International Journal of Computer Applications Technology and Research
Volume 11–Issue 08, 337-340, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1108.1007

Blockchain as a Solution of Information Security and Data Privacy Issues: Review

Ndung'u Rachael Njeri
Department of Information Technology
Murang'a University of Technology
Kenya

Abstract: The growth of technology has seen development of smart devices that are connected to each other giving rise to device-mesh technology. This has given rise to many owners of these devices sharing data through various web applications such as online marketplaces. The protection of data is paramount for every organization dealing with such data. An evaluation of Blockchain technology as a solution to data privacy is studied. The study concludes that though blockchain is the technology to pursue for securing and protection data, it has numerous challenges and limitations towards data privacy. More research is needed to guarantee an absolute data privacy protection.

Keywords: Blockchain; Information security; Data privacy; Issues of privacy; Personal Identifying Information

1. INTRODUCTION

Today's technology advancement has evolved tremendously with Artificial Intelligence (AI) taking the lead. In the IEEE computer society's top ten technology predictions for 2021, machine learning, robotics and industrial Internet of Things (IoT) were seen as the technologies that would highly hold disruptive potential by 2021 going forward, amongst other top technologies. Smart devices evolution brings with it issues of mass data collection of very high magnitude bringing the term 'big data' in the fore. The big data has issues with its management, analytics and data privacy issues [1].

[2] when writing for CPO Magazine on data privacy in the era of the Internet of Things noted that new smart home devices like the Amazon Echo and Google Home were raising numerous legal and data privacy issues, primarily because these IoT devices were recording conversations that were held in daily life. Smart toys powered by AI are everywhere playing with kids, giving them company and socializing with kids. But for them to get answers asked by the kids they have to be connected to the internet. Hackers take advantage and use such toys to infiltrate to confidential data by monitoring or illegally spying on children. A major concern with such data sourced from these smart devices is their security and privacy. How these data are used could compromise the data privacy regulations, which can result to compromising the processed information. It's the responsibility of organizations developing smart devices to take responsibility and ensure that their products are protected from unauthorized access.

Protection of information and data access is paramount for every organization dealing with huge amounts of company and personnel data. Data management is becoming an important frontier in many organizations, which are dealing with data providers, data collectors and data processors. Blockchain technology is among the latest strategies of decentralized data storage that reduces to the minimum unauthorized access of stored data through trust. Some research works [1] [3] have concentrated on bid data privacy challenges. From an information security viewpoint, many systems have been developed aiming on privacy of personal data, such as data anonymization that protect personal identifiable information (PII), differential privacy technique that adds noise to the computational procedure before data sharing, and encryption schemes that allow processing of encrypted data [4]. This paper aims to assess the solutions on data privacy issues using blockchain technology, as a recent emerging technology. Description of data privacy will be assessed and blockchain technology explained. Critical examination of blockchain will be done in light of establishing how it can be used as a solution for the data privacy.

2. DATA PRIVACY

Many literatures have been written trying to understand the term data privacy. There are many viewpoints to evaluate data privacy. No complete and comprehensive description has been given to it since many researchers evaluate privacy issues depending on where they stand. Issues of privacy can be based on the users where their personal private data is disclosed to those who are not merited and without the users' consent. Organization's data privacy would mean securing confidential data about the organization from its competitors and once such data is leaked, data privacy is compromised.

Data privacy relates to control of the distribution and use of consumer information including and not limited to statistical features of human populace such as age, income, used to identify individuals, search history and personal profile information [5] [6] [7].

Privacy is multi-dimensional that requires profound address of the issues necessary to fully understand this important issue. Many database handlers provide privacy controls based on how they understand privacy issues. Data privacy taxonomy, can be considered in four technical dimensions; data providers - individuals or organizations that provide data to be stored, data collectors - individuals or organizations that initially collects, uses and stores from the providers, data user - individual or organizations that solicits for acquired data as third parties and data warehouse- the data store itself as key players in matters data privacy. Data privacy is guided by basic tenets such as need to specify the purpose, need to acquire consent, limiting data collection and use to only what is required, data disclosure and its retention is limited as possible, data collected is accurate and security safeguards are assured and that data is open to provider who verifies compliance to these tenets. They identified four dimensions to

understand data privacy as purpose, visibility, granularity and retention [8].

