Social media users privacy protection from social surveillance using Blockchain Technology

Ms. Purnima Ahirao
Research Scholar
Computer Engineering
Mukesh Patel School of Technology Management & Engineering
Shirpur, India
ahirao.purnima002@nmims.edu.in

Dr.Shubham Joshi
Assistant Professor
Computer Engineering
Mukesh Patel School of Technology Management & Engineering
Shirpur, India
shubhamjoshi@ieee.org

Abstract*— Due to rapid technology improvements, an increasing number of people are being connected to the digital world. Considering other parts of life, the internet has grown increasingly vital to Indians. More than 4.39 billion people use the internet, and almost 70% of them use social media on smartphones, tablets, laptops, and other computers. Management, staff, and users all play an important role in information security. As a result, the human has become the weakest point in the digital environment. Human understanding and behavior are essential for successful and efficient usage of technology. The human aspect can be divided into two categories: one in which humans are directly involved in the system in some way, and the other in which they are not. The end users' lack of understanding, belief, conduct, and inappropriate use of technology are the other specific variables. Users desire security, flexibility, and simplicity of use all at the same time. Finding a balance between all these criteria is extremely difficult for any business or service provider. Users must be willing to give all information, including personal and sensitive information, in order to be online. The future is expected to be data-centric and data-driven. Data, according to researchers, is the new fuel that will drive technology forward. As a result, service providers have become accustomed to organizing the data for computational and surveillance purposes. It is time for data to be protected and confidentiality to be respected. The most significant aspect of this issue is the user's concern in deciding whether to share the data. If exchanging data is required, should the customer be able to track who has accessed the data? After that, the user should have a say in who gets access to the information and who does not. So, these are the numerous issues that require research and the development of a robust solution that places data control in the hands of the user. One of the regulations that deals with data protection and user privacy is the GDPR. This regulation must be made mandatory in all countries, including India. Using blockchain technology, the authors discuss various approaches for protecting user privacy and controlling data access. The authors also attempt to propose a different strategy to resolving the privacy and security issues using blockchain technology in a modified and enhanced manner.

I. INTRODUCTION

Online social media has become a daily routine for common man. More and more users tend to be using one or many social media platforms. The technological advancement in terms of connecting people through online social media is like two sides of the same coin. It has its advantages and disadvantages as investigated through different perspectives. Researchers are working on the various aspects of online social media platforms such as what are the different behavioral patterns and related privacy issues of users on an online social network (OSN). How the behaviors get transformed into personal information disclosure is the research question raised [1], New user account gets added daily and many users get enrolled to the social network and start their journey of share, like and comment unaware of the harmful reactions in the form of security and privacy concerns on the way back. Research is needed to understand the different human behavior and their related actions on online social network. Research focusing on the human behavior based on some dimension and establishes the relation between the individual's behavior and the privacy factors compromised on OSN. The mapping gives the understanding how personal information disclosure is easy and makes users ultimate target for the online social fraud. Further the authors observes that fear of missing out "(FOMO) is the perception of OSN users, which is triggered when the individual observes their peers' activities on the social media therein getting noticed and gaining popularity among other users." (Viana, Anxieties can be developed in users and so they tend to use more and more social media without thinking about the privacy and security compromises. Talking about the various situations and behaviors the research concludes that different individual behaves in different manner given the similar situations. To solve the issues, a model having multidimensional perspective such as of involvement level, perception breadth and action height is proposed to investigate the effect of behavior on privacy and security issues on OSN. This multidimensional analysis of user behaviors can support creation of pid, a list containing different behaviors with added dimensions. Though this list is limited, it supports the study on user behaviors while on OSNs and its importance on privacy and security aspects. The different dimensions of user behavior such as (depth of) involvement for example free-value feeling, wherein the user is ready to compromise his privacy for getting some free products. Next major user behavior directions are "FOMO (fear of missing out)" need to considered as a part of collective self-esteem, which makes to be (width of) perception, such as they before me, collective perception and (height of) action simple privacy self-presentation or privacy laziness. Online social networks have been growing exponentially. Next comes the concept of trust on online social media. As discussed by authors [2], For analyzing trust, the inspection depends on level of interactions proportional to increase social network sustainability. Secondly it also instigates the proportional relation between the total members and their corresponding interactions along with maintaining the privacy factor

