale quite well up to the 24 cores we tested; achieving a speedup of 22 with 24 cores. Our multi-core scoring and alignment algorithms reduce the running time by as much as 61.4% and 47.3% relative to multi-core versions of the Smith-Waterman scoring algorithm and Myers and Miller's alignment algorithm, respectively.

1 Introduction

Sequence alignment is a fundamental and well-studied problem in the biological sciences. In this problem, we are given two sequences $A[1:m] = a_1a_2\cdots a_m$ and $B[1:n] = b_1b_2\cdots b_n$ and we are required to find the score of the best alignment and possibly also an alignment with this best score. When aligning two sequences, we may insert gaps into the sequences. The score of an alignment is determined using a matching (or scoring) matrix that assigns a score to each pair of characters from the alphabet in use as well as a gap penalty model that determines the penalty associated with a gap sequence. In the linear gap penalty model, the penalty for a gap sequence of length $k > 0$ is kg, where g is some constant while in the affine model this penalty is $g_{open} + (k-1) * g_{ext}$. The affine model more accurately reflects the fact that opening a gap is more expensive than extending one. Two versions of sequence alignment–global and local–are of interest. In global alignment, the entire A sequence is to be aligned with the entire B sequence while in local alignment, we wish to find a substring of A and B that have the highest alignment score. The alphabet for DNA, RNA, and protein sequences is, respectively, {A, T, G, C}, {A, U, G, C}, and {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}.

Figure 1 illustrates these concepts using the DNA sequences $A[1:8]$ = {AGTACGCA} and $B[1:5]$ = {TATGC}. The symbol '_' denotes the gap character. The alignment of Figure 1(a) is a global alignment and that of Figure 1(b) is a local one. To score the alignments, we have used the linear penalty model with $g = -2$ and the scores for pairs of aligned characters, which are taken from BLOSUM62 matrix in [1], are $c(T,T) = 5$, $c(A,A) = 4$, $c(C,C) = 9$, $c(G,G) = 6$, and $c(C,T) = -1$. The score for the shown global alignment is 17 while that for the shown local alignment is 23. If we were using an affine penalty model with $g_{open} = -4$ and $g_{ext} = -2$, then the penalty for each of the gaps in positions 1 and 8 of the global alignment would be -4 and the overall score for the global alignment would be

*Corresponding Author: Chunchun Zhao, czhao@cise.ufl.edu

https://dx.doi.org/10.25046/aj030236

13.

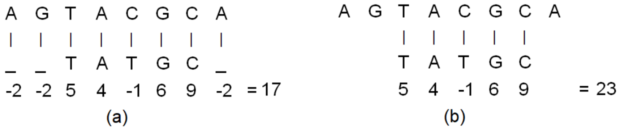


Figure 1: Example alignments using the linear gap penalty model. (a) Global alignment (b) Local alignment

In [2], the authors first proposed an $O(mn)$ time algorithm, called Needleman-Wunsch(NW) algorithm, for global alignment using the linear gap model. This algorithm requires $O(n)$ space when only the score of the best alignment is to be determined and $O(mn)$ space when the best alignment is also to be determined. In [3], the authors proposed a new algorithm called Smith-Waterman(SW) algorithm, which modified the NW algorithm so as to determine the best local alignment. In [4], the author proposed a dynamic programming algorithm called Gotoh algorithm, for sequence alignment using an affine gap penalty model. The asymptotic complexity of the SW and Gotoh algorithms is the same as that of the NW algorithm.

When mn is large and the best alignment is sought, the space, $O(mn)$, required by the algorithms of NW, SW and Gotoh exceeds what is available on most computers. The best alignment for these large instances can be found in [5] using sequence alignment algorithms derived from Hirschberg's linear space divide-and-conquer algorithm for the longest common subsequence problem. In [6], the authors developed a Myers-Miller alignment algorithm. It is the linear space $O(mn)$ time version of Hirschberg's algorithm for global sequence alignment using an affine gap penalty model. And in [7], the authors do this for local alignment.

In an effort to speed sequence alignment, fast sequence-alignment heuristics have been developed. As in [8, 9, 10],BLAST, FASTA, and Sim2 are a few examples of software systems that employ sequence alignment heuristics. Another direction of research, also aimed at speeding sequence alignment, has been the development of parallel algorithms. Parallel algorithms for sequence alignment may be found in [11]-[19], for example.

In this paper, we focus on reducing the number of cache misses that occur in the computation of the score of the best alignment as well as in determining the best alignment. Although we explicitly consider only the linear gap penalty model, our methods readily extend to the affine gap penalty model. Our interest in cache misses stems from two observations–(1) the time required to service a last-level-cache (LLC) miss is typically 2 to 3 orders of magnitude more than the time for an arithmetic operation and (2) the energy required to fetch data from main memory is typically between 60 to 600 times that needed when the data is on the chip. As a result of observation (1), cache misses dominate the overall running time of applications for which the hardware/software cache prefetch modules on the target computer are ineffective in predicting future cache misses. The effectiveness of hardware/software cache prefetch mechanisms varies with application, computer, and compiler. So, if we are writing code that is to be used on a variety of computer platforms, it is desirable to write cache-efficient code rather than to rely exclusively on the cache prefetching of the target platform. Even when the hardware/software prefetch mechanism of the target platform is very effective in hiding memory latency, observation (2) implies excessive energy use when there are many cache misses.

This paper is an extension of work originally in [20], which has been presented by us in the 2015 IEEE 5th international conference on Computational Advances in Bio and Medical Sciences (ICCABS). The main contributions are

1. cache efficient single-core and multi-core algorithms to determine the score of the best alignment;

2. cache efficient single-core and multi-core algorithms to determine the best alignment.

The rest of the paper is organized in the following way. In Section 2, we describe our cache model. Our cache-efficient algorithms for scoring and alignment are developed and analyzed in Section 3. Experimental results are presented in Section 4. In Section 5, we present a discussion of these results and in Section 6, we present the limitations of our work. Finally, we conclude in Section 7.

2 Cache Model

For simplicity in the analysis, we assume a single cache comprised of s lines of size w words (a word is large enough to hold a piece of data, typically 4 bytes) each. So, the total cache capacity is sw words. The main memory is partitioned into blocks also of size w words each. When the program needs to read a word that is not in the cache, a cache miss occurs. To service this cache miss, the block of main memory that includes the needed word is fetched and stored in a cache line, which is selected using the LRU (least recently used) rule. Until this block of main memory is evicted from this cache line, its words may be read without additional cache misses. We assume the cache is written back with write allocate. Write allocate means that when the program needs to write a word of data, a write miss occurs if the block corresponding to the main memory is not currently in cache. To service the write miss, the corresponding block of main memory is fetched and stored in a cache line. Write back means that the word is written to the appropriate cache line only. A cache line with changed content is written back to the main memory when it is about to be overwritten by a new block from main memory.

Rather than directly assess the number of read and write misses incurred by an algorithm, we shall count the number of read and write accesses to main memory.

Every read and write miss makes a read access. A read and write miss also makes a write access when the data in the replacement cache line is written to main memory.

We emphasize that the described cache model is a very simplified model. In practice, modern computers commonly have two or three levels of cache and employ sophisticated adaptive cache replacement strategies rather than the LRU strategy described above. Further, hardware and software cache prefetch mechanisms are often deployed to hide the latency involved in servicing a cache miss. These mechanisms may, for example, attempt to learn the memory access pattern of the current application and then predict the future need for blocks of main memory. The predicted blocks are brought into cache before the program actually tries to read/write from/into those blocks thereby avoiding (or reducing) the delay involved in servicing a cache miss. Actual performance is also influenced by the compiler used and the compiler options in effect at the time of compilation. As a result, actual performance may bear little relationship to the analytical results obtained for our simple cache model. Despite this, we believe the simple cache model serves a useful purpose in directing the quest for cache-efficient algorithms that eventually need to be validated experimentally.

3 Cache Efficient Algorithms

3.1 Scoring Algorithms

3.1.1 Needleman-Wunsch and Smith-Waterman algorithm

Let H_{ij} be the score of the best global alignment for $A[1:i]$ and $B[1:j]$. We wish to determine H_{mn}. In [2], the authors derived the following dynamic programming equations for H using the linear gap penalty model. These equations may be used to compute H_{mn}.

$$H_{i,0} = -i * g \;\; H_{0,j} = -j * g, \;\; 0 \le i \le m, \;\; 0 \le j \le n \tag{1}$$

When $i > 0$ and $j > 0$,

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + c(a_i, b_j) \\ H_{i,j-1} + c(\_, b_j) = H_{i,j-1} - g \\ H_{i-1,j} + c(a_i, \_) = H_{i-1,j} - g \end{cases} \tag{2}$$

where $c(a_i, b_j)$ is the match score between characters a_i and b_j and g is the gap penalty.

For local alignment, H_{ij} denotes the score of the best local alignment for $A[1:i]$ and $B[1:j]$. In [3], the Smith-Waterman equations for local alignment using the linear gap penalty model are:

$$H_{i,0} = 0, \;\; H_{0,j} = 0, \;\; 0 \le i \le m, \;\; 0 \le j \le n \tag{3}$$

When $i > 0$ and $j > 0$,

$$H_{i,j} = \max \begin{cases} 0 \\ H_{i-1,j-1} + c(a_i, b_j) \\ H_{i,j-1} + c(\_, b_j) = H_{i,j-1} - g \\ H_{i-1,j} + c(a_i, \_) = H_{i-1,j} - g \end{cases} \tag{4}$$

Several authors (in [5, 6], for example) have observed that the score of the best local alignment may be determined using a single array $H[0:n]$ as in algorithm $Score$(Algorithm 1.)

Algorithm 1 Smith-Waterman scoring algorithm

```
$Score(A[1:m], B[1:n])$
for $j \leftarrow 0$ to $n$ do
    $H[j] \leftarrow 0$ //Initialize row 0
end for
for $i \leftarrow 1$ to $m$ do
    $diag \leftarrow 0$ // Compute row i
    for $j \leftarrow 1$ to $n$ do
        $nextdiag \leftarrow H[j]$
        $H[j] \leftarrow \max\{0, diag + c(A[i], B[j]), H[j-1] - g, H[j] - g\}$
        $diag \leftarrow nextdiag$
    end for
end for
return $H[n]$
```

The scoring algorithm for the Needleman and Wunsch algorithm is similar. It is easy to see that the time complexity of the algorithm of Algorithm 1 is $O(mn)$ and its space complexity is $O(n)$.

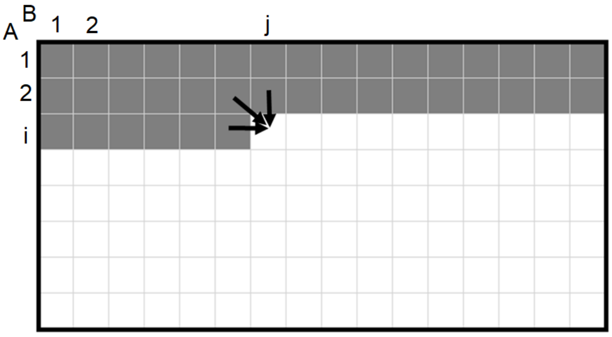


Figure 2: Memory access pattern for Score algorithm(Algorithm 1).

For the (data) cache miss analysis, we focus on read and write misses of the array H and ignore misses due to the reads of the sequences A and B as well as of the scoring matrix c (notice that there are no write misses for A, B, and c). Figure 2 shows the memory access pattern for H by algorithm $Score$. The first row denotes the initialization of H and subsequent rows denote access for different value of i (i.e., different iterations of the `for i` loop). The computation of H_{ij} is done using a single one-dimensional array $H[]$, following the i'th iteration of the `for i` loop, $H[j] = H_{ij}$. In each iteration of this loop, the elements of $H[]$ are accessed left-to-right. During the initialization loop, H is brought into the cache in blocks of size w. Assume that n is sufficiently large so that $H[]$ does not entirely fit into the cache. Hence, at some value of j, the cache capacity is reached and further progress of the initialization loop causes the least recently used blocks of $H[]$

(i.e., blocks from left to right) to be evicted from the cache. The evicted blocks are written to main memory as they have been updated. So, the initialization loop results in n/w read accesses and (approximately) n/w write accesses (the number of write accesses is actually $n/w - s$). Since the left part of $H[]$ has been evicted from the cache by the time we start the computation for row i ($i > 0$), each iteration of the `for i` loop also results in n/w read accesses and approximately n/w write accesses. So, the total number of read accesses is $(m+1)n/w \approx mn/w$ and the number of write accesses is also $\approx mn/w$. The number of read and write accesses is $\approx 2mn/w$, when n is large.

We note, however, that when n is sufficiently small that $H[]$ fits into the cache, the number of read accesses is n/w (all occur in the initialization loop) and there are no write accesses. In practice, especially in the case of local alignment involving a very long sequence, one of the two sequences A and B is small enough to fit in the cache while the other may not fit in the cache. So, in these cases, it is desirable to ensure that A is the longer sequence and B is the shorter one so that H fits in the cache entirely. This is accomplished by swap A and B sequences.

When $m < n$ and $H[1:m]$ fits into the cache and $H[1:n]$ does not, algorithm *Score* incurs $O(mn/w)$ read/write accesses, while swap A and B incurs $O(m/w)$ read/write accesses.

3.1.2 Diagonal Algorithm

An alternative to computing the score by rows is to compute by diagonals. While this uses two one-dimensional arrays rather than one, it is more readily parallelized than *Score* as all elements on a diagonal can be computed at the same time; elements on a row need to be computed in sequence.

Algorithm 2 Diagonal scoring algorithm

```
$Diagonal(A[1:m], B[1:n])$
$d2[0] \leftarrow d1[0] \leftarrow d1[1] \leftarrow 0$
for $d \leftarrow 2$ to $m+n$ do
    $x \leftarrow (d <= m?0 : d-m); y \leftarrow (d <= n?d : n);$
    for $i \leftarrow$ x to y do
        $j \leftarrow d - i$
        $diag \leftarrow next + c(A[i] + B[j])$
        $left \leftarrow d1[i-1] - g$
        $up \leftarrow d1[i] - g$
        $next \leftarrow d2[i]$
        $d2[i] \leftarrow max\{0, diag, left, up\}$
    end for
    $swap(d1, d2)$
end for
return $d1[n]$
```

Algorithm *Diagonal* (Algorithm 2) uses two one-dimensional arrays $d1[]$ and $d2[]$, where $d2[]$ stores the scores for the $(d-2)$th diagonal data and $d1[]$ stores them for the $(d-1)$th diagonal. When we compute the element $H_{i,j}$ in the dth diagonal, the previous diagonal element $H_{i-1,j-1}$ is fetched from $d2[]$ and the previous left element $H_{i-1,j}$ and previous upper element $H_{i,j-1}$ are fetched from $d1[]$. The calculated $H_{i,j}$ overwrites the old value in $d2[]$.

The total number of read accesses is mn/w for each diagonal array and the total number of write accesses is mn/w for both arrays combined. The number of cache misses for *Diagonal* is approximately $3mn/w$ when n is large.

3.1.3 Strip Algorithm

When neither $H[1:m]$ nor $H[1:n]$ fits into the cache, accesses to main memory may be reduced by computing H_{ij} by strips of width q such that q consecutive elements of $H[]$ fit into the cache. Specifically, we partition $H[1:n]$ into n/q strips of size q (except possibly the last strip whose size may be smaller than q) as in Figure 3. First, all H_{ij} in strip 0 are computed, then those in strip 1, and so on. When computing the values in a strip, we need those in the rightmost column of the preceding strip. So, we save these rightmost values in a one-dimensional array $strip[0:m]$. The algorithm is given in Algorithm 3. We note that sequence alignment by strips has been considered before. For example, in [12], the authors using the similar approach in their GPU algorithm. Their use differs from ours in that they compute the strips in pipeline fashion with each strip assigned to a different pipeline stage in round robin fashion and within a strip, the computation is done by anti-diagonals in parallel. On the other hand, we do not pipeline the computation among strips and within a strip, our computation is by rows.

Algorithm 3 Strip scoring algorithm

```
$Strip(A[1:m], B[1:n])$
for $j \leftarrow 1$ to $m$ do
    $strip[j] \leftarrow 0$ //leftmost strip
end for
for $t \leftarrow 1$ to $n/q$ do
    for $j \leftarrow t*q$ to $t*q+q-1$ do
        $H[j] \leftarrow 0$ //Initialize first row
    end for
    for $i \leftarrow 1$ to $m$ do
        $diag \leftarrow strip[i-1]$
        $H[t*q-1] \leftarrow strip[i]$
        for $j \leftarrow t*q$ to $t*q+q-1$ do
            $nextdiag \leftarrow H[j]$
            $H[j] \leftarrow \max\{0, diag + c(A[i], B[j]), H[j-1] - g, H[j] - g\}$
            $diag \leftarrow nextdiag$
        end for
        $strip[i] \leftarrow H[t*q+q-1]$
    end for
end for
return $H[n]$
```

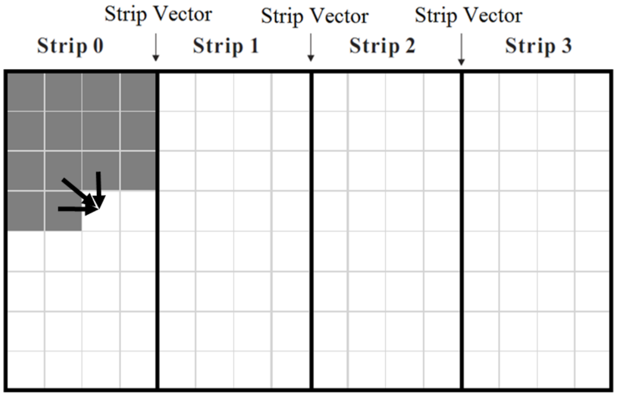


Figure 3: Memory access pattern for Strip algorithm (Algorithm 3).

It is easy to see that the time complexity of algorithm *Strip* is $O(mn)$ and that its space complexity is $O(m+n)$. For the cache misses, we focus on those resulting from the reads and writes of $H[]$ and *strip*[]. The initialization of *strip* results in m/w read accesses and approximately the same number of write accesses. The computation of each strip makes the following accesses to main memory:

1. q/w read accesses for the appropriate set of q entries of H for the current strip and q/w write accesses for the cache lines whose data are replaced by these H values. The write accesses are, however, not made for the first strip.

2. m/w read accesses for *strip* and m/w write accesses. The number of write accesses is less by s for the last strip.

So, the overall number of read accesses is $m/w + (q/w + m/w) * n/q = m/w + n/w + mn/(wq)$ and the number of write accesses is approximately the same as this. So, the total number of main memory accesses is $\approx 2mn/(wq)$ when m and n are large.