The explicit responsibility of data usage, access and providence of secure safeguards that will ultimately secure that data. The data collectors are also guided on how long they can store the data and to whom they can disclose it to through consent given by the data provider. Many at times data sources/owners have no control of how their data, i.e. how it will be used and for how long. Majority of data controllers will share data at their disposal without any regard of privacy regulations. Any technology that give the data owner control of his data, how it will be used by firms and authorities without compromising security and limiting them the capability to personalize services, will go a long way to provide technical solutions to data privacy.

Smart devices that are been employed by many organizations are collecting personal data and storing them in the databases. The IoT objects are connected to each to other and are provided with internet. These objects accumulate a lot of information from their surroundings and share with each other through the internet connection over software system. They produce a lot of information that are used by reliant services such as online marketing. This raised data privacy issues since these objects distribute personal data that would reveal identities of their owners [9].

Organizations on the other hand are collecting information, combining facts from separate sources, merging and swapping them using smart software, and then selling them as merchandise compromising with the very rule of consent from the data providers. This is calling for privacy protection from the governments. Therefore, issues of user personal data disclosure and misuse is threatening the very core of information security in organizations. Many firms are concerned on how stolen data can harm their businesses and how it can be used to compromise how they do business. Wrongly acquired data can harm the reputation of individuals or organizations, which can make them lose self-image and business partners. This would be very critical and getting how to prevent and safely secure such data is among the high priorities that government, firms and individuals are seeking protection for. Though there are many legal, ethical and policy avenues advanced for data privacy protection, they are beyond the scope of this paper. This paper seeks to examine privacy protection through technical projects, and as was mentioned earlier, Blockchain will form the basis of such venture discussed herein.

3. BLOCKCHAIN TECHNOLOGY

The blockchain is a decentralized, digitized, highly distributed public ledger or record that can be accessed by anyone via the Internet, or privately through a restricted network. It works like an enormous virtual accounting book, which records details of every single transaction of data between two parties. It's replicated, shared and coordinated digital data structure that is maintained by consensus protocol and it spreads across many establishments, countries and several sites. It's under peer-to-peer network with no central control. Blockchain transactions uses cryptographic keys where public keys are used as an address to the system and the private key used for signing transactions.

The digital data is packaged into blocks, which are connected together sequentially, in a manner that makes such data immutable once recorded, unless collusion of member nodes on the peer-to-peer network to alter such data. Each block contains a hash pointer as a link to a previous block, a timestamp of when the block was created and the data being transacted. There are numerous blockchains and each blockchain serves a different purpose. Blockchains can be open and public, or they can be privately run by enterprises or even individuals. There many varieties of blockchains thus lacking single set standards, complicating how they are implemented [10] [11]. There are many participants that may qualify as data controllers- those who determine the purpose and manner of processing- for themselves and data processors - those who process on behalf of data controllers must be guided by applicable data privacy laws [12] [17].

3.1 How blockchain and data privacy works

[12] [17] considered the blockchain technology and how it works. He deduced that for blockchain to work effectively the members participating on the chain must be on the same level of knowledge about the blockchain principles. Though many users can upload any kind of data, including personal data, many generic blockchain may not be able to provide data privacy protections thus calls for the users to uphold data protection laws as they upload data onto the blockchain since data transmitted through the chain is visible to every participant though the information cannot be removed from the chain.

