Blockchain Technology: Decentralized Storage to Protect Privacy

R. Thaney[1], V. Mishra[2], G. Mishra[3, #]
[1]*Bharati Vidyapeeth College of Engineering, New Delhi*
[2]*Indian Institute of Technology, Delhi*
[3]*Guru Gobind Singh Indraprastha University, New Delhi*

[#] Corresponding Author, Email: *gargi.mishra1983@gmail.com*

***Abstract*— The lately witnessed increase in incidents of surveillance and security breaches, compromising user's privacy and personal data and information question the current model of organizations, in which third-parties collect and control tremendous amounts of a person's personal data. Bitcoin and other popular blockchains like Ethereum have proved that trusted, auditable computing is possible in the financial space using a decentralized network of peers accompanied by a public ledger. This paper describes a decentralized personal data storage system based on blockchain which includes the proof-of-work model that ensures that the end user owns and controls personal data without compromising security and privacy. A protocol that turns a blockchain into an automated access control manager that does not require the use of a third party is implemented and analysed. Unlike Bitcoin, transactions in this system are not strictly financial but instead they are used to carry instructions, such as storing and sharing data. In this paper, possible future uses of blockchain that could manifest its use into a well versed and robust solution for trusted computing problems that are still prevalent in our society are also covered.**

***Keywords*— Bitcoin, Blockchains, Decentralized storage, Data privacy, Auditable computing, Financial transaction.**

1. INTRODUCTION

The amount of data and its quantity are increasing every second and more data has been generated in the past couple of years since the inception of the human race. It is projected that 25% of the worlds data has been collected till now [1]. Three hundred petabytes of personal data is solely collected by Facebook, which is hailed as the biggest online social-networking site and there are many other such internet giants. Most of the innovation and economic growth that is happening today is due to Big Data and data is being collected continuously and is also being analysed. Many of the organizations use the data collected by them to personalize services, to help them make better decisions, and predict the trends of the future [2]-[3]. Today, the most valued asset in our economy is data [4]. While we all are reaping the benefits of a data-driven society, there has been a rising public concern about user privacy. Both private and public centralized organizations take advantage of large amounts of private and susceptible data. People have little to no control over how their data is stored and how it is used. Recently, the media has frequently covered contentious incidents related to privacy. Various attempts have been made by organizations to address the privacy concerns, on both governmental perspective [5] and technological standpoint. The proprietary verification software used by most of the significant companies in the industry is implemented based on the OAuth protocol [6], which makes them act as centralized and trusted authorities. Many methods have been developed by researchers based on the security viewpoint, which target privacy issues and pay attention to personal information [7]. Data anonymization techniques endeavour to guard individually restricted information. Privacy-preserving techniques consist of differential privacy, a method that perturbs information or adds

disturbances to the computational process prior to sharing the data [8], and encryption methods which allow computations and queries to be run over protected data. Predominantly, the fully homomorphic encryption [9] methods permit any calculation to run over encrypted data, but are too incompetent to be used as mainstream technology currently. Recently, a new set of explicable systems have been invented. The system of this kind is Bitcoin, which gives users the functionality to transfer currencies (bitcoins) securely without a central supervisory body, using a publicly certifiable open ledger (or blockchain) [10]. Since then, other projects [11] are also established which focus on applicability of blockchains for serving other functionalities which require trusted computing [12]-[13].

In this paper, blockchain [14] is utilized to construct a personal data storage service with the focus on decentralization of user data and privacy. Further, the proof-of-work model in the blockchain is implemented to authenticate the authenticity of transactions on each block. The scope of future improvements is also discussed which can be made to the underlying technology. This paper showcase that Blockchain services have the potential to become a critical resource in trusted computing. The remaining paper is organized as follows. Section 2 describes the methodology and problems that exist with the current structure that is centralized in nature, Subsection 2.1 proposes the solution as the storage using blockchain and throws light on future scope of blockchains, Section 3 discusses the results obtained and in section 4 paper is concluded.

2. METHODOLOGY

In this section, visual representation of blockchain elements and code flow is given. Fig. (1) shows the elements of block chain. The flow of code is described using Fig. (2). The implementation of block chain technology can be presented in three steps: In the first logical step, Block Structure is decided. In this paper, only the most essential blocks are included to keep the implementation simple. The blocks included in this implementation are Index, Timestamp, Data, Hash and Previous hash. Block Hashing In the second step, hashing of blocks is implemented. The integrity of data is protected by block hashing. The contents of the block are taken over by a SHA-256. Third step belongs to block generation. While generating a new block, hash of the previous block must be known. Rest of the required contents like index, hash, data and timestamp should be created after that. The end-user can then provide the block data.