978-1-6654-9291-1/22/$31.00 ©2022 IEEE.

involved, the authors observe that trust community plays a critical role to evaluate the openness of the social network and subsequently protect users' personal data protection concern [2]. A trust community is one that fosters an atmosphere where its members can openly and honestly discuss their thoughts, opinions, and experiences without worrying about their privacy or worrying about being criticized. It is believed that social trust can develop trust communities, which further will ensure trust in an online social network. This is indeed a way to attract more members and increase their frequency of interactions. The authors have developed a social trust model to build trust communities based on social capital [2]. The richness of member relationships, or the interactions from which members benefit, is referred to as social capital. Author [Surya Nepal et al., 2013] investigates Popularity Trust (PopTrust) and Engagement Trust (EngTrust). "Popularity trust [Caverlee et al., 2010] refers to the acceptance and approval of a member by others in the community, while engagement trust captures the involvement of someone in the community". This real dataset is then used to prove the presence of people having separate roles in a community, and make it believable that differentiating these roles are beneficial. The impact is measured by addressing the question of sustainability in the community. This is done by filtering out highly trusted members of the community. The filtering process indeed helps community developers to figure out the most involved target user group that finds a special place and are much needed in the community. The authors focused on two main objectives firstly validate the separation of engagement and popularity trusts, and then to utilize the model to understand whether the method is the sustainable? The authors observed that many unique members get categorized into the two trust models. This result leads to validating that the popularity trust and engagement trust are indeed separate heads. The analysis also shows that the uniqueness in highly interactive community numbers is proportionally lower as compared to less interactive community members. Secondly discussing sustainability factor of social networks through the model, the authors observed that social capital level decreased when highly trusted members were removed. The goal was to identify number of highly trusted members in minimum can be possible to retain the sustainability of the community. Popularity and engagement trusts were measured for all members. It was observed that removal of highly trusted members at intervals of 5% of the total population shows a decrease in the ratio of the social capital by maximum 50%. This ratio further decreases up to 80% with the elimination of 15% of highly trusted members. The analysis provides a clearly indication that the ratio of total interactions and unique interactions is indeed more in the highly interactive community and it decreases in case of low interactive community [2]. Considering data set like Facebook, the experiment was performed involving the members of the community and their different interactions. Two data sets having more members and less interactions and having less members and more interactions are used. Filtering is done based on highly influenced and active members of the community. Result shows that the sustainability of the network is challenged if the highly trusted and active members are removed from the network. Researchers are trying to understand the social media platform as a use case in order to evaluate the effectiveness of the engagement of the users. The question raised in [3] was which social media platforms can be used for analyzing given a particular application? And at the same time, focuses on various analysis and users' engagement measures on online social media. The authors observes that while using social media, users produce posts, photos, and various content which are original in nature. Network connections gets created due to interaction with other people due to exchange and sharing of comments, reactions, etc. All the content are time-stamped and creates digital archives of user activity in a combined manner. The time-stamped data provides base for researchers to observe various responses of the connections in the real time. For example, the participant tends to make a personal engagement while they post or comment, give reactions during their stay in social media platform. This personal engagement can be measured in the form of degree of their involvement. The other parameters identified by the authors are the degree of the forms of posts involving prompt comments which induces various engaging content. Further, considering multiple treatment groups, the authors also focus on other parameters like uniqueness of the posts and their relative count. Also, the count of viewed photos or videos can act as a degree of engagement at the group-level. These Metrics helped in quantification and describing the nature of engagement, as well as act as a factor to evaluate relationships between intervention engagement, participant satisfaction and thereafter retention too [3].

The user's privacy issue is of concern as many frauds can take place on online social media platforms. The question arises in the form of what are the key ideas and concepts of social network analysis that can help in fraud detection [4]. The study inferred that SNA's visual and analytical features can help to detect as well as prevent web-based attacks by effectively running the fraud prevention function. "A social network is made up of nodes (vertices) that are connected to other related nodes by edges (relationships)" [Kirchner et al.]. Kirchner in [4] considers Density, it is the overall degree of connectivity between social network nodes. It is described as the ratio of the edges in each section of a social network to the total number of edges that the social network may possibly include. This measure is extremely instrumental in determining potential fraud hotspots in financial outlets involving most of the transactional accounts and corresponding control measures. Monitoring "Credit card transaction and money-laundering" are the major areas where density metrics could be used for deeper investigations. Next kichner in [4] also considers Centrality, It indicates the structural significance of a node in the network and is a measure of a node's proximity to the center(s) of high activity in a network. Centrality value can be used in node identification that are instrumental in fraud applications and, in turn, use resources on investigating these highly probable suspects. Authors in [4] recommend to create broad subsets of the network maintaining the logical linkages of the independent networks with the help of domain experts. This is proposed to carry out by selecting already flagged nodes and recreating the probable network connections specifically considering the flagged nodes. Building a cluster (s) using the snowball method is the next step as proposed by the authors. visual patterns/links can be

used to generate metrics for network features. Further authors propose that mitigation measures can be applied as per the requirement based on the inference from SNA. The method was applied to a banking institution giving car loans, considering the suspect dealers as nodes, and were selected based on some pre-defined criteria and were recursive expansion of their linkages using the snowball method including all transaction accounts and the cars loan details. This Density measure helped in identifying the fraudulent relationships between dealers and accounts nodes which further led to identification of clusters. In the further step centrality measures was applied to the identified clusters and key actors within each cluster were circled. A proactive step was taken by the bank to zero down the target dealers and break up the identified fraud clusters. This helped banks in denying financial credit to all vehicles that followed the above pattern in future. Once the list of fraud nodes was available, the bank listed the genuine dealer and their transactions that fell into the same category. The retrospective nature of SNA, i.e., one can only react to a fraudulent instance after it has occurred acted as a major hurdle in the process as per the authors observation. Adding SNA process in association with a runtime and real transaction decreases the overall processing time to a larger extent, which is not desirable for most organizations. Hence the author recommend that the visual observations and the pattern identified through expert knowledge of associations in an existing fraud network can act as a knowledge input to create improved analytical models [4].

II. SOCIAL SURVEILLANCE PERSPECTIVE

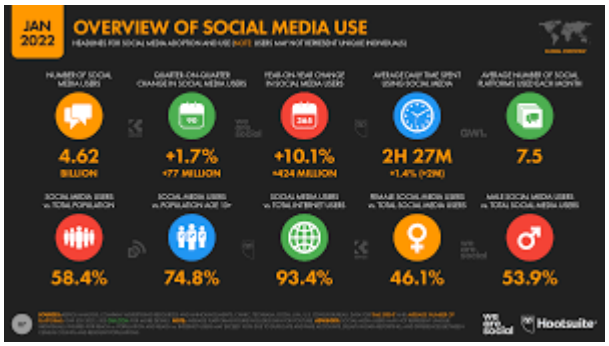

Fig. 1. Social Media Usage Statistics

As shown in the figure[11], online social media platforms is indeed an useful tool which serves the purpose of connecting people sitting at across far ends of the globe. But it surely has a major disadvantage in terms of data privacy and confidentiality which is the basis of any information system. The online social media platform providers are the service providers and they have got the pool of data with them. They have understood the potential of the data pool and then using technologies like AI, ML including neural network user profiling is achieved. This process is without the knowledge of the user. Figure2[14] shows the how social media surveillance overrides the privacy of the user. That is where the concern comes in, researchers are trying to raise this as an objection. The user should have the right to its own information. Likewise, they have the right to share the data in the offline mode, they know whom they have submitted the data and for what purpose wherein the online data collection does not follow any norms of informing the user about exactly what is the purposes of data collection and how it is going to be stored and what processes the data is going to be subjected to. This is like right to information for citizens as exercised by the government of India. Many obligations and legal perspective have been deliberated with respect to this issue with the OSN platform Facebook. So, the authors propose that the user should have all the control with them and access rights to share or not to share their personal as well as sensitive data. Surely the service provider after getting the required concern from the user can go ahead and use the data but without permission exploitation of the data privacy is a major issue that needs immediate attention.