3.2 Alignment Algorithms

In this section, we examine algorithms that compute the alignment that results in the best score rather than just the best score. While in the previous section we explicitly considered local alignment and remarked that the results readily extend to global alignment, in this section we explicitly consider global alignment and remark that the methods extend to local alignment.

3.2.1 Myers and Miller's Algorithm

When aligning very long sequences, the $O(mn)$ space requirement of the full-matrix algorithm exceeds the available memory on most computers. For these instances, we need a more memory-efficient alignment algorithm. In [6], Myers and Miller have adapted Hirschberg's linear space algorithm for the longest common subsequence problem to find the best global alignment in linear space. Its time complexity is $O(mn)$. However, this linear space adaptation performs about twice as many operations as does the full-matrix algorithm. In [11], the authors have developed a hybrid algorithm, FastLSA, whose memory requirement adapts to the amount of memory available on the target computing platform. In this section and the next, we focus on the adaptation of Myers-Miller algorithm.

It is easy to see that an optimal (global) alignment is comprised of an optimal alignment of $A[1 : m/2]$ and $B[1 : j]$ and an optimal alignment of $A[m : m/2 + 1]$ ($A[m : i]$ is the reverse of $A[i : m]$) and $B[n : j + 1]$ for some j, $1 \leq j \leq n$. The value of j for which this is true is called the *optimal crossover point*. Myers and Miller's linear space algorithm for alignment determines the optimal alignment by first determining the optimal crossover point (si, sj) where $si = m/2$, and then recursively aligning $A[1 : m/2]$ and $B[1 : sj]$ as well as $A[m : m/2 + 1]$ and $B[n : sj + 1]$. Equivalently, an optimal alignment of $A[1 : m]$ and $B[1 : n]$ is an optimal alignment of $A[1 : m/2]$ and $B[1 : sj]$ concatenated with the reverse of an optimal alignment of $A[m : m/2 + 1]$ and $B[n : sj + 1]$. Hence, an optimal alignment is comprised of a sequence of optimal crossover points. This is depicted visually in Figure 4. Figure 4(a) shows alignments using 3 possible crossover points at row $m/2$ of H. Figure 4(b) shows the partitioning of the alignment problem into 2 smaller alignment problems (shaded rectangles) using the optimal crossover point (meeting point of the 2 shaded rectangles) at row $m/2$. Figure 4(c) shows the partitioning of each of the 2 subproblems of Figure 4(b) using the optimal crossover points for these subproblems (note that these crossovers take place at rows $m/4$ and $3m/4$, respectively). Figure 4(d) shows the constructed optimal alignment, which is presently comprised of the 3 determined optimal crossover points.

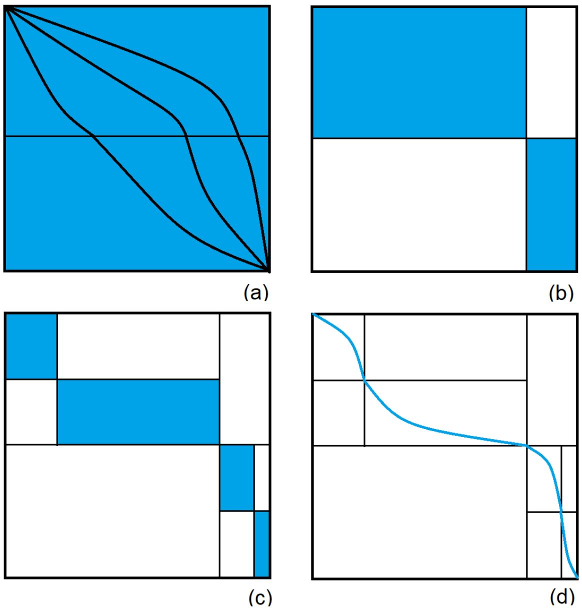


Figure 4: An alignment as crossover points. (a) Three alignments (b) Optimal crossover for m/2 (c)Optimal crossover points at m/4 and 3m/4 (d) Best alignment

Algorithm 4 Myers and Miller algorithm

```
$MM(A[1:m], B[1:n], path[1:m])$
if $m <= 1$ or $n <= 1$ then
    $Linear\ search\ to\ find\ optimal\ crossover\ point\ [si, sj]$
    $path[m] \leftarrow [si, sj]$
else
    $H_{top} \leftarrow MScore(A[1:\frac{m}{2}], B[1:n])$
    $H_{bot} \leftarrow MScore(A[m:\frac{m}{2}+1], B[n:1])$
    $Linear\ search\ to\ find\ optimal\ crossover\ point\ [si, sj]$
    $MM(A[1:\frac{m}{2}], B[1:sj], path[1:\frac{m}{2}])$
    $path[\frac{m}{2}] \leftarrow [si, sj]$
    $MM(A[\frac{m}{2}+1:m], B[sj+1:n], path[\frac{m}{2}+1:m])$
end if
```

The Myers and Miller algorithm, MM (Algorithm 4), uses a modified version of the linear space scoring algorithm $Score$ (Algorithm 1) to obtain the scores for the best alignments of $A[1:i]$ and $B[1:j]$, $1 \leq i \leq m/2$, $1 \leq j \leq n$ as well as for the best alignments of $A[m:i]$ and $B[m:j]$, $m/2 < i \leq m$, $1 \leq j \leq n$. This modified version $MScore$ differs from $Score$ only in that $MScore$ returns the entire array H rather than just $H[n]$. Using the returned H arrays for the forward and reverse alignments, the optimal crossover point for the best alignment is computed as in algorithm MM (Algorithm 4). Once the optimal crossover point is known, two recursive calls are made to optimally align the top and bottom halves of A with left and right parts of B. The approximately time complexity for iteration k is $O(2mn/2^k)$, hence the total time complexity is rough $2mn$.

In each level of recursion, the number of main memory accesses is dominated by those made in the calls to $MScore$. From the analysis for $Score$, it follows that when n is large, the number of accesses to main memory is $\approx 2mn/w(1 + 1/2 + 1/4 + \cdots) \approx 4mn/w$.

3.2.2 Diagonal Myers and Miller Algorithm

Let $Mdiagonal$ be algorithm $Diagonal$ (Algorithm 2) modified to return the entire H array rather than just $H[n]$. Our diagonal Myers and Miller algorithm ($MMDiagonal$) replaces the two statements in algorithm MM (Algorithm 4) that invoke $MScore$ with a test that causes $Mdiagonal$ to be used in place of $MScore$ when both of m and n are sufficiently long.

From the analysis for $Diagonal$, it follows that when m and n are large, the number of accesses to main memory is $\approx 3mn/w(1 + 1/2 + 1/4 + \cdots) \approx 6mn/w$.

3.2.3 Striped Myers and Miller Algorithm

Let $MStrip$ be algorithm $Strip$ (Algorithm 3) modified to return the entire H array rather than just $H[n]$. Our striped Myers and Miller algorithm ($MMStrip$) replaces the two statements in algorithm MM (Algorithm 4) that invoke $MScore$ with a test that causes $MStrip$ to be used in place of $MScore$ when both of m and n are sufficiently long.

From the analysis for $Strip$, it follows that when m and n are large, the number of accesses to main memory is $\approx 2mn/(wq)(1 + 1/2 + 1/4 + \cdots) \approx 4mn/(wq)$.

3.3 Parallel Scoring Algorithms

3.3.1 Parallel Score Algorithm

As remarked earlier in $Score$ algorithm, the elements in a row of the score matrix need to be computed sequentially from left to right because of data dependencies. So, we are unable to parallelize the inner **for** loop of $Score$ (Algorithm 1). Instead, we adopt the unusual approach of parallelizing the outer **for** loop while computing the inner loop sequentially using a single processor. Initially, processor s is assigned to do the outer loop computation for $i = s$, $1 \leq i \leq p$, where p is the number of processors. Processor s begins after a suitable time lag relative to the start of processor $s - 1$ so that the data it needs for its computation has already been computed by processor $s - 1$. That is, processor 1 begins the inner loop computation for $i = 1$ at time 0, then, with a suitable time lag, processor 2 begins the outer loop computation for $i = 2$, then, with a further lag, processor 3 begins the $i = 3$ computation and so on. When a processor has finished with its iteration i computation, it starts on iteration $i + p$ of the outer loop. Synchronization primitives are used to ensure suitable time lags. The time complexity of the resulting p-core algorithm PP_Score is $O(mn/p)$.

3.3.2 Parallel Diagonal Algorithm

The inner **for** loop of $Diagonal$ (Algorithm 2) is easily parallelized as the elements on a diagonal are independent and may be computed simultaneously. So, in our parallel version, we divide the diagonal d into p blocks, where p is the number of processors. We assign a block to each processor from left to right as in Figure 5. The time complexity of the resulting p-core algorithm $PP\_Diagonal$ is $O(mn/p)$.

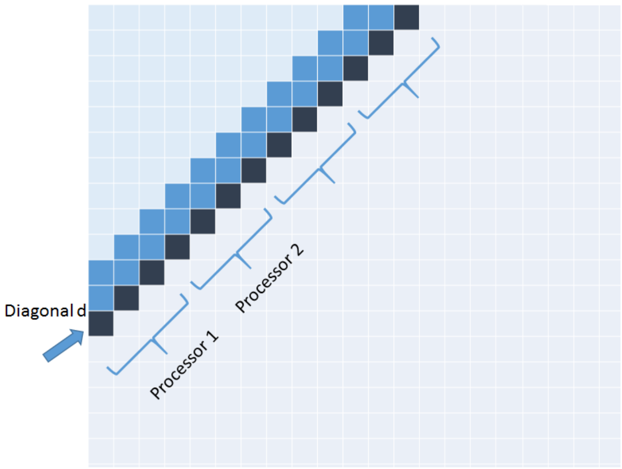


Figure 5: Parallel Diagonal algorithm.

3.3.3 Parallel Strip Algorithm

In the *Strip* scoring algorithm, we partition the score matrix H into n/q strips of size q (Figure 3) and compute the strips one at a time from left to right. Inside a strip, scores are computed row by row from top to bottom. We see that the computation of one strip can begin once the first row of the previous strip has been computed. In our parallel version of this algorithm, processor i is initially assigned to compute strip i, $1 \le i \le p$. When computing a value in its assigned strip, a processor needs to wait until the values (if any) needed from the strip to its left have been computed. When a processor completes the computation of strip j, it proceeds with the computation of strip $j+p$. Figure 6 shows a possible state in the described parallel strip computation strategy. We maintain an array $signal[]$ such that $signal[r] = s+1$ iff the row r computation for strips 1 through s has been completed. This array enables the processor working on the strip to its right to determine when it can begin the computation of its r'th row. The time complexity of the resulting parallel strip algorithm, PP_Strip, is $O(mn/p)$.

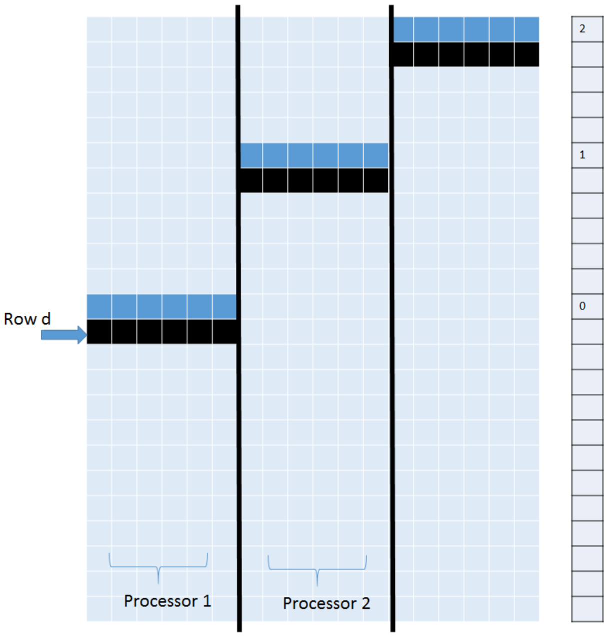


Figure 6: Parallel Strip algorithm.

3.4 Parallel Alignment Algorithms

In the single-core implementation, we divide the H matrix into two equal size parts and apply the scoring algorithm to each part. Then, we determine the optimal crossover point where the sum of the scores from both directions is maximum. This crossover point is used to divide the matrix into two smaller score matrices to which this decomposition strategy is recursively applied. The first application of this strategy yields two independent subproblems and following an application of the strategy to each of these subproblems, we have 4 even smaller subproblems. Following k rounds, we have 2^k independent subproblems.

For the parallel version of alignment algorithms, we employ the following strategies:

- When the number of independent matrices is small, each matrix is computed using the parallel version of score algorithms PP_Score, $PP\_Diagonal$ and PP_Strip; where p processors are assigned to the parallel computation. In other words, the matrices are computed in sequence.

- When the number of independent matrices is large, each matrix is computed using the single-core algorithms *Score*, *Diagonal* and *Strip*. Now, p matrices are concurrently computed.

Let PP_MM, $PP\_MMDiagonal$ and $PP\_MMStrip$, respectively, denote the parallel versions of MM, $MMDiagonal$ and $MMStrip$.

4 Results

4.1 Experimental Settings and Test Data

We implemented the single-core scoring and alignment algorithms in C and the multi-core scoring and alignment algorithms in C and OpenMP. The relative performance of these algorithms was measured on the following platforms:

1. Intel Xeon CPU E5-2603 v2 Quad-Core processor 1.8GHz with 10MB cache.

2. Intel I7-x980 Six-Core processor 3.33GHz with 12MB LLC cache.

3. Intel Xeon CPU E5-2695 v2 2xTwelve-Core processors 2.40GHz with 30MB cache.

For convenience, we will, at times, refer to these platforms as Xeon4, Xeon6, and Xeon24 (i.e., the number of cores is appended to the name Xeon).

All codes were compiled using the gcc compiler with the O2 option. On our Xeon4 platform, we used the “perf” [21] software to measure energy usage through the RAPL interface. So, for this platform, we report cache misses and energy consumption as well as running time. For the Xeon6 and Xeon24 platforms, we provide the running time only.

For test data, we used randomly generated protein sequences as well as real protein sequences obtained from the Globin Gene Server[22] and DNA/RNA/protein sequences from the National Center for Biotechnology Information (NCBI) database [23]. We used the BLOSUM62[1] scoring matrix for all our experiments. The results for our randomly generated protein sequences were comparable to those for similarly sized sequences used from the two databases [22] and [23]. So, we present only the results for the latter data sets here.

4.2 Xeon E5-2603 (Xeon4)

4.2.1 Score Algorithms

Figure 7 and Table 1 give the number of cache misses on our Xeon4 platform for different sequence sizes. The last two columns of Table 1 gives the percent reduction in the observed cache miss count of *Strip* relative to *Score* and *Diagonal*. *Strip* has the fewest cache misses followed by *Score* and *Diagonal* (in this order). *Strip* reduces cache misses by up to 86.2% relative to *Score* and by up to 92.3% relative to *Diagonal*.

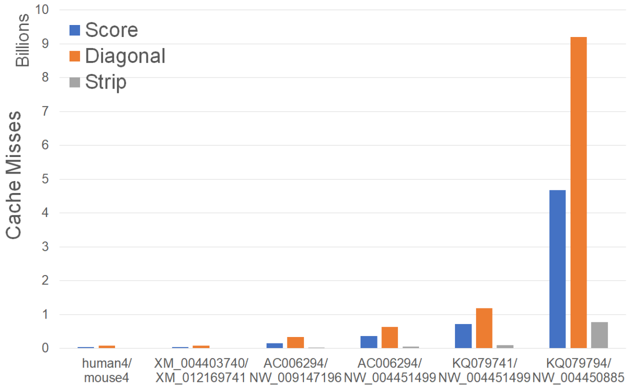


Figure 7: Cache misses of scoring algorithms, in billions, on Xeon4.

Figure 8 and Table 2 give the running times of our scoring algorithms on our Xeon4 platform. In the figure, the time is in seconds while in the table, the time is given using the format $hh:mm:ss$. The table also gives the percent reduction in running time achieved by *Strip* relative to *Score* and *Diagonal*.

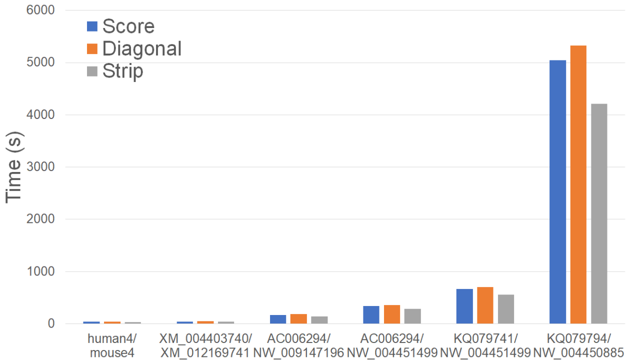


Figure 8: Run time of scoring algorithms, in seconds, on Xeon4.

As can be seen, on our Xeon4 platform, *Strip* is the fastest followed by *Score* and *Diagonal* (in this order). *Strip* reduces the running time by up to 17.5% relative to *Score* and by up to 22.8% relative to *Diagonal*. The reduction in running time, while significant, isn't as much as the reduction in cache misses possibly due to the effect of cache prefetching, which reduces cache induced computational delays.

Figures 9 and Tables 3 give the CPU and cache energy consumed, in joules, by our Xeon4 platform.

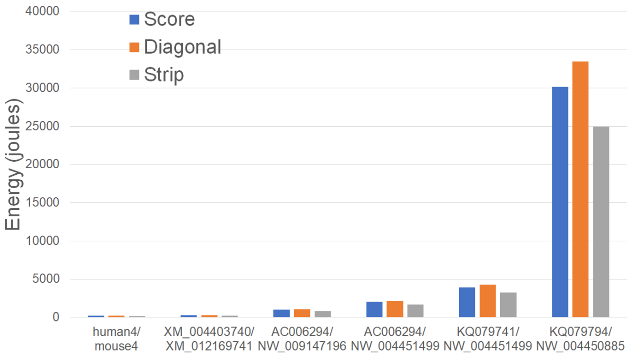


Figure 9: CPU and cache energy consumption of scoring algorithms, in joules, on Xeon4.

On our datasets, *Strip* required up to 18.5% less CPU and cache energy than *Score* and up to 25.5% less than *Diagonal*. It is interesting to note that the energy reduction is comparable to the reduction in running time suggesting a close relationship between running time and energy consumption for this application.

