

Crowdsourcing Security - Opportunities and Challenges

Lalit Mohan S

International Institute of Information Technology
Hyderabad, India
lalit.mohan@research.iiit.ac.in

YR Reddy

International Institute of Information Technology
Hyderabad, India
raghu.reddy@iiit.ac.in

ABSTRACT

With improved digital literacy and cheaper data connections, crowdsourcing has become more prevalent. Crowdsourcing tasks such as image tagging, language translation, etc. have contributors with varied skill level. Question and answers, software development/testing and product design are more serious tasks and require expert contributions. With increasing security threats, crowdsourcing in 'security' domain is seen as an upcoming trend. In the survey, opportunities and challenges in crowdsourcing security are analyzed based on a survey conducted with 229 participants. The results provide insights on motivation, quality mechanisms and improvement areas for increased participation. 'Question and Answers' that includes information sharing and 'software development and testing' that includes Bug Bounty are identified as top crowdsourcing tasks by security specialists although they are concerned about security and privacy of information. Among quality attributes analyzed, contributor 'credibility' is identified as a key area of improvement for increased participation in crowdsourcing.

CCS CONCEPTS

• **Crowdsourcing** → *Security; Opportunities; Challenges;*

KEYWORDS

Crowdsourcing, Security, Opportunities, Challenges

ACM Reference Format:

Lalit Mohan S and YR Reddy. 2018. Crowdsourcing Security - Opportunities and Challenges. In *Proceedings of CHASE'18:IEEE/ACM 11th International Workshop on Cooperative and Human Aspects of Software (CHASE'18)*. , 4 pages. <https://doi.org/10.1145/3195836.3195862>

1 INTRODUCTION

Outsourcing an activity to an undefined set of people in the form of an open call over internet is termed as crowdsourcing [9]. With improved digital literacy, better internet connectivity and affordable computing devices, crowdsourcing contributions are becoming part of the mainstream work. To exploit the opportunities, startups and enterprises are developing crowdsourcing platforms [15] with some of them focussing on

niche tasks. The micro tasks that include image tagging, transcription, survey participation, language translation and others require mental re-collection only to complete the task quickly. Amazon MTurk, CrowdFlower and SurveyMonkey are some of the popular crowdsourcing platforms for the micro tasks. Macro tasks such as research and prototyping, software development and testing and Product design require deeper thinking and are effort intensive. Organizations such as 99designs, InnoCentive and others deal with creative and research oriented work. eYeka [21], a crowdsourcing research organization has published in their report that creative contribution is increasing year-on-year across industry sectors. Freelancer and Upwork are some of the primary crowdsourcing platforms for skilled IT resources to complete software development and testing tasks. Wikipedia, Quora, StackExchange and others that involve content management by crowd are also seeing increased adoption. Seeing these benefits, nations and enterprises (Dell's Ideastorm, Barclays and others) are also building or using platforms [18] to encourage internal employees for idea generation, operational excellence, policy making, etc. MyGov¹ is one of the world's largest crowdsourcing platforms from Government of India seeking citizen contributions on policy making and other related tasks.

The motivation to participate on crowdsourcing platforms can be intrinsic (internal satisfaction) or extrinsic (praise, fame, money and grades). With these increased contributions and varying motivation, quality has been an area of concern [2]. Based on the task type (micro or macro), platforms have adopted different quality mechanisms such as statistical methods and machine learning, expert rating and voting, game theory and others. Some platforms filter contributors before task assignment and some do it after the contribution. Some of the leading platforms for micro tasks use quality of past contributions and expert rating for quality control. Software crowdsourcing platforms use technology skill assessment mechanisms to identify quality contributors. Q&A crowdsourcing platforms leverage user-topic follow graph, votes, likes and views [19] for quality assessment.