Fig. 2. Social Media Surveillance targeting privacy

Table 1 provides comparison between the various social media platforms used widely and their way of handling the personal identifiable information.

TABLE 1

Social Surveillance Criteria	Social media Platforms			
	Facebook	Twitter	Instagram	whatsapp
Free	yes	yes	Yes	yes
Personal Data Collection	Yes, further used for prediction through artificial intelligence	Yes, through posted tweets	Yes	yes
Location Tracking	Yes, for nearby purchase areas	Yes, from where as well as which device one is using	Yes	yes
Targeting based on your shared interests while	Dynamic ads created to their individual profile	Targeted advertising through personal data collection	Yes	no

using the platform	and behavioria l patterns	for revenue generatio n		
Privacy concern taken	Yes, but in a way that the user is not explicitly made aware of	Yes, in a default setting which the user is again not aware of	Yes	yes
Post sharing	By default public	By default public	Yes	Yes end to end encrypte d
Tracking of browsing activity outside the platform	No	Yes, by giving concern for it	Yes	no
Sharing of data to third party	Yes, to advertisin g platforms	Yes, to advertisin g platforms through API	Yes	yes
Tracking of Non-Users	No	Yes, if the applicatio n has an agreemen t with twitter	Yes, without taking concern	No

III. BLOCKCHAIN PERSPECTIVE OF DATA PROTECTION AND PRIVACY

Ownership of the content is important and can be done with the help of Ushare application [5] based on blockchain technology. Authors focus on the blockchain technology as it can be applied to any digital asset transaction that are traded online. key features of blockchains include data confidentiality, availability, and integrity. Programming code, a storage file, and an account value are all components of smart contracts, a blockchain application. It enables any user to publish a transaction and form a contract. This account remains unaltered over the long run. The authors in [5] propose ushare as a unique blockchain that describes assets as data shared or broadcasted to the network. In order to share a transaction with its circle by the user, the following steps are performed. First the personal certificate authority encrypts the data and stores it on hash table. Secondly the hash id of the encrypted data is shared by the user with their members. A token value is associated with each transaction, the data owner decides the token value. It signifies the highest allowable number of shares of the data. This enables traceability and control over shareability of the data. Lastly the transactions are recorded on the blockchain with the user identity and the data hash id to record the trails. The application works in a way like first a transaction is broadcasted and the relationship System is called to verify that the data share can be allowed. This is done by checking the value of the token. Token value is decremented for future shares by user's circle members. The relation system as proposed by the authors denies any transactions of the data item are once the token value becomes zero [5]. The authors in [6] proposed the BPP framework, it is a blockchain-based privacy-preserving system for online social networks. The framework aims to provide solution to combat social surveillance as it focusses to carry out data management in a secured manner. The system aims to achieve data movement over the social network considering privacy concerns of user's personal data. The system enables the data querier either user or service provider to use the social network services in an efficient and in a full fledge privacy preserved environment. This is achieved through SNSP, a cloud server which adds flexibility and scalability to the system. Smart contract is used in Ethereum blockchain to protect the query privacy. Ethereum consensus mechanism is used for verifying the query output for its authenticity and completeness. The system works on a cloud computing environment where the data publisher outsources his/her data and can further share her/his data with many different data queriers in a flexible and scalable manner. This data is stored on the SNSP server in an encrypted form thus providing privacy protected data usage to online social network users. The indexes of the corresponding outsourced files are encrypted and recorded on the blockchain. Further execution of the keyword search for the data querier can be performed in a secure and efficient manner through the smart contract [6]. Authors in [7] discusses distributed online social network to overcome privacy issues prevailing in traditional OSN's. The proposed system works in contrast to what is followed in traditional OSNs, where the services are divided into control service and data storage. Control service handles identity management and information lookup (such as user and content information), whereas Storage service handles data storage and accessibility. The proposed system [7] arranges the two services into three layers. First is the blockchain layer which acts as control service, Second the storage layer for specifically focusing on storage service, and third is the application layer which provides the user interface to interact with the upper two layers [7]. Discussing further about the applications available on mobile platforms, where users need to install third-party services. Such services collect personal data as well as credentials without the knowledge and control of the user. To address the privacy concerns of such users the authors propose to develop data management platform which focusses on privacy protection at the personal level [8]. The system involves traditional blockchain and any other database management system also known as offblockchain storage. The system is meant for mobile users, who download and use mobile apps and services without realizing the dangers of their actions. The system involves nodes acting as entities in the blockchain and a distributed private key-value pair for storage purpose. In the blockchain the users enjoy (pseudo) anonymity, only their profiles based on services opted can be recorded on the blockchain for identity verification. The system works in the following manner, firstly the user needs to sign up, following the signup a new set of (user, service) identity is generated along with its associated permissions and