4.2.2 Parallel Scoring Algorithms

Figure 10 and Table 4 give the number of cache misses on our Xeon4 platform for our parallel scoring algorithms. *PP_Strip* has the fewest cache misses followed by *PP_Score* and *PP_Diagonal* (in this order). *PP_Strip* reduces cache misses by up to 98.1% relative to *PP_Score* and by up to 99.1% relative to *PP_Diagonal*. We observe also that the total cache misses for *PP_Score* is slightly higher than for *Score* for smaller instances and lower for larger instances. *PP_Diagonal*, on the other hand, consistently has more cache misses than *Diagonal*. *PP_Strip* exhibits a significant reduction in cache misses. This is because we chose the strip width to be such that p strip rows fit in this cache. Most of the cache misses in the *Strip* are from the vector that transfers boundary results from one strip to the next. When p strips are being worked on simultaneously, the inter-strip data that is to be transferred is often in the cache and so many of the cache misses incurred by the single-core algorithm are saved. The remaining two algorithms do not allow this flexibility in choosing the segment size a processor works on; this size is fixed at $O(n/p)$.

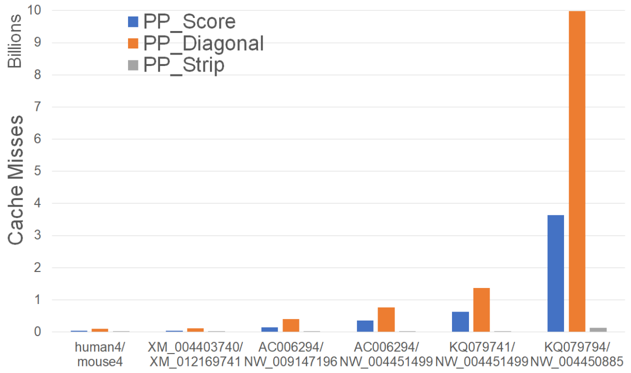


Figure 10: Cache misses of parallel scoring algorithms, in billions, on Xeon4.

Table 1: Cache misses of scoring algorithms, in millions, on Xeon4.

A	|A|	B	|B|	Score	Diagonal	Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	33	74	6	82.1%	92.1%
XM_004403740	104,267	XM_012169741	103,004	38	86	7	81.8%	91.9%
AC006294	200,000	NW_009147196	200,000	144	342	26	81.9%	92.3%
AC006294	200,000	NW_004451499	398,273	363	630	52	85.6%	91.7%
KQ079741	392,981	NW_004451499	398,273	712	1,190	98	86.2%	91.7%
KQ079794	1,083,068	NW_004450885	1,098,196	4,676	9,200	772	83.5%	91.6%

Table 2: Run time of scoring algorithms, in hh:mm:ss, on Xeon4.

A	|A|	B	|B|	Score	Diagonal	Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:39	0:00:42	0:00:33	16.2%	21.6%
XM_004403740	104,267	XM_012169741	103,004	0:00:45	0:00:49	0:00:38	16.2%	22.8%
AC006294	200,000	NW_009147196	200,000	0:02:49	0:03:02	0:02:22	16.2%	22.1%
AC006294	200,000	NW_004451499	398,273	0:05:42	0:05:58	0:04:42	17.5%	21.3%
KQ079741	392,981	NW_004451499	398,273	0:11:03	0:11:44	0:09:14	16.4%	21.4%
KQ079794	1,083,068	NW_004450885	1,098,196	1:24:06	1:28:45	1:10:10	16.6%	20.9%

Figure 11 and Table 5 give the running times for our parallel scoring algorithms on our Xeon4 platform. In the figure, the time is in seconds while in the table, the time is given using the format $hh : mm : ss$. As in the table, PP_Strip is the fastest algorithm in practice, which is up to 40.0% faster than PP_Score and up to 38.4% faster than $PP\_Diagonal$.

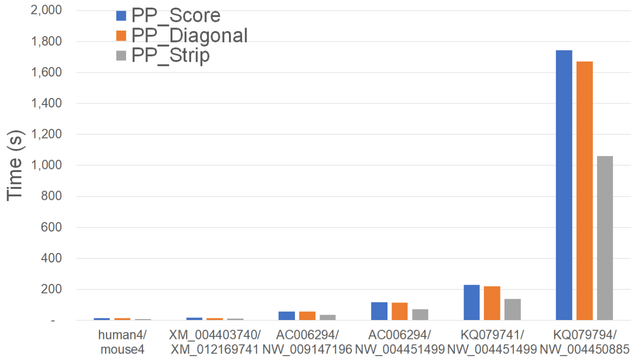


Figure 11: Run time of parallel scoring algorithms, in seconds, on Xeon4.

Table 6 gives the speedup of each of our parallel scoring algorithms relative to their sequential counterparts. As can be seen, the speedup of PP_Strip (i.e., $Strip/PP\_Strip$) is between 3.92 and 3.98, which is quite close to the number of cores (4) on our Xeon4 platform. PP_Score achieves a speedup in the range 2.82 to 2.94 and the speedup for $PP\_Diagonal$ is in the range 3.12 to 3.21.

The excellent speedup exhibited by PP_Strip is due largely to our ability to greatly reduce cache misses for this algorithm.

Figures 12 and Tables 7 give the CPU and cache energy consumed, in joules, by our Xeon4 platform. On our datasets, PP_Strip required up to 41.2% less CPU and cache energy than PP_Score and up to 45.5% less than $PP\_Diagonal$.

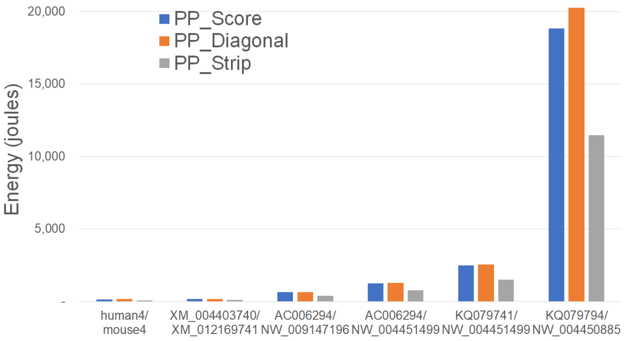


Figure 12: CPU and cache energy consumption of parallel scoring algorithms, in joules, on Xeon4.

Compared to the sequential scoring algorithms, the multi-core algorithms use higher CPU power but less running time. Since the power increase is less than the decrease in running time, energy consumption is reduced.

4.2.3 Alignment Algorithms

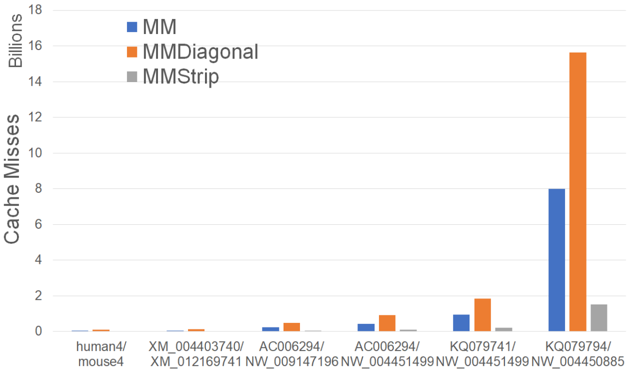


Figure 13: Cache misses for alignment algorithms, in billions, on Xeon4.

Figure 13 and Table 8 give the number of cache misses of our single-core alignment algorithms on our Xeon4 plat-

Table 3: CPU and cache energy consumption of scoring algorithms, in joules, on Xeon4.

A	|A|	B	|B|	Score	Diagonal	Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	230.67	252.29	190.6	17.4%	24.5%
XM_004403740	104,267	XM_012169741	103,004	268.29	297.28	221.49	17.4%	25.5%
AC006294	200,000	NW_009147196	200,000	997.88	1100.6	829.36	16.9%	24.6%
AC006294	200,000	NW_004451499	398,273	2026.26	2178.97	1651.46	18.5%	24.2%
KQ079741	392,981	NW_004451499	398,273	3944.3	4300.59	3253.46	17.5%	24.3%
KQ079794	1,083,068	NW_004450885	1,098,196	30125.93	33472.81	24980.35	17.1%	25.4%

Table 4: Cache misses of parallel scoring algorithms, in millions, on Xeon4.

A	|A|	B	|B|	PP_Score	PP_Diagonal	PP_Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	34	102	1	96.7%	98.9%
XM_004403740	104,267	XM_012169741	103,004	39	115	1	96.8%	98.9%
AC006294	200,000	NW_009147196	200,000	146	398	4	96.9%	98.9%
AC006294	200,000	NW_004451499	398,273	362	768	7	98.1%	99.1%
KQ079741	392,981	NW_004451499	398,273	628	1,373	14	97.7%	99.0%
KQ079794	1,083,068	NW_004450885	1,098,196	3,642	9,976	121	96.7%	98.8%

form. $MMStrip$ has the fewest number of cache misses followed by MM and $MMDiagonal$ (in this order). $MMStrip$ reduces cache misses by up to 81.0% relative to MM and by up to 90.3% relative to $MMDiagonal$.

Figure 14 and Table 9 give the running times of our single-core alignment algorithms on our Xeon4 platform. As can be seen, $MMStrip$ is the fastest followed by MM and $MMDiagonal$ (in this order). $MMStrip$ reduces running time by up to 15.0% relative to MM, by up to 13.4% relative to $MMDiagonal$. As was the case with our scoring algorithms, the reduction in running time, while significant, isn't as much as the reduction in cache misses.

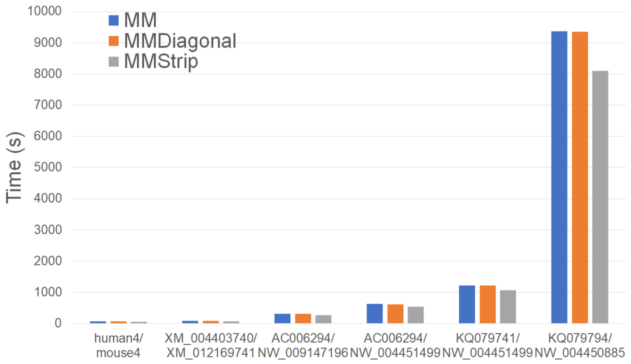


Figure 14: Run time of alignment algorithms, in Seconds, on Xeon4.

Figures 15 and Tables 10 give the CPU and cache energy consumption, in joules, by our single-core alignment algorithms. On our datasets, $MMStrip$ reduced up to 17.5% less CPU and cache energy than MM and up to 18.7% less than $MMDiagonal$. Once again, the energy reduction is comparable to the reduction in running time suggesting a close relationship between running time and energy consumption for this application.

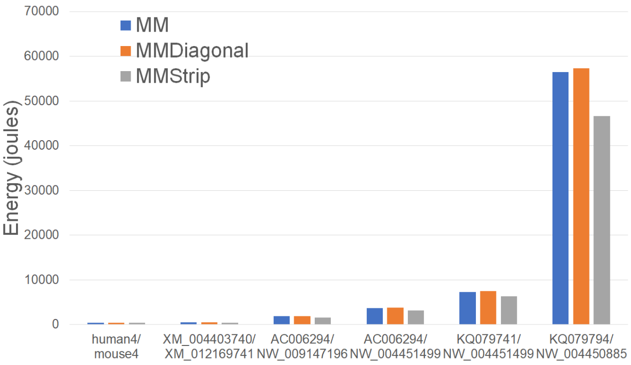


Figure 15: CPU and cache energy consumption of alignment algorithms, in joules, on Xeon4.

4.2.4 Parallel Alignment Algorithms

Figure 16 and Table 11 give the number of cache misses of our multi-core alignment algorithms on our Xeon4 platform.

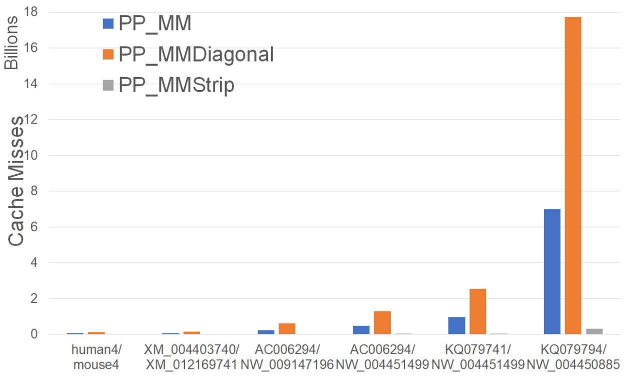


Figure 16: Cache misses for parallel alignment algorithms, in billions, on Xeon4.

$PP\_MMStrip$ has the fewest number of cache misses followed by PP_MM and $PP\_MMDiagonal$ (in this order). $PP\_MMStrip$ reduces cache misses by

Table 5: Run time of parallel scoring algorithms on Xeon4.

A	|A|	B	|B|	PP_Score	PP_Diagonal	PP_Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:14	0:00:13	0:00:08	40.0%	37.7%
XM_004403740	104,267	XM_012169741	103,004	0:00:16	0:00:15	0:00:10	39.8%	37.1%
AC006294	200,000	NW_009147196	200,000	0:00:58	0:00:58	0:00:36	39.0%	38.4%
AC006294	200,000	NW_004451499	398,273	0:01:56	0:01:53	0:01:11	39.0%	37.0%
KQ079741	392,981	NW_004451499	398,273	0:03:49	0:03:39	0:02:19	39.3%	36.6%
KQ079794	1,083,068	NW_004450885	1,098,196	0:29:02	0:27:52	0:17:39	39.2%	36.7%

Table 6: Speedup of parallel scoring algorithms on Xeon4.

A	|A|	B	|B|	Score/PP	Diagonal/PP	Strip/PP
human4	97,634	mouse4	94,647	2.82	3.12	3.93
XM_004403740	104,267	XM_012169741	103,004	2.82	3.19	3.92
AC006294	200,000	NW_009147196	200,000	2.89	3.14	3.97
AC006294	200,000	NW_004451499	398,273	2.94	3.18	3.97
KQ079741	392,981	NW_004451499	398,273	2.89	3.21	3.98
KQ079794	1,083,068	NW_004450885	1,098,196	2.90	3.18	3.98

up to 95.5% relative to PP_MM and by up to 98.2% relative to $PP\_MMDiagonal$.

Figure 17 and Table 12 give the running times for our parallel alignment algorithms on the Xeon4 platform. $PP\_MMStrip$ is faster than PP_MM by up to 37.4% and faster than $PP\_MMDiagonal$ by up to 40.3%.

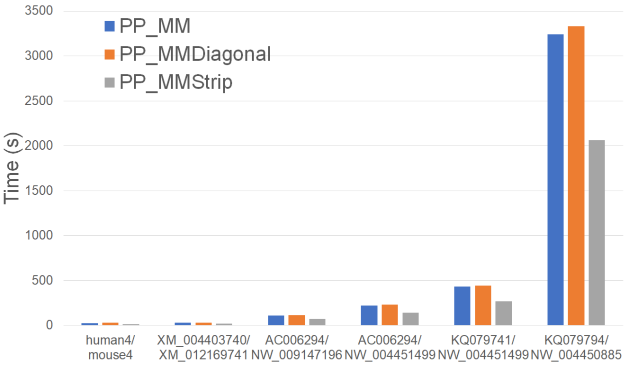


Figure 17: Run time of parallel alignment algorithms, in seconds, on Xeon4.

Table 13 gives the speedup of each parallel alignment algorithm relative to its single-core counterpart. The speedup achieved by $PP\_MMStrip$ (relative to $MMStrip$) ranges from 3.56 to 3.94 while that for PP_MM is in the range 2.77 to 2.88 and that for $PP\_MMDiagonal$ is in the range 2.53 to 2.81.

Figures 18 and Tables 14 give the CPU and cache energy consumption, in joules, by our multi-core alignment algorithms. On our datasets, $PP\_MMStrip$ required up to 29.9% less CPU and cache energy than PP_MM and up to 42.1% less than $PP\_MMDiagonal$. Once again, the energy reduction is comparable to the reduction in running time suggesting a close relationship between running time and energy consumption for this application.

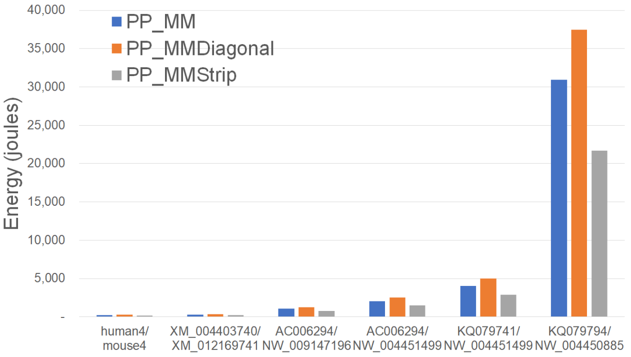


Figure 18: CPU and cache energy consumption of parallel alignment algorithms, in joules, on Xeon4.

4.3 I7-x980 (Xeon6)

4.3.1 Scoring Algorithms

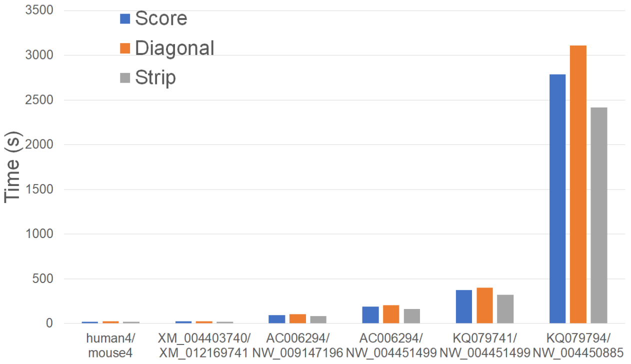


Figure 19: Run time of scoring algorithms, in seconds, on Xeon6

Figure 19 and Table 15 give the running times of our single-core scoring algorithms on our Xeon6 platform. As can be seen, *Strip* is the fastest followed by *Score* and *Diagonal* (in this order). *Strip* reduces running time by up to 14.3% relative to *Score* and by up to 22.4% relative to *Diagonal*.

Table 7: CPU and cache energy consumption of parallel scoring algorithms on Xeon4.

A	|A|	B	|B|	PP_Score	PP_Diagonal	PP_Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	150.12	161.77	88.23	41.2%	45.5%
XM_004403740	104,267	XM_012169741	103,004	174.30	175.68	102.76	41.0%	41.5%
AC006294	200,000	NW_009147196	200,000	631.46	661.29	381.02	39.7%	42.4%
AC006294	200,000	NW_004451499	398,273	1,256.11	1,294.55	760.57	39.5%	41.2%
KQ079741	392,981	NW_004451499	398,273	2,471.51	2,535.50	1,493.38	39.6%	41.1%
KQ079794	1,083,068	NW_004450885	1,098,196	18,830.03	20,253.39	11,477.30	39.0%	43.3%

Table 8: Cache misses for alignment algorithms, in millions, on Xeon4.

