IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 2)
Available online at: www.ijariit.com

Implementing decentralized storage using blockchain to address data ownership and privacy concerns

Ojus Kapoor
ojuskapoor@gmail.com
Bharati Vidyapeeth's College of Engineering,
Paschim Vihar, Delhi

Suyash Arora
arora.suyash@gmail.com
Bharati Vidyapeeth's College of Engineering,
Paschim Vihar, Delhi

Shivam Gupta
gupta96.shivam@gmail.com
Bharati Vidyapeeth's College of Engineering,
Paschim Vihar, Delhi

Siddharth Gorey
sid.gorey@gmail.com
Bharati Vidyapeeth's College of Engineering,
Paschim Vihar, Delhi

Silica Kole
silica.kole@bharatividyapeeth.edu
Bharati Vidyapeeth's College of Engineering,
Paschim Vihar, Delhi

ABSTRACT

The currently seen tremendous increase in events of surveillance, security breaches and unauthorized access intervening people's privacy, information, personal data requires the need to doubt and put the present model into question which seems dubious, in which third parties control as well as collect tremendous amounts of a person's personal data. Bitcoin and other popular crypto currencies like ethereum have proved in the financial sector that reliable, trustworthy computing is feasible by use of a decentralized network of peers accompanied by a public ledger. We present a decentralized personal data storage based on blockchain and also including proof-of-work that ensures users themselves own and control their data without compromise of security and privacy. We introduce a roll-back mechanism to the blockchain and by implementing an algorithm that converts a blockchain into an automatic access control manager that does not require any trust in a third-party source. The transactions in our implementation are not financial like in the case of cryptocurrencies, rather they are used for instructions, like storing and sharing data. Finally, we discuss the probable future extensions to the blockchain technology that could harness them into an exceptional solution for eliminating the trust related issues prevailing in our society and internet.

Keywords: *Blockchain, Storage, Privacy, Proof-of-work, Roll-back, Peer-to-peer, Decentralized.*

1. INTRODUCTION

The amount of data and its volume are tremendous and more data has been created in the past two years than entire history of humanity.

The estimates say that most of the data that exists has been collected and stored in the last two years. The data collected amounts to a huge number and all of it of personal data and images of one user or the another [2]. That is just one internet giant. In this era of Big Data and Data Mining, data has become among the most valuable resource and its being constantly collected and analyzed. Large internet giants [6] and the companies who collect, purchase, sell our data use it to gain insights on user behavior, group sentiments, and trends to predict the next right step for the future and growth of the company or the client. The recently witnessed Facebook data breach by Cambridge Analytics and the doubt that it was a part of election meddling makes data and user privacy a far more serious problem [3].

No doubt that the current data rich society has multifold benefits and conveniences, there is an ever-growing public outrage and concern about their own privacy. Centralized organizations have been known to misuse and even sell huge amounts of personal and private information for their own profits. Users have no control whatsoever over the data that is being stored [12] about them and the way that data is being used. In recent months, media too has awakened to this "data" crises and hence a search for an appropriate solution is underway.

2. RELATED WORK

There have been numerous attempts to address these data and user privacy issues, from legal perspectives [8], as well as from a technological and developer's point. Taking the perspective of user security, the developers have proposed multiple techniques. The various attempts at data anonymization try to secure the user identifiable data. The various methods to protect privacy has *differential privacy*, a simple technique that simply adds rubbish to the data, also called noise, before transmitting it over a server [11], and there exists encryption methods that require some mathematical computations to be done before the data could be read. Full encryption [4] schemes allow processing to be performed over the encrypted data directly but in the current phase of development are very inefficient.

In this decade, a whole new technology, the blockchain, of truly accountable systems has been introduced. The first such implementation of the blockchain was the Bitcoin [14], which is digital currency that is decentralized [15] and no central authority has any rights over it and all of the transactions of that currency are truly transparent and full proof as they are maintained in a public ledger file.