'Security' (protecting information present in computing devices) is identified as an upcoming crowdsourcing task by CrowdsourcingWeek². With increasing internet access and IT dependency, cyber security³ vulnerabilities, threats and incidents have increased manifold. It is becoming difficult in terms of time, money, effort and skill for an organization to source security related information single-handedly and protect its assets. Facebook ThreatExchange [5] and AlienVault [1] are some of the emerging crowdsourcing platforms for security information exchange. StackExchange and Quora that have questions and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHASE'18, May 27, 2018, Gothenburg, Sweden

© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5725-8/18/05...\$15.00
<https://doi.org/10.1145/3195836.3195862>

¹<https://www.mygov.in/>

²<http://crowdsourcingweek.com/blog/crowdsourcing-cybersecurity/>

³<https://tinyurl.com/PWCSecurity>

answers on Security related topics provide the benefit of collective intelligence groupthink or crowd wisdom). Considering this potential, it is important to understand the opportunities and challenges of requester and contributor on crowdsourcing in security domain. While there has been research in identifying crowdsourcing challenges and opportunities in various domains such as SMEs [14], Biomedicine [11] and Heterogeneous networked environments [12], the research in 'crowdsourcing security' is limited. Like medicine for individual health, security is very critical to the health of the organization. The actions and attitude of security teams are typically different as they are expected to doubt and verify every action to ensure confidentiality, integrity and availability of information. Our research study on crowdsourcing in security is to address the following questions -

- **RQ1** What crowdsourcing tasks/contributions interests security specialists?
- **RQ2** What factors influence security specialists contributions in crowdsourcing?

By stating 'Security Specialists', we mean individuals (industry, researchers, faculty and students) working on network security, performing vulnerability and penetration testing, security auditors, Chief Information Security Officers and others with focus on Cyber and Information Security. In the following sections of the paper, we discuss our research method that includes survey questions; Analysis of the survey results followed by conclusion.

2 RESEARCH METHOD

To understand opportunities and challenges in crowdsourcing security, we conducted an online survey⁴ using Social media channels such as Facebook, LinkedIn groups ('ISACA' and 'Security Network Group') and mailing lists (such as CISOForum of Banks). The survey link was also shared with members of Data Security Council of India⁵, an organization containing 'Security Specialists' as its members from India and Singapore. The 11 survey questions contains 3 questions to capture demographics of survey participants followed by questions on various crowdsourcing attributes. The questions on crowdsourcing were broadly classified on tasks, quality concerns and opportunities for improvements. 5 of the 8 questions are on crowdsourcing with multiple options (checkboxes) to select as participants may have multiple reasons on crowdsourcing opportunities and challenges. Based on the available literature [4] [9] [16], following are the survey questions on crowdsourcing -

- **Q4** What crowdsourcing tasks are you contributing and/or find more related to your work?
The multiple choice options provided are 'Software Development and Testing including Bug Bounty programs' (*Software*), 'Language Translation' (*Lang Tran*), 'Question and Answers including Threat Information Exchange' (*Q&A*), 'Streets and Maps Identification' (*Street*), 'Image Tagging' (*Image Tag*), 'Talent Acquisition' (*Talent*), 'Design - Logos, Business Cards' (*Design*) and Others. These tasks are based on the widely accepted crowdsourcing tasks [4] classification.