A	|A|	B	|B|	MM	MMDiagonal	MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	42	96	12	71.6%	87.6%
XM_004403740	104,267	XM_012169741	103,004	54	119	14	74.5%	88.5%
AC006294	200,000	NW_009147196	200,000	227	488	48	78.6%	90.1%
AC006294	200,000	NW_004451499	398,273	441	926	95	78.4%	89.7%
KQ079741	392,981	NW_004451499	398,273	959	1,851	200	79.1%	89.2%
KQ079794	1,083,068	NW_004450885	1,098,196	7,994	15,642	1,521	81.0%	90.3%

4.3.2 Parallel Scoring Algorithms

Figure 20 and Table 16 give the running times for our parallel scoring algorithms on our Xeon6 platform. As with *Xeon*4, *PP_Strip* is faster than *PP_Score* and *PP_Diagonal* and reduces the running time by up to 42.5% and 55.6%, respectively.

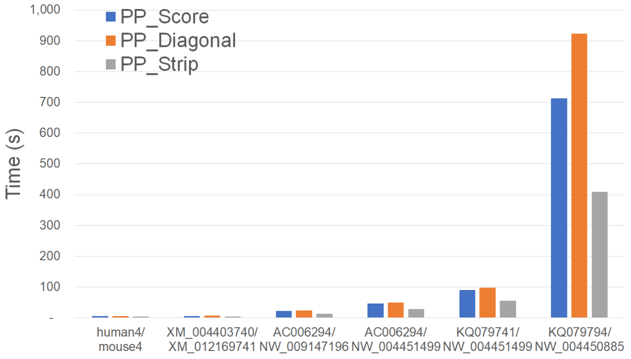


Figure 20: Run time of parallel scoring algorithms, in seconds, on Xeon6.

Table 18 gives the speedup of each of our parallel algorithms relative to their single-core counterparts. *PP_Strip* achieves a speedup of up to 5.89, which is very close to the number of cores. The maximum speedup achieved by *PP_Score* and *PP_Diagonal* was 4.09 and 4.25, respectively.

4.3.3 Alignment Algorithms

Figure 21 and Table 17 give the running times of our parallel scoring algorithms on the Xeon6 platform. As can be seen, *MMStrip* is the fastest followed by *MM* and *MMDiagonal* (in this order). *MMStrip* reduces running time by up to 12.6% relative to *MM* and by up to 14.2% relative to *MMDiagonal*.

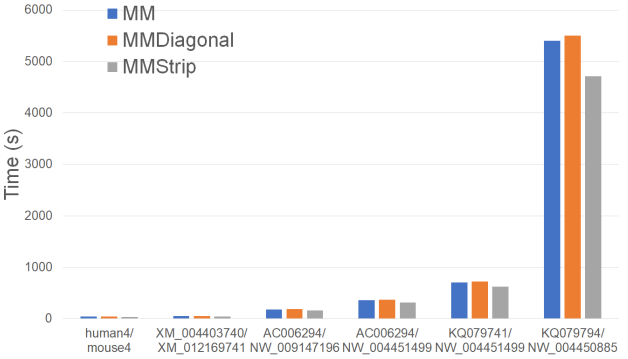


Figure 21: Run time of alignment algorithms, in seconds, on Xeon6.

4.3.4 Parallel Alignment Algorithms

Figure 22 and Table 20 give the running times of our parallel alignment algorithms on the Xeon6. *PP_MMStrip* is faster than *PP_MM* and *PP_MMDiagonal* and reduces the running time by up to 39.9% and 44.8%, respectively.

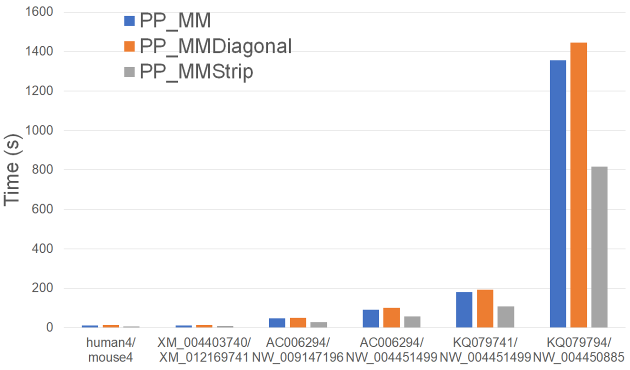


Figure 22: Run time of parallel alignment algorithms, in seconds, on Xeon6.

Table 21 gives the speedup of each of our parallel algorithms relative to their single-core counter-

Table 9: Run time of alignment algorithms on Xeon4.

A	|A|	B	|B|	MM	MMDiagonal	MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:01:12	0:01:13	0:01:03	12.6%	12.7%
XM_004403740	104,267	XM_012169741	103,004	0:01:27	0:01:24	0:01:14	15.0%	12.8%
AC006294	200,000	NW_009147196	200,000	0:05:16	0:05:14	0:04:33	13.6%	13.0%
AC006294	200,000	NW_004451499	398,273	0:10:25	0:10:18	0:09:04	13.0%	12.0%
KQ079741	392,981	NW_004451499	398,273	0:20:27	0:20:28	0:17:47	13.0%	13.1%
KQ079794	1,083,068	NW_004450885	1,098,196	2:36:12	2:35:53	2:14:59	13.6%	13.4%

Table 10: CPU and cache energy consumption of alignment algorithms on Xeon4.

A	|A|	B	|B|	MM	MMDiagonal	MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	426.38	438.07	369.94	13.2%	15.6%
XM_004403740	104,267	XM_012169741	103,004	509.66	511.32	431.59	15.3%	15.6%
AC006294	200,000	NW_009147196	200,000	1873.99	1911.35	1600.77	14.6%	16.2%
AC006294	200,000	NW_004451499	398,273	3707.55	3772.38	3189.38	14.0%	15.5%
KQ079741	392,981	NW_004451499	398,273	7313.49	7512.26	6278.63	14.2%	16.4%
KQ079794	1,083,068	NW_004450885	1,098,196	56478.59	57339.29	46589.55	17.5%	18.7%

parts. *PP_MMStrip* achieves a speedup of up to 5.78, which is very close to the number of cores. The maximum speedup achieved by *PP_MM* and *PP_MMDiagonal* was 3.98 and 3.80, respectively.

4.4 Xeon E5-2695 (Xeon24)

4.4.1 Scoring Algorithms

Figure 23 and Table 19 give the running times of our single-core scoring algorithms on our Xeon24 platform. As was the case on our other test platforms, *Strip* is the fastest followed by *Score* and *Diagonal* (in this order). *Strip* reduces running time by up to 19.7% relative to *Score* and by up to 35.1% relative to *Diagonal*.

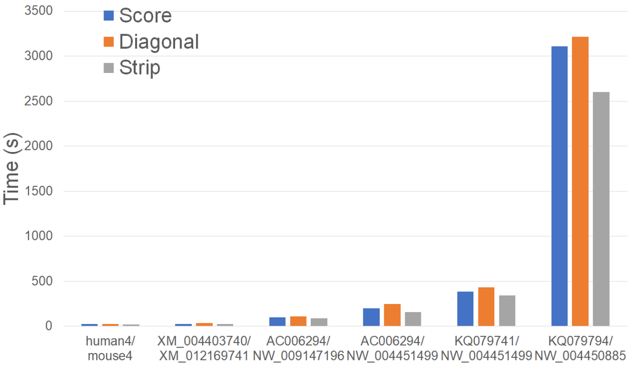


Figure 23: Run time of scoring algorithms, in seconds, on Xeon24.

4.4.2 Parallel Scoring Algorithms

Figure 24 and Table 22 give the running times for our parallel scoring algorithms on our Xeon24 platform. *PP_Strip* is faster than *PP_Score* and *PP_Diagonal* and reduces the running time by up to 61.4% and 76.2%, respectively.

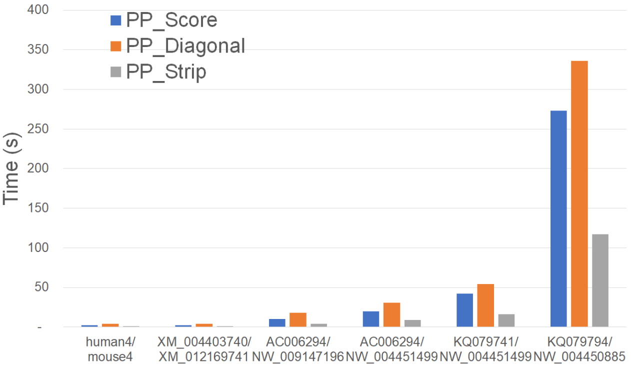


Figure 24: Run time of parallel scoring algorithms, in seconds, on Xeon24.

Table 23 gives the achieved speedup. *PP_Strip* scales quite well and results in a speedup of up to 22.22. The maximum speedups provided by *PP_Score* and *PP_Diagonal* are 11.36 and 9.56, respectively.

4.4.3 Alignment Algorithms

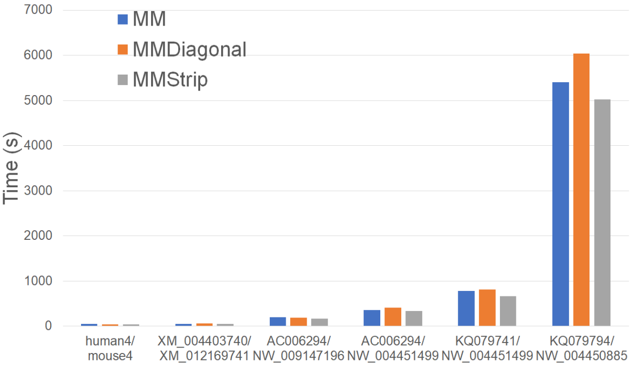


Figure 25: Run time of alignment algorithms, in seconds, on Xeon24.

Figure 25 and Table 24 give the running times of our single-core alignment algorithms on our Xeon24 plat-

Table 11: Cache misses for parallel alignment algorithms, in millions, on Xeon4.

A	|A|	B	|B|	PP_MM	PP_MMDiagonal	PP_MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	64	137	9	85.3%	93.1%
XM_004403740	104,267	XM_012169741	103,004	71	161	17	75.8%	89.4%
AC006294	200,000	NW_009147196	200,000	251	625	23	91.0%	96.4%
AC006294	200,000	NW_004451499	398,273	495	1,296	39	92.2%	97.0%
KQ079741	392,981	NW_004451499	398,273	975	2,558	52	94.6%	98.0%
KQ079794	1,083,068	NW_004450885	1,098,196	7,001	17,721	314	95.5%	98.2%

Table 12: Run time of parallel alignment algorithms on Xeon4.

A	|A|	B	|B|	PP_MM	PP_MMDiagonal	PP_MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:26	0:00:29	0:00:17	34.5%	40.3%
XM_004403740	104,267	XM_012169741	103,004	0:00:30	0:00:33	0:00:21	31.6%	37.3%
AC006294	200,000	NW_009147196	200,000	0:01:52	0:01:56	0:01:12	36.1%	38.5%
AC006294	200,000	NW_004451499	398,273	0:03:43	0:03:50	0:02:20	37.4%	39.1%
KQ079741	392,981	NW_004451499	398,273	0:07:12	0:07:23	0:04:31	37.2%	38.8%
KQ079794	1,083,068	NW_004450885	1,098,196	0:54:03	0:55:33	0:34:20	36.5%	38.2%

form. *MMStrip* is the fastest followed by *MM* and *MMDiagonal* (in this order). *MMStrip* reduces running time by up to 17.1% relative to *MM* and by up to 16.8% relative to *MMDiagonal*.

4.4.4 Parallel Alignment Algorithms

Figure 26 and Table 25 give the running times of our parallel alignment algorithms on Xeon24. As can be seen, *PP_MMStrip* is faster than *PP_MM* and *PP_MMDiagonal*. It reduces the running time by up to 47.3% and 84.6%, respectively.

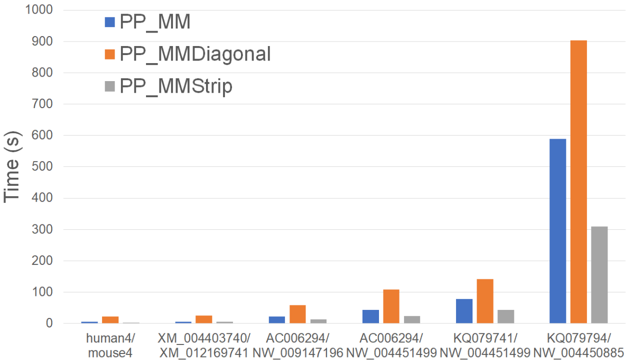


Figure 26: Run time of parallel alignment algorithms, in seconds, on Xeon24.

Table 26 gives the speedup of our parallel algorithms. *PP_MMStrip* achieves a speedup of up to 16.2 while *PP_MM* and *PP_MMDiagonal* have maximum speedups of 9.79 and 6.58.

5 Discussion

By accounting for the presence of caches in modern computers, we are able to arrive at sequence alignment algorithms that are considerably faster than those that do not take advantage of computer caches. Our benchmarking demonstrates the value of optimizing cache usage. Our cache-efficient algorithms *Strip* and *MMStrip* were the best-performing single-core algorithms and their parallel counterparts were the best-performing parallel algorithms. *Strip* reduced running time by as much as 19.7% relative to the classical scoring algorithm *Score* due to Smith and Waterman and *MM_Strip* reduced running time by as much as 17.1% relative to the alignment algorithm of Myers and Miller. Neither the algorithm of Smith and Waterman nor that of Myers and Miller optimize cache utilization. The parallel versions of *Strip* and *MM_Strip* were up to 61.4% and 47.3% faster than the parallel versions of the Smith and Waterman and the Myers and Miller algorithms, respectively.

6 Limitations

Our cache miss analyses assume a simple cache model in which there is a single LRU cache. In practice, computers have multiple levels of cache and employ sophisticated and proprietary cache replacement strategies. Despite the use of a simplified cache model for analysis, the developed cache-efficient algorithms perform very well in practice.

7 Conclusion

The main contributions of this papers are

1. cache efficient single-core and multi-core algorithms to determine the score of the best alignment;
2. cache efficient single-core and multi-core algorithms to determine the best alignment.

The effectiveness of our cache-efficient algorithms has been demonstrated experimentally using three computational platforms. Future work includes developing the cache-efficient algorithms for other problems in computational biology.

Conflict of Interest The authors declare no conflict of interest.

Table 13: Speedup of parallel alignment algorithms on Xeon4.

A	|A|	B	|B|	MM/PP	MMDiagonal/PP	MMStrip/PP
human4	97,634	mouse4	94,647	2.77	2.53	3.70
XM_004403740	104,267	XM_012169741	103,004	2.78	2.56	3.56
AC006294	200,000	NW_009147196	200,000	2.80	2.70	3.81
AC006294	200,000	NW_004451499	398,273	2.84	2.69	3.89
KQ079741	392,981	NW_004451499	398,273	2.84	2.78	3.94
KQ079794	1,083,068	NW_004450885	1,098,196	2.88	2.81	3.93

Table 14: CPU and cache energy consumption of parallel alignment algorithms on Xeon4.

A	|A|	B	|B|	PP_MM	PP_MMDiagonal	PP_MMStrip	Imp1	Imp2
human4	97634	mouse4	94647	236.98	305.14	181.6	23.4%	40.5%
XM_004403740	104267	XM_012169741	103004	272.26	352.56	216.83	20.4%	38.5%
AC006294	200000	NW_009147196	200000	1044.86	1279.77	747.86	28.4%	41.6%
AC006294	200000	NW_004451499	398273	2039.55	2540.52	1483.8	27.2%	41.6%
KQ079741	392981	NW_004451499	398273	4020.48	4979.17	2905.32	27.7%	41.7%
KQ079794	1083068	NW_004450885	1098196	30964.02	37461.35	21703.65	29.9%	42.1%

Acknowledgment This work was supported, in part, by the National Science Foundation under award NSF 1447711.

References

[1] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proc Natl Acad Sci U S A*, vol. 89, pp. 10 915–10 919, 1992.

[2] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, pp. 443–453, 1970.

[3] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, pp. 195–197, 1981.

[4] O. Gotoh, "An improved algorithm for matching biological sequences," *Journal of Molecular Biology*, vol. 162, pp. 705–708, 1982.

[5] D. S. Hirschberg, "A linear space algorithm for computing longest common subsequences," *Communications of the ACM*, vol. 18, pp. 341–343, 1975.

[6] E. Myers and W. Miller, "Optimal alignments in linear space," *Computer Applications in the Biosciences(CABIOS)*, vol. 4, pp. 11–17, 1988.

[7] X. Huang, R. Hardison, and W. Miller, "A space-efficient algorithm for local similarities," *Comput Appl Biosci*, vol. 6, p. 373âĂŞ381, 1990.

[8] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, pp. 403–410, 1990.

[9] W. Pearson and D. Lipman, "Improved tools for biological sequence comparison," *Proceedings of the National Academy of Sciences USA*, vol. 85, pp. 2444–2448, 1988.

[10] K. Chao, J. Zhang, J. Ostell, and W. Miller, "A local alignment tool for very long dna sequences," *Comput Appl Biosci*, vol. 11, pp. 147–153, 1995.

[11] A. Driga, P. Lu, J. Schaeffer, D. Szafron, K. Charter, and I. Parsons, "Fastlsa: a fast, linear-space, parallel and sequential algorithm for sequence alignment," *Algorithmica*, vol. 45, p. 337âĂŞ375, 2006.

[12] J. Li, S. Ranka, and S. Sahni, "Pairwise sequence alignment for very long sequences on gpus," *IEEE 2nd International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, 2012.

[13] E. O. Sandes and A. C. M. A. Melo, "Smith-waterman alignment of huge sequences with gpu in linear space," *IEEE International Symposium on Parallel and Distributed Processing (IPDPS)*, pp. 1199–1211, 2011.

[14] S. Aluru and N. Jammula, "A review of hardware acceleration for computational genomics," *IEEE Design and Test*, vol. 31, pp. 19–30, 2014.

[15] S. Rajko and S. Aluru, "Space and time optimal parallel sequence alignments," *IEEETPDS:IEEE Transactions on Parallel and Distributed Systems*, vol. 15, 2004.

[16] A. Khajeh-Saeed, S. Poole, and J. B. Perot, "Acceleration of the smithâĂŞwaterman algorithm using single and multiple graphics processors," *Journal of Computational Physics*, p. 4247âĂŞ4258, 2010.