The data providers, data controllers, and data processors must be guided by governance agreements among the participants in data privacy. Due to the fact that blockchain deals with data which is distributed in nature, many issues emerge on who is responsible to control what data, thus posing a challenge on which data privacy laws to be followed. To achieve data privacy on blockchain, use of hash personally identifiable information is applied serving as locus link to an off- chain data store. The degree of anonymity that can be provided by blockchain is that of pseudonomity, where virtual identity is required for transactions, while integrity of blockchain largely depends on the complex proof-of-work protocol and largely on honest miners. To manage data privacy, decentralizing data over private- by design systems is way to go. The benefit of adopting such a technology is that of decentralized records that are not controlled centrally and that contains immutable transactions. These transactions are registered and authenticated thus ensuring detection of misuse and tampering of data. This can be achieved by adoption of peer-to-peer systems which blockchain technology employs [9] [12] [17].

Blockchain is a decentralized (not centrally controlled) peer-to-peer system using proof-of-work consensus algorithm that relies upon cooperation among individual nodes to carry out information transmission operations. To add more security and integrity of information while using blockchain, cryptographic public key is used for authentication. This lessens over reliance of third-parties and offer more security to transactions and identity privacy. Like in bitcoin systems, which implements cryptographic Proof of Work (PoW) together with nested chain of hashed addresses to remove the need for third party providing security and privacy even when dealing with unknown nodes. Bit-message used offers anonymity in a trustless network through transmitting encrypted messages in messaging streams. The identities of data owners and their profiles are kept private by using trustless system [13] [11]. Therefore, blockchain can be used

International Journal of Computer Applications Technology and Research
Volume 11–Issue 08, 337-340, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1108.1007

to automatically control decisions about collecting, storing and sharing sensitive data by making the ledger act as a legal confirmation for accessing or storing data since its immutable.

Similarly, researchers are working on a protocol that would sit on top of an existing blockchain that promises 'secrets contracts' as opposed to 'smart contracts' which will make the nodes on the blockchain be able to compute data without 'seeing' it. This would allow users to maintain control over personal data through preventing its monetization by online platforms. This would enhance their trustworthiness without having to give individuals access to their specific personal data.

3.2 Blockchain challenges

Blockchain technology is not a technology without issues but why is it deemed to offer security of data for organizations is due to the very nature of the blockchain, anybody wanting to tamper with it would have to make changes to records which are stored on multiple computers, or use a lot of computing power to mine a new branch of the blockchain. Many establishments and researchers are wondering whether blockchain technology would provide data privacy, integrity and anonymity as secure means to deal with data on information systems. When reviewing about blockchain technology, [14] found out that the technology has faults when it comes to offering privacy. Its anonymity element could be traced through tracing some subset of addresses that could be mapped to an IP address by simply observing the transaction relay traffic making the transaction linking possible. This would greatly compromise the data privacy since the data owners could be identified. The proposed technique from that review was transaction mixing technique in order to achieve increased privacy. A mixing transaction technique would allow the participants on the blockchain to move from one user address to another without a clear trace linking between the addresses. Such transactions could act as basic to aid improve anonymity when transaction linking becomes more puzzling [14].

The degree of anonymity that can be provided by blockchain is that of pseudonomity, where virtual identity is required for transactions, while integrity of blockchain largely depends on the complex proof-of-work protocol and largely on honest miners. This pseudonomity means that the block can be traced back to its source thus it doesn't have an exclusive anonymous status. With PoW, the majority attack, which is given at 51% is common. That is the likelihood of mining a block depends on the work done by the miner. This mechanism will make people want to link together to mine more blocks contributing to mining groups. When such a group holds 51% computing power, it can take control the blockchain. Ostensibly, it can cause security issues [15]. Other technical challenges and limitations of the blockchain technology was that of throughput, latency, size and bandwidth, versioning, hard forks, scalability and multiple chains. Very few literatures were found dealing with these challenges and this portend a rich ground of further research [14].

[16] reviewed literature on blockchain and they identified three algorithms that were designed to maintain how the blockchains could remain verifiable as distributed ledgers, one with pattern clustering algorithm that would cluster nodes into different clusters while using sequence data to represent how nodes behave, and uses similarity metrics and utilizes Euclidean distance. Another algorithm they identified was one using post quantum that ensured security against external threats thus upholding blockchain authentication, where secret keys are generated using lattice basis delegation. The third algorithm discoursed was a multilink integrated factor that would shorten time for validation while improving the dependability of the communication within the blockchain utilizing node link number that would identify those nodes with high capacity of trust and communication. However, all the three algorithms were not without issues, even though they were with high privacy and high secure metrics, malicious nodes were would still exist.