- **Q5** What is your motivation to participate in crowdsourcing activities?
Either 'Intrinsic' or 'Extrinsic' option is available to select. A short text clarifying intrinsic as contributing for internal satisfaction and extrinsic as contributing for praise, fame, money and grades is provided.
- **Q6** How often do you contribute in crowdsourcing tasks?
The following options are given for selection - 'Every day', 'Once a week', 'Rarely' and 'Not yet'. These options provide insight into the frequency of participation and also the level of engagement for a task type.
- **Q7** What concerns have you experienced in the adoption of crowdsourcing?
The multiple choice options are 'Not many respondents' (*Not Many*), 'Delay in responses' (*Resp Delay*), 'Relevance of response' (*Relevance*), 'Data Security and Privacy' (*Sec & Priv*) apart from an option to express Other concerns.
- **Q8** What percentage of responses you felt are relevant in crowdsourcing?
The options provided are $\leq 10\%$, $> 10\%$ and $\leq 50\%$, $> 50\%$ and $\leq 75\%$ and $> 75\%$. The responses to the question provides an understanding on the concern severity.
- **Q9** Why is relevance of response important in crowdsourcing?
The options are 'Improves adoption of crowdsourcing and thus leading to benefits of crowd wisdom' (*Improv Cont*), 'Reduces cost' (*Costs Reduc*), 'Leads to out-of-box thinking and surface new perspectives' (*Out of Box*) and Others. These options are suggested based on the factors that could lead to rise of crowdsourcing [9].
- **Q10** What existing relevance mechanisms you find most useful in crowdsourcing?
Multiple choice options are 'Voting/Rating - Majority consensus' (*Vote/Rate*), 'Expert Review/Contributor's quality - Manual' (*Expert Inv*), 'Automated analysis and processing of response text' (*Auto*) and Others.
- **Q11** What can be done to improve relevance of response in crowdsourcing?
The options are 'use existing and relevant information for providing responses without waiting for crowd respondents' (*Know Base*), 'Credibility (positive and negative) of responder' (*Credibility*), 'Responses should be coherent' (*Coherence*), 'Responses should be inline with question type' (*Que Align*) and Others.

The multiple choice options on relevance concerns, mechanisms and areas of improvement are based on the quality measures adopted by crowdsourcing platforms and the available literature [16]. Q4 and Q6 survey questions address RQ1 research question while survey questions Q5, Q7-Q11 address RQ2 research question.

3 RESULTS AND ANALYSIS

The survey had N = 229 participants responding to 11 questions and we did not have a target population as we depended on email distribution lists, social media and other free channels. Based on the correlation between person's age and evolution of Computer, Internet and Social media [17], the adaptation to technology is

⁴<https://goo.gl/forms/BkocNa3XxQvAeK9l1>

⁵<https://www.dsci.in/members-and-associates>

different. Hence, we broadly divided the participants into 4 age groups (<25 - mostly students, 25-40 - working class, 41-60 - tenured employees in the working class and >60 - retired/on the verge retirement from employment) to understand crowdsourcing attributes. We had participation from 10 different countries, however, 88% of the participants are from India. 87% of the participants have or pursuing professional qualifications such as Engineering, Masters of Business Administration (MBA), etc. and the remaining are Doctorates (PhD). About 79% of the participants⁶ are in the age group of 25-60 years. The age group can be considered as participants that already have an employment and are purpose oriented in their contributions. When we looked at their motivation reasons for participation, they are all not necessarily intrinsic. However, 77% (139 participants) of 79% (181 participants from age group - 25-60 years) stated intrinsic motivation as the reason for participation contradicting with the findings of Kaufmann et al. [10]. The reason for difference in motivation could be that most our participants are already employed or students with higher education as qualification. Our survey results confirm with the work of Yuxiang et al. [3] in the recent times that states extrinsic motivation may not necessarily bring longevity to crowdsourcing. As we shared the survey link through various online platforms, we also had participation from members of Academics (students and faculty), Industry (IT and other profiles such as Administration that are not IT oriented) apart from Security specialists. The mix of participants helped us to compare and contrast the opportunities and challenges of Security Specialists vis-a-vis Others on crowdsourcing.