Problems with the centralized system: This technology addresses the privacy concerns that are faced by users when third-party services are involved. It focuses exclusively on mobile platforms, in which services deploy applications to be installed by the users. These applications continuously collect personal data of the user without their consent. In this analysis, it is assumed that the services are honest-but-curious. It can be renowned that other data privacy concerns can also be addressed by the same method [15], such as therapeutic data shared by patients for methodical research while having the resources to scrutinize how it is used and the capability to immediately stop taking part. Contrary to this, the system described in this paper protects the end user from the following widespread confidentiality issues. ***Data possession:*** The proposed solution focuses on making sure that the users are given the ownership and control of their own private data. In practice, the system gives the ownership of the data to the users and the services are treated as guests with delegated permissions. ***Data lucidity:*** Every user has absolute lucidity about the kind of information being collected about

them and how it is being used. ***Fine-grained access control:*** One chief concern is that it is compulsory for users to endow a list of permissions upon signing up on these mobile applications. These permissions are approved for an indefinite period and the only way to change the contract is by opting-out. Contrary to that, in this structure, the user has the privilege to change the list of permissions and invalidate access to the data collected before at any given time. This system can be applied to improve the existing permissions message box in mobile applications without changing the GUI. The access control policies would instead be stored on a blockchain, and only the user would have the right to change them.

2.1 Blockchain

In this paper we proposed a solution that offers the user to store personal data on a blockchain. Blockchain is decentralized by definition and it has no central authority. It's decentralized nature and absence of central authority (which can abuse and misuse the user's data and information) helps retaining user privacy.

Fig. 1: Block Chain Elements

A blockchain, is a type of data structure that allows sharing data across a distributed system of computers after its transactions are identified and tracked digitally and this ends up creating a distributed network of trusted nodes. The transmission of possessions can be tracked in a transparent and secure way using the distributed ledger technology [16] offered by the blockchain. Important features of Blockchain technology are discussed here.