[17] S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg, "Versatile and open software for comparing large genomes," *Genome Biol*, vol. 5, 2004.

[18] T. Almeida and N. Roma, "A parallel programming framework for multi-core dna sequence alignment," *Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on*, pp. 907 – 912, 2010.

[19] K. Hamidouche, F. M. Mendonca, J. Falcou, A. C. M. A. Melo, and D. Etiemble, "Parallel smith-waterman comparison on multicore and manycore computing platforms with bsp++," *International Journal of Parallel Programming*, vol. 41, pp. 1110–136, 2013.

[20] C. Zhao and S. Sahni, "Cache and energy efficient alignment of very long sequences," *2015 IEEE 5th international conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, 2015.

[21] "Perf tool," `https://perf.wiki.kernel.org/index.php/Main_Page`.

[22] "Globin gene server," `http://globin.cse.psu.edu/globin/html/pip/examples.html`.

[23] "Ncbi database," `http://www.ncbi.nlm.nih.gov/gquery`.

Table 15: Run time of scoring algorithms, in hh:mm:ss, on Xeon6.

A	|A|	B	|B|	Score	Diagonal	Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:21	0:00:24	0:00:19	13.2%	21.1%
XM_004403740	104,267	XM_012169741	103,004	0:00:25	0:00:27	0:00:22	13.2%	20.2%
AC006294	200,000	NW_009147196	200,000	0:01:33	0:01:43	0:01:21	13.2%	21.7%
AC006294	200,000	NW_004451499	398,273	0:03:08	0:03:25	0:02:41	14.3%	21.4%
KQ079741	392,981	NW_004451499	398,273	0:06:13	0:06:43	0:05:19	14.3%	20.8%
KQ079794	1,083,068	NW_004450885	1,098,196	0:46:27	0:51:51	0:40:16	13.3%	22.4%

Table 16: Run time of parallel scoring algorithms, in hh:mm:ss, on Xeon6.

A	|A|	B	|B|	PP_Score	PP_Diagonal	PP_Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:06	0:00:06	0:00:04	27.9%	30.4%
XM_004403740	104,267	XM_012169741	103,004	0:00:06	0:00:07	0:00:04	37.4%	40.6%
AC006294	200,000	NW_009147196	200,000	0:00:23	0:00:24	0:00:14	38.8%	41.2%
AC006294	200,000	NW_004451499	398,273	0:00:47	0:00:49	0:00:28	40.6%	43.0%
KQ079741	392,981	NW_004451499	398,273	0:01:31	0:01:38	0:00:55	39.0%	43.4%
KQ079794	1,083,068	NW_004450885	1,098,196	0:11:53	0:15:23	0:06:50	42.5%	55.6%

Table 17: Run time of alignment algorithms, in hh:mm:ss, on Xeon6.

A	|A|	B	|B|	MM	MMDiagonal	MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:41	0:00:42	0:00:37	11.2%	12.7%
XM_004403740	104,267	XM_012169741	103,004	0:00:48	0:00:49	0:00:43	11.3%	12.8%
AC006294	200,000	NW_009147196	200,000	0:03:00	0:03:05	0:02:39	11.7%	14.0%
AC006294	200,000	NW_004451499	398,273	0:05:59	0:06:06	0:05:16	11.8%	13.7%
KQ079741	392,981	NW_004451499	398,273	0:11:45	0:12:01	0:10:22	11.8%	13.7%
KQ079794	1,083,068	NW_004450885	1,098,196	1:29:59	1:31:40	1:18:37	12.6%	14.2%

Table 18: Speedup of parallel scoring algorithms on Xeon6.

A	|A|	B	|B|	Score/PP	Diagonal/PP	Strip/PP
human4	97,634	mouse4	94,647	3.91	4.15	4.70
XM_004403740	104,267	XM_012169741	103,004	3.99	4.12	5.54
AC006294	200,000	NW_009147196	200,000	3.99	4.25	5.66
AC006294	200,000	NW_004451499	398,273	3.98	4.17	5.74
KQ079741	392,981	NW_004451499	398,273	4.09	4.11	5.76
KQ079794	1,083,068	NW_004450885	1,098,196	3.91	3.37	5.89

Table 19: Run time of scoring algorithms, in hh:mm:ss, on Xeon24.

A	|A|	B	|B|	Score	Diagonal	Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:23	0:00:26	0:00:20	12.1%	20.9%
XM_004403740	104,267	XM_012169741	103,004	0:00:26	0:00:35	0:00:24	8.9%	32.9%
AC006294	200,000	NW_009147196	200,000	0:01:41	0:01:51	0:01:28	13.5%	21.2%
AC006294	200,000	NW_004451499	398,273	0:03:18	0:04:05	0:02:39	19.7%	35.1%
KQ079741	392,981	NW_004451499	398,273	0:06:23	0:07:14	0:05:44	10.2%	20.7%
KQ079794	1,083,068	NW_004450885	1,098,196	0:51:48	0:53:34	0:43:19	16.4%	19.1%

Table 20: Run time of parallel alignment algorithms on Xeon6.

A	|A|	B	|B|	PP_MM	PP_MMDiagonal	PP_MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:11	0:00:13	0:00:07	33.6%	44.8%
XM_004403740	104,267	XM_012169741	103,004	0:00:12	0:00:15	0:00:08	33.4%	44.2%
AC006294	200,000	NW_009147196	200,000	0:00:48	0:00:51	0:00:29	39.2%	42.9%
AC006294	200,000	NW_004451499	398,273	0:01:32	0:01:42	0:00:58	37.5%	43.7%
KQ079741	392,981	NW_004451499	398,273	0:03:01	0:03:14	0:01:49	39.7%	43.8%
KQ079794	1,083,068	NW_004450885	1,098,196	0:22:37	0:24:06	0:13:36	39.9%	43.6%

Table 21: Speedup of parallel alignment algorithms on Xeon6.

A	|A|	B	|B|	MM/PP	MMDiagonal/PP	MMStrip/PP
human4	97,634	mouse4	94,647	3.81	3.22	5.10
XM_004403740	104,267	XM_012169741	103,004	3.88	3.31	5.17
AC006294	200,000	NW_009147196	200,000	3.72	3.59	5.41
AC006294	200,000	NW_004451499	398,273	3.88	3.58	5.48
KQ079741	392,981	NW_004451499	398,273	3.90	3.71	5.70
KQ079794	1,083,068	NW_004450885	1,098,196	3.98	3.80	5.78

Table 22: Run time of parallel scoring algorithms on Xeon24.

A	|A|	B	|B|	PP_Score	PP_Diagonal	PP_Strip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:02	0:00:04	0:00:01	52.7%	72.8%
XM_004403740	104,267	XM_012169741	103,004	0:00:02	0:00:04	0:00:01	46.3%	68.6%
AC006294	200,000	NW_009147196	200,000	0:00:10	0:00:18	0:00:04	56.5%	76.2%
AC006294	200,000	NW_004451499	398,273	0:00:20	0:00:31	0:00:09	56.2%	72.0%
KQ079741	392,981	NW_004451499	398,273	0:00:42	0:00:54	0:00:16	61.4%	69.8%
KQ079794	1,083,068	NW_004450885	1,098,196	0:04:33	0:05:36	0:01:57	57.2%	65.2%

Table 23: Speedup of parallel alignment algorithms on Xeon24.

A	|A|	B	|B|	Score/PP	Diagonal/PP	Strip/PP
human4	97,634	mouse4	94,647	10.17	6.50	18.90
XM_004403740	104,267	XM_012169741	103,004	10.63	8.45	18.05
AC006294	200,000	NW_009147196	200,000	10.03	6.03	19.94
AC006294	200,000	NW_004451499	398,273	9.98	7.90	18.32
KQ079741	392,981	NW_004451499	398,273	9.07	8.04	21.10
KQ079794	1,083,068	NW_004450885	1,098,196	11.36	9.56	22.22

Table 24: Run time of alignment algorithms, in hh:mm:ss, on Xeon24.

A	|A|	B	|B|	MM	MMDiagonal	MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:46	0:00:44	0:00:40	11.9%	7.6%
XM_004403740	104,267	XM_012169741	103,004	0:00:50	0:00:56	0:00:46	7.4%	16.7%
AC006294	200,000	NW_009147196	200,000	0:03:19	0:03:09	0:02:45	17.1%	12.5%
AC006294	200,000	NW_004451499	398,273	0:06:01	0:06:46	0:05:41	5.8%	16.2%
KQ079741	392,981	NW_004451499	398,273	0:12:56	0:13:30	0:11:09	13.8%	17.4%
KQ079794	1,083,068	NW_004450885	1,098,196	1:30:06	1:40:36	1:23:45	7.1%	16.8%

Table 25: Run time of parallel alignment algorithms on Xeon24.

A	|A|	B	|B|	PP_MM	PP_MMDiagonal	PP_MMStrip	Imp1	Imp2
human4	97,634	mouse4	94,647	0:00:05	0:00:22	0:00:03	28.9%	84.6%
XM_004403740	104,267	XM_012169741	103,004	0:00:06	0:00:25	0:00:05	29.4%	81.9%
AC006294	200,000	NW_009147196	200,000	0:00:23	0:00:58	0:00:14	41.3%	76.3%
AC006294	200,000	NW_004451499	398,273	0:00:43	0:01:48	0:00:24	43.6%	77.6%
KQ079741	392,981	NW_004451499	398,273	0:01:19	0:02:22	0:00:44	44.4%	69.0%
KQ079794	1,083,068	NW_004450885	1,098,196	0:09:49	0:15:04	0:05:10	47.3%	65.7%

Table 26: Speedup of alignment algorithms on Xeon24.

A	|A|	B	|B|	MM/PP	MMDiagonal/PP	MMStrip/PP
human4	97,634	mouse4	94,647	9.68	1.99	12.00
XM_004403740	104,267	XM_012169741	103,004	7.81	2.23	10.24
AC006294	200,000	NW_009147196	200,000	8.56	3.27	12.08
AC006294	200,000	NW_004451499	398,273	8.39	3.75	14.02
KQ079741	392,981	NW_004451499	398,273	9.79	5.70	15.17
KQ079794	1,083,068	NW_004450885	1,098,196	9.18	6.68	16.20

Advances in Science, Technology and Engineering Systems Journal Vol. 3, No. 2, 345-355 (2018)

www.astesj.com

Special Issue on Advancement in Engineering Technology

ASTESJ

ISSN: 2415-6698

Direct Torque Control Strategy Based on the Emulation of Six-Switch Inverter Operation by a Four-Switch Inverter Using an Adaptive Fuzzy Controller

Salma Charmi*, Bassem El Badsi, Abderrazak Yangui

University of Sfax, Electrical Department, Sfax Engineering National School, P.O. Box 1173, 3038 Sfax, Tunisia

ARTICLE INFO

Article history:
Received: 06 March, 2018
Accepted: 15 April, 2018
Online: 30 April, 2018

Keywords:
Interior permanent magnet synchronous machine
Unbalanced voltage vectors
Balanced voltage vectors
Four-switch three-phase inverter
Six-switch three-phase inverter
Fuzzy logic control

ABSTRACT

This paper presents a novel direct torque control (DTC) strategy aimed to four-switch three-phase (FSTP) inverter-fed an interior permanent magnet synchronous machine (IPMSM), using a fuzzy logic toolbox in speed control loop. In fact, the introduced DTC approach is based on the emulation of the operation of the standard six-switch three-phase (SSTP) inverter. This fact has been produced thanks to suitable combinations of four unbalanced voltage vectors intrinsically generated by the FSTPI, leading to the synthesis of six balanced voltage vectors yielded by the SSTPI. It has been found from the simulation results that the adaptive fuzzy speed controller implemented for basic and proposed DTC strategies dedicated to FSTPI-fed an IPMSM drives, exhibits interesting performances over different operating conditions, more robustness and less steady-state error especially when there exist motor parameter uncertainties and unexpected load changes occur, compared to the ones yielded by the conventional proportional-integral controller.

1. Introduction

In recent years, direct torque control strategy is proposed by Takahashi and Depenbrock [1, 2] in the middle of 1980's. This strategy is increasingly applied for induction machines thanks to its several advantages such as (i) simple control scheme which makes it possible rapid real-time implementation, (ii) fast dynamic torque response and (iii) high robustness and stability against the load torque variations [3, 4] and reference mechanical speed changes. The presented strategy has been successively extended to different kinds of AC machines in various applications [5, 6], including variable reluctance machines [7] and permanent magnet synchronous machines [8], which is becoming popular for variable speed drive systems due to its high efficiency, high power factor, and more robustness. Since then, numerous several investigations carried out in order to improve the performance of the classical DTC strategy. The major focused features are the uncontrolled switching frequency of the inverter and the high torque ripple resulting from using of the stator flux and the electromagnetic torque hysteresis controllers. Commonly, the voltage source inverter (VSI) feeding an IPMSM drives under direct torque control approach is a conventional six-switch three phase inverter which is employed for high efficiency and performance operating of the motor drives. In contrast to this, for economic reasons, reducing the cost of the inverter topology with reduced number of inverter switching devices is still under investigation and has been suggested in [9, 10].

A DTC strategy dedicated to a FSTPI-fed induction machine (IM) drives has been proposed in [11]. In a FSTPI drive system, only two phases of the machine are controlled by power switching devices and the remaining phase is connected directly to the middle point of dc-bus voltage. The resulting modification reduces the number of power switches from six in a SSTPI to four, as in a FSTPI. For the proposed purpose, high-performance in terms of total harmonic distortion reduction allied to control of the inverter switching losses are proven.

Despite a lot of yielded high-performances and cost inverter reduction, the stator phase currents are unidirectional, and hence, this topology is limited to particular industrial applications in the field of motion control. On the other hand, the proposed strategy is penalized by the low dynamic and high torque ripple which is particularly caused by the application of unbalanced voltage vectors to control the stator flux and the torque with a subdivision of the $\alpha\beta$ plane limited to four sectors.

[*] Salma Charmi, Sfax Engineering National School, charmi.salma@yahoo.fr

https://dx.doi.org/10.25046/aj030237

An attempt to discard the previously described disadvantages has been proposed where the vector selection table for DTC of IPMSM driven by a FSTPI has been implemented by using of the new approach based on the principle of similarity between FSTPI and SSTPI [12]. The $\alpha\beta$ plane is traditionally divided into six sectors and the formation of the required reference space voltage vectors is done in the same way as for conventional SSTPI via employing of the effective vectors.

A further problem of this standard control strategy is that the proportional-integral (PI)-controller with constant parameters can't easily achieve swift response, small overshooting and fine speed control precision in a wide speed range, especially when there exist motor parameter uncertainties and unexpected load changes occur. More of the past research on variable speed IPMSM drives mainly concentrated on the development of the efficient control algorithms for high-performance drives.

To come up with above inherent drawbacks associated with PI-controller, some advanced techniques in artificial-intelligence based control such as: (i) fuzzy logic control (FLC), (ii) neural network control, (iii) sliding mode control, and (iv) robust control have been developed to achieve high-performance speed control of voltage source inverter feeding-IPMSM drives under direct torque control strategy.

In such a case, this paper proposes a new direct torque control strategy aimed to four-switch three-phase inverter-fed an interior permanent magnet synchronous machine drives, where the speed closed-loop regulator used an adaptive fuzzy logic controller. The introduced strategy is based on the emulation of the six-switch inverter operation owing to the synthesis of an appropriate vector selection table, which is traditionally addressed by two level hysteresis controllers for the stator flux and electromagnetic torque.

A complete comparative study between the conventional proportional-integral controller and the fuzzy logic controller is investigated, considering both transient and steady-state operations and different operating conditions. Basic principles of the introduced approaches are presented and some features illustrating the performance of the developed algorithm are verified and proven.

2. DTC Strategy of FSTPI-Fed an IPMSM Drives

2.1. Direct Torque Control Background

As well known that direct torque control strategy, whose overall block diagram is illustrated in Figure 1, basically consists to control directly and independently the stator flux linkage and electromagnetic torque through an appropriate selection of the inverter control signals, in order to fulfill the requirements as whether the controlled variables need to be increased, decreased, or maintained. In the standard version of DTC scheme, this trend is carried out in accordance with the output of hysteresis controllers of stator flux c_ϕ , c_τ of electromagnetic torque and θ_s angular displacement of the stator flux vector ϕ_s in the $\alpha\beta$ plane.

The dynamic of the stator flux vector is governed by the stator voltage equation expressed in the stationary reference frame as follows in (1):

$$\frac{d\phi_s}{dt} = V_{s-}r_s I_s \tag{1}$$

Where V_s , I_s and r_s are the stator voltage vector, stator current vector, and stator resistance, respectively.

Neglecting the voltage drop ($r_s I_s$) across the stator resistance and taking into account that the voltage vector is constant in each sampling period (T_s), the stator flux vector variation turns to be proportional to the applied voltage vector. The expression of ϕ_s turns to be as in (2):

$$\Delta\phi_s = \phi_s^{(k+1)} - \phi_s^{(k)} = V_s T_s \tag{2}$$

2.2. Conventional DTC Strategy of Four-Switch Inverter

As mentioned earlier, with using of a four-switch inverter system and as shown in Figure 1, two among the three phases of the motor drives are controlled by four insulated-gate bipolar transistors (IGBTs) of the four-switch inverter (S_1 to S_4) and the third one is connected directly to the middle point of the dc-bus voltage. According to (3), the motor's stator voltage vectors (abc) are expressed in terms of binary variables S_1 and S_2 of the upper power switching devices as follow:

$$\begin{pmatrix} V_{as} \\ V_{bs} \\ V_{cs} \end{pmatrix} = \frac{V_{dc}}{6}\begin{pmatrix} 4 & -2 & -1 \\ -2 & 4 & -1 \\ -2 & -2 & 2 \end{pmatrix}\begin{pmatrix} S_1 \, S_2 1 \end{pmatrix} \tag{3}$$

Table 1: Vector selection table of the conventional DTC strategy

c_ϕ , c_τ	+1, +1	+1, -1	-1, +1	-1, -1
I	V_3	V_2	V_4	V_1
II	V_4	V_3	V_1	V_2
III	V_1	V_4	V_2	V_3
IV	V_2	V_1	V_3	V_4

The components $\alpha\beta$ of the stator voltage vectors are gained from (abc) ones using Clarke's transformation are shown in (4):

$$\begin{cases} V_{\alpha s} = \sqrt{\dfrac{2}{3}}(V_{as} - \dfrac{V_{bs}}{2} - \dfrac{V_{cs}}{2}) \\ V\,\beta s = \sqrt{\dfrac{2}{3}}(\dfrac{\sqrt{3}}{2}V_{bs} - \dfrac{\sqrt{3}}{2}V_{cs}) \end{cases} \tag{4}$$

The four active voltage vectors (V_1 V_4

), are generated from four available combinations of the states of the upper IGBTs. As shown in Figure 2, the former have unbalanced amplitudes and are spaced by $\frac{\pi}{2}$, in such away V_1

V_3 have an amplitude of $\frac{V_{dc}}{\sqrt{6}}$ and the remaining ones have equally an amplitude of $\frac{V_{dc}}{\sqrt{2}}$.