Our Contribution

1) We use blockchain technology [1] along with combining it with storage provisions to construct a personal data storage system focused on user data decentralization and privacy. 2) We implement proof-of-work in our blockchain to verify the legitimacy of transactions in each block. 3) We incorporate roll-back mechanism in the blockchain in case of infected node. 4) We lay down the possible future extensions of this this technology to address the rising user data privacy concerns.

3. THE PROBLEM WITH CENTRALIZED SYSTEM

In this paper, we have tried to address major privacy issues, users have to face while using services which are offered for free. We have extended the major part of research on various sources which tend to collect user's personal data including mobile applications, websites collecting cookies from the web browser, the social media platform and email services etc. These collect data continuously and in this the user has no say as the permissions of data access has to be granted to enable the usability of the application. In our work, we consider that the services offered to us are honest but with the doubt of uncertainty. We may note here that this same system could be used for data concerns in various other fields like medical, voting where this could serve as a transparent and reliable system to cast and track votes count.

The major factors are:

Data Ownership

Our proposed solution is focused on providing the users with the ownership and control over their personal data that is being tracked by various services being used by us.

Data Transparency and Auditability

While using the third-party services, it is essential for a user to know about the extent of data that is being collected. Even after collecting the data, collected data should be handled in a secure way.

Additionally, this solution extends its usability by providing a suitable data storage option and aspects of reusability at the same time

Provides the suitable data metrics

To get insights about the audiences, places, and all such information about them is solely dependent on data, most of the sectors including transport, infrastructure and more of like these. Also, all the latest technologies like machine learning, Ai etc. are data-driven, and these are further improved with the inclusion of each bit of data.

So, it is equally essential to managing the data. Managing the data includes Deciding the strategy of data determining the objective. Integration of data since data is collected from various sources, it needs proper unification. Standardization of data since data is growing enormously, it is better to compress and sort the data for its efficient use Blockchain based data structure provides the best possible way to solve the problem as it provides hack-proof data storage, easy data recovery, and transparency as some of the advantages.

4. SOLUTION: BLOCKCHAIN

The proposed solution offers user's personal data storage over a blockchain structure which is decentralized by definition and no central authority is present which may hence be involved in exploiting and misusing the user data and information hence retaining user privacy to the fullest.

By definition, a blockchain is just a type of data structure implementation that facilitates identifying and tracking transactions digitally using the common ledger file and broadcasting all this data and information across a decentralized network of nodes or computers, making it a tamperproof network. The distributed ledger technology which is used in blockchain provides complete and outmost transparency over the data and also provides a protected way for discovering the owner of data and also helps in tracking down the movement of assets within the network.

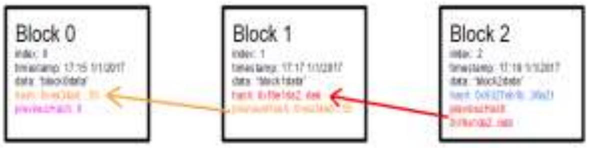


Fig. 1 Blockchain Elements

Data Security

The concern of data security is addressed as using the blockchain for storing various kinds of data like financial transactions [19] etc. The system is practically tamperproof and impossible to hack.

In the case of tampering, just one server is not enough to manipulate the data and for unauthorized access, altering the ledger file or the data is only possible when acceptance and conformation from more than half of the network nodes is received by tampering them too i.e. 51% of the same is hacked. Tampering with multiple servers can prove to be extremely challenging, even for the cybercriminals belonging to the top of the pyramid with great skills try to do that.

Network

Considering the fact here, as told before, in order to tamper the data in the blockchain, one needs to hack 51% of the chain which requires very high computing function to be solved which is close to impossible while talking about an established network.

Proof-of-work

The work done in order to mine the block i.e. adding a new block to the blockchain requires a proof that it is added legally by solving the computing function.

Also, it is made difficult each time a new block is added. So, it is essential for this process to asymmetric i.e. The work must be difficult while mining but easier to verify that the block is correctly mined. It is different if we compare it with CAPTