International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 11, November 2023)

Privacy in the Crowd: A Review of Tools and Techniques in Crowdsourcing Platforms

Santosh Kumar[1], Dr. Mohammad Faisal[2]
[1]*Research Scholar,* [2]*Professor & Head,Department of Computer Application, Integral University, Lucknow, India*
[1]sanb2@rediffmail.com, [2]mdfaisal@iul.ac.in

***Abstract*-- This review critically examines the interplay between privacy and crowdsourcing platforms, addressing the escalating concerns surrounding the protection of individual privacy as crowdsourcing extends its reach from commercial to non-commercial domains and public policy-making. The analysis evaluates the efficacy of privacy-preserving measures within crowdsourcing, encompassing anonymization strategies, access controls, cryptographic protocols, and differential privacy frameworks. Emphasizing the ethical, legal, and societal implications, the review explores the challenges posed by the open nature of crowdsourcing, where contributors voluntarily share insights, raising questions about personal information protection and responsible content use. A comprehensive literature review further enriches our understanding of privacy in various contexts, including ubiquitous computing, game theory in cybersecurity, and specific crowdsourcing applications in policy-making and geography. The study highlights the vital role of Privacy Enhancing Technologies (PETs) and presents frameworks for systematically addressing privacy concerns. This synthesis serves as a valuable resource for researchers, policymakers, and practitioners, inspiring advancements in privacy safeguards within the dynamic landscape of crowdsourcing platforms. As the intersection of collective intelligence and individual privacy continues to shape discourse, ongoing efforts are imperative to strike a delicate balance in this complex and evolving landscape.**

***Keywords*-- Crowdsourcing, Privacy, Privacy Enhancing Technologies (PETs), Data Protection, User Privacy.**

I. INTRODUCTION

Crowdsourcing is widespread! According to (Lehdonvirta and Bright, 2015): "Today, if elections were to be invented, the term for them would be Crowdsourcing the Government." Organisations that engage a relatively large number of individuals for the provision of labour, ideas, expertise, or opinions are now loosely applying the term "crowdsourcing," which is undergoing rapid development(Lehdonvirta and Bright, 2015). Currently, crowdsourcing is being applied to domains other than business and consumer. Additionally, citizen empowerment can be promoted and transparency enhanced through increased citizen participation in policymaking (Fischer, 1993)(Prpić, Taeihagh and Melton, 2015)(Liu, 2017).

A number of policy domains have begun to use crowdsourcing techniques, including transportation(Seltzer and Mahmoudi, 2013)(Aitamurto *et al.*, 2016), urban planning, global governance, and state and federal policy. It has also been shown that crowdsourcing can assist with the current problems with collecting data and making decisions for policy analysis and design.

Despite recent advancements in crowdsourcing's use in the public sector, extremely few studies have investigated its function in the policy cycle. There has been little use of crowdsourcing in the policy cycle thus far, and its potential is still untapped, despite its meteoric rise(Prpić, Taeihagh and Melton, 2015). Ahead of the policy evaluation, agenda-setting, and problem-defining stages, researchers have mostly used Open Collaboration (OC) platforms. However, with a few of notable exceptions, methods such as Virtual Labour Markets (VLM)and Tournament Crowdsourcing (TC) have mainly been disregarded.

This review paper examines the methods and techniques used to address privacy challenges in crowdsourcing platforms, taking a comprehensive look at the delicate relationship between the two. Crowdsourcing raises several important problems around data privacy, the accidental revelation of sensitive information, and the appropriate use of user-generated content due to the transparent nature of the process, in which many people willingly offer their opinions.

The primary objective of this investigation is to examine the current methods employed by crowdsourcing platforms in safeguarding privacy. The objective of the review is to identify areas of deficiency or potential enhancement by assessing the effectiveness of existing tools and procedures. Anonymization techniques, access controls, cryptographic protocols, and differential privacy frameworks are all components of the range of measures used to preserve privacy in the analysis.

The paper explores the shifting dynamics of crowdsourcing and thelegal, ethical, and sociological issues surrounding the collection and use of user-generated data. It also showcases effective privacy strategies.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 11, November 2023)

II. Literature Review

The topic of privacy protection in ubiquitous computing has been the subject of much research(Agarwal and Hall, 2013)(Lin *et al.*, 2012)(Yang *et al.*, 2012). One research(Agarwal and Hall, 2013)introduced a system that can identify when users access private information. The plan is to let people rate their privacy settings on various mobile apps using a crowdsourced recommendation engine. Several studies have shown security risks associated with crowdsourcing, and the connection between security and privacy is obvious(Yang *et al.*, 2015). The primary use case for such systems is monitoring and analysing data usage.

It was utilized to examine information security concerns associated with the collection of data from citizens participating in the crowdsourcing smart city initiative for the purpose of ensuring public safety(Cilliers and Flowerday, 2014). Mobile phone sensing apps can find the locations of crowdsourcing participants, which lets people report strange public safety events(Christin *et al.*, 2011).

Regarding crowdsourcing systems generally, a number of survey papers were given in order to define the types and traits of applications using crowdsourcing(Yuen, King and Leung, 2011), evaluate a crowdsourcing system, and offer remedies to deal with the problems associated with crowdsourcing systems(Yin *et al.*, 2014).

In this work, (Vlachos, Stamatiou and Nikoletseas, 2023)presented the Privacy Flag Observatory, a platform that serves as a central tool for the research project Privacy Flag, which is supported by the European Union. The objective of this project is to make Europeans more aware of the possible privacy risks that include the daily-used software and services, such as phones and websites. Part of what made the project's goals possible was the Privacy Flag Observatory. That's the goal of this real-time platform for monitoring security and privacy threats: to gather, store, analyse, and share information about these threats with both regular people and experts. Although other components, such as the mobile phone add-on, are touched upon briefly in this paper, the focus is squarely on the observatory platform. The platform calls upon crowdsourcing to collect data and directs it to other components either installed on users' devices or in remote servers and databases.

The objective of the research conducted by (Yahuza et al., 2021) was to evaluate the contemporary patterns pertaining to privacy concerns impacting the "Internet of Devices" (IoD) network. The amount of security and privacy vulnerabilities posed by different kinds of drones was examined by researchers.

The authors subsequently emphasized the need of a secure IoD architecture and put out a proposed solution. Additionally, a thorough classification of the assaults on the "Internet of Things" (IoT) network was provided by the researchers. Furthermore, the researchers conducted an analysis of the most current strategies used in mitigating attacks on the Internet of Things (IoT) devices. In addition, the strategies also provide an overview of the performance assessment methodologies and performance indicators used. Ultimately, the researchers provided a framework to guide future investigations, enabling scholars to discern the most recent prospects in the realm of IoT study.

According to the research conducted by (Kaaniche, Laurent and Belguith, 2020) , the primary aim was to determine whether "Privacy Enhancing Technologies" (PETs) has the capability to simultaneously address the often conflicting objectives of economic and ethical considerations. This research presents a framework that categorizes eight types of "Privacy Enhancing Technologies" (PETs) into three distinct groups. Additionally, to enhance clarity, the study includes an examination of three kinds of tailored services. Following a comprehensive definition and presentation of the key characteristics of PETs, this research continues to identify and highlight the specific PETs that are most suitable for each customized service category, supported by relevant and instructive examples. The next part examines the interdisciplinary privacy difficulties that might potentially hinder the widespread use of these approaches. These challenges include technological, social, legal and economic considerations. Ultimately, the study offers suggestions and underscores many avenues for further research.

In this research, game theory was used to model defensive deception for privacy and cybersecurity. Specifically, the researchers (Pawlick, Colbert and Zhu, 2019) conducted a survey of 24 publications published between 2008 and 2018 that focus on this topic. Subsequently, a taxonomy was put out by the researchers, which establishes six distinct categories of deception: Changes in the target, changing target defense, deception, mixing, honey-x, and a hacker contact. These categories are classified according to their information structures, actions, agents,and durations, all of which are concepts encompassed by game theory. The objectives of this study are to precisely delineate several forms of defensive deception, to provide a comprehensive overview of the existing body of literature, to offer a range of models that may be used for practical research purposes and to pinpoint areas that hold potential for further investigation.

The classification presented in this study establishes a systematic framework for comprehending various forms of defensive deception that are often encountered in the domains of cybersecurity and privacy.

According to the study conducted by (Yuan et al., 2020) The problem of privacy leaking is a significant concern in geographical crowdsourcing across several circumstances. The researchers investigate the topic of privacy preservation in the context of geographical crowdsourcing. The primary obstacle is in the effective allocation of jobs to proximate employees, without necessitating precise knowledge of the specific locations of those tasks and workers. In order to tackle this issue, the researchers put up a system that ensures privacy while eliminating the need for online trustworthy third parties. The researchers developed a location protection approach based on a grid system. This method ensures the confidentiality of worker and task locations, while yet allowing for the retention of distance-aware information pertaining to these protected locations. Consequently, it becomes possible to measure the distance between tasks and employees. The authors present an efficient method for task assignment that is capable of promptly assigning jobs to nearby employees while operating on encrypted data. In order to safeguard the integrity of the task content, the use of attribute-based encryption and symmetric-key encryption is employed. This approach allows the establishment of secure channels among servers, hence guaranteeing the secure and precise delivery of the work, even in the presence of untrusted servers. Furthermore, the researchers conducted an analysis of the security aspects associated with their proposed methodology. Real experiments have been done on datasets derived from real-world scenarios. The experimental findings demonstrate that the strategy used by the researchers surpasses the performance of previously established methodologies.

In (Xia and McKernan, 2020)investigated the difficulties in confidentiality that are connected to the features of crowdsourcing tasks, platforms, requesters, and workers in the crowd. This discussion and categorization of these privacy challenges into theoretical and practical ones also occurs. Based on the review and discussion, this paper proposes strategies to better understand and address many crowdsourcing privacy threats and challenges.

According to the study conducted by (Niksefat, Kaghazgaran and Sadeghiyan, 2017) This work presents a classification of privacy concerns in Intrusion Detection Systems (IDSs) and then uses it to identify emerging difficulties and issues within the domain.

Within this taxonomy, privacy-sensitive IDS data is categorized into three distinct types: input data, built-in data and produced data. After that, the established taxonomy is used to compare and evaluate the research prototypes. The study examines and evaluates the privacy mechanisms used in the systems under investigation, focusing on their impact on the performance and accuracy of the Intrusion Detection System. Ultimately, the use of “taxonomy and survey” methodologies serves to identify many potential avenues for future study.

The increasing number of smartphones and other mobile devices with numerous sensors has coincided with the rise of mobile crowdsourcing, or MCS, as a viable option for data collection and processing. The advantages of MCS over conventional wireless sensor networks include mobility, scalability, cost-effectiveness, and artificial intelligence. Despite this, MCS continues to struggle with a great deal of difficulties in terms of trust, privacy, and security.(Feng et al., 2018)offered an overview of these difficulties and addressed possible ways to overcome them. In order to create a trustworthy, private, and secure MCS system, researchers examined its features, found its security vulnerabilities, and listed the most important requirements. In addition, they evaluate the benefits and drawbacks of pre-existing solutions in light of these needs.