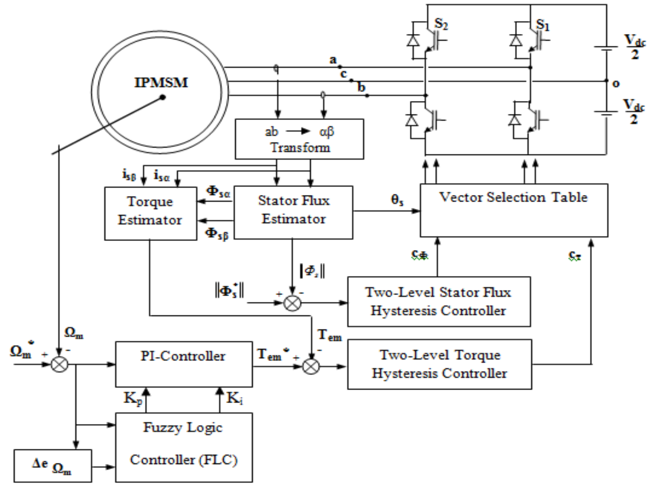


Figure 1. Overall block diagram of direct torque control strategy dedicated to a FSTPI-fed an IPMSM drives.

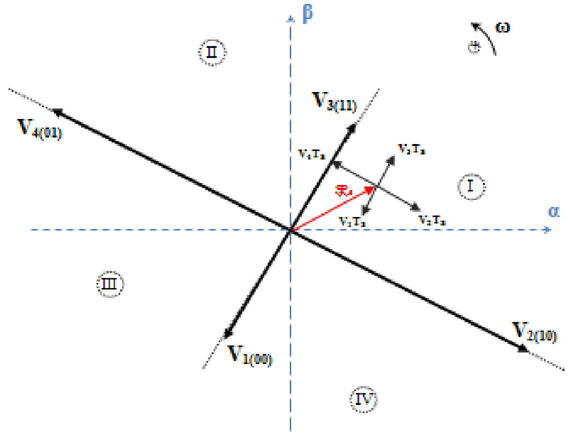


Figure 2. Unbalanced active voltage vectors generated by a FSTPI.

2.3. *Vector Selection Table Corresponding to the Proposed DTC Strategy of Four-Switch Three-Phase Inverter*

To reduce the speed and torque ripple yielded by the classical four-switch three-phase inverter topology, an optimized vector selection table corresponding to the introduced DTC strategy of FSTPI is originally established for induction machine in 2013 [6] similarly to the basic SSTPI switching table. As reported in the literature, the proposed vector selection table must be defined based on the output states of the instantaneous stator flux and electromagnetic torque two-level hysteresis controllers, together, with the equivalent sector in which the instantaneous stator flux vector ϕ_S is located.

inspired from the earlier one introduced by Takahashi is based on the emulation of SSTPI operation by the FSTPI. This trend is achieved through the generation of six balanced voltage vectors employing the four intrinsic unbalanced ones of FSTPI $(V_1, V_3, V_{34H}, V_{41H}, V_{23H}, V_{12H})$, where subscript H indicates the half of the corresponding voltage vector (effective).

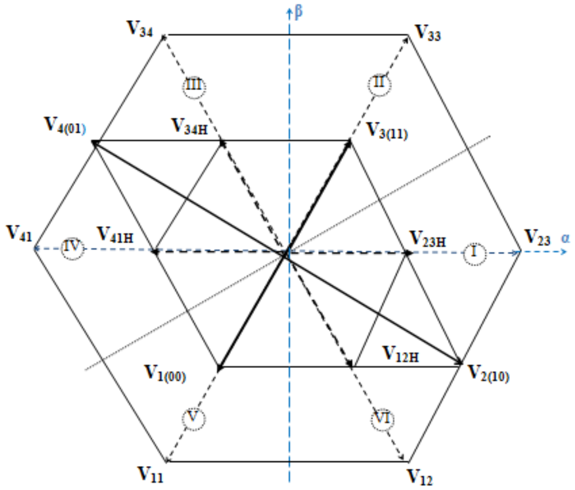


Figure 3. Generation of SSTPI active voltage vectors using four unbalanced voltage.

As shown in Figure 3, these voltage vectors are limited by six symmetric sectors in the Clarke plane and they have the same amplitude of $\sqrt{\frac{2}{3}}V_{dc}$ and equally shifted by $\frac{\pi}{3}$, like the case of the conventional six-switch three-phase inverter topology. Taking into account of the symmetry of the six sectors, the following analysis of the torque and stator flux variations will be limited to sector I. In order to highlight the appropriateness of the application of the emulation of SSTPI operation, the output variables c_ϕ and c_τ of two-level hysteresis comparators have been kept unchanged during two successive sampling periods $2T_S$.

Table 2: Vector selection table of proposed direct torque control strategy

c_ϕ , c_τ	+1, +1	+1, -1	-1, +1	-1, -1
I	V_3	V_{12H}	V_{34H}	V_1
II	V_{34H}	V_{23H}	V_{41H}	V_{12H}
III	V_{41H}	V_3	V_1	V_{23H}
IV	V_1	V_{34H}	V_{12H}	V_3
V	V_{12H}	V_{41H}	V_{23H}	V_{34H}
VI	V_{23H}	V_1	V_3	V_{41H}

During such sampling period, a half of corresponding voltage vector is applied. Inspired from the approach cited in [6], the adopted control laws to synthesize the corresponding vector selection [Table 2] are defined as:

V_3 (V_3 then $V_3 \Rightarrow V_{33}$) achieves the control combination $(c_\phi = +1, c_\tau = +1)$,

- the application of V_{12H} (V_1 then V_2) achieves the control combination $(c_\phi = +1, c_\tau = -1)$, the application of V_{34H} (V_3 then V_4) achieves the control combination $(c_\phi = -1, c_\tau = +1)$,
- the application of V_1 (V_1 then$V_1 \Rightarrow V_{11}$) achieves the control combination $(c_\phi = -1, c_\tau = -1)$.

3. Modeling of Speed Controller-Based an Adaptive Fuzzy Logic

A fuzzy logic controller (FLC) looks at the world in imprecise terms, similar to how a human being perceives information [13]. It converts a linguistic control strategy into an automatic control strategy [14] and fuzzy rules are constructed by expert knowledge or experience database. Unlike the conventional PI-controller, the FLC modeling doesn't depend on the rating parameters of the motor. Consequently, the FLC offers robust performance under sudden change in command speed and/or load torque disturbances. Besides, the concept of fuzzy set is made precise through the membership functions. The input membership functions transform input analog signals into fuzzy numbers. These latter can be combined through fuzzy rules to generate specific actions.

As shown in Figure 1, the fuzzy logic controller adjusts and optimizes in real time the PI-controller gains (k_p andk_i) through fuzzy inference mechanism. As well known that the FLC rule base design involves defining rules that relate the input variables to the output model properties [15]. The designed controller presents two equally inputs and outputs, in this presented work. The input variables of the developed speed controller are: (i) the speed error ($e_{\Omega m}$), and (ii) the time speed error variation ($\Delta e_{\Omega m}$), which are calculated at every sampling instant and are given in (5):

$$\begin{cases} e_{\Omega m}(k) = \Omega_m^*(k) - \Omega_m(k) \\ \Delta e_{\Omega m}(k) = e_{\Omega m}(k) - e_{\Omega m}(k-1) \end{cases} \tag{5}$$

Where $\Omega_m^*(k)$ is the reference mechanical speed, $\Omega_m(k)$ is the actual mechanical speed, and $e_{\Omega m}(k-1)$ is the value of speed error at previous sampling time.

The output variables for the illustrated fuzzy logic controller are $k_p^{'}$ and $k_i^{'}$. It is well known that the fuzzy controller toolbox generally consists of three main parts: fuzzification process, linguistic rule base, and defuzzification process.

3.1. Fuzzification Process

The input linguistic variables $e_{\Omega m}(k)$ and $\Delta e_{\Omega m}(k)$ are converted into fuzzy variables $E_{\Omega m}$ and $\Delta E_{\Omega m}$ respectively, that can be identified by the level of membership functions in the fuzzy set. The input fuzzy sets ($E_{\Omega m}$ and$\Delta E_{\Omega m}$) are defined graphically by membership functions which are represented by triangular shapes and with 50% overlapping as shown in Figure 4. The universe of discourse of $E_{\Omega m}$ and $\Delta E_{\Omega m}$ are divided into seven overlapping fuzzy sets, namely given as: NB (Negative Big), NM (Negative Medium), NS (Negative Small), ZE (Zero), PS (Positive Small), PM (Positive Medium), and PB (Positive Big). Each fuzzy variable is a member of the subsets with a degree of membership (μ) varying between 0 and 1.

The both output linguistic variables include two fuzzy subsets S (Small) and B (Big). Figure 5 illustrates the membership functions of the output linguistic variables $k_p^{'}$ and $k_i^{'}$ respectively, which are designed with standard trapezoidal shapes and with 50% overlapping.

Table 3 : Fuzzy control rules table of $k_p^{'}$ and $k_i^{'}$ respectively

$\Delta E_{\Omega m}$ $E_{\Omega m}$	NB	NM	NS	ZE	PS	PM	PB
NB	B/B	B/B	B/B	B/B	B/B	B/B	B/B
NM	S/B	B/B	B/S	B/S	B/S	B/B	B/B
NS	S/B	S/B	B/B	B/S	B/B	S/B	S/B
ZE	S/B	S/B	S/B	B/S	S/B	S/B	S/B
PS	S/B	S/B	B/B	B/S	B/B	S/B	S/B
PM	S/B	B/B	B/S	B/S	B/S	B/B	S/B
PB	B/B	B/B	B/B	B/B	B/B	B/B	S/B

3.2. Linguistic Rule Base

The fuzzy rules are conditional statements that use fuzzy operators and membership functions to make control decisions. The reasoning control rules in the system are expressed in "if-then" format. The i^{th} rules R_i can be written as: If $E_{\Omega m}$ is A_i and $\Delta E_{\Omega m}$ is B_i, then $k_p^{'}$ is C_i and $k_i^{'}$ is D_i. Where A_i, B_i, C_i and D_i denote the fuzzy sets and with i = 1 to 49.

In this step, the input control variables $E_{\Omega m}$ and $\Delta E_{\Omega m}$ are processed by a fuzzy inference engine that has the capability of simulating human decision-making based of fuzzy concepts and of inferring fuzzy control actions employing fuzzy implication and the rules of inference. Depending on inherent law of input and output variables, this system has 49 fuzzy inference rules. According to the speed error $E_{\Omega m}$ and its rate of change $\Delta E_{\Omega m}$, the fuzzy logic rule bases of $k_p^{'}$ and $k_i^{'}$ at different states can be acquired as shown in Table 3.

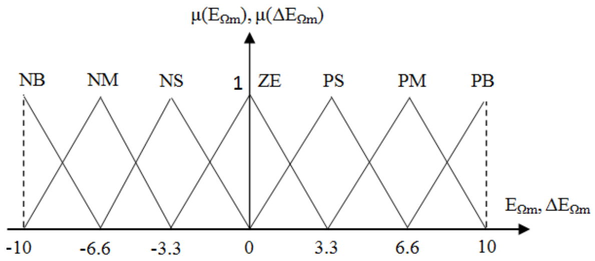


Figure 4. Membership functions of input variables of FLC.

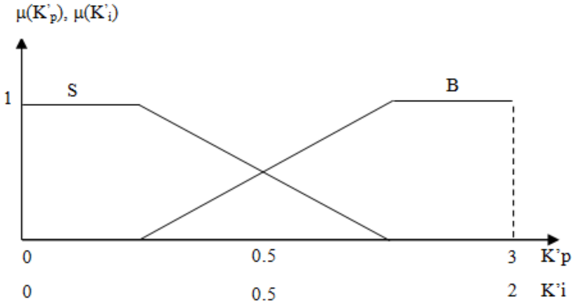


Figure 5. Membership functions of output variables of FLC.

3.3. Defuzzification Process

Once the fuzzy inference results are derived, the control outputs can be acquired from the defuzzification process, to get the crisp values of outputs. In general, the inferred fuzzy action is converted to a crisp value, X, through the widely used weighted average, which is equivalent to center of area (COA) method [16] to yield (6):

$$X = \frac{\sum_{i=1}^{49}[\mu(x_i)x_i]}{\sum_{i=1}^{49}\mu(x_i)} \qquad (6)$$

Where X is a grade value of $x_i(t)$ (it denotes the output variable $k_p^{'}$ and/ or $k_i^{'}$) and $\mu(x_i)$ is a compatibility (weighing factor; derived by using Mamdani's minimum fuzzy implication rule). The final actual values can be obtained by using linear transform to the output values. The linear transform formulas of proportional coefficient k_p and integral coefficient k_i are given in (7):

$$\begin{cases} k_p = (k_{p\max} - k_{p\min})k_p^{'} + k_{p\min} \\ k_i = (k_{i\max} - k_{i\min})k_i^{'} + k_{i\min} \end{cases}$$
$$(7)$$

In a direct torque control strategy, the discrete expression of the reference electromagnetic torque is expressed as:

$$T_{em}^{*}(k) = k_p e_{\Omega m}(k) + k_i \int_0^{T_s} e_{\Omega m}(k)dt \qquad (8)$$

As shown in (8), the control effects can easily be acquired at different speed demand by dynamically adjusting k_p and k_i in accordance with the speed errors. In the novel fuzzy logic system, the Mamdani's minimum operation rule is employed for the optimal fuzzy reasoning algorithm.

4. Simulation Based Investigation of Performance of an Adaptive Fuzzy Speed Controller

In order to verify the effectiveness of the proposed adaptive speed controller and to achieve a satisfactory performance of the IPMSM drives fed by a FSTPI for basic and proposed DTC strategies at different dynamic operating conditions, a numerical simulation has been carried out using MATLAB/SIMULINK software.

The ratings and parameters of the IPMSM drives, used in the simulation works, are provided in Table 4. The sampling period T_s is equal to 50μs. The reference stator flux is equal to $\sqrt{3}$ times its rated value within a dc-bus voltage equal to 400V. The bandwidths used in the stator flux controller and in the electromagnetic torque one are equal to ±0.01Wb. The fixed gains of the standard PI-controller used in the speed control loop are k_p = 2 and k_i = 0.53.

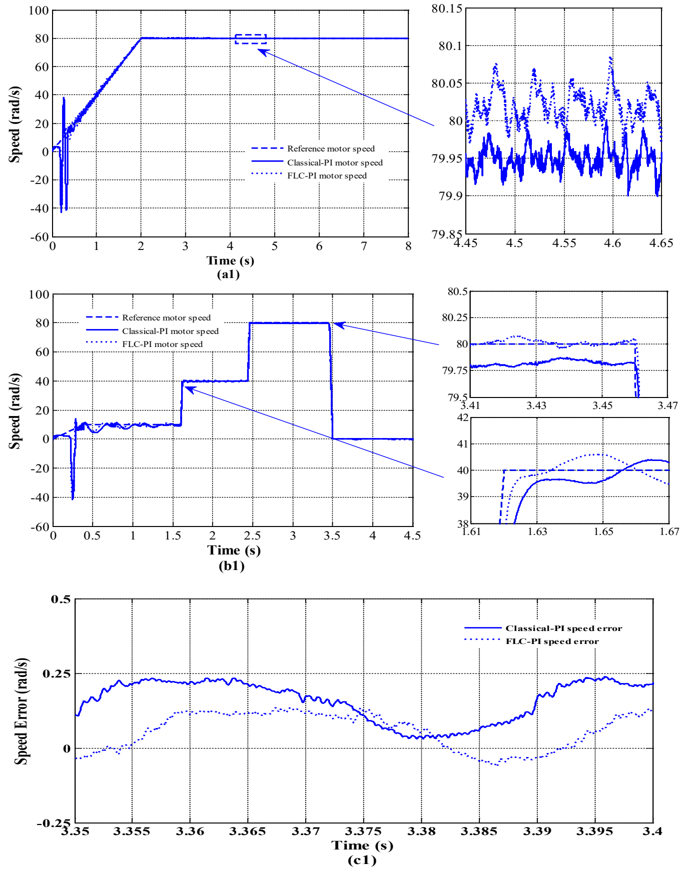
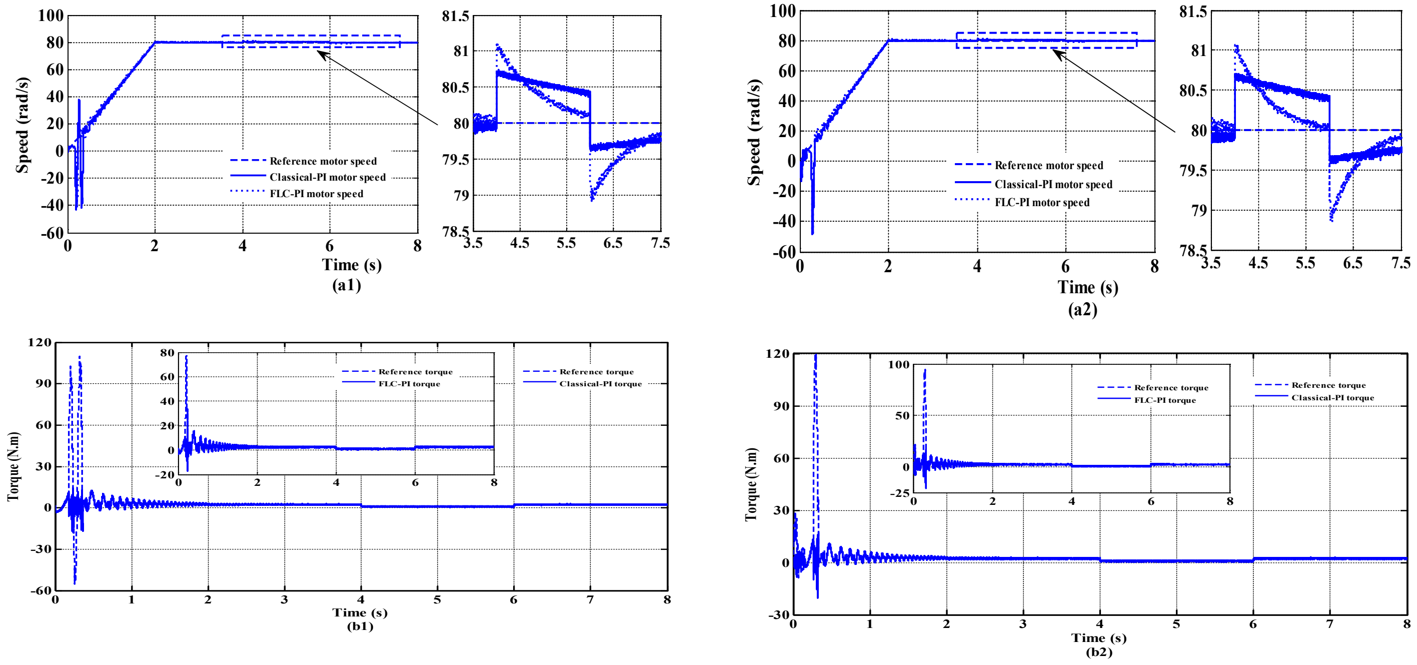


Figure 6. Simulated speed response yielded by conventional DTC strategy at forward motoring operation under constant load torque. Legend (a): motor speeds and its reference in the case of constant mechanical speed, (b) and (c): speed response with its error under variable reference speed.