3.1 Adoption Opportunities and Concerns

The 229 (N) participants provided 624 (R) responses to Q4 survey question on crowdsourcing tasks. While the intrinsic motivation (Q5) in response across profiles is 79%, 93% of security specialists are intrinsically motivated⁷. Security specialists and faculty are more intrinsically motivated, the reasons could be their nature of work, education, age and also that their participation/interest is more in serious tasks. From the survey results, it was identified that the mean participation/interest is 2.72 (R/N) crowdsourcing tasks per participant with student community having more involvement at 3.05 tasks and lowest for non-IT work profile community at 2.29 tasks. Higher participation from student community could be attributed to their higher levels of enthusiasm and affinity to internet world considering that their all born in digital age. Security specialists have an average participation of 2.75 tasks suggest that they are open for contribution in various crowdsourcing tasks. Also, a one-way ANOVA test on Security specialists contributions with others is performed, the F-score (0.001) is less than FCritical (4.6) score suggesting that the participation levels of security specialists is not different from all others. Based on the tasks participation/interest level and ANOVA scores, we conclude that security specialists are interested and ready to contribute in crowdsourcing. In the reminder of the paper, we focus on security specialists concerns and suggestions on Q&A and software tasks as they were identified as the key interests.

Profile	Not Many	Resp Delay	Relevance	Sec & Priv	Others	Mean
All	31%	45%	57%	35%	3%	34%
Security	33%	30%	54%	48%	7%	34%
Q&A - All	25%	32%	43%	24%	1%	25%
Q&A - Security	22%	24%	42%	30%	1%	24%
Software - All	14%	22%	27%	18%	0%	16%
Software - Security	21%	16%	22%	30%	0%	18%
Mean	25%	28%	41%	31%	2%	25%

Figure 1: Concerns in Crowdsourcing Adoption

While we have participants with different profiles, we restricted our comparison with security specialists versus others (Others = 229 (N) - security specialists, i.e., faculty, IT staff, student and other profiles such as Administration that are not IT oriented). The work of Kucherbaev et al. [13] and Hetmank [7] discusses that crowdsourcing tasks are beyond microtasks and that there is growing participation across organizations, sectors and communities for non microtasks, supporting our observation on the interests of security specialists. The Q7 survey question on concerns in crowdsourcing had 383 responses. From these selected options (represented in Figure 1), it is evident that security specialists are relatively more concerned about 'Data Security and Privacy' while leveraging crowdsourcing as it could expose vulnerabilities of their systems. As observed from survey results and the available literature [23] [8], data security & privacy is a concern that needs to be addressed to improve adoption of crowdsourcing. Reasons such as 'delay in responses' and 'lack of participation' do not seem to be a concern for security specialists. Pareto analysis on concerns indicate that relevance and data security & privacy are the common concerns for security specialists and others. Relevance remains a concerning factor for Q&A and software tasks as well. As we are using relevance and quality interchangeably, we looked at the existing literature on quality concerns [2] [20] in crowdsourcing and relate it with the participants responses.

3.2 Quality Control in Crowdsourcing

To Q8 survey question, a majority (<=75%) of participants stated the quality of crowdsourcing contributions is <=50% as shown in figure 2. However, the need for improved quality is different for security specialists vs others. To Q9 survey question on quality importance that had 372 responses, security specialists view improved quality can further enhance contributions in crowdsourcing and out-of-box thinking, leading to the benefits of groupthink [6] or crowd wisdom. This provides an insight that security specialists consider crowdsourcing for more serious tasks and also have intrinsic reasons for participation. This establishes that if quality concerns are addressed, security specialists will adopt crowdsourcing with an intent for a longer journey rather than for immediate/trivial gains. Based on this, we analyze the current mechanisms for quality control and areas for improvement in crowdsourcing quality. Based on the Q10 survey question that had 345 responses, it is evident that majority of the participants expect the involvement of 'Expert/Moderation' in the quality assessment. While participants are concerned on quality, it is interesting to note that they find the current 'Voting/Rating' as

⁶<https://tinyurl.com/CrowdDemo>

⁷<https://tinyurl.com/CrowdMot>

Profile	Quality				Benefits of Improved Quality			
	<10%	10-50%	>50 - 75%	>75%	Improve Cont	Cost Reduc	Out of Box	Others
All	20%	55%	19%	6%	67%	48%	46%	1%
Security	13%	63%	18%	6%	67%	42%	57%	1%
Q&A - All	12%	36%	16%	5%	49%	34%	36%	0%
Q&A - Security	6%	49%	10%	4%	48%	28%	43%	1%
Software - All	10%	23%	9%	2%	35%	27%	22%	0%
Software - Security	7%	34%	10%	0%	36%	21%	31%	0%
Mean	12%	44%	14%	4%	50%	33%	39%	1%