In their research (Paulin and Haythornthwaite, 2016), emphasized the significance of including open, peer-generated resources. This inclusion is leading to a transformation in the dynamics of learning, specifically in terms of who acquires knowledge, from whom they acquire it and the methods through which they acquire it. These modifications result in a transfer of duties, whereby the responsibility now lies on the motivated and self-directed learner-participants rather than the course creator. Considerable scholarly inquiry has been dedicated to investigating the obstacles associated with establishing and maintaining interactive learning environments in the field of virtual education. However, the emergence of massive open online courses necessitates the exploration of novel methodologies that surpass the current body of study pertaining to participation settings inside academically structured courses. Instead of institutionally defined classes, the authors of this study want to study open virtual communities and online crowds. The authors used existing scholarly works on online organizing, learning science and developing educational practice as a basis for examining the influence of collaboration and peer production on the learning process, specifically in relation to the concept of "crowdsourcing the curriculum."

The study conducted by (Heurix *et al.*, 2015) tried to address this research gap by presenting an extensive classification of Privacy Enhancing Technologies (PETs). The taxonomy was designed to involve various dimensions and properties related to privacy, including user privacy and data privacy. This is particularly important as existing taxonomies primarily focus on aspects related to information security, often overlooking privacy-specific characteristics. The text offers the reader a mechanism for the methodical evaluation of various privacy-enhancing technologies (PETs). This aids in the identification of constraints within current PETs, supplementary technologies and prospective avenues for further study. In order to showcase its practicality, the suggested taxonomy is implemented on a collection of pivotal technologies that span several fields, including data anonymization, privacy-preserving data searching, communication protection and identity concealment.

In (Poblet, García-Cuesta and Casanovas, 2014)proposed a taxonomy for classifying a suite of web-based tools and platforms that have been put into use in the disaster management field in the last several years. Crowdsourcing or automated tools are needed to identify and analyze such events because it is difficult. While crowdsourcing relies on human resources to generate, compile, or filter initial data, automated tools analyse publicly available information by means of information retrieval techniques.

III. CONCLUSION

Even though crowdsourcing platforms are experiencing rapid user and data growth, there is clear evidence of privacy gaps that have not been adequately addressed in the research literature.This review article has undertaken a meticulous exploration of the intricate relationship between privacy concerns and the evolving landscape of crowdsourcing platforms. As crowdsourcing continues to permeate various domains, from commercial enterprises to non-commercial realms and policy-making initiatives, the inherent tension between harnessing collective intelligence and safeguarding individual privacy becomes increasingly evident.

The investigation has illuminated the complex difficulties presented by the unrestricted nature of crowdsourcing, where a varied array of participants actively contribute their perspectives, concepts, and work. The evaluation has emphasised the urgent requirement for strong privacy measures to tackle concerns over the safeguarding of personal information, unintentional exposure of sensitive data, and ethical utilisation of user-generated content.

The review has analysed the existing privacy-preserving tools and techniques in crowdsourcing platforms. It has found a range of measures, including anonymization schemes, access limits, cryptographic protocols, and differential privacy frameworks. The review acknowledges the successful implementations, but also highlights the ongoing ethical, legal, and societal concerns related to the gathering and use of user-generated data in these interactive collaborative spaces.

The literature review section enhanced our comprehension of the wider privacy landscape, encompassing diverse studies on privacy preservation in ubiquitous computing, the application of game theory in cybersecurity, and the specific utilisation of crowdsourcing in geographical and policy-making domains. The inclusion of many viewpoints enhances the overall understanding of privacy issues and possible remedies within the framework of crowdsourcing.

As we explore the complex landscape of privacy in public settings, it becomes clear that continuous research and advancement in Privacy Enhancing Technologies (PETs) are essential. The review has identified classifications and frameworks established by scholars to methodically assess and tackle privacy concerns in many contexts, ranging from geographical crowdsourcing to intrusion detection systems.

This review functions as a complete reference for scholars, policymakers, and practitioners in the domains of crowdsourcing, privacy, and technology. The objective of this study is to stimulate more research and progress in privacy protection on crowdsourcing platforms by carefully examining current tools and methodologies, highlighting areas that need improvement, and providing a wide range of relevant literature. As we consider the future, the convergence of collective intelligence and individual privacy will surely continue to influence the conversation, requiring continual efforts to achieve a careful equilibrium in this intricate environment.