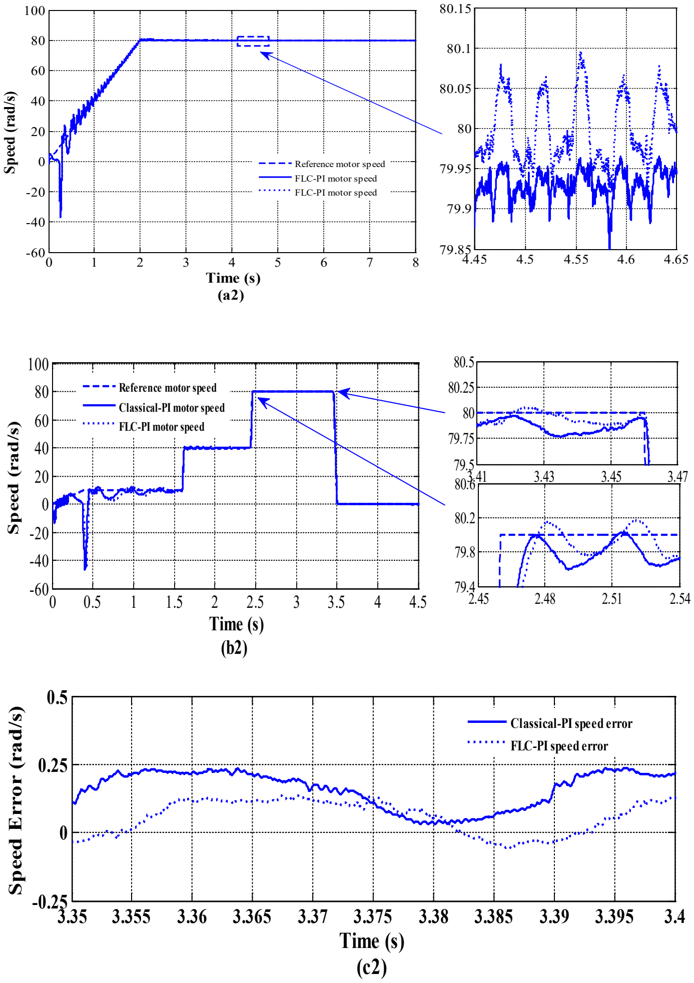


Figure 7. Simulated speed response yielded by proposed DTC strategy at forward motoring operation under constant load torque. Legend (a): motor speed and its reference in the case of constant mechanical speed, (b) and (c): speed response with its error under variable reference speed.

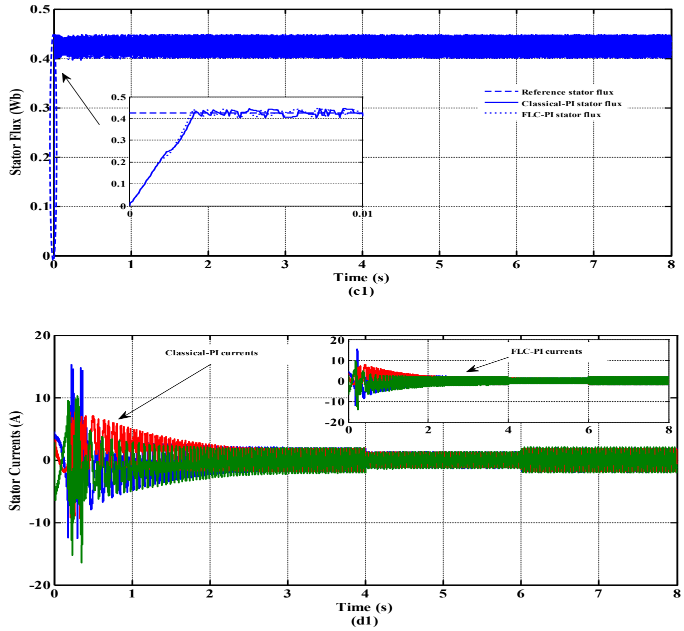


Figure 8. Simulated waveforms of IPMSM drives offered by classical DTC strategy at forward motoring operation under variable load torque. Legend (a1): motor speed and its reference, (b1): electromagnetic torque with its reference, (c1): stator flux with its reference, and (d1): motor stator currents.

Table 4 : IPMSM specifications

Rated voltage	208V	Frequency	60Hz
Rated torque	2N.m	p	2
Φm	0.25Wb	rs	1.93Ω
Ld	44.42mH	Lq	79.57mH
J	3g.m2	f	0.8mN.m.s

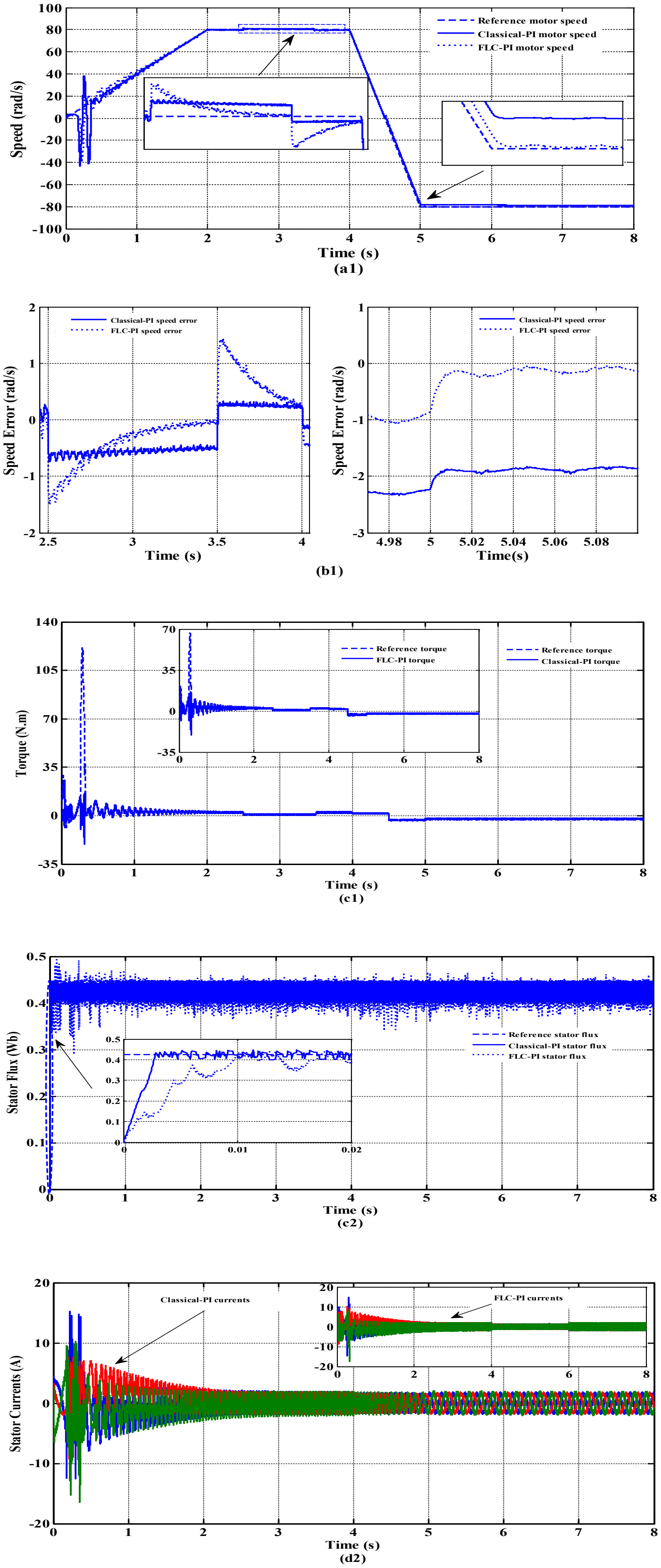


Figure 9. Simulated waveforms of IPMSM drives offered by introduced DTC strategy at forward motoring operation under variable load torque. Legend (a2): motor speed and its reference, (b2): electromagnetic torque with its reference, (c2): stator flux with its reference, and (d2): motor stator currents.

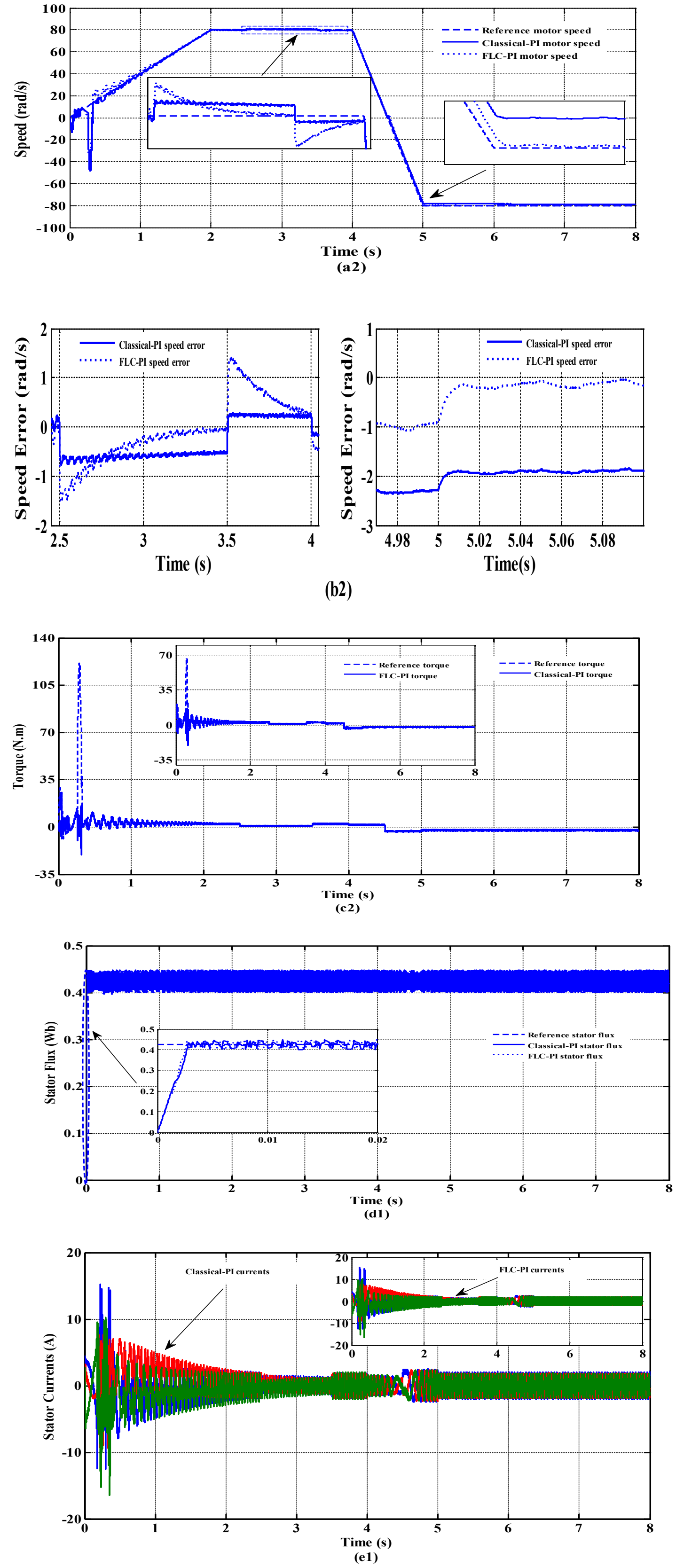


Figure 10. Simulated waveforms of IPMSM drives offered by classical DTC strategy at forward and reversal motoring operation conditions under proportional load torque. Legend (a1): motor speed and its reference, (b1): motor speed error, (c1): electromagnetic torque with its reference, (d1): stator flux with its reference, and (e1): motor stator currents.

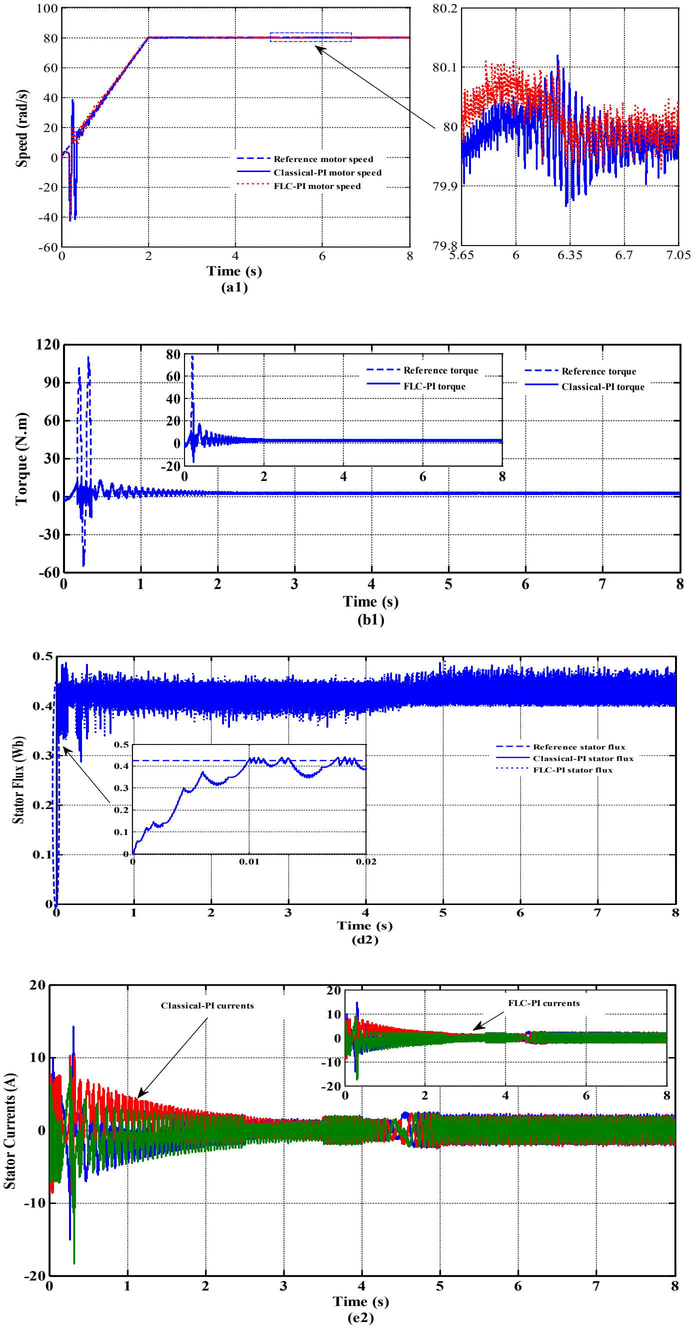


Figure 11. Simulated waveforms of IPMSM drives offered by introduced DTC strategy at forward and reversal motoring operation conditions under proportional load torque. Legend (a2): motor speed and its reference, (b2): motor speed error, (c2): electromagnetic torque with its reference, (d2): stator flux with its reference, and (e2): motor stator currents

The realized works are focused on the motor speed error under transient and steady-state operations, considering different operating conditions, such as: (i) variations of the step reference speed, (ii) sudden change of a load torque, and (iii) variation of the stator resistance, over basic and proposed DTC strategies (subscripts 1 and 2, respectively) using both speed controllers.

Let us consider the start-up of the IPMSM within a ramp-shape reference speed during 2s to reach a constant value of $\Omega_m = 80 rad/s$ for a constant load torque $T_l = 1.5 N.m$ and under forward motoring operation of the machine. Referring to Figure. 6(a1) and Figure. 7(a2), one can clearly notice that the developed controllers implemented in both direct torque control strategies (basic and proposed ones) yield practically a high dynamic behaviour at transient-state and there is approximately no difference between them at stabile-state operation. Otherwise, the adaptive PI-controller offers quick and smooth speed response with a negligible error during both transient and steady-state operations. Conversely, the conventional PI-controller exhibits a significant speed error especially under transient-state operation for both stratgies and the motor speed can’t reach its reference one at all range time.