The review of blockchain technology by [14] summarized their findings, which are demonstrated on figure 1 below

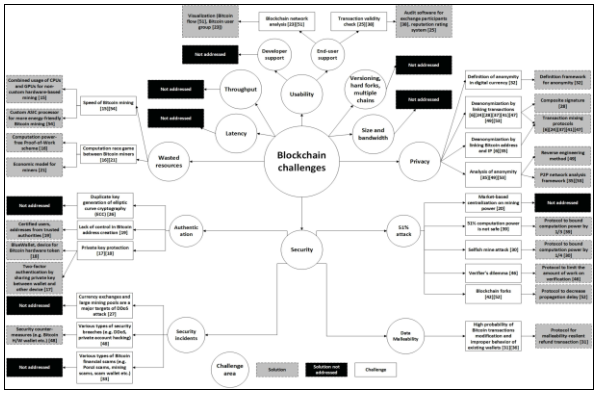

Figure 1. Summary of the identified challenges and solutions of Blockchain

3.3 Solutions for Data Anonymity

The implemented blockchain-based methods that improve on the anonymity, a data privacy concept has been applied on technologies such as Bitcoin. Bitcoin is an online payment system, which is publicly viewed as a means of sending monies anonymously. Users may take too lightly the amount of Personally Identifying Information (PII) obviously linked to this digital currency. To conform to Anti- Money Laundering and Know Your Customer regulations based on the online marketplaces, online marketplaces monitor user activity and collect PII from the bank accounts and credit cards used to purchase Bitcoins. A discoverable link between real-world identities and online Bitcoin transactions can be eventually created from that information. To eliminate this anonymity threat, privacy-conscious users rely on Bitcoin mixing services to remove identity-based connections from their coins. Several mixing schemes are been used with the bitcoin transactions such as Bitcoin Mixer, where mixing service's inclination to reprocess shared pots for storing and redistributing Bitcoins makes it readily identifiable in the blockchain and makes it easier to monitor for addresses depositing to and receiving coins from the service, Bit Launder that complicates analysis by using unpredictable payout timing. The mixing service uses repetitive extracting addresses, which makes it easier to monitor the blockchain for questionable receiving addresses, and to calculate possible originating addresses through balance variances, Bitcoin Blender where timing analysis reliability when tested was reduced due to variable time differences recorded between deposits and withdrawals across trials. However, research showed unique blockchain mixing characteristics can be used to mark each mixing service.

Nevertheless, these techniques have their limitations and therefore they are limited to offer the desired anonymity when transacting using bitcoins. To improve on anonymity on data privacy, such mixing techniques could be improved by eliminating their various limitations and either extending or remodeling to have a technique that can offer better degree of anonymity as a way to conceal personal identity while using blockchain transactions.

3.4 Blockchain Opportunities

Despite the blockchain challenges, there is a sea of opportunities for data privacy protection. Personal data can be collected into personal identifiable information (PII) and stored in distributed shared ledgers. Data privacy is about maintaining data integrity, confidentiality and availability, while blockchain might not enforce confidentiality, it enforces strong integrity and availability owning its decentralized nature in that data within it are transparent to members of the blockchain network. The ledger technology has an opportunity in cybersecurity for securing personal data to transactional data through its encrypted form of data transmission. This technology can be utilized to offer regulation relating to personal identifiable information and data privacy, organizations should embrace the technology and work towards integrating it with data privacy's regulation of the organization or country so as to achieve the best from both worlds.

4. CONCLUSION

Blockchain technology is in the recent past been one of the most reputed disruptive technologies that deemed the ultimate information security and data privacy breach solution. Though in its very early stages of development, many governments and reputed organizations are on research to seek how this technology could help to secure data such as land records, patient's records in the health environments, supply chain markets data and even the data privacy for online platforms. Blockchain is at infancy since its implementation is still very low. This paper however showcases the benefits of blockchain, its various challenges especially on data privacy and could be solutions for those challenges.