REFERENCES

[1] Agarwal, Y. and Hall, M. (2013) 'ProtectMyPrivacy: Detecting and mitigating privacy leaks on iOS devices using crowdsourcing', in MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services. Available at: https://doi.org/10.1145/2462456.2464460.

[2] Aitamurto, T. et al. (2016) 'Civic CrowdAnalytics: Making sense of crowdsourced civic input with big data tools', in AcademicMindtrek 2016 - Proceedings of the 20th International Academic Mindtrek Conference. Available at: https://doi.org/10.1145/2994310.2994366.

[3] Christin, D. et al. (2011) 'A survey on privacy in mobile participatory sensing applications', Journal of Systems and Software [Preprint]. Available at: https://doi.org/10.1016/j.jss.2011.06.073.

[4] Cilliers, L. and Flowerday, S. (2014) 'Information security in a public safety, participatory crowdsourcing smart city project', in 2014 World Congress on Internet Security, WorldCIS 2014. Available at: https://doi.org/10.1109/WorldCIS.2014.7028163.

[5] Feng, W. et al. (2018) 'A survey on security, privacy, and trust in mobile crowdsourcing', IEEE Internet of Things Journal [Preprint]. Available at: https://doi.org/10.1109/JIOT.2017.2765699.

[6] Fischer, F. (1993) 'Citizen participation and the democratization of policy expertise: From theoretical inquiry to practical cases', Policy Sciences [Preprint]. Available at: https://doi.org/10.1007/BF00999715.

[7] Ankur Pandey; Ankur Chaturvedi; Manish Gupta; Praveen Kumar Mannepalli, An Automated Face Mask Detection System using Deep CNN on AWS Cloud Infrastructure , 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), DOI: 10.1109/ICESC57686.2023.10193710

[8] Heurix, J. et al. (2015) 'A taxonomy for privacy enhancing technologies', Computers and Security, 53.

[9] Kaaniche, N., Laurent, M. and Belguith, S. (2020) 'Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey', Journal of Network and Computer Applications, 171.

[10] Lehdonvirta, V. and Bright, J. (2015) 'Crowdsourcing for public policy and government', in Policy and Internet. Available at: https://doi.org/10.1002/poi3.103.

[11] Rashmi Singh; Praveen Kumar Mannepalli, Invasive Weed optimization Algorithm Based Trained Neural Network for Cloud Malicious Threat Detection, 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES), 10.1109/TRIBES52498.2021.9751674

[12] Lin, J. et al. (2012) 'Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing', in UbiComp'12 - Proceedings of the 2012 ACM Conference on Ubiquitous Computing. Available at: https://doi.org/10.1145/2370216.2370290.

[13] K. Anil; Praveen Kumar Mannepalli, Achieving Effective Secrecy based on Blockchain and Data Sharing in Cloud Computing, 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), DOI: 10.1109/CSNT51715.2021.9509663.

[14] Praveen Kumar Mannepalli; Devendra Singh Kushwaha, Face Recognition Based on Cascade Classifier Using Deep Learning, 2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP), DOI: 10.1109/IHCSP56702.2023.10127172

[15] Praveen Kumar Mannepalli; Parcha Kalyani; An Early Detection of Pneumonia in CXR Images using Deep Learning Techniques, 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), DOI: 10.1109/ICIDCA56705.2023.10100230

[16] Liu, H.K. (2017) 'Crowdsourcing designs: A synthesis of literatures', in Proceedings of the Annual Hawaii International Conference on System Sciences. Available at: https://doi.org/10.24251/hicss.2017.332.

[17] Niksefat, S., Kaghazgaran, P. and Sadeghiyan, B. (2017) 'Privacy issues in intrusion detection systems: A taxonomy, survey and future directions', Computer Science Review, 25.

