ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 1784-1798 (2017)

www.astesj.com

Special Issue on Recent Advances in Engineering Systems

ASTESJ

ISSN: 2415-6698

Investigating the Expertise Indicators of Vulnerability Discovery Professionals

Mortada Al-Banna*, Boualem Benatallah, Moshe Chai Barukh

UNSW Australia, Computer Science and Engineering, Sydney, Australia

ARTICLE INFO

Article history:
Received: 15 June, 2017
Accepted: 29 August, 2017
Online: 23 September, 2017

Keywords:
Crowdsourcing
Expertise
Vulnerability Discovery
Professionals

ABSTRACT

In crowdsourcing, selecting the person with suitable expertise is very important; especially since the task requester is not always in direct contact with the worker. Recently, this has become increasingly important particularly when the crowdsourced tasks are complex and require skillful workers (e.g. software development, software testing, vulnerability discovery, and open innovation). In this paper, we investigate the expertise indicators of vulnerability discovery professionals in a crowdsourcing vulnerability discovery platform. We conduct a systematic literature review, we review online contents, conduct interviews with domain experts, and survey vulnerability discovery professionals involved in the task of vulnerability discovery. We discuss the indicators we have found, and we provide some recommendations to help improve the process of selecting vulnerability discovery professionals to perform crowd tasks related to vulnerability discovery.

1. Introduction

The complexity of software-based systems is increasing dramatically as development becomes even more distributed across multiple heterogeneous, autonomous, and evolving cloud services. More specifically, the increased reliance on third-party software-based systems (e.g., cloud services, open APIs, external programming libraries and black-box software packages) makes it very difficult for in-house IT experts to deal with the inherent risks of using external software. In order to overcome potential vulnerability issues, several organizations outsource tasks such as vulnerability discovery to third-party providers. More recently, the approach of crowdsourcing vulnerability discovery has emerged. Companies like Google, Facebook, and Microsoft have their own crowdsourced vulnerability discovery programs (aka vulnerability reward programs). At present, even the government sector is adopting this approach (e.g., United States Department of Defense). In Addition, several crowdsourcing platforms for vulnerability discovery have emerged (e.g., Bugcrowd 'Bugcrowd.com', HackerOne 'hackerone.com', Synack 'synack.com', and Cobalt 'cobalt.io'). Similarly, competitions have been conducted for discovering vulnerabilities (e.g., Pwn2Own). This paper is an extension of work originally presented in IEEE 2nd International Conference on Collaboration and Internet Computing (CIC) [1].

Crowdsourcing vulnerability discovery allows organization to benefit from the merits of crowdsourcing (e.g., diversity, scale and speed). Alongside the benefits of crowdsourcing, some quality control concerns have emerged (e.g., clarity of the outcome, trustworthiness of the crowd) [2]. Facebook, in their vulnerability discovery program, highlights the following : "We received 13,233 submissions in 2015. Of these, only 526 were valid and eligible to receive rewards". Similarly, Bugcrowd mentions in the state of bug bounty report for 2016 [3], the platform received from January 1, 2013 to March 31, 2016, a total of 54,114 submissions, of which only 9,963 contained valid vulnerabilities. Although the variation in expertise is considered desirable for discovering vulnerabilities [4], organizations may be overwhelmed by the number of contributions and hence need to adopt techniques to ensure the quality of submitted vulnerability reports. One technique used in crowdsourcing to mitigate quality concerns is worker selection. Research shows that recruiting suitable workers enhances the quality of contributions [2,5].

In this respect, we are not advocating that limiting vulnerability discovery to selected vulnerability discovery professionals from the crowd is superior over open programs. The scope of this paper is to examine how Vulnerability Discovery Professionals (henceforth referred to as VDPros) are selected in crowdsourcing vulnerability discovery platforms. Organizations that crowdsource the task of vulnerability discovery are not always in direct contact with VDPros whom perform the task. Additionally, traditional methods of recruitments are not always feasible (e.g., interviews, probation periods). Hence, understanding expertise indicators of VDPros participating in the crowdsourced task of vulnerability discovery becomes even more important. We aim to understand how to measure the expertise of VDPros in order to make sure that the selected VDPro would yield a high-quality outcome for the vulnerability discovery task. In practice this is critical, since poorly

*Corresponding Author: Mortada Al-Banna, UNSW Australia, Computer Science and Engineering, Sydney, Australia | Email: m.al-banna@unsw.edu.au

managed vulnerabilities could cause huge losses to organizations whether financially or to the reputation . Our study aims to answer the following research questions:

– RQ1: What are the perceived indicators of expertise for selecting VDPros?

– RQ2: Which sources are accepted as reliable to extract these indicators of expertise for selecting VDPros?

While there is a large body of research for worker selection in crowdsourcing [6], to the best of our knowledge none yet addresses the selection of VDPros in crowdsourced vulnerability discovery. Giboney et al. developed a conceptual-expertise-based tool that can be used to discriminate between novice and expert VDPros [7]. They rely on self-reported skills and testing the knowledge of the VDPros to measure their expertise. This could be suitable in traditional recruitment but for crowdsourcing systems it might have some shortcomings (e.g., recurrent questions, automating simple question-answering) [8]. On the other hand, Hafiz and Fang performed an empirical study on VDPros who have disclosed vulnerabilities [9]. They investigated methods and tools used by VDPros to discover the vulnerabilities and how the community of VDPros tend to focus on certain vulnerability types. While they have touched some aspects in regard to the expertise of VDPros involved in discovering vulnerabilities, they have not explored the criterion for indicators of expertise with respect to VDPros.

Accordingly, this paper makes the following main contributions: we have conducted an extensive review of literature using a "systematic literature review" methodology [10], along with investigating a diverse range of online contents (e.g., blogs, technical reports and articles). We have then adopted "exploratory research methods" [11,12], where we relied on the insights we gained from the literature and the online contents, to conduct semi-structured interviews with domain experts. We also conducted an online survey for the VDPros involved in the task of vulnerability discovery to validate our findings. We employed "qualitative data analysis" guidelines [12,13] to analyse our data collected and scientifically present our findings. And finally, based on our findings we propose recommendations that we believe will pave the way for further advancement in the effective selection of security experts in vulnerability discovery crowdsourcing. While arguably personality traits are important characteristic in the selection of VDPros, in this paper we only focus on the indicators to expertise. An indicator to expertise is a signal about the knowledge of the VDPros in a certain area and their competency to solve problems within that area.

The rest of this paper is organized as follows: In Section 2, we provide some background knowledge. In Section 3 we discuss our review of literature and online contents. The methodology we adopt in this work is thereby illustrated in Section 4. In Section 5, we outline our findings; and in Section 6 we provide recommendations to help improve selecting VDPros for vulnerability discovery tasks. We conclude with a discussion and directions for future work in Section 7.

2. Background

In this section, we examine the topic of crowdsourcing and its application to vulnerability discovery. We provide an illustrative example to further clarify the research problem.

2.1. Crowdsourcing Vulnerability Discovery

A vulnerability is a security flaw, which arises from system design, implementation, or maintenance (e.g. SQL injection vulnerability). By exploiting these vulnerabilities, malicious parties could gain unauthorized access to protected resources. The field of vulnerability discovery and repair has been very active in recent years, especially with the increase in number of security threats and incidents [14].

Crowdsourcing (also known as human computation) harnesses the wisdom of large groups and communities working independently to solve problems, much as open source does for software development. From the service customers' perspective, crowdsourcing is a form of digital service (notwithstanding its essentially human infrastructure), yet its unique value draws from its effective way to perform tasks that remain difficult for, or even beyond the reach of machine computation [15] (e.g., image tagging, natural language translation, and transcription). Several organizations including DARPA, NASA, Honda and various other organizations use platforms such as MTurk and Ushahidi to crowdsource information gathering. Other commercial crowdsourcing platforms have recently emerged alongside MTurk; platforms like Innocentive, TopCoder, Kaggle, and uTest use crowdsourcing for tasks that require skilled workers (Web design, testing, Web development tasks, R&D challenges).

Crowdsourced vulnerability discovery programs allow vulnerability discovery professionals to play the role of an attacker to discover vulnerabilities. In a crowdsourced vulnerability discovery program, software providers submit vulnerability discovery tasks to a community of VDPros. This approach is gaining increasing popularity recently [1] . Crowdsourcing vulnerability discovery could be in the form of an open call and managed directly by an organization (e.g., Facebook VRP, and Google VRP), or directed toward members of specialized platform for crowdsourcing vulnerability discovery (e.g., Bugcrowd, Cobalt, HackerOne, or Synack). Crowdsourced vulnerability discovery programs could be publicly available for everyone and VDPros self-select and decide to participate according to their preference (e.g., Uber program in HackerOne[2]), or could be private where the organization invite only selected VDPros to participate according to a specific criterion (e.g., LinkedIn Private program[3]). There are also unregulated methods for the crowd to submit discovered vulnerability information (e.g., black and grey vulnerability markets[16], and VDPros disclosing vulnerabilities publicly[4]).

Several researchers investigated crowdsourced vulnerability discovery from different angles. Laszka et al investigated the problem with invalid submitted vulnerability reports and their relation to incentives misalignment [17]. Zhao et al. investigated the web vulnerability discovery echo system and analyse vulnerability trends, response behaviour and reward structures [18]. Finifter et al. investigated two crowdsourced vulnerability discovery programs [19]. The authors illustrated that crowdsourced vulnerability discovery programs are economically efficient and explored the difference in the reward strategies and the effect on the engagement with VDPros.