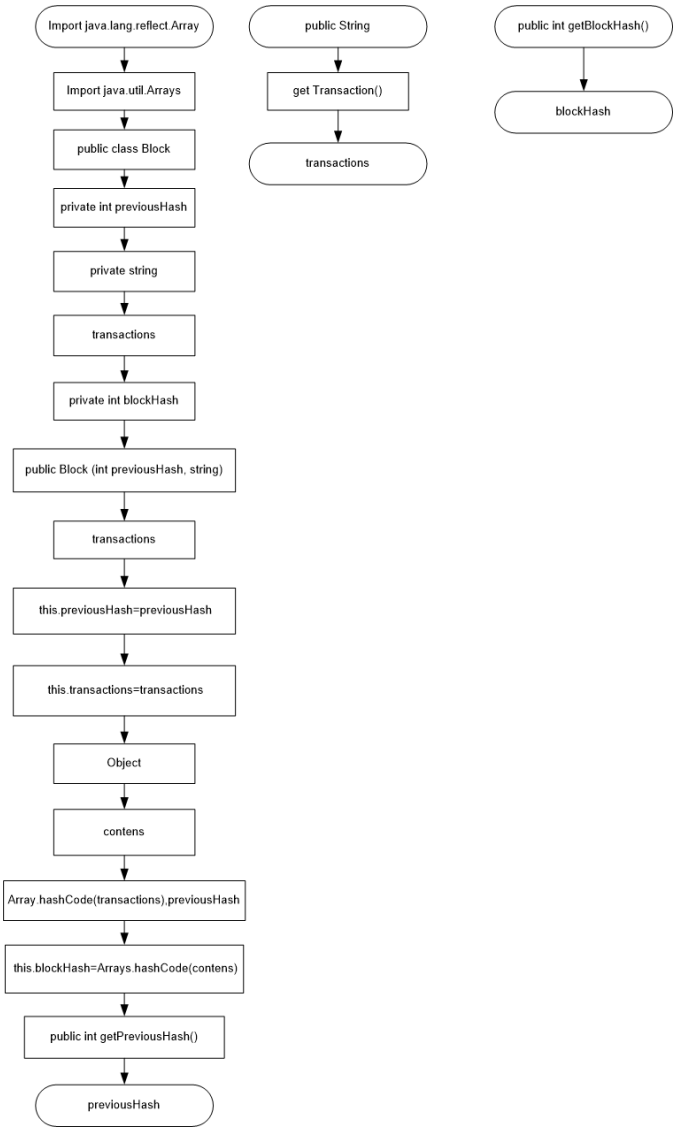

Fig. 2: Flow of Code

Data security: Since the financial information is stored across a network of computers, the task of snooping data becomes much more intricate for hackers. In centralized systems, breaking through only one server is quite easier. In contrast to centralized systems, falsifying a balance or making a fraudulent transaction on a blockchain can only be achieved if the bulk of the network is compromised [17]. Hacking a single block would require a lot of computing power which can only be provided by many supercomputers chained together and since many people can't afford that kind of technology, it is extremely difficult for even the most accomplished cyber-criminals to hack a single block. Being able to break through

enough servers to forge records on the blockchain is practically impossible, especially as hackers would need to commit a breach on each node simultaneously. ***Proof-of-work:*** A key feature of this method is its asymmetry which implies that the work should be reasonably difficult on the side of the requester but should easily check the service supplier. This scheme is also known as a CPU cost function and it is different from CAPTCHA. ***Practically un-hackable:*** As the Blockchain does not require the permission of a human, it is extremely hard to hack as it will require control over most of the nodes and possess CPU power equivalent to 51% of the total processing power of all nodes on a blockchain combined, which is practically not feasible. ***Network:*** Every new transaction is broadcasted over the nodes. It is then stored on a block by the node which then works on generating a new proof-of-work timestamp for the block previously used. When a proof-of-work method is found by a node, the block is broadcasted to all the remaining nodes. The block is recognized by the node if all transactions in it are legitimate and not previously used up. The recognition of the block is done when nodes start working on generating the block after that on the existing chain by using the hash of the accepted block as the previous hash. The longest chain is deemed to be correct one by the nodes and they keep working on extending it. The end-user can then provide the block data.

Validating the integrity of the block: The integrity of each block or chain is validated at every instant of time [10]. This is shown in Fig. (3) and it is true particularly when we want to check whether to accept the new blocks received from other nodes or not. Once the blockchain has been structured and built, next step is building the decentralized application. The idea of our application is to build a secure database that is based on this blockchain so that it can be used to store precious amount of data that can be researched further. So, the basic idea of building a secure database is to create a file which consists of data and indexes of all the blocks present on the blockchain. In this implementation Ecmascript6 and NodeJS are used for building the application. And one more advantage of this technique is that encryption of the files is not required.

Fig. 3: Validating the blockchain

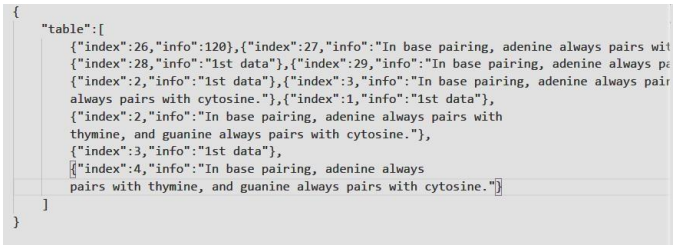

Fig. 4: Contents of the file stored

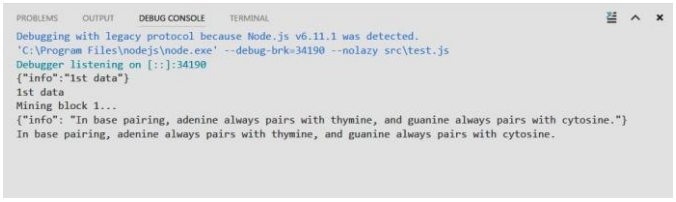

Fig. 5: Mining a block

The data present in file will be available to all the verified users of the application, so that they can store and share valuable information. Whenever a block is initialized in the blockchain, the file selector reads the file and parse the JSON and append the data of the block to the JSON object. This whole process is executed whenever the block is created or initialized. The major advantage of this technology is that the data collected in the text file as JSON can be researched further for the good of society and would be freely available and more over it is based on blockchain so it cannot be hacked or tampered.