recorded to the blockchain as a Taccess transaction. Secondly the data collected (for eg., location) is stored on the blockchain as a Tdata transaction in an encrypted form. It uses shared encryption key which further gets stored on an off-blockchain key-value store. During this process only a hashed pointer to the data is preserved on the public ledger. Once this process gets completed, now the user can use Tdata transaction alongwith the pointer (key) associated to it to get the data. At this stage the consensus mechanism of blockchain is used to verify the authenticity of the user or service by checking the respective digital signature. For the service, its data access permissions are verified as well. The user can invoke the Taccess transaction in order to modify permissions availed for a particular service at any time and a new set of permissions are facilitated. This functionality also facilitates revoking access to previously stored data. The authors in [8] also propose developing a web-based (or mobile) dashboard for user's data access statistics and update permissions functionality which is in like centralized-wallets feature, such as Coinbase for Bitcoin1. The same system is further improved by MIT Media Lab [9] as considering the "Decentralized Privacy". The extension intended to give more weight to trusted nodes and efficient block computation. The MIT Media Lab defined a new trust parameter based on node behavior instead of the Proof-of-work mechanism. Proof-of work puzzle aims in solving a computationally heavy puzzle by the participants whenever a new block is to be appended to the blockchain. However, significant number of resources gets wasted in solving proof-of-work puzzles. To overcome this gap, another method for securing the blockchain, The Bitcoin community first adopted the proof-of-stake algorithm in 2011. The resource for proof-of-stake is the number of coins that are held, but with proof-of-work, the likelihood of mining a block depends on the effort done by the miner (who generates a block). An attacker can successfully complete an attack on the blockchain, in a case where he controls more than 50 percent of the entire network resources. This contrasts with the proof-of-stake mechanism, where in an attacker when tries to monopolize the coins, value of the coins held will be decreased and the network participants will be able to detect the malicious action. This works as a deterrence against attacks.

After reviewing different proposals, it is observed that blockchain can surely provide protection to personal data roaming around social media platforms. Table 2 provides overview of various existing solutions for privacy protection using blockchain technology.It is observed that one of the very important feature of data privacy is facilitating the user to forget the data after specific interval. We propose to improvise data protection over the other features of the reviewed proposals by using a hybrid blockchain i.e Public and Private blockchain. The proposed system will have all the basic requirements such as data access, data sharing, data modification and data removal to satisfy privacy features of online social media users.

TABLE 2

Data Privacy Features	Blockchain based Applications for Privacy Preserving				
	Ushare[5]	BPP[6]	Distributed[7] OSN	Decentralized Blockchain[8]	Decentralized Blockchain by MIT lab
Data Encryption	Yes	Yes	No	Yes	Yes
Access to Data	Yes	Yes	Yes	Yes	Yes
Right to share	Yes, with the help of token	Yes, through cloud based query mechanism	Yes, through layered structure	Yes, through use of Blockchain and off blockchain storage	Yes, through trusted nodes
Right to modify	No	No	No	Yes	Yes
Right to forget	NO	NO	No	No	No