1 www.vulnerability-lab.com/list-of-bug-bounty-programs.php

2 hackerone.com/uber

3 security.linkedin.com/vulnerability-disclosure

4 schneier.com/essays/archives/2007/01/schneier_full_disclo.html

2.2. Illustrative Example

Let's assume that a company called AwsomeAPI has decided to delegate the vulnerability testing for its new API to a crowdsourcing platform for vulnerability discovery. AwsomeAPI is concerned about the quality of the output of the task. As the quality of the output from the crowdsourced task is related to the expertise of the crowd worker [2], AwsomeAPI has thereby decided to limit participation of the program to only expert VDPros. AwsomeAPIs has decided that the scope of the program is to protect its API from the injection attacks (e.g. SQL injection, and script injection) [20]. The internal team in AwsomeAPI has identified that in order for the VDPro to be competent in the vulnerability discovery task, they need to have: knowledge about the technology (e.g., the API to be tested, and protocols such as http, and SSL/TLS); programming skills (e.g., low level programming, and scripting); knowledge about security practices (e.g. security controls such as authentication and authorization, and software vulnerabilities).

AwsomeAPI is facing some difficulties in regard to selecting VDPros with the required level of expertise, and how can they assess the expertise of the VDPro.

3. Expertise Indicators from Literature and Online Contents

We reviewed literature using a systematic literature review (SLR) approach where we followed the guidelines illustrated in [10,21]. As an additional source of information, we have also examined a range of online content, such as companies' blogs, and technological articles, and rely on the online research methods [22]. We provide a brief description and summarize our findings in this section.

3.1. Steps of the Systematic Literature Review

Various approaches amongst literature have been adopted to determine expertise. Hence the use of literature review helps in identifying the indicators that have been relied on in these approaches and how relevant they are in measuring the expertise of VDPros involved in the task of vulnerability discovery. We adopt the methods of literature review on the recommendations of Kitchenham [10] and Okoli et al. [21].

I. Selecting the Articles

We searched for relevant articles addressing people selection according to expertise. We included both journal and conference articles as the conference articles have been found to be of important role in the field of information systems [23]. The search was conducted for the period between January 2009 and January 2016 for articles in IEEE Digital Library, ACM Digital Library, and Scopus databases. We focused on articles available in full text and in English language. We used the following keywords for the search: "Expert" OR "Expertise" OR "Reputation", "Selection" OR "Finding", "Crowd" OR "Crowdsourcing". We checked that searches using similar keywords like 'crowd-source', 'crowdsourced', and 'crowdsource', identify the same corpus and therefore can be discarded.

The result of the search was 172 articles. After removing the duplicates, editorial introductions, posters, tutorials, workshop summaries, and articles that just mention the search keywords and that are not relevant to discovering expertise, a total of 96 articles remain in the initial pool of resources.

II. Filtering Articles and Additional Search

We applied the following inclusion criteria: (i) articles concerned with selecting experts to perform a task; (ii) articles measuring expertise of workers in a platform or members in a community; and (iii) articles measuring reputation for workers in a platform or members in a community. We also applied the exclusion criteria for articles with the main focus of relying on self-assessments (e.g., self-reported skills or qualities).

By applying the inclusion and exclusion criterion, the number of articles becomes 23. We performed additional literature search in a less systematic approach, focusing on articles that we were already aware of ourselves, articles found by a full text search using Google scholar, and relying on back snowballing where we looked at references cited in the articles we have. We believe that this addition has provided a more comprehensive review of scientific contributions in this area. The final number of articles that were reviewed was 34.

III. Data Extraction and Synthesis

We reviewed the articles in order to identify the indicators of expertise relied on in literature. The first author codified the review articles and the other authors checked the extracted data and modifications were made according to the comments. In any cases of disagreement on data extraction, the three researchers discussed until reaching consensus.

We synthesized the extracted data in order to answer the research question. We reviewed the extracted data and built a list of the indicators of expertise that were relied on in each article. We then merged the indicators that were related together and narrowed down our list of indicators to five indicators.

3.2. Emerging Indicators from the Literature Review and Online Contents Investigation

Five indicators to expertise have emerged from the literature (illustrated in Fig. 1). In the following, we briefly discuss each of these indicators, and we explain the resources that motivated us to derive them as appropriate indicators. Table 1 illustrate a summary of the indicators and the supporting research work.

Figure 1. Selecting Vulnerability Discovery Professionals (Expertise Indicators)

I. Certification

This is one of the most known indicators to knowledge in both educational bodies (academic record certification) and industrial bodies (industrial professional certifications). Relying on a proof issued by a trusted party (whether being tested or not) the crowdsourced task owner will know that the VDPro have the required knowledge to perform the task of discovering

vulnerabilities. Crowdsourcing platforms like MTurk allow the task requester to issue a qualification test for each of the crowd workers to assert their eligibility to perform the task. Online freelancers market UpWork (previously known as oDesk) rely on certifications to assess the competency of the freelancer along with online testing for the skills. Similarly, Synack which is a crowdsourcing vulnerability discovery platform, administer to VDPros a written and practical evaluation to certify that they are eligible to join the platform [24].

Ejiaku et al. in their work discussed the importance of Information technology certification and its impact on the recruitment process [25]. In a similar context, Bishop and Frincke in their work discussed general professional certifications and compared them with academic certifications. They also looked at how each of these certifications reflect the knowledge required for certain jobs and their effect in increasing the possibility of success on the job [26]. On the other hand, McGill et al. addressed the importance of recertification in order to ensure the validity of the certification and express the true knowledge and expertise of the certification holder [27].

Kanij et al. in their work surveyed several software testers for factors they considered most important to the task of software testing [28]. In their study, 56.7% of the respondents to the questionnaire indicated that they had done training or certification in software testing in the last five years and the majority of the participants found certification and training useful. In a similar context but in different field, Botella et al. identified certification as a key requirement for persons who aim to become an expert in usability evaluation [29].

II. Referrals

Referrals or recommendations are also well known in the domain of recruitment as a way to gather and confirm expertise. Many organizations rely on referrals and encourage the employees to recommend potential candidates for employment. Salesforce has indicated in its official blog that its main strategy for recruitment for 2015 is based on referrals [30].

Many online professional communities and professional social platforms have recognized the importance of referrals in determining expertise. For instance, professional social media platform LinkedIn facilitates endorsements in order to tag a specific skill to the professional [31]. CoderWall allow users to endorse and assert any skill that the profile owner has mentioned in her profile [32]. Repcoin a novel reputation market platform rely solely on recommendations to build up the users' reputation of having expertise in a certain field, and as it is built on top of the bitcoin technology the platform facilitates having the reputation as a portable accreditation that could be presented anywhere [33].

Some research has addressed the impact of referral as an indicator to expertise. Schall et al. in their research introduced a context-sensitive trust-based algorithm called ExpertHITS inspired by the concept of hubs and authorities in Web-based environments [34]. They rely mostly on referrals from experts to delegate a task to another expert in her network. In the same context, Zhang et al. proposed a mechanism for finding experts by a referral chain from the initiator to the expert [35]. Similarly, the work of Pushpa et al. has addressed searching for expertise in a network of co-authorship where a query to find an expert is seen by an agent that manages the expertise model of the users. It thereby makes a decision to forward the request to relevant experts in the neighbourhood of co-authors if the query is not matched [36].

Within the enterprise work environment, finding an expert to perform certain task has been a problem addressed by several researchers. Braub et al. have proposed people tagging where they rely on employees to tag certain skill or knowledge to their colleagues [37]. Ghosh et al. have proposed relying on user curated Twitter lists to determine topic experts, since users include experts on topics that interest them in their lists [38].

III. Performance Review

Reviewing the performance of crowd workers has been relied on as an indicator for knowledge and expertise in many crowdsourcing platforms [39,40]. Similarly, the performance of VDPros within the crowdsourcing vulnerability discovery platform could be used to assert their level of expertise (reflected by a high reputation score). BugCrowd calculates a reputation score for each VDPro according to the number of discovered vulnerabilities, the speed of discovery, and abiding by the rules of the vulnerability discovery program [41]. Similarly, Cobalt (former known as Crowdcurity) relies on the same aspects but they add to them the feedback from the task requester and the quality of the vulnerability discovery report [42]. Hackerone in addition to relying on the number of the vulnerabilities discovered, speed and abiding by the rules, they add extra points for VDPros who acquire a bonus credit from the task requester (e.g., for clever vulnerabilities, or critical ones) [43].

Zhao et al in their work highlighted the performance of VDPros in a web vulnerability disclosure program as a measure of their expertise [44]. Similarly, Algarni and Malaiya in their work relied on the performance of VDPros to define the top discoverers of vulnerabilities [45]. Anvik et al. proposed an approach that apply machine learning algorithm to an open bug repository to recommend the suitable developer to fix a software bug according to their past performance [46]. In the same context, Hassan and Curry in their work proposed extracting the crowd worker expertise from the previously accomplished tasks of certain type (e.g. image tagging) in order to predict the performance of the crowd worker in tasks of different types (e.g. image recognition) [47]. Sarasua and Matthias however, proposed the use of a cross-platform Curriculum Vitae in their work to record the crowd worker expertise and skills according to the performance of the crowd worker in several crowdsourcing platforms [48].

Ebay, as an online market, set reputation for each of the sellers according to the quality of the goods they provide and their performance (e.g. speed of delivery, how they handle returns) and accordingly they categorize the seller as being expert [49]. Resnick and Zeckhauser in their work looked at the reputation system in eBay and how predictive it is for future performance of the sellers [50]. Similarly, UpWork[5] rely on the freelancer's job success rate and client feedback to assess her expertise.

IV. Artefacts

The artefact could be in the form of a software code in GitHub, an answer to a question in a Q&A platform like Stackoverflow, or a published article. Vasilescu et al. have looked at the relation

[5] www.upwork.com/i/how-it-works/client/

between the activity of users in Stackoverflow and GitHub, they have discovered that active GitHub committers ask fewer questions and provide more answers than others in Stackoverflow [51]. In the same context, Hauff et al. in their work proposed an approach for matching job advertisements with expert developers according to their artefacts contributed in GitHub [52].