Figure 2: Benefits of Improved Quality

Profile	Current Quality Mechanisms				Improvements in Quality			
	Auto	Vote/Rate	Expert Inv	Others	Know Base	Credibility	Coherence	Que Align
All	32%	58%	61%	1%	39%	63%	41%	41%
Security	24%	63%	63%	1%	40%	61%	37%	43%
Q&A - All	25%	45%	43%	0%	31%	44%	33%	31%
Q&A - Security	15%	49%	45%	1%	28%	42%	27%	40%
Software - All	18%	26%	31%	0%	20%	31%	20%	21%
Software - Security	15%	34%	36%	1%	24%	36%	22%	18%
Mean	22%	46%	47%	1%	30%	46%	30%	32%
Standard Deviation	6%	13%	12%	0%	7%	12%	8%	10%

Figure 3: Quality Mechanisms and Improvements

useful along with 'Expert Moderation'. But as 75% participants find that the current quality is <50%, we also had a survey question (Q11) to understand on the areas of improvement in quality. A higher number of 426 responses to Q11 as compared to previous question (Q10) with 345 responses allude that the participants are concerned on existing quality mechanisms and are exploring methods for quality improvements. It is evident from the participants response that the credibility of contributor is given more importance. The work of Han et al. [22] and others [16] also state the need for trust and credibility as critical factors for the success of crowdsourcing. For Q&A - Others are more open to automated analysis using knowledge base as compared to security specialists. On software crowdsourcing task (Q4) - credibility is also given more importance by security specialists as participants might have thought that they could be exposing system vulnerabilities into wrong hands. By credibility, it is not necessarily the pedigree of educational degrees or organizations but credibility of performances[9]. While participants did not give much importance to the current auto mechanisms for quality assessments, they are relatively more open to the use knowledge base, question-response type validation for enhancing quality. In the case of software tasks, the importance of credibility can be related to the approach that Freelancer, UpWork and other software crowdsourcing platforms use for contributor technical skill assessment for task allocation.

4 CONCLUSION

The survey analysis helped us to understand that security specialists consider crowdsourcing as a solution for strengthening security. They are intrinsically motivated to contribute on crowdsourcing platforms and are concerned about the quality of contributions. They view credibility as a major factor for assessing the quality of contribution co-relating with the type of tasks. Q&A platforms including threat exchange, Software Development including Bug bounty programs are reiterated as major opportunities for collaboration in crowdsourcing security. Though we understand that most of the survey participants are from India,

security as a domain has no geographical boundaries and outsourcing security work is done for lack of in-house skills rather reduce the costs. Based on some these insights, we see that there can be a separate platform or section in crowdsourcing for security specialists and can include even domain search capabilities. There can also be a quality metric for assessing the crowdsourcing contributions in domain with flexibility to make it configurable based on the domain and the task requirements.