3. RESULT AND DISCUSSION

The results are recorded at every step while implementation of Blockchain and shown in Fig. (3), Fig. (4) and Fig. (5). This paper highlights the potential future applications of blockchain. It could play a significant role in shaping more distributed state-of-the-art trusted computing platforms, compared to current systems. The advanced approach of Blockchain is to never let a service scrutinize the raw data, but instead, it allows to run computations directly on the network for obtaining the final results. If we split data into shares [18], we could then use the secure Multi-party Computation method to securely assess any function. Blockchains in general, assume that all nodes are equally trusted and their contribution in the collective decision-making process is solely based on their computational resources (known as the Proof-of-work algorithm) [19]-[20]. In other words, for every node n, $trust_n \propto$ resources(n) (probabilistically). This relation decides the node's weight in votes which leads to adverse effects, most notably vulnerability to attacks, excessive energy consumption and high-latency.

4. CONCLUSIONS

In this paper, a method to implement a decentralized storage system is proposed, in which the network has potential to give more value to the nodes that are trusted and is able to effectively compute more blocks. The model is resistant to attacks since it requires time for trust to be earned with in the system. Though this mechanism has the potential to draw some other types of attacks (one possibility is when the status of the nodes starts to increase and they start acting maliciously later during the operation), it can be mitigated by selecting several nodes at random based on their trust, to vote on each block, then taking the vote which is in equally-weighted majority. Single actors irrespective of their trust-level are prohibited from having too much power this way.

REFERENCES

[1] Schwab K, Marcus A, Oyola JO, Hoffman W, Luzi M. Personal data: The emergence of a new asset class. In An Initiative of the World Economic Forum, 2011.

[2] Wilkinson S, Boshevski T, Brandoff J, Buterin V. Storj a peer-to-peer cloud storage network, 2014.

[3] Shannon PT, Reiss DJ, Bonneau R, Baliga NS. The Gaggle: an open-source software system for integrating bioinformatics software and data sources. BMC bioinformatics, 7(1), 176, 2006.

[4] European Commission: Proposal of comprehensive reform of data protection rules to increase users control of their data to cut costs for business, 2012.

[5] Xing Q, Wang B, Wang X. POSTER: BGPCoin: A Trustworthy Blockchain-based Resource Management Solution for BGP Security. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2591-2593, 2017.

[6] MandjeeT. Bitcoin, its legal classification and its regulatory framework. Journal of Business Securities Law, 15(2), 157, 2015.

[7] Shamir A. How to share a secret. Communications of the ACM, 22(11), 612-613, 1979.

[8] Gentry C, Halevi. Implementing gentrys fully-homomorphic encryption scheme. In Annual International Conference on the Theory and Applications of Cryptographic Techniques S. Springer, 129-148, 2011.

[9] Ali M, Nelson JC, Shea R, Freedman MJ. Blockstack: A Global Naming and Storage System Secured by Blockchains. In USENIX Annual Technical Conference, 181-194, 2016.

[10] Barber S, Boyen X, Shi E, Uzun E. Bitter to better how to make bitcoin a better currency. In International Conference on Financial Cryptography and Data Security, Springer 399-414, 2012.

[11] Cooley R, Mobasher B, Srivastava J. Web mining: Information and pattern discovery on the world wide web. In Tools with Artificial Intelligence, 1997.Proceedings. Ninth IEEE International Conference, 558-567, 1997.

[12] Kosba A, Miller A, Shi E, Wen Z, Papamanthou C. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In Security and Privacy (SP), IEEE Symposium, 839-858, 2016.

[13] Bamert T, Decker C, Elsen L, Wattenhofer R, Welten S. Have a snack, pay with Bitcoins. In Peer to-Peer Computing (P2P), IEEE Thirteenth International Conference, 1-5, 2013.

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

[15] Vagata P, Wilfong K. Scaling the Facebook data warehouse to 300 PB. Facebook, heattu, 30, 2014.
[16] Miers I, Garman C, Green M, Rubin AD. Zerocoin: Anonymous distributed e-cash from bitcoin. In Security and Privacy (SP), IEEE Symposium, 397-411, 2013.
[17] Atzei N, Bartoletti M, Cimoli T. A survey of attacks on Ethereum smart contracts (SoK). In International Conference on Principles of Security and Trust, Springer, 164-186, 2017.
[18] Croman K, Decker C, Eyal I, Gencer AE, Juels A, Kosba A, Song D. On scaling decentralized blockchains. In International Conference on Financial Cryptography and Data Security, Springer, 106-125, 2016.
[19] Mizrahi IBCLA, Rosenfeld M. Proof of Activity: Extending Bitcoins Proof of Work via Proof of Stake, 2014.
[20] Andrychowicz M, Dziembowski, S, Malinowski D, Mazurek L. Secure multiparty computations on bitcoin. In Security and Privacy (SP), IEEE Symposium, 443-458, 2014.