Table 1. Indicators and Supporting Articles

Indicator	Supporting Articles
Certification	[9] [11] [21] [34] [46] [62]
Referral	[12] [26] [44] [52] [56] [71] [75] [76]
Performance Review	[2] [5] [28] [32] [33] [53] [54] [57] [64] [69] [72]
Artefacts	[8] [15] [16] [23] [30] [37] [50] [63] [65]
Association	[25] [43] [54] [58] [61] [68] [70] [74]

Kolari et al. in their work looked at blogs related to certain enterprise as a source of evidence of that expertise for potential employees. The authors investigated whether depending on blogs has similar effect as depending on emails but with less privacy concerns and they even have the added value of allowing implicit voting via comments by the community [53]. Balog et al. have presented strategies for finding experts relying on document repositories in the enterprise [54]. In their work, they have proposed two approaches, the first determines the employee expertise based on the documents (e.g., reports, manuals) that they are associated with, whilst the second locates documents based on topic, and then finds associated experts.

Pelechrinis et al. [55] proposed an approach to infer expertise and reliability of the answers provided in a Q&A platform depending on the pattern of the authors answering to other questions. The authors investigated users' consistent responses to questions related to particular topic and their expertise in that topic, as well as determining whether the user is answering too many unrelated topics, it would therefore be safe to assume they are an amateur in each of those areas.

Demartini investigated whether in order to find experts with certain knowledge, it is better to look at where knowledge is created. The author proposed an algorithm to find experts according to their contribution in Wikipedia [56]. Chang et al. in their reputation based access control for crowdsourcing platform relies on the quality of the artefacts submitted by the crowd worker in order to assert the quality of the worker herself [57].

In the academic field, academics have been evaluated by their publications such as the number and quality of publications, number of citations, and the h-index. Tang et al. have presented an academic search system, called ArnetMiner which relies on the publication of academics as an indicator to their knowledge and expertise in a particular field amongst other sources [58]. Similarly, Freund et al proposed an expert-finding system in the academic domain where they also rely on the publications of academics as one of the measures of expertise [59].

V. Association

Relying on associations to determine expertise is an approach that is recently emerging. Fu et al. in their work proposed strategies to discover the associations amongst people from emails and Web pages. Relying on these associations, a candidate can acquire extra expertise probability from a reliable expert who has strong relationship with the candidate [60]. Li et al investigated documents' co-authorship as a relation between experts, and they assessed if the probability of being an expert increases if they have co-authored a topic relevant document together with a well-known expert [61]. Similarly Zhang et al. proposed a propagation-based approach to expert finding in a social network where they relied on local information (e.g. published articles in certain topic) and then propagated possible expertise of a candidate according to their relationships (e.g. co-authored or supervised) [62].

Daniel Schall introduced a ranking approach called Dynamic Socially-Aware PageRank where the expert ranking depends on context aware interactions. As an example, an expert of certain topic may attract a large number of help requests in regard to that topic [63]. In the same context, in [64] the authors investigated the usage of social relationships in a microblogging platform as a potential evidence that a person is a real expert. In the sense that if user A who is an expert follow user B, then user B is most likely to be an expert too. Similarly, Zhao et al. in their work have addressed the problem of lacking information about the crowd worker's past performance (cold-start worker). They proposed an approach to determine the expertise of the worker from social networks, such as, if the cold-start worker was associated with another worker who is considered to be an expert then the cold start worker is also to be considered and expert [65].

Pujol et al. in their work proposed an algorithm called NodeRanking, which rely on modelling the community as a social graph relying on relationships. Such as, being mentioned in the user's website or having email exchange between the two users. Their approach is to infer expertise according to the degree of connectedness of each of the nodes in the graph [66]. Suryanto et al. looked at Q&A platforms and proposed that users who ask high quality questions that attract experts are considered experts and users who provide answers to questions asked by experts are considers experts as well [67].

4. Methodology

In our study we rely on exploratory research methods [11,12]. Our study consists of three phases of data collection: (i) Literature review and online content investigation; (ii) Interviews; and (iii) Online surveys. Figure 2 illustrates the phases of the study. Our methodology consisted of data collection and iterative phases of data analysis.

In the first phase all the authors contributed to the discussion and analysis of the literature, and online content following the SLR guidelines [10]. For the second phase, the interview questions were discussed by the authors and refined through multiple iterations. The interviews were conducted by one of the authors. Summaries of the interviews were discussed by the authors to draw insights and identify the key emerging themes. For the third phase, the survey questions were designed by the same author who conducted the interviews. The survey questions were revised through multiple iterations of refinement by the authors and three external VDPros before being published. This helped us refine the terminology used in the survey questions and to ensure the study was clear and

motivating to VDPros. The results of the survey were summarized and discussed by authors. In this section, we discuss the research questions, study design, data collection methods and data analysis approach.

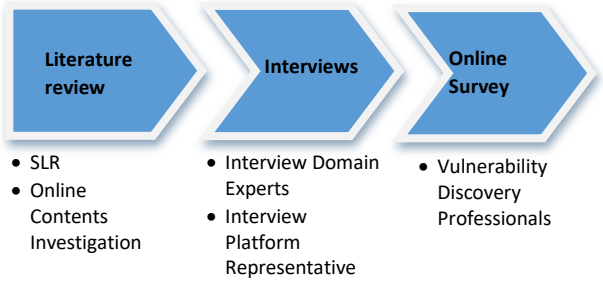

Figure 2. Stages of the Study

4.1. Study Design

In the first phase of our study, we reviewed scientific literature using systematic literature review methodology [10]. As mentioned above, we also investigated a diverse range of online contents (e.g., companies' blogs, experts' blogs, and technological articles) [22]. We utilized the findings in this initial phase to refine our interview and survey questions for the second and third phases. The results of the initial phase were already presented in the previous section.

In the second phase of our study, we interviewed 32 participants. We also approached one of the platforms for crowdsourcing vulnerability discovery for additional insights. Through iterative analysis of the data collected from this phase we were able to gain better understanding of what domain experts perceived as indicators to measure expertise in VDPros and what are the reliable sources to extract them.

In the third phase, by relying on the results of previous phases, we approached VDPros involved in the task of vulnerability discovery with an online survey. The purpose of this survey was to acquire the opinion of VDPros who are involved in the task of vulnerability discovery – and to thereby validate our findings from previous phases.

4.2. Study Design

In the first phase of our study, we reviewed scientific literature using systematic literature review methodology [10]. As mentioned above, we also investigated a diverse range of online contents (e.g., companies' blogs, experts' blogs, and technological articles) [22]. We utilized the findings in this initial phase to refine our interview and survey questions for the second and third phases. The results of the initial phase were already presented in the previous section.

In the second phase of our study, we interviewed 32 participants. We also approached one of the platforms for crowdsourcing vulnerability discovery for additional insights. Through iterative analysis of the data collected from this phase we were able to gain better understanding of what domain experts perceived as indicators to measure expertise in VDPros and what are the reliable sources to extract them.

In the third phase, by relying on the results of previous phases, we approached VDPros involved in the task of vulnerability discovery with an online survey. The purpose of this survey was to acquire the opinion of VDPros who are involved in the task of vulnerability discovery – and to thereby validate our findings from previous phases.

4.3. Data Collection

For the interviews, we invited participants via e-mail invitation, LinkedIn messages, and advertising our study in information security groups on LinkedIn (e.g., Information Security Community [6]). We also leveraged snowballing to get recommendation from people about their colleagues that would also be interested in participating [68]. Snowballing was the most effective approach since direct approaches may be considered spamming – especially as spamming attacks are well known by security experts who are much more cautious to accept invitations from an unknown party. The interviews lasted between 30-60 minutes and were conducted via Skype, Google Hangouts, telepresence systems, phone calls, and e-mail exchange. Audio was recorded when the interviewee consented and notes were taken. The interviews were semi-structured based on the findings of Phase 1 of the study, and the interviewer asked additional questions when appropriate. Summary of interview questions are outlined in Table 2 and the full interview guide is published online[7].

We also interviewed a senior representative from one of the platforms that supports crowdsourcing vulnerability discovery. This interview provided us with indispensable insight from a typical platform that is responsible for managing vulnerability discovery programs. Hence, they provided very valuable input in accordance to their experience from interacting with over 250 clients and thousands of VDPros.

Finally, we targeted VDPros involved in the task of vulnerability discovery with an online survey. We designed the questions of the survey based on the results from Phase 1 and 2. We shared the preliminary survey with three of the VDPros to acquire feedback before releasing it. The feedback helped us to modify some of the terminology used and modify some of the questions that were considered vague. The survey included both closed and open questions [69]. While closed-form questions are easier to summarize, open questions allowed respondents to include their remarks and comments.

We analysed the responses by applying structural codes and then aggregating the results for common trends [70]. Most closed-form questions used a Likert scale [71] with five possible responses from "Very Important" to "Not Important". The actual form of the survey is published online[8]. We sought participants via direct e-mail or tagging on twitter, LinkedIn messages, posting in Bugcrowd forum[9], Offensive Community site[10], and LinkedIn information security groups.

[6] www.linkedin.com/groups/38412

[7] Interview Guiding Questions: https://goo.gl/jVyOCe

[8] Survey Questions: https://goo.gl/forms/kXjIHjAEkZdwUvIt2

[9] forum.bugcrowd.com

[10] offensivecommunity.net/

Table 2. Summarized Interview Question for the Domain Experts

Characterization

1. Tell us about your organization (size, business area)?
2. What is your role in the organization?
3. Are you familiar with the approaches for vulnerability discovery (management, practices, decision making)?

Indicators to Expertise

1. What are the Indicators you rely on to measure expertise in vulnerability discovery professionals involved in the task of vulnerability discovery?
2. How much do you value certification as an indicator that could be relied on to measure the expertise in Vulnerability discovery professionals?
3. How much do you value referrals to measure the expertise of vulnerability discovery professionals?
4. How much do you value performance as an indicator to the expertise of vulnerability discovery professionals?
5. How much do you value artefacts generated by the vulnerability discovery professionals as an indicator to their expertise?
6. How much do you value association networks as an indicator to expertise for vulnerability discovery professionals?