The key point of this paper was to explore whether this technology can be used as solution to data privacy. In above literature, it was pointed out that among major blockchain challenges was that of anonymity since the kind provided by blockchain is pseudonomity where one would be traced back through IP address mapping to identify the owner of transaction, yet the anonymity envisaged is that regardless of any trace, owners of transactions cannot be identified.

Several literatures studied herein reports how blockchain was envisaged to work but very few shows how it's been implemented and the success about its implementation. The various blockchain-based techniques used to provide data privacy by enhancing anonymity were without their challenges.

It's the researchers' assertion that as much as blockchain technology is looked upon to be a solution of data privacy, it's still yet to be demonstrated as the ultimate solution of data privacy. The researcher feels that more research is needed to model data privacy enhancing techniques or methods, which can guarantee better data protection and privacy based on blockchain technology. This doesn't mean that blockchain is not worthy technology as far as security of data is concerned. The literature shows how data recorded on blockchain would be difficult to interfere with since it required consensus of participants to change any details and this is achieved through rigorous proof-of-work algorithms thus immutability of data is achieved, which strongly provides data security. Therefore, blockchain is commendable technology as far as data security is concerned though much more research is required to improve on its data privacy provision.

5. REFERENCES

[1] S. Yu, "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data," in IEEE Access, vol. 4, pp. 2751-2763, 2016, doi: 10.1109/ACCESS.2016.2577036.

[2] Nicole Lindsey, March 17, 2017 - https://www.cpomagazine.com/data-privacy/data-privacy-era-internet-of-things/

[3] Soria-Comas, J., Domingo-Ferrer, J. Big Data Privacy: Challenges to Privacy Principles and Models. *Data Sci. Eng.* **1,** 21–28 (2016). https://doi.org/10.1007/s41019-015-0001-x

[4] Zyskind, G., & Nathan, O. (2015, May). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops* (pp. 180-184). IEEE.

[5] Foxman, E. R., & Kilcoyne, P. (1993). Information technology, marketing practice, and consumer privacy: Ethical issues. *Journal of Public Policy & Marketing*, *12*(1), 106-119.

[6] IEEE TRENDS- https://www.computer.org/press-room/2017-news/top-technology-trends-2018

[7] Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. Journal of the Academy of Marketing Science, 45(2), 135-155.

[8] Barker, K., Askari, M., Banerjee, M., Ghazinour, K., Mackas, B., Majedi, M., ... & Williams, A.(2009, July). A data privacy taxonomy. In British National Conference on Databases (pp. 42-54). Springer, Berlin, Heidelberg.

[9] Conoscenti, Marco; Vetrò, Antonio; De Martin, Juan Carlos (2016). Blockchain for the Internet of Things: a Systematic Literature Review. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir (MAR), Nov. 29 2016-Dec. 2 2016. pp. 1-6

[10] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[11] Aitzhan, N. Z., & Svetinovic, D. (2016). Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Transactions on Dependable and Secure Computing*.

[12] McMyn, A., & Sim, M. (2017). R3 Reports with Hogan Lovells.

[13] J. Warren, "Bitmessage: A peer-to-peer message authentication and delivery system," white paper (27 November 2012), https://bitmessage. org/bitmessage. pdf, 2012.

[14] Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where is current research on blockchain technology?—a systematic review. *PloS one*, *11*(10), e0163477.

[15] Lin, I. C., & Liao, T. C. (2017). A survey of blockchain security issues and challenges. *Int. J. Netw. Secur.*, *19*(5), 653-659.

[16] Gorkhali, A., Li, L., & Shrestha, A. (2020). Blockchain: A literature review. *Journal of Management Analytics*, *7*(3), 321-343.

[17] https://www.hlengage.com/_uploads/downloads/5425GuidetoblockchainV9FORWEB.pdf