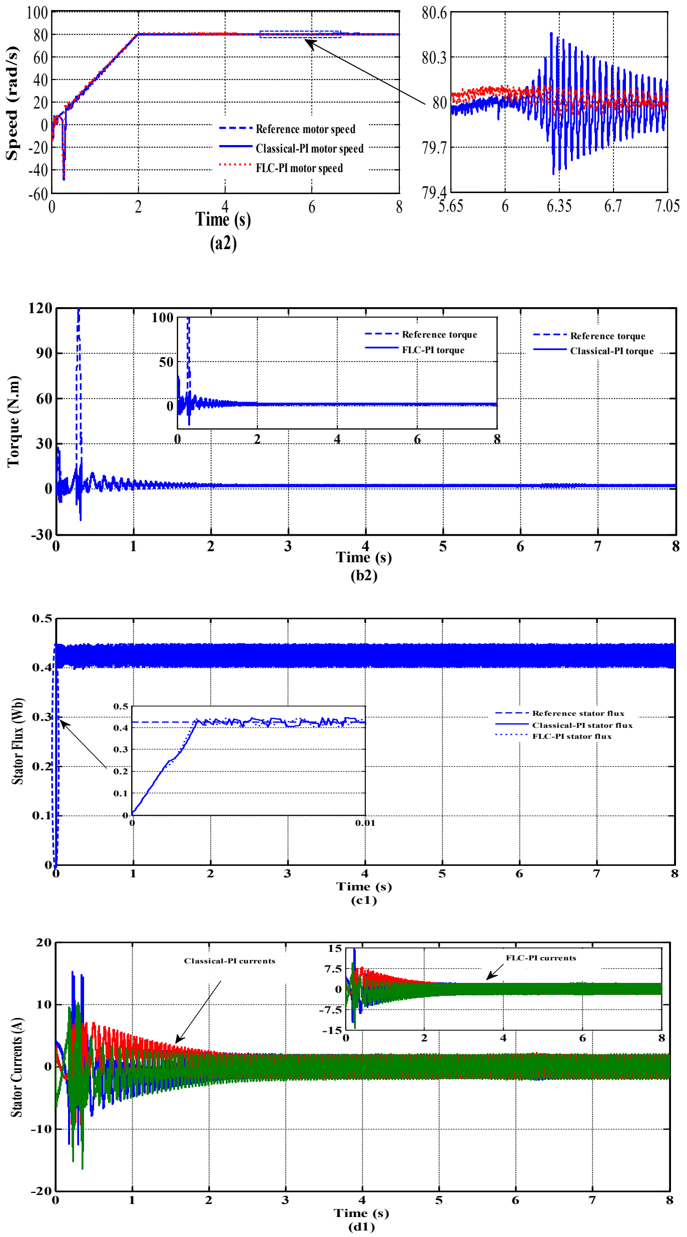


Figure 12. Simulated waveforms of IPMSM drives exhibited by basic DTC strategy at forward motoring operation conditions under stator resistance changes. Legend (a1): motor speed and its reference, (b1): electromagnetic torque with its reference, (c1): stator flux with its reference, and (d1): motor stator currents.

So as to extend an equitable comparison between the designed speed controllers considering the case of sudden changes of reference mechanical speed and for a constant load torque $T_l = 1.5N.m$, the simulated starting responses of the four-switch three-phase inverter based IPMSM drives are illustrated in Figures 6.(b1 and c1) and in Figures 7.(b2 and c2). Moreover, the reference mechanical speed increases from 10rad/s to 40rad/s at 1.6s. After 840ms, it increases from its previous value to reach its rated value $\Omega_m = 80rad/s$. Finally, at 3.46s, the speed decreases until it vanishes.

According to Figure 6.(b1) and for the first reference speed change (10rad/s to 40rad/s), with using of basic DTC strategy, the necessary time for mechanical speed to attain its reference value is almost about 45ms for the conventional PI-controller, while with the adaptive fuzzy one, the time response is almost about 24ms.

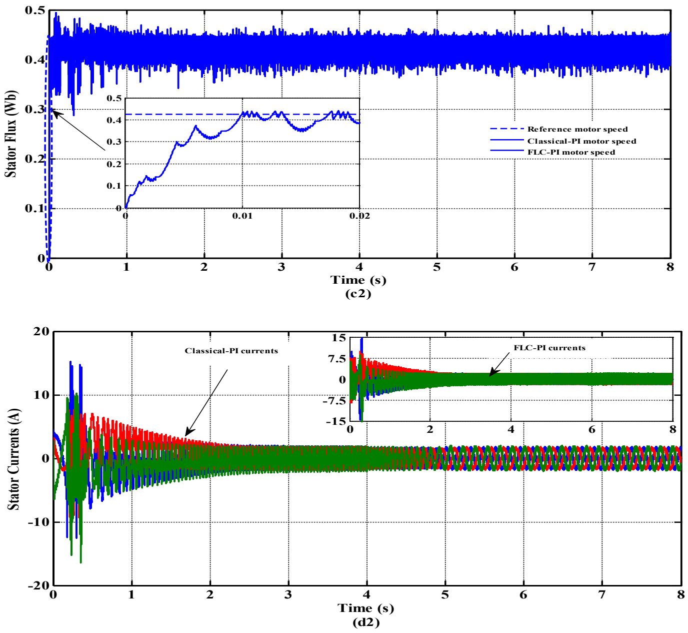


Figure 13. Simulated waveforms of IPMSM drives exhibited by proposed DTC strategy at forward motoring operation conditions under stator resistance changes. Legend (a2): motor speed and its reference, (b2): electromagnetic torque with its reference, (c2): stator flux with its reference, and (d2): motor stator currents.

For second speed change shown in Figure 7.(b2) and with designed FLC regulator implemented in the introduced DTC approach, the speed needs no more than 27ms to reach its desired value, in contrast to the conventional one where the speed remains 63ms to attain the desired one. Thus, as shown in zoomed views of motor speed, one can conclude that the speed response is quick enough under adaptive PI-controller compared to that with the conventional one.

Additionally, as shown in zoom-in view of motor speed responses for both DTC strategies, one can notice clearly that, for the third speed change, the speed has smooth and quick trajectory with FLC system at speed change transition, compared to that exhibited by the classical speed controller. Thus, a good tracking performance has been gained with the developed adaptive PI-controller under application of variable mechanical speed.

Figures 6.(b1 and c1) and Figures 7.(b2 and c2) show that the proposed fuzzy regulator offers a negligible speed error, compared to that provided one by the conventional one which exhibits a higher steady-state error and the motor speed can't follow its reference value at all range time. So that in fact demonstrates the efficiency of FLC in the speed control loop in terms of swiftness and robustness of system drive when there exist changes of reference speed steps.

Furthermore, so as to highlight the performance and effectiveness gained by the introduced FLC as a speed regulator in DTC scheme, sudden abrupt variations of load torque is applied. Let us consider the start-up of the motor drives within a ramp-shape reference speed during 2s to reach a constant value $\Omega_m = 80rad/s$, at forward motoring operation and for a constant load torque $T_l = 1.5N.m$. At 4s, the load torque varied by a reduction of 30% of its initial value and after 2s, it removed to its starting value.

From analysis of Figure 8.(a1) and Figure 9.(a2), one can observe that the steady-state gained by both basic and proposed DTC strategies, is likewise achieved within very short duration for the load torque variations under the proposed adaptive PI-controller. Furthermore, the zoom views of motor speed shown in presented figures prove that the conventional PI-controller provides a higher steady-state error and the motor speed can't clearly proceed its reference one at all range time. Thereby, the classical PI-controller is more sensitive to load disturbances counter to the expanded adaptive fuzzy one. As a deduction, less speed error and better robustness as regards reference command speed and load torque changes are the most frequently adduced advantages of the adaptive PI-controller over the classical one.

Figures 8.(b1, c1 and d1) and Figures 9.(b2, c2 and d2) illustrate the waveforms of the electromagnetic torque, stator flux and stator currents of the motor drives. Accordingly, one can be seen that both controllers have same motor torque behaviors and stator currents responses which are pure steady-state sine-wave with fewer harmonic at steady-state operation. But it appears that with the using of fuzzy logic system, the motor can start almost with smooth reference torque and low peak thinks to the introduced FLC speed regulator, compared to the classical one.

From analysis of Figure 8.(c1) and Figure 9.(c2), one can clearly notice that DTC strategy using both developed speed controllers in the standard $\alpha\beta$ plane leads to high dynamic and reduced amplitude of the stator flux. Furthermore, same remarks can be concluded from the following operation conditions to these mentioned variables. Thus, a good tracking performance has been gained with the developed adaptive PI-controller.

Figures 10.(a1 and b1) and Figures 11.(a2 and b2) show the case of forward and reversal motoring of the IPMSM drives under a proportional load torque. It should be noted that, for positive and negative speeds, the motor speed reaches rapidly and smoothly the reference value using the adaptive PI-controller with negligible steady-state error, which confirms the superiority of the proposed speed controller over the conventional one.

Referring to Figures. 10(c1, d1 and e1) and Figures. 11(c2, d2 and e2), it can be seen that the FLC speed regulator offers reduced amplitude and high dynamic of the stator flux, damped peak of the reference electromagnetic torque thinks to the adapting gains and better forms of the motor stator currents, which confirms the superiority of this designed controller over the classical one, especially for the introduced DTC strategy.

As it is well known that the speed error is caused not only by abrupt change of load torque but also by sudden variation of stator resistance, as shown in Figure 12 and Figure 13. As it is concerned previously, the motor drives is initially accelerated within same ramp-shape of reference speed during 2s to reach a constant value of $\Omega_m = 80 rad/s$, with its rated stator resistance $r_s = 1.93\Omega$ and under a constant load torque $T_l = 1.5 N.m$.

As it is illustrated in Figure 14, the first change of stator resistance from its nominal value 1.93 to its half one (0.965) is applied at 5s. And after 0.5s, the latter value of r_s is raised to reach 2.895. Finally, the last change of stator resistance from 2.895 to its rated value is applied at 6.3s. In all, the suggested value of the stator resistance can be established in (9) as follow:

$$r_s^* = r_s(1 \pm 50\%) \tag{9}$$

According to Figure 12.(a1) and Figure 13.(a2), it is distinctly seen that the designed fuzzy controller leads to better and high steady-state performance than that yielded by the conventional one. Like where in different cases discussed earlier, the motor speed is nearly stable and it reaches rapidly and accurately its reference value with a negligible steady-state error, though the stator resistance exceeds its nominal value. Nevertheless, the conventional PI-controller which exhibits considerable and significant overshoots of motor speed, when the variation of stator resistance occurs. Same remarks, which are depicted for the case of the application of a variable load torque, can be noticed here for the obtained curves of the electromagnetic torque, the stator flux and the motor stator currents under the case of variation of the stator resistance of the motor [see Figures. 12(b1, c1 and d1) and Figures. 13(b2, c2 and d2)]. So that validates the merit of using of an adaptive fuzzy speed controller.

As a summary, for both classical and proposed DTC strategies, the developed fuzzy speed controller leads to reduced ripple of the reference electromagnetic torque during both transient and steady-states operations of the motor, to rapid and high stator flux response and to more balanced and sinusoidal motor stator currents.

Figure 15.(a) illustrates the comparison of total harmonic distortion values of the a-phase stator current (THD) and the electromagnetic torque ripple yielded by the developed strategies under conventional-PI and fuzzy logic controllers in the standard (four-sectors [see Figure 16.(a)]) and proposed (six-sectors [see Figure 16.(b)]) αβ planes (4SPI, 6SPI, 4SFLC and 6SFLC), considering a range of the stator frequency varying from 1.58Hz to 25Hz. The electromagnetic torque ripple is determined by the equations (10) and (11), presented below. According to Figure 15.(a), one can clearly observe that, over the whole stator frequency range, the proposed DTC strategy using both classical and adaptive fuzzy speed controllers generally leads to lower current THD values at all motor speeds, than those offered by basic DTC approach.

On the other hand, one can clearly notice that the proposed adaptive fuzzy controller exhibits the best steady-state performance in terms of current THD values compared to that offered by the conventional PI-controller. So the quantitative comparison confirms the superiority of the adaptive fuzzy controller not only during the dynamic response, but also in the steady-state performance. The electromagnetic torque ripple of developed strategies is calculated using standard deviation function and it is expressed as follow:

$$T_e^{ripple} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(T_e(i) - T_e^{av})^2} \tag{10}$$

where

$$T_e^{av} = \frac{1}{N}\sum_{i=1}^{i=N} T_e(i) \tag{11}$$

Where N is the sampling number during one short period, $T_e(i)$ is actual electromagnetic torque, and T_e^{av} is average electromagnetic torque.

Referring to Figure 15.(b), it should be noted that, over the whole stator frequency range, the electromagnetic torque ripple of the introduced DTC strategy is lower than that offered by the basic DTC one, even if the adaptive fuzzy speed controller or the conventional one is employed. At high stator frequency above 24Hz, the torque ripple of FLC regulator is similar to that of classical PI-controller, under both basic and proposed DTC strategies.

5. Conclusions

This paper dealt with the introduced direct torque control strategy (emulation of SSTPI) incorporating an adaptive fuzzy logic in speed control loop of IPMSM-fed by a four-switch three-phase inverter. According to speed error and its first time derivative, the proportional and integral gains of the proportional-integral speed controller are online adjusted.

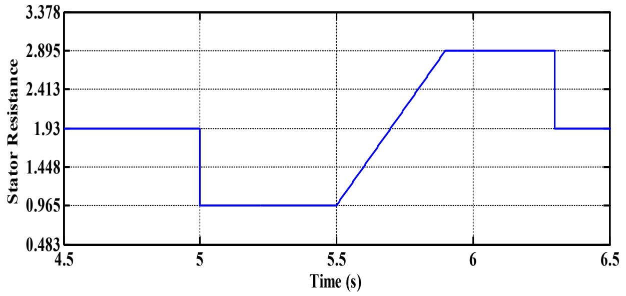


Figure 14. Stator resistance variation.

Simulation-based comparative study between the performances of the developed fuzzy logic controller and the conventional one has been carried out, considering different cases such as: (i) variations of the step reference speed, (ii) sudden change of a load torque, and (iii) variations of the stator resistance.

It has been found that the speed controller-based fuzzy logic toolbox ensures fast dynamic response, less harmonic distortion of the a-phase stator current, reduced torque ripple, more robustness and less steady-state error when there exist variations of the motor parameters and load torque disturbances.

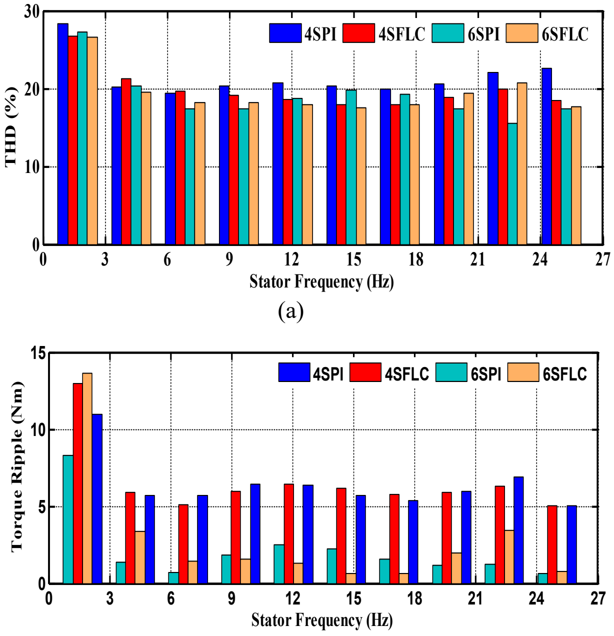


Figure 15. Quantitative steady-state performance comparison yielded by basic DTC strategy and proposed one over conventional-PI and adaptive fuzzy-PI controllers for a constant load torque $T_l = 1.5N.m$, at various reference speed levels. (a): THD values, (b): torque ripple.

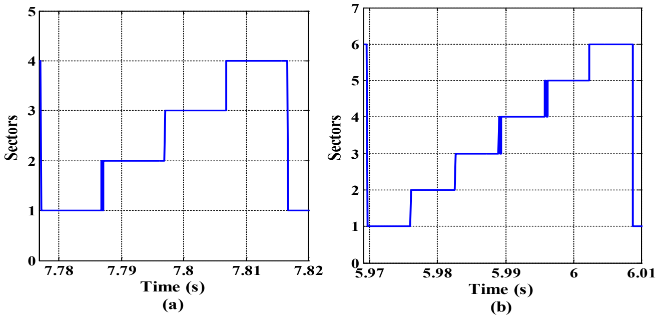


Figure 16. Stationary Clarke plane. (a): four-sectors (4S), (b): six-sectors (6S).

References

[1] M. Depenbrock, "Direct self controlled (DSC) of inverter-fed induction machines" IEEE Trans. Power Electronics., **3**(4), October. 1988.

[2] I. Takahashi and T. Noguchi, "A new quick-response and high-efficiency control strategy of an induction motor" IEEE Trans. Ind. Appl., **22**(5), 820-827, 1986.

[3] G. C. D. Sousa, B. K. Bose, and J. G. Cleland, "A fuzzy logic based on-line efficiency optimization control of an indirect vector-controlled induction motor drive" IEEE Trans. Ind. Electron., **42**(2), 192-198, Apr. 1995.

[4] Y. Chen, B. Yang, X. Gu, and S. Xing, "Novel fuzzy control strategy of IPMSM drive system with voltage booster" in Proc. 6th World Congr. Intell on Control Autom, **2**, pp. 8084-8087, Jun. 21-23, 2006.

[5] K.B. Mohanty, "A direct torque controlled induction motor with variable hysteresis band" 11th International Conference on Computer Modeling and Simulation, 2009.

[6] B. El Badsi, B. Bouzidi, and A. Masmoudi, "DTC scheme for a four-switch inverter-fed induction motor emulating the six-switch inverter operation" IEEE Trans. Power Electron., **28**(7), 3528-3538, Jul. 2013.

[7] R. B. Inderka and R. W. De Doncker, "DITC-Direct instantaneous torque control of switched reluctance drives" IEEE Trans. Ind. Applicat., **39**, pp. 1046-1051, July/Aug. 2003.

[8] L. Zhong, M. Rahman, W. Hu, and K. Lim, "Analysis of direct torque control in permanent magnet synchronous motor drives" IEEE Trans. Power Electron., **12**(3), 528-536, May. 1997.

[9] B. A. Welchko and T. A. Lipo, "A novel variable-frequency three-phase induction motor drive system using only three controlled switches" IEEE Trans. Ind. Appl., **37**(6), 1739-1745, Nov/Dec. 2001.

[10] H. W. vander Broeck and J. D. van Wyk, "A comparative investigation of a three-phase induction machine drive with a component minimized voltage-fed inverter under different control options" IEEE Trans. Ind. Appl., IA-**20**(2), 309-320, Mar/Apr. 1984.

[11] M. Azab and A. L. Orille, "Novel flux and torque control of induction motor drive using four switch three phase inverter" in Proc. IEEE Annu. Conf. Ind. Electron. Soc, Denver, CO., **2**, pp. 1268-1273, Nov/Dec. 2001.

[12] Bassem El Badsi, Badii Bouzidi, and Ahmed Masmoudi, "DTC Scheme for a Four-Switch Inverter-Fed Induction Motor Emulating the Six-Switch Inverter Operation" IEEE Trans. Power Electronic., **28**(7), 3528-3538, July. 2013.

[13] L. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes" IEEE Trans. Syst. Man and Cybern., **3**, pp. 28-44, 1973.

[14] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic control-part 2" IEEE Transaction on Systems, Man and Cybernetics., **20**(2), Marc W, pp. 419-435, April. 1990.

[15] Y. Luo and W. Chen, "Sensorless stator field orientation controlled induction motor drive with a fuzzy speed controller" Computer and Mathematics with Appl., **64**, pp. 1206-1216, 2012.

[16] A. Arias, "Improvements in direct torque control of induction motors", Ph.D. dissertation, Electron. Eng. Dept., Technical Univ. Catalonia, Terrassa, Spain, 2000.