Reliable Sources

1. What sources of certification you consider reliable to indicate expertise of vulnerability discovery professionals?
2. What forms of referrals do you rely on to measure expertise in vulnerability discovery professionals?
3. How much do you value past performance as an indicator to measure expertise in vulnerability discovery professionals?
4. What are the artefacts you think would be a reliable source to express expertise?
5. What forms of association do you think are reliable to express expertise?

4.4. Participants

Due to the time constrains and the exploratory nature of the study, we followed the recommendations of Adler and Adler and aimed for a sample size between 12 and 60 [72]. We reached saturation in the opinions of the interviewees and this gave us confidence that we captured sufficient depth to enable us to assert the solidity of our findings. We designed a comprehensive set of questions for the interviews and surveys relying on the literature review and online contents.

I. Interview Participants

Table 3 shows the details about the domain experts who have participated in the interviews. We refer by domain experts to personnel who oversee the security posture of organizations. Although they do not perform vulnerability discovery tasks as their daily job, they nonetheless influence (directly or indirectly) setting the criterion for selecting VDPros to perform the task.

For the purpose of analysis, participants are referred to by an anonymous identifier (P#). We invited participants from different disciplines (e.g., finance, entertainment, telecoms, etc.), so that we capture a diverse set of views sourced from different approaches for measuring expertise. Among the interviewees P3, P11, P13, P14, P17, P24, and P29 have experience in crowdsourcing vulnerability discovery. Furthermore, we interviewed a representative of one of the crowdsourcing vulnerability discovery platforms.

Table 3. Information about Interview Participants

P#	Industry	Size of Company (# Employees)	Position
1	Entertainment	450-600	CISO
2	Finance	350-500	CIO
3	Security Consultancy	30-50	Senior Security Consultant
4	Backup solutions	30-50	Quality Assurance manager
5	Networking solution	Global company	Senior Security Analyst
6	Multi-disciplinary	Global company	Security Testing Lead
7	Telecommunication	350-500	Information Security Expert
8	Telecommunication	350-500	Security Analyst
9	Payment systems	100-150	CISO
10	SaaS provider	30-50	Security Team Leader
11	Multi-media	100-150	Platform Security Lead
12	Security Consultation	50-100	Senior Security Engineer
13	Charity Foundation	15-30	Internal Security Assessor
14	Mobile Gaming	10-15	Security Lead
15	Video production	50-100	Principle Security Advisor
16	SaaS provider	100-150	Senior Security Engineer
17	Social Media	50-100	Security Lead
18	Financial Institute	150-250	CIO
19	Security Consultation	25-50	Penetration Testing Lead
20	Security Software	15-30	CTO
21	Medical	250-300	CISO
22	Education	300-500	IT Manager
23	Security Consultation	30-50	Senior Consultant
24	Software Provider	450-500	Security Architect
25	Financial Institute	250-300	Security & System Engineer
26	SaaS provider	250-300	Senior Security Analyst
27	Telecom Equipment	Global Company	Security Evaluation lead
28	Transportation	Global Company	Security Engineering Mngr
29	Bitcoin Exchange	150-200	Product Manager
30	SaaS Provider	25-50	Cyber Security Lead
31	Software Provider	400-450	CSO
32	SaaS Provider	600-700	CTO

II. Online Survey Respondents

We targeted VDPros who have been involved in the task of vulnerability discovery. We also targeted members of crowdsourcing vulnerability discovery platforms (Hackerone, Bugcrowd, and Cobalt), members of online cyber-security community along with relying on snowballing as well. The number of participants was 78 but we excluded 12 responses. In particular, 8 responses were submitted by participants that were not directly involved in vulnerability discovery; we used test questions in the

survey to confirm that. Another 4 responses were apparently artificial, since we noticed that these participants selected the first choice of all the questions, which would otherwise create incorrect inconsistencies in the data.

For the demographics of the participants in the online survey, 45% of the participants were aged 18-25 years old, 30% were aged 25-32, 13% were aged more than 33 years old, and the rest were aged less than 18 years old. With respect to location, the majority of participants were from the United States 37%, 21% were from India, 12% were from Pakistan, 9% were from Europe, 7% from Egypt, and the rest did not share their country of origin. Furthermore, 45% of the participants work in the security industry, 30% were students, 10% work in system administration, and the rest work in software development.

4.5. Data Analysis

We analysed the data following the qualitative data analysis guidelines [12,13]. The following stages were included in the analysis: (i) Transcription of the data recorded; (ii) Organization of the transcripts data and notes taken into easily retrievable sections; (iii) Getting more familiar with the data through reading and re-reading and writing down notes and summaries; (iv) Labelling data segments (we relied on structural coding to annotate the responses[70]); and (v) contrasting different views of the interviewees and re-coding to develop a more refined understanding of how the indicators are valued by the participants. As part of the analysis, similarities and differences of the compiled codes were clustered together in order to create categories. Conceptual saturation was reached when no new categories were generated.

5. Findings

In this section, we address the research questions and present the themes that emerged from our study. We also discuss the findings from the interviews and online survey, and annotate with selected quotes to help appreciate how we derived our conclusions.

5.1. The Interviews of Domain Experts

I. Indicators to the expertise of Vulnerability Discovery Professionals as Perceived by Domain Experts

There are four indicators for expertise that our study participants agreed on: (i) Certification; (ii) Referral; (iii) Performance review; and (iv) Artefacts. Relatively, there is weak support in regard to relying on association (see Fig 3).

Certification as an Indicator to Expertise: 75% of the interviewees value certification as an important indicator of expertise. One of the interviewees described it as a clear indicator to commitment "To get a formal education in the field from an academic body or being certified to have knowledge in the field via acquiring an industrial certification, it shows that the VDPro is passionate and committed enough to invest this much time and effort" [P1]. The remaining 25% of the interviewees did not consider certification as an indicator to expertise for VDPros. One of the interviewees stated that "Some skills required for the VDPro involved in the task of vulnerability discovery are not taught in universities or professional courses" [P19].

Referral as an Indicator to Expertise: Referral is highly valued by 90% of the interviewees since it provides attestation for the expertise of the VDPro from a person who they have known or worked with. One of the interviewees mentioned "Knowing that the person doing the referral has interacted with and observed the VDPro is a great way to measure expertise" [P1]. One of the interviewees indicated that when they trust a person doing the referral it would give very valuable credit to the VDPro especially since roles in the security field are sensitive ones. "Knowing and trusting the person doing the referral would mean minimized risks for us" [P8]. 57% of the interviewees mentioned that even if they do not know the person making the referral, if s/he is well-known in the industry, this would add substantial credit to the VDPro.

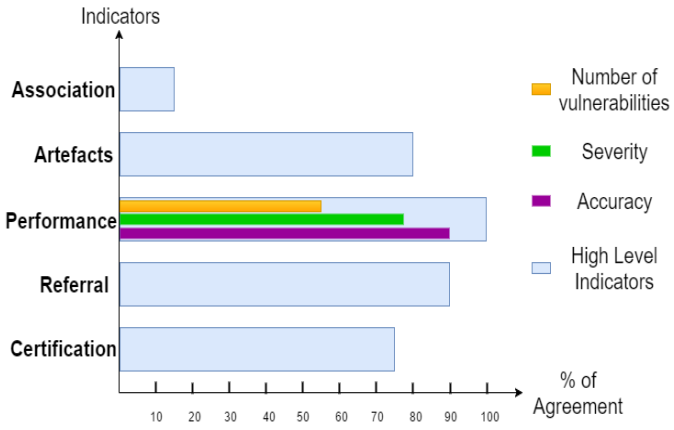

Figure 3. Indicators of Expertise (Domain Experts' Perspective)

Performance Review as an Indicator to Expertise: We found 84% of the interviewees assigned performance review as the top indicator while the remaining 16% considered it of high value. An interviewee stated, "If the VDPro has discovered vulnerabilities successfully many times before then why would s/he stop doing that now?" [P7]. Some considered it of high importance since it means the VDPro have used their knowledge in performing the task. One interviewee said, "Showing knowledge is good, but showing how the knowledge is implemented is more impressive" [P9].

We investigated further and asked about how the performance of VDPros is considered within the task of vulnerability discovery. Several themes have emerged. The first theme is considering the severity of the vulnerabilities discovered. 78% of the interviewees agreed that the severity of the discovered vulnerabilities is the best measure for performance. One of the interviewees mentioned "Discovering a critical vulnerability, that could cause a great loss or embarrassment to the company, is highly appreciated by any company and it means that the VDPro is an expert in what s/he is doing" [P15]. While the remaining 24% of interviewees considered it a good measure. Another theme is considering the number of vulnerabilities discovered by the VDPro. 56% of the interviewees considered it an important measure to performance. While 24% considered it somewhat an important measure. 19% considered it useful only when a good number of vulnerabilities discovered were of high severity. An interviewee has mentioned "If the VDPro has discovered thousands of vulnerabilities that of very low severity for me it does not tell they are good in what they are doing. May be they just have a good automated tool" [P12]. The last theme that has emerged is considering the accuracy in discovering vulnerability (i.e., the ratio of the valid vulnerabilities reported over all the vulnerability reports submitted by the VDPro) as a measure for performance. 90% of the interviewees agreed that accuracy is an important measure for the performance of the

VDPros. One of the interviewees mentioned "That is why we target VDPros with good expertise, we need to minimize the false positives since they consume a lot of resources" [P14].

Artefacts as an Indicator to Expertise: 81% of the interviewees agreed that contributed artefacts are important indicator for expertise. There was a contrast in opinions about the type of artefacts. 40% of the interviewees considered code artefacts and contribution to open-source projects to be the most valuable artefacts. An interviewee stated, "Developing your own testing tools is a clear indicator that you deeply understand the concepts" [P4]. Another 24% of the interviewees considered published articles, and blogging about vulnerabilities as a more important type of artefacts. One of the interviewees said "VDPros are supposed to be good in breaking systems and understanding how they work. I would consider a VDPro with good publications about security as an expert in the field" [P6]. While the opinion of the rest of the interviewees is that it is more effective to look at all artefacts and assess the quality as whole. An interviewee mentioned "Whether it is a publication, source code, or a blog, if I see many experts citing the work or using the open-source tool then it means the VDPro is not generating useless artefacts and that s/he is acknowledged for the contribution" [P4].