REFERENCES

- [1] Aliant Vault. 2017 (accessed 19-Mar-2018). *Open Threat Exchange*. <https://www.alienvault.com/>.
- [2] Mohammad Allahbakhsh, Boualem Benattallah, Aleksandar Ignjatovic, Hamid Reza Motehari-Nezhad, Elisa Bertino, and Schahram Dustdar. 2013. Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing* 17, 2 (2013), 76–81.
- [3] Yuxiang Chris Zhao and Qinghua Zhu. 2014. Effects of Extrinsic and Intrinsic Motivation on Participation in Crowdsourcing Contest: A Perspective of Self-Determination Theory. *Online Information Review* 38, 7 (2014), 896–917.
- [4] Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. 2012. Towards an Integrated Crowdsourcing Definition. *Journal of Information science* 38, 2 (2012), 189–200.
- [5] Facebook. 2017 (accessed 19-Mar-2018). *Threat Exchange*. <https://developers.facebook.com/products/threat-exchange>.
- [6] Samuel Greengard. 2011. Following the Crowd. *Commun. ACM* 54, 2 (2011), 20–22.
- [7] Lars Hetmank. 2013. Components and Functions of Crowdsourcing Systems - A Systematic Literature Review. *Wirtschaftsinformatik* 4 (2013).
- [8] Mokter Hossain and Ilkka Kauranen. 2015. Crowdsourcing: A Comprehensive Literature Review. *Strategic Outsourcing: An International Journal* 8, 1 (2015), 2–22.
- [9] Jeff Howe. 2006. The Rise of Crowdsourcing. *Wired magazine* 14, 6 (2006), 1–4.
- [10] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money: Worker Motivation in Crowdsourcing - A Study on Mechanical Turk. In *AMCIS*, Vol. 11. 1–11.
- [11] Ritu Khare, Benjamin M Good, Robert Leaman, Andrew I Su, and Zhiyong Lu. 2015. Crowdsourcing in Biomedicine: Challenges and Opportunities. *Briefings in Bioinformatics* 17, 1 (2015), 23–32.
- [12] Axel Korthaus and Wei Dai. 2012. Crowdsourcing in Heterogeneous Networked Environments-Opportunities and Challenges. In *15th International Conference on Network-Based Information Systems (NBIS)*. IEEE, 483–488.
- [13] Pavel Kucherbaev, Florian Daniel, Stefano Tranquillini, and Maurizio Marchese. 2016. Crowdsourcing Processes: A survey of Approaches and Opportunities. *IEEE Internet Computing* 20, 2 (2016), 50–56.
- [14] R Maiolini and R Naggi. 2011. Crowdsourcing and SMEs: Opportunities and Challenges. In *Information Technology and Innovation Trends in Organizations*. Springer, 399–406.
- [15] Ethan Seltzer and Dillon Mahmoudi. 2013. Citizen Participation, Open innovation, and Crowdsourcing : Challenges and Opportunities for Planning. *CPL bibliography* 28, 1 (2013), 3–18.
- [16] Chun How Tan, Eugene Agichtein, Panos Ipeirotis, and Evgeniy Gabrilovich. 2014. Trust, but Verify: Predicting Contribution Quality for Knowledge Base Construction and Curation. In *Proceedings of the 7th ACM international conference on Web search and Data Mining*. ACM, 553–562.
- [17] Alexander Van Deursen and Jan Van Dijk. 2011. Internet Skills and the Digital Divide. *New media & society* 13, 6 (2011), 893–911.
- [18] Maja Vukovic and Claudio Bartolini. 2010. Towards a Research Agenda for Enterprise Crowdsourcing. *Leveraging Applications of Formal methods, Verification, and Validation* (2010), 425–434.
- [19] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y Zhao. 2013. Wisdom in the Social Crowd: an Analysis of Quora. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1341–1352.
- [20] Jing Wang and Panagiotis Ipeirotis. 2013. A Framework for Quality Assurance in Crowdsourcing. (2013).
- [21] Yaoqi Lai. 2017 (accessed 19-Mar-2018). *Creative Crowdsourcing Leader eYeka Releases 2017 Industry Report and Welcomes The Age of Ideation*. <http://tinyurl.com/eYeka2017>.
- [22] Han Yu, Zhiqi Shen, Chunyan Miao, and Bo An. 2012. Challenges and Opportunities for Trust Management in Crowdsourcing. In *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 02*. IEEE Computer Society, 486–493.
- [23] Yuxiang Zhao and Qinghua Zhu. 2014. Evaluation on Crowdsourcing Research: Current status and Future Direction. *Information Systems Frontiers* 16, 3 (2014), 417–434.