IV. Conclusion

Online social network cannot be avoided due to continuous digital advancement. Users are driven by these networks or platforms to a larger extent. The service providers are collecting users' data using computational intelligence for and improvising user experience. At the same time this data is also being used for surveillance purposes. All this is done without following adequate regulations related to data privacy and protection such as GDPR. The paper thus reviewed various aspects of online social network users including different behavioral patterns and various parameters that are being materialized by the service providers for advancements in the usability of the services. It also discussed about the privacy challenges of the users personal data. Thereafter the paper discussed about the use of Blockchain Technology to cater to the needs of data privacy in an efficient manner. Different techniques have been discussed and reviewed by the authors which provides better privacy control of the user data. Finally the authors recommend the use of hybrid blockchain to facilitate all the requirements of the Data Protection while using Online social network.

REFERENCES

[1]. Viana, Thiago , Hemns, Jessica and Paterson, Julie (2021) "Towards a Multidimensional Model to Represent Human Behavior on Online Social Networks" International Journal of Cyber-Security and Digital Forensics, 10 (3). pp. 91-99.

[2]. Nepal, Surya & Paris, Cécile & Bista, Sanat & Sherchan, Wanita. (2013), "A Trust Model Based Analysis of Social Networks", Int. J. of Trust Management in Computing and Communications. 1. 10.1504/IJTMCC.2013.052522.

[3]. Lindsay E. Young, Stephanie Soliz, Jackie Jingyi Xu, Sean D. Young, "A review of social media analytic tools and their applications to evaluate activity and engagement in online sexual health interventions, Preventive Medicine Reports", Volume 19, 2020, 101158, ISSN 2211-3355.

[4]. Kirchner, Implementing social network analysis for fraud prevention, Business solutions through information technology, 2011 CGI Group Inc.

[5]. Antorweep Chakravorty and Chunming Rong. 2017. Ushare: user controlled social media based on blockchain. In Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication (IMCOM '17). Association for Computing Machinery, New York, NY, USA, Article 99, 1–6. https://doi.org/10.1145/3022227.3022325.

[6]. Shiwen Zhang, Tingting Yao, Voundi Koe Arthur Sandor, Tien-Hsiung Weng, Wei Liang & Jinshu Su (2021) A novel blockchain-based privacy-preserving framework for online social networks, Connection Science, 33:3, 555-575, DOI: 10.1080/09540091.2020.1854181

[7]. L. Jiang and X. Zhang, "BCOSN: A Blockchain-Based Decentralized Online Social Network," in IEEE Transactions on Computational Social Systems, vol. 6, no. 6, pp. 1454-1466, Dec. 2019, doi: 10.1109/TCSS.2019.2941650.

[8]. G. Zyskind, O. Nathan and A. '. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," 2015 IEEE Security and Privacy Workshops, 2015, pp. 180-184, doi: 10.1109/SPW.2015.27.

[9]. Dongqi Fu and Liri Fang, "Blockchain-based trusted computing in social network," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), 2016, pp. 19-22, doi: 10.1109/CompComm.2016.7924656.

[10]. Huichuan Liu, "Internet marketing, consumer surveillance and personal privacy: social exchange or panoptic control?," University as a Bridge from Technology to Society. IEEE International Symposium on Technology and Society (Cat. No.00CH37043), 2000, pp. 275-278, doi: 10.1109/ISTAS.2000.915646.

[11]. https://datareportal.com/reports/digital-2022-global-overview-report

[12]. Kathryn C. Montgomery, Youth and surveillance in the Facebook era: Policy interventions and social implications, Telecommunications Policy, Volume 39, Issue 9, 2015, Pages 771-786, ISSN 0308-5961, https://doi.org/10.1016/j.telpol.2014.12.006.

[13]. https://vpnoverview.com/privacy/social-media/what-does-twitter-know-about-me/

[14]. https://freedomhouse.org/report/freedom-on-the-net/2019/the-crisis-of-social-media/social-media-surveillance

[15]. https://www.digitalinformationworld.com/2021/08/instagram-amid-rising-privacy-concerns.html