Association as an Indicator to Expertise: 84% of the interviewees considered association as a less important indicator of expertise of a VDPro. Sources like membership to a community, affiliation with a group, professional relationships with an expert were considered as indicators of passion and interest rather than expertise. One of the interviewees mentioned "Being affiliated with a group focused on security, indicates passion but not necessarily expertise" [P3]. Another interviewee shared with us "if the VDPro is well connected it could be useful since it indicates that the VDPro could reach out to experts if needed, but I would not say it is a must have" [P6]. The remaining 16% of the interviewees considered it a good indicator to expertise. An interviewee stated, "It is definitely not enough to see the number of connections or followers to determine expertise, but when I see interactions such as a post in LinkedIn being discussed or a tweet being retweeted by experts in the field then it must be that they acknowledge what the VDPro has produced" [P10].

II. Perceived reliable Sources to Extract the Expertise Indicators of Vulnerability Discovery Professionals?

As sources of reliable certifications, 87% of the interviewees valued certification from industrial bodies (e.g., $(ISC)^2$) the highest, followed by certifications from academic bodies (e.g., Universities) which was selected by 72% of the interviewees. On the other hand, accredited MOOCs and online learning were considered good to show eagerness to learn but not as much credible by 75% of the participants in the study.

As sources of referral, all interviewees valued direct referral the most (i.e. a direct form of communication). We asked the interviewees if they would value indirect forms like LinkedIn endorsements or LinkedIn recommendations and 85% of the interviewees did not consider LinkedIn endorsements as reliable. While 62% of the interviewees thought that recommendations in LinkedIn could be useful. One of the interviewees shared with us "People in LinkedIn could just be nice and click 'endorse' for certain skills that appeared under the name of the VDPro, but making some effort and writing some detailed recommendation could be worthwhile to look at" [P4].

For performance review, 75% of the interviewees considered vulnerability databases to be the most reliable source. A vulnerability database is a platform to collect and disseminate information about discovered vulnerabilities (e.g., the National Vulnerability Database). Vulnerability Databases are considered an important source especially since they provide details of the vulnerabilities discovered (e.g., description of the vulnerability discovered, vulnerability severity score, references for security advisories related to the vulnerability). Furthermore, 83% of the participants in the study considered being mentioned in the hall of fame or 'thank-you' pages (e.g., Facebook white hat thanks page) is of high value. An interviewee said, "If the VDPro is mentioned in the hall of fame of Facebook or Google then they pack enough skills to be an expert" [P5]. Also 83% of the interviewees considered the measured performance within the crowdsourcing vulnerability discovery platform as a good source to extract expertise. While the rest agreed that it is a good source only if the crowdsourcing vulnerability discovery platform discloses the details of the discovered vulnerabilities by the VDPros. One of the interviewees stated, "With vulnerability databases, I can have a complete idea about the vulnerability discovered and its impact, but sometimes with the vulnerabilities discovered within a crowdsourcing vulnerability discovery platform we can only see limited information" [P16].

As reliable sources for artefacts, 39% considered GitHub to be the most reliable source. 60% considered security conferences, security forum, and blogs to be the most reliable sources. On another hand, 75% of the interviewees considered Q&A platforms (e.g., Security StackExchange) not very important sources. Additionally, 72% of the interviewees mentioned that they consider publishing exploits for vulnerability as a high-level indicator of expertise. An interviewee mentioned "By showing that you know how to exploit vulnerabilities, you show that you deeply understand how the vulnerability works and what its impact is" [P13].

5.2. The Perspective of Crowdsourcing Vulnerability Discovery Platforms

To gain additional insights into the perspective of the crowdsourcing vulnerability discovery platforms, we interviewed a representative of one of the platforms. The interview allowed us to gain some valuable insights about the programs the platform is managing and the importance of selecting VDPros. The platform's representative mentioned "Almost all of our customers now start with a private program with vetted VDPros before moving to a public program. It's only a small fraction of companies are ready to work with thousands of VDPros". About the effect of selecting VDPros in minimizing the false positives in reported vulnerabilities the platform's representative mentioned, "Public bug bounties is 95% noise and 5% signal. Unless you have a very mature security organization it is not recommended to use public bounties. Private bounties with vetted VDPros are the way to go"

Regarding the indicators of expertise, the platform relies on the past performance, artefacts and will also look at professional social network. It is quoted that: "We rely on past performance in private or public bounties, on our platform or other public bounty programs (Facebook/Google). Also, we check LinkedIn, GitHub profiles as well as conference talks"

When we asked for elaboration about the sources the platform rely on, the platform's representative mentioned: "We look at LinkedIn, GitHub, and Conferences. This is a manual process and it is on a case-by-case basis. E.g. a researcher might have given a

presentation on Blackhat but have no connections on LinkedIn etc. And the other way around".

About certification and whether clients ask for specific requirements, the platforms' representatives mentioned: "We typically don't see requirements like "university degree," but occasionally we will get requests based on geographical region, or professional accreditation (CISSP[11], CISA[12], etc.)".

5.3. The Online Survey of Vulnerability Discovery Professionals

In this section, we briefly discuss the results of the online survey. Figure 4 illustrates the high-level results of the online survey. 60% of the participants agreed that certification is an important indicator, while the rest considered it between somewhat important and useful. We asked about the sources of certification. 62% of the participants thought industrial certification were important, while only 30% considered certification from academic body as important. When we asked the participants about the importance of referral as a practice in the security domain, 72% considered it important while 11% considered it useful and the rest considered it not important. We also asked the participants their opinion about relying on artefacts as indicators to expertise: 83% considered it an important indicator. 74% considered publishing vulnerability exploits as an important indicator to expertise. Followed by, publishing security articles and presentations or videos both at 63%. They considered the least important source of artefacts is answering questions in Q&A platforms (e.g., Security StackExchange) where only 30% considered it "somewhat important". Regarding performance review, all participants considered it very important. 84% of the participants considered the number of vulnerabilities discovered by a VDPro as a very important measure for performance. While 75% considered the severity of vulnerability discovered as an important measure to performance. On the other hand, 56% considered the accuracy in discovering vulnerability as an important measure to performance, 36% considered it useful but not important, and 8% considered it not important. Finally, we asked the participants about how much they value associations and being part of communities or groups as an indicator to expertise. 81% of the participants considered it an important indicator, and the rest considered it somewhat important.

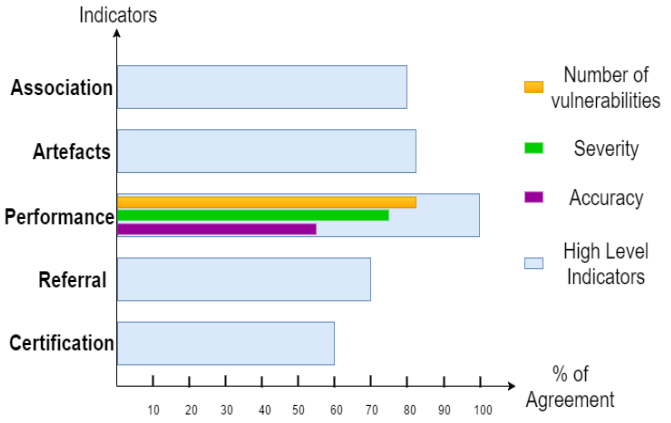

Figure 4. Indicators of Expertise (Vulnerability Discovery Professionals' Perspective)

5.4. Analyzing Public Data about Vulnerability Discovery Professionals

As an additional experiment, we have looked at public information of 100 VDPros from the leader boards (among the top 100 ranked in each platform) of three different crowdsourcing vulnerability discovery platforms: Bugcrowd, Cobalt, and Hackerone. For each VDPro we aggregated additional publicly available data from several external sources (LinkedIn, Twitter, GitHub, StackExchange (SE), blogs, and personal/professional websites). In this section, we discuss the result of the experiment in the light of the indicators we have discovered and illustrate some of the insights we have acquired. As we have relied only on the public data, there are possible threats to validity. The vulnerability discovery professionals could be providing falsified data to the public or hiding some information. We believe that these cases do not have much impact on the analysis; VDPros in the mentioned platforms seek to maintain good reputation to the potential business partners. Additionally, we aggregate the information from multiple sources to minimize this concern.