[18] Paulin, D. and Haythornthwaite, C. (2016) 'Crowdsourcing the curriculum: Redefining e-learning practices through peer-generated approaches', The Information Society, 32(2).

[19] Pawlick, J., Colbert, E. and Zhu, Q. (2019) 'A Game-theoretic Taxonomy and Survey of Defensive Deception for Cybersecurity and Privacy', ACM Computing Surveys, 52(4).

[20] Deepak Kumar Rathore;Praveen Kumar MannepalliA Review of Machine Learning Techniques and Applications for Health Care2021 International Conference on Advances in Technology, Management & Education (ICATME), DOI: 10.1109/ICATME50232.2021.9732761

[21] Poblet, M., García-Cuesta, E. and Casanovas, P. (2014) 'Crowdsourcing tools for disaster management: A review of platforms and methods', Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [Preprint]. Available at: https://doi.org/10.1007/978-3-662-45960-7.

[22] Prpić, J., Taeihagh, A. and Melton, J. (2015) 'The fundamentals of policy crowdsourcing', Policy and Internet [Preprint]. Available at: https://doi.org/10.1002/poi3.102.

[23] Seltzer, E. and Mahmoudi, D. (2013) 'Citizen Participation, Open Innovation, and Crowdsourcing: Challenges and Opportunities for Planning', Journal of Planning Literature [Preprint]. Available at: https://doi.org/10.1177/0885412212469112.

[24] Vlachos, V., Stamatiou, Y.C. and Nikoletseas, S. (2023) 'The Privacy Flag Observatory: A Crowdsourcing Tool for Real Time Privacy Threats Evaluation', Journal of Cybersecurity and Privacy [Preprint]. Available at: https://doi.org/10.3390/jcp3010003.

[25] Xia, H. and McKernan, B. (2020) 'Privacy in Crowdsourcing: a Review of the Threats and Challenges', Computer Supported Cooperative Work: CSCW: An International Journal [Preprint]. Available at: https://doi.org/10.1007/s10606-020-09374-0.

[26] Yahuza, M. et al. (2021) 'Internet of Drones Security and Privacy Issues: Taxonomy and Open Challenges', IEEE Access, 9.

[27] R Singh, PK Mannepalli - … Journal of Advanced Research Engineering and …, 2020, Cloud Malicious Threat Detection By Features From Intelligent Water Drop Set And EBPN

[28] DK Rathore, PK MannepalliA Review of Machine Learning Techniques and Applications for Health Care, - 2021 International Conference on Advances.

[29] Mannepalli, P.K., Kulurkar, P.,An Enhanced Classification Model for Depression Detection Based on Machine Learning with Feature Selection TechniqueDOB :10.1109/ISCON52037.2021.9702492

[30] Yang, D. et al. (2012) 'Crowdsourcing to Smartphones : Incentive Mechanism Design for Mobile Phone Sensing Categories and Subject Descriptors', Mobicom [Preprint].

[31] Yang, K. et al. (2015) 'Security and privacy in mobile crowdsourcing networks: Challenges and opportunities', IEEE Communications Magazine [Preprint]. Available at: https://doi.org/10.1109/MCOM.2015.7180511.

[32] Yin, X. et al. (2014) 'What? How? Where? A survey of crowdsourcing', in Lecture Notes in Electrical Engineering. Available at: https://doi.org/10.1007/978-94-007-7618-0_22.

[33] Yuan, D. et al. (2020) 'PriRadar: A Privacy-Preserving Framework for Spatial Crowdsourcing', IEEE Transactions on Information Forensics and Security [Preprint]. Available at: https://doi.org/10.1109/TIFS.2019.2913232.

[34] Yuen, M.C., King, I. and Leung, K.S. (2011) 'A survey of crowdsourcing systems', in Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011. Available at: https://doi.org/10.1109/PASSAT/SocialCom.2011.36.