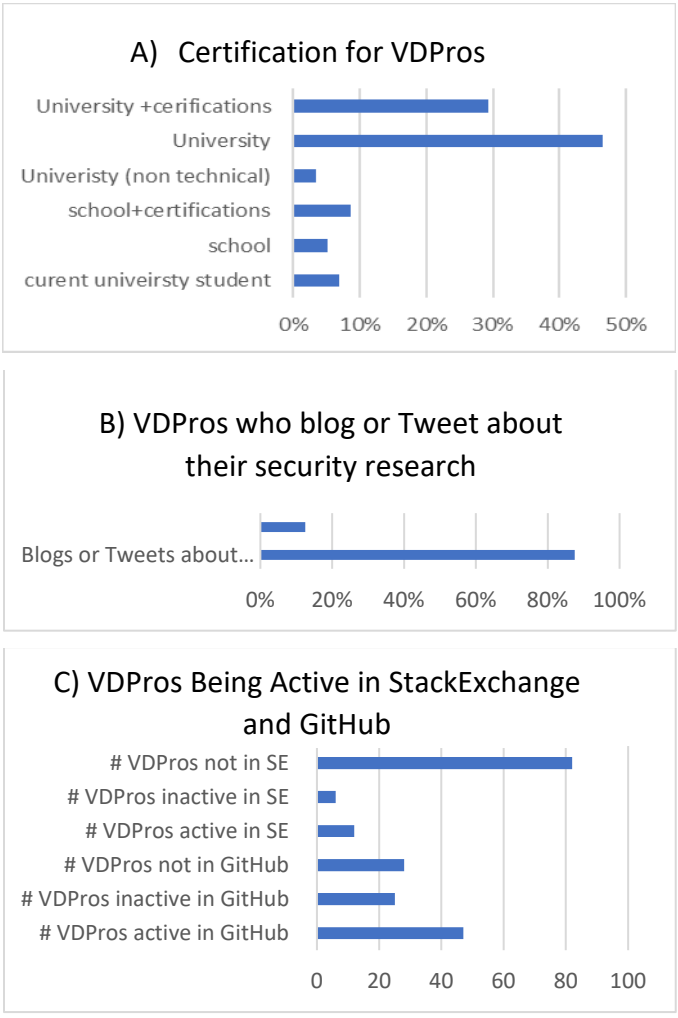

Figure 5. Insights from the public information about Vulnerability Discovery Professionals

[11] www.isc2.org/cissp/default.aspx

[12] http://www.isaca.org/certification/cisa-certified-information-systems-auditor/pages/default.aspx

We have noticed in this analysis of public data, 76% of the VDPros are university graduates of computer and technology fields. In addition (42%) of the VDPros have certifications in the security field (e.g., Certified Ethical hacker). We also noticed that 88% of the VDPros tend to publicly publish information about the vulnerabilities they discover, methods of discovery, and exploitation of vulnerabilities. They rely on their personal blogs or Websites in addition to twitter. For artefacts created in platforms like GitHub or StackExchange, we discovered that 82% of the VDPros have no profiles in this platform or have inactive accounts. While for GitHub, 73% of the VDPros have profiles in Github and 47% are active. Figure 5 illustrates the insights from the public data about the VDPros.

6. Discussion

In this section, we: (i) summarize our findings; (ii) provide some recommendations based on these findings; and (vi) discuss potential limitations of our study.

6.1. Analysis of our Findings

It is clear from our findings that domain experts and VDPros disagree in regard to association being an indicator to expertise. A possible reason is that the cyber security community is a strong community where teaming with other VDPros and pair practice is considered one of the factors to success[13]. So, in the opinion of VDPros, being associated with expert VDPros reflects expertise.

We observed from the results of the interviews that participants in our study valued certification from industrial parties more than from academic bodies. According to the input from the representative of the crowdsourcing vulnerability discovery platform we have interviewed, the platform has received requests to involve VDPros with industrial certifications before, but never a request to specify academic certification. A possible reason for this is that some academic bodies are not progressing as fast as the industrial bodies in the field of information security. One of the interviewees mentioned "some universities are not really up to date with the requirements of the security industry" [P5].

For measuring the performance of VDPros within the task of vulnerability discovery there are contrasting views in the way the number of vulnerabilities and accuracy thereof are valued. A possible reason for this contrast could be that domain experts look at how to utilize the resources to verify and mitigate the vulnerabilities. If VDPros are submitting a high number of vulnerability reports with only few of them being valid vulnerabilities, this defeats one of the purposes in selecting VDPros with the required expertise. On the other hand, VDPros may have the impression that all the vulnerability reports they submit (even if the task requester considered them invalid for being duplicates, or out of scope) illustrate their expertise. The reason for this is because they have already demonstrated what is required to discover the vulnerabilities.

6.2. Recommendations

I. Vulnerability Discovery Professionals Analytics

Information about VDPros are scattered within different sources mentioned earlier (e.g., crowdsourcing vulnerability discovery platforms, professional communities, research articles, social coding platforms, Q&A sites). Therefore, in order to facilitate effective selection of VDPros to perform the task of vulnerability discovery, we recommend leveraging VDPros analytics. Analytics rely on characterizing features, which are variables that grasp and encode information from raw or curated data, thereby enabling to derive meaningful inferences from data. Security expertise features could be extracted from within vulnerability discovery platforms (e.g., number of vulnerabilities discovered in the platform) and external sources (e.g., number of contribution to security related projects in GitHub, number of contributed records in ExploitDB[14]). It is possible to extract security expertise features using crowdsourcing (e.g., crowdsourcing tasks to judge the quality of comments, questions, and answers by VDPro in a security forum like Bugcrowd Forum[15]), as well as automated extraction of security expertise features (e.g., relying on GitHub API calls to extract the number of security related code repositories). Properly leveraged efforts in vulnerability discovery professional and content analytics [73,74] could potentially allow improving the selection of VDPro, ensure compliance with governance rules, and provide insights into the task of crowdsourcing vulnerability discovery (e.g., work effort estimation and time tracking, VDPro interests, or malicious VDPros). In addition, we recommend providing task requesters with a customizable language to specify rules (i.e., selection expressions) based on indicators they deem relevant to include in the selection process, as well as indicator aggregation strategies. As an example, let us assume Organization-A has a business rule stating that when contracting external parties, they prioritize certification from academic bodies. The security officer responsible for the task considers the exploits published in ExploitDB along with the number of repositories in GitHub as indicators to expertise. In this case, the task requester may specify the following rule to acquire a ranked list of VDPros: (Have University degree) AND (Number of records in ExploitDB) AND (Number of repositories in GitHub)). Moreover, we expect new indicators to be incrementally introduced and customized. Security is a complex area and there could be several different types of indicators, from generic to more specific knowledge (e.g., operating systems, databases, devices, browsers, and frameworks). This rule-based language will allow for the specification of fine-grained selection indicators and hence will improve achieving a high probability in matching the task with the VDPro.

II. Harvesting Expertise Indicators Systematically

It is clear from our findings that it is difficult to homogenize the knowledge acquired from different sources (domain experts, VDPros, crowdsourcing vulnerability discovery platforms) in order to determine the indictors to measure expertise in VDPros. We propose to consider techniques that systematically harvest the knowledge from these different perspectives. We believe this could be done by leveraging a knowledge-driven approach to create a security expertise feature graph. We adopt the notion of a feature from the field of machine learning and data science [75]. The security expertise feature graph will serve to store available features for reuse as part of future curation processes. Organizing features into high-level indictors (e.g., certifications, performance review) and low-level features (e.g., security course, number of vulnerabilities discovered) will help simplify their representation,

[13] hackerone.com/blog/what-great-hackers-share

[14] A database of vulnerability exploits (www.exploit-db.com).

[15] forum.bugcrowd.com

and aid searching and using features and indicators. Experts in the security community would be invited to curate this graph by incrementally and collectively: adding new features (e.g., a sub-feature of "artefacts" could be the number of published security articles); vote up/down; or add resources and comments as annotations to these features. Moderators may be assigned to verify amendments to the graph, reject, or accept new features. A similar approach has been proposed by Mark Klien which is a large-scale argumentation system for decision making in large communities [76].

III. Cyber Security Competitions

We observed that there are contrasting views between the opinions of domain experts in industry and VDPros involved in the task of vulnerability discovery. In order to minimize this contrast, we recommend for industrial organizations and crowdsourcing vulnerability discovery platforms to strengthen their ties with VDPros. A possible step to achieve this is to rely on cyber security competitions (e.g., Google capture the flag[16], Cyber Security Challenge Australia[17]) while emphasizing on educating VDPros about important practices that are relevant (e.g., providing high rewards or bonus to the VDPros with higher accuracy, providing higher rewards for high quality vulnerability reports). Similarly, involving universities in these competitions will be more beneficial. This would not only attract perspective students to pursue careers in the security industry and educate them about the rights skills and practices, it would also increase the focus of academic bodies to progress further in the field. Bridging the gap between what students is being offered at universities in the field of cyber security and what the security industry needs.

6.3. Limitations and Threats to Validity

When investigating online content, it is possible to come across material that is biased or inaccurate (since blogs and online articles are not peer reviewed). In order to minimize any bias or inaccuracies, we checked the background of the author to ensure are based on expert opinions. We also excluded articles that are marketing motivated (i.e. promoting or demoting certain technique or approach for personal gain).

For the interviews, the number of participants is not high so it is possible that we missed some insights. We tried to minimize this threat by targeting various industries in order to obtain a broader view of possible opinions. We also considered only one participant from each organization. The strong background we acquired from performing the literature review and online content investigation helped us to formulate clear questions for the interviews and online survey. We observed saturation in the opinions of domain experts, and this indicated we have reached a proper depth to draw solid conclusions.

For the online survey, although we obtained responses from the VDPros, we cannot assume that this number covers all the possible opinions from the security community. One of the authors attended security conferences (e.g., Ruxcon[18], BSides[19]) to discuss with VDPros directly. While some feedback was acquired about the study from VDPros, it was mostly an off-the-record approach. We also reached out to some of the VDPros for clarification of their answers in the survey, and this provided us with some additional insights.

As the questions from the interview and the online survey relied on the results from surveying the literature and online contents, another threat of validity is having the study suffer from confirmatory bias. In order to minimize this threat, we started each interview with the open question of what the interviewees rely on to measure the expertise of VDPros and we also closed the interview with asking if there are any additional indictors we missed. Similarly, for the online survey we deliberately added some open-ended questions to allow the VDPros to add any additional indicators that were not mentioned in the survey. Some VDPros added comments but they fall within the same indicators that we have already included (e.g., publishing exploits, which we already considered part of the artefacts generated by the VDPro).

Another important aspect to consider in the experiment we conducted for analysing public profiles of VDPro is time constrains. As a result, we were only able to look at a limited number of vulnerability discovery professionals; there could thus be insights we did not acquire due to this limitation. We intend to do a large-scale analysis and survey as future work.

7. Conclusion and Future Work

The contribution of this research is threefold: (i) Exploring the indicators to measure expertise of VDPros involved in the crowdsourced task of vulnerability discovery; (ii) Discovering the reliable sources that can be used to extract information about these indicators; and (iii) Recommendations to improve the selection process for VDPros involved in tasks related to discovering vulnerabilities. Our findings will help in the process of selecting VDPros that are more likely to contribute higher quality outcomes in this type of task.

We observed that domain experts and vulnerability discovery professionals, who have participated in our study, agree unanimously on certification, referral, performance review, and artefacts as good indicators to measure expertise in VDPros. There is some conflict about the importance of association as an indicator of expertise. We also observed that the opinions of domain experts and VDPros involved in the task of vulnerability discovery diverge in some areas. Being aware of this divergence is an important step to build a better ecosystem.

This is the first step for investigating the problem of selecting VDPros in Vulnerability Discovery Platforms. As future work, an additional aspect we will pursue is identifying malicious VDPros, and methods to successfully block them from participating in the task. We will develop techniques to model and capture VDPros behaviour patterns and abstract them into meaningful concepts (e.g., honest or malicious). We will investigate indicators and powerful aggregation functions and operators to characterize and identify malicious VDPros.

Acknowledgment

We greatly acknowledge all participants that provided valuable feedback and insight in interviews and online survey. A special appreciation to Jacob Hansen, Jeffrey Stapleton, Troy Hunt,

[16] https://capturetheflag.withgoogle.com/

[17] https://www.cyberchallenge.com.au/

[18] ruxcon.org.au/

[19] www.securitybsides.com/w/page/12194156/FrontPage

Dimitrios Stergiou, and Roee Hay for their valuable input and help in reaching out to members of the security community.

References

[1] Mortada Al-Banna, Boualem Benatallah, and Moshe Chai Barukh. 2016. Software Security Professionals: Expertise Indicators. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, 139–148. DOI:https://doi.org/10.1109/CIC.2016.030

[2] Mohammad Allahbakhsh, Boualem Benatallah, Aleksandar Ignjatovic, Hamid Reza Motahari-Nezhad, Elisa Bertino, and Schahram Dustdar. 2013. Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Comput.* 17, 2 (2013), 76–81. DOI:https://doi.org/10.1109/MIC.2013.20

[3] BugCrowd. 2017. *2017 State of Bug Bounty Report*. Retrieved July 26, 2017 from https://pages.bugcrowd.com/hubfs/Bugcrowd-2017-State-of-Bug-Bounty-Report.pdf

[4] Anne Edmundson, Brian Holtkamp, Emanuel Rivera, Matthew Finifter, Adrian Mettler, and David Wagner. 2013. An empirical study on the effectiveness of security code review. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 197–212. DOI:https://doi.org/10.1007/978-3-642-36563-8_14

[5] Edith Law, Conner Dalton, Nick Merrill, Albert Young, and Krzysztof Z. Gajos. 2013. Curio: A Platform for Supporting Mixed-Expertise Crowdsourcing. *First AAAI Conference on Human Computation and Crowdsourcing*. Retrieved from http://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7534

[6] LE Celis, SP Reddy, IP Singh, and S Vaya. 2016. Assignment Techniques for Crowdsourcing Sensitive Tasks. *19th ACM Conf. ...* (2016). Retrieved September 3, 2016 from http://dl.acm.org/citation.cfm?id=2835202

[7] Justin Scott Giboney, Jeffrey Gainer Proudfoot, Sanjay Goel, and Joseph S. Valacich. 2016. The Security Expertise Assessment Measure (SEAM): Developing a scale for hacker expertise. *Comput. Secur.* 60, (2016), 37–51. DOI:https://doi.org/10.1016/j.cose.2016.04.001

[8] DE Difallah, G Demartini, and P Cudré-Mauroux. 2012. Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms. *CrowdSearch* (2012). Retrieved September 5, 2016 from http://ftp.exascale.info/sites/default/files/Difallah_CrowdSearch_2012.pdf

[9] Munawar Hafiz and Ming Fang. 2016. Game of detections: how are security vulnerabilities discovered in the wild? *Empir. Softw. Eng.* 21, 5 (October 2016), 1920–1959. DOI:https://doi.org/10.1007/s10664-015-9403-7

[10] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.* 33, TR/SE-0401 (2004), 28. DOI:https://doi.org/10.1.1.122.3308

[11] Steve Easterbrook, Janice Singer, Margaret-Anne Storey, and Daniela Damian. 2008. Selecting Empirical Methods for Software Engineering Research. *Guid. to Adv. Empir. Softw. Eng.* (2008), 285–311. DOI:https://doi.org/10.1007/978-1-84800-044-5_11

[12] C.B. Seaman. 1999. Qualitative methods in empirical studies of software engineering. *IEEE Trans. Softw. Eng.* 25, 4 (1999), 557–572. DOI:https://doi.org/10.1109/32.799955

[13] Anne Lacey and Donna Luff. 2007. Qualitative data analysis. *Trowbridge, Wiltsh. Cromwell ...* (2007), 13–14.

[14] An ISACA and RSA Conference Survey. 2015. *State of Cybersecurity: Implications for 2015*. Retrieved August 22, 2016 from http://www.isaca.org/cyber/Documents/State-of-Cybersecurity_Res_Eng_0415.pdf

[15] Thomas W. Malone, Robert Laubacher, and Chrysanthos Dellarocas. 2010. The collective intelligence genome. *IEEE Engineering Management Review 38*, 38. DOI:https://doi.org/10.1109/EMR.2010.5559142

[16] Andy Ozment. 2004. Bug auctions: Vulnerability markets reconsidered. In *Workshop on Economics of Information Security (WEIS)*, 1–23. Retrieved from http://www.andyozment.com/papers/weis04-ozment-bugauc-slides.pdf

[17] Aron Laszka, Mingyi Zhao, and Jens Grossklags. 2016. Banishing Misaligned Incentives for Validating Reports in Bug-Bounty Platforms. . Springer, Cham, 161–178. DOI:https://doi.org/10.1007/978-3-319-45741-3_9

[18] Mingyi Zhao, Jens Grossklags, and Peng Liu. 2015. An Empirical Study of Web Vulnerability Discovery Ecosystems. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, 1105–1117. DOI:https://doi.org/10.1145/2810103.2813704

[19] Matthew Finifter, Devdatta Akhawe, and David Wagner. An Empirical Study of Vulnerability Rewards Programs.

[20] Mitre. CAPEC - CAPEC-152: Inject Unexpected Items (Version 2.9). Retrieved February 22, 2017 from https://capec.mitre.org/data/definitions/152.html

[21] C Okoli and K Schabram. 2010. A guide to conducting a systematic literature review of information systems research. *Sprouts Work. Pap. Inf. Syst* (2010). Retrieved February 20, 2017 from http://www.academia.edu/download/3250666/OkoliSchabram2010SproutsLitReviewGuide.pdf

[22] Nina Wakeford and Kris Cohen. 2008. Fieldnotes in Public: Using Blogs for Research. In *The SAGE Handbook of Online Research Methods*. 307–327. DOI:https://doi.org/10.4135/9780857020055

[23] J Freyne, L Coyle, and B Smyth. 2010. Relative status of journal and conference publications in computer science. *Commun. ACM* 53.11, (2010), 124–132. Retrieved February 21, 2017 from http://dl.acm.org/citation.cfm?id=1839701

[24] Synack. Join a Global Hacker Team | Synack Red Team. Retrieved February 22, 2017 from https://www.synack.com/red-team/

[25] SA Ejiaku and MA Badamas. 2010. An examination of information technology certification: A measure of professional qualification. *Bus. Res. Yearb.* 17, (2010), 119–125. Retrieved February 22, 2017 from http://lakishasimmons.com/images/publications/bothsimmonsandaiken_2010v1bry.pdf#page=132

[26] Matt Bishop and Deborah Frincke. 2004. Academic degrees and professional certification. *IEEE Security and Privacy 2*, 56–58. DOI:https://doi.org/10.1109/MSP.2004.91

[27] Tanya McGill and Michael Dixon. 2013. An investigation of the impact of recertification requirements on recertification decisions. In *Proceedings of the 2013 annual conference on Computers and people research - SIGMIS-CPR '13*, 79. DOI:https://doi.org/10.1145/2487294.2487310

[28] Tanjila Kanij, Robert Merkel, and John Grundy. 2014. A preliminary survey of factors affecting software testers. In *Proceedings of the Australian Software Engineering Conference, ASWEC*, 180–189. DOI:https://doi.org/10.1109/ASWEC.2014.32

[29] Federico Botella, Eloy Alarcon, and Antonio Peñalver. 2014. How to classify to experts in usability evaluation. In *Proceedings of the XV International Conference on Human Computer Interaction - Interacción '14*, 1–4. DOI:https://doi.org/10.1145/2662253.2662278

[30] Sarah Boutin. 2015. Behind the Scenes at Salesforce: Our #1 Recruiting Secret - Salesforce Blog. Jan. Retrieved February 22, 2017 from https://www.salesforce.com/blog/2015/01/behind-scenes-salesforce-our-1-recruiting-secret.html

[31] LinkedIn. Skill Endorsements - Overview | LinkedIn Help. Retrieved February 22, 2017 from https://www.linkedin.com/help/linkedin/answer/31888?lang=en

[32] A community of great programmers and their programming tips. Retrieved February 22, 2017 from https://coderwall.com/

[33] Repcoin: Find your expert. Retrieved June 12, 2015 from http://www.repcoin.com

[34] Daniel Schall, Florian Skopik, and Schahram Dustdar. 2012. Expert Discovery and Interactions in Mixed Service-Oriented Systems. *IEEE Trans. Serv. Comput.* 5, 2 (April 2012), 233–245. DOI:https://doi.org/10.1109/TSC.2011.2

[35] Lan Zhang, Xiang Yang Li, Yunhao Liu, QiuYuan Huang, and Shaojie Tang. 2012. Mechanism design for finding experts using locally constructed social referral web. In *Proceedings - IEEE INFOCOM*, 2896–2900. DOI:https://doi.org/10.1109/INFCOM.2012.6195723

[36] S. Pushpa, K. S. Easwarakumar, Susan Elias, and Zakaria Maamar. 2010. Referral based expertise search system in a time evolving social network. *Proc. Third Annu. ACM Bangalore Conf. - Comput. '10* (2010), 1–8. DOI:https://doi.org/10.1145/1754288.1754294

[37] Simone Braun, Christine Kunzmann, and Andreas Schmidt. 2010. People tagging and ontology maturing: Toward collaborative competence

management. In *From CSCW to Web 2.0: European Developments in Collaborative Design - Selected Papers from COOP 2008*, 133–154. DOI:https://doi.org/10.1007/978-1-84882-965-7_7

[38] Saptarshi Ghosh, Naveen Sharma, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. 2012. Cognos: crowdsourcing search for topic experts in microblogs. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, 575. DOI:https://doi.org/10.1145/2348283.2348361

[39] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. 2012. Task recommendation in crowdsourcing systems. In *Proceedings of the First International Workshop on Crowdsourcing and Data Mining - CrowdKDD '12*, 22–26. DOI:https://doi.org/10.1145/2442657.2442661

[40] Hyun Joon Jung and Hyun Joon. 2014. Quality assurance in crowdsourcing via matrix factorization based task routing. In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 3–8. DOI:https://doi.org/10.1145/2567948.2567951

[41] Kymberlee Price. 2015. How We Measure Crowd Performance. Retrieved February 22, 2017 from https://blog.bugcrowd.com/measure-crowd-performance/

[42] Jacob Hansen. 2015. Introducing Rep. Retrieved February 22, 2017 from https://blog.cobalt.io/introducing-rep-ecbacb04fa96#.sm5eltcs7

[43] Hackerone. 2014. Introducing Reputation. Retrieved February 22, 2017 from https://www.hackerone.com/blog/introducing-reputation

[44] Mingyi Zhao, Jens Grossklags, and Kai Chen. 2014. An Exploratory Study of White Hat Behaviors in a Web Vulnerability Disclosure Program. In *Proceedings of the 2014 ACM Workshop on Security Information Workers - SIW '14*, 51–58. DOI:https://doi.org/10.1145/2663887.2663906

[45] A Algarni and Y Malaiya. Software vulnerability markets: Discoverers and buyers. *waset.org*. Retrieved May 27, 2016 from http://www.waset.org/publications/9997791

[46] John Anvik, Lyndon Hiew, and Gail C. Murphy. 2006. Who should fix this bug? *Proceeding 28th Int. Conf. Softw. Eng. - ICSE '06* 2006, (2006), 361. DOI:https://doi.org/10.1145/1134285.1134336

[47] Umair ul Hassan, Edward Curry, U. Hassan, and Edward Curry. 2013. A Capability Requirements Approach for Predicting Worker Performance in Crowdsourcing. *Proc. 9th IEEE Int. Conf. Collab. Comput. Networking, Appl. Work.* (2013), 429–437. DOI:https://doi.org/10.4108/icst.collaboratecom.2013.254181

[48] Cristina Sarasua and Matthias Thimm. 2013. Microtask available, send us your CV! In *Proceedings - 2013 IEEE 3rd International Conference on Cloud and Green Computing, CGC 2013 and 2013 IEEE 3rd International Conference on Social Computing and Its Applications, SCA 2013*, 521–524. DOI:https://doi.org/10.1109/CGC.2013.87

[49] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Exp. Econ.* 9, 2 (June 2006), 79–101. DOI:https://doi.org/10.1007/s10683-006-4309-2

[50] Paul Resnick and Richard Zeckhauser. 2002. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *Adv. Appl. ...* 11 (2002), 127–157. DOI:https://doi.org/10.1016/S0278-0984(02)11030-3

[51] Bogdan Vasilescu, Vladimir Filkov, and Alexander Serebrenik. 2013. StackOverflow and GitHub: Associations between Software Development and Crowdsourced Knowledge. In *2013 International Conference on Social Computing*, 188–195. DOI:https://doi.org/10.1109/SocialCom.2013.35

[52] Claudia Hauff and Georgios Gousios. 2015. Matching GitHub developer profiles to job advertisements. *Proceedings of the 12th Working Conference on Mining Software Repositories*, 362–366.

[53] Pranam Kolari and Kelly Lyons. 2008. Expert Search using Internal Corporate Blogs. *Sigir* (2008), 2–5.

[54] Krisztian Balog, Leif Azzopardi, and Maarten de Rijke. 2006. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '06*, 43. DOI:https://doi.org/10.1145/1148170.1148181

[55] Konstantinos Pelechrinis, Vladimir Zadorozhny, and Vladimir Oleshchuk. 2011. A Cognitive-based scheme for user reliability and expertise assessment in Q&A social networks. In *2011 IEEE International Conference on Information Reuse & Integration*, 545–550. DOI:https://doi.org/10.1109/IRI.2011.6009614

[56] Demartini and Gianluca. 2007. Finding experts using Wikipedia. *Proceedings of the 2nd International Conference on Finding Experts on the Web with Semantics - Volume 290*, 33–41.

[57] Jian Chang, Peter Gebhard, Andreas Haeberlen, Zack Ives, Insup Lee, Oleg Sokolsky, and Krishna K. Venkatasubramanian. 2013. TrustForge: Flexible access control for collaborative crowd-sourced environment. In *2013 11th Annual Conference on Privacy, Security and Trust, PST 2013*, 291–300. DOI:https://doi.org/10.1109/PST.2013.6596065

[58] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 990. DOI:https://doi.org/10.1145/1401890.1402008

[59] Luanne Freund, Kristof Kessler, Michael Huggett, and Edie Rasmussen. Exposing and Exploring Academic Expertise with Virtu.

[60] Yupeng Fu, Rongjing Xiang, Yiqun Liu, Min Zhang, and Shaoping Ma. 2007. Finding Experts Using Social Network Analysis. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*, 77–80. DOI:https://doi.org/10.1109/WI.2007.14

[61] Juanzi Li, Jie Tang, Jing Zhang, and Q Luo. 2007. EoS: expertise oriented search using social networks. *Proc. 16th Int. Conf. World Wide Web* (2007), 1271–1272. DOI:https://doi.org/10.1145/1242572.1242803

[62] Jing Zhang, Jie Tang, and Juanzi Li. 2007. Expert Finding in a Social Network. In *Advances in Databases: Concepts, Systems and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1066–1069. DOI:https://doi.org/10.1007/978-3-540-71703-4_106

[63] Daniel Schall and Daniel. 2012. Expertise ranking using activity and contextual link measures. *Data Knowl. Eng.* 71, 1 (January 2012), 92–113. DOI:https://doi.org/10.1016/j.datak.2011.08.001

[64] Xiu Xiu Li, Jianguo Jianguo Ma, Yujiu Yujiu Yang, and Dongzhi Dongzhi Wang. 2013. A Service Mode of Expert Finding in Social Network. In *2013 International Conference on Service Sciences (ICSS)*, 220–223. DOI:https://doi.org/10.1109/ICSS.2013.48

[65] Zhou Zhao, James Cheng, Furu Wei, Ming Zhou, Wilfred Ng, and Yingjun Wu. 2014. SocialTransfer. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14*, 779–788. DOI:https://doi.org/10.1145/2661829.2661871

[66] Josep M. Pujol, Ramon Sangüesa, and Jordi Delgado. 2002. Extracting reputation in multi agent systems by means of social network topology. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems part 1 - AAMAS '02*, 467. DOI:https://doi.org/10.1145/544741.544853

[67] Maggy Anastasia Suryanto, Ee Peng Lim, Aixin Sun, and Roger H. L. Chiang. 2009. Quality-aware collaborative question answering. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining - WSDM '09*, 142. DOI:https://doi.org/10.1145/1498759.1498820

[68] R Atkinson and J Flint. 2001. Accessing Hidden and Hard-to-reach Populations: Snowball Research Strategies. *Soc. Res. Updat.* (2001). Retrieved September 9, 2017 from http://eprints.gla.ac.uk/37493/

[69] Martyn Denscombe. 2010. *The good research guide for small-scale social research projects*.

[70] J Saldaña. 2015. *The coding manual for qualitative researchers*. Retrieved September 3, 2016 from https://books.google.com.au/books?hl=en&lr=&id=ZhxiCgAAQBAJ&oi=fnd&pg=PP1&dq=The+Coding+Manual+for+Qualitative+Researchers&ots=yHYb8GPZjU&sig=7ZFrhcKUzg7vhCugj7NHPMKCmHc

[71] Clare M. Lewandowski, New Co-investigator, and Clare M. Lewandowski. 2015. *Statistics In a Nutshell: A Desktop Quick Reference 2th*. DOI:https://doi.org/10.1017/CBO9781107415324.004

[72] Sarah Elsie Baker and Rosalind Edwards. 2012. How many qualitative interviews is enough: Expert voices and early career reflections on sampling and cases in qualitative research. *Natl. Cent. Res. Methods Rev. Pap.* (2012), 1–43. DOI:https://doi.org/10.1177/1525822X05279903

[73] Tamsyn P Waterhouse, Spear Street, and San Francisco. 2013. Pay by the Bit : An Information-Theoretic Metric for Collective Human Judgment. *Comput. Support. Coop. Work* (2013), 623–637. DOI:https://doi.org/10.1145/2441776.2441846

[74] Jane Hunter, Abdulmonem Alabri, and Catharine van Ingen. 2013. Assessing the quality and trustworthiness of citizen science data. *Concurr. Comput.*

Pract. Exp. 25, FEBRUARY 2013 (2013), 454–466. DOI:https://doi.org/10.1002/cpe.2923

[75] Avrim L. Blum and Pat Langley. 1997. Selection of relevant features and examples in machine learning. *Artif. Intell.* 97, 1 (1997), 245–271. DOI:https://doi.org/10.1016/S0004-3702(97)00063-5

[76] Mark Klein. 2012. How to Harvest Collective Wisdom on Complex Problems: An Introduction to the MIT Deliberatorium. *Cent. Collect. Intell. Work. Pap.* (2012), 15. Retrieved from http://cci.mit.edu/docs/working_papers_2012_2013/kleinwp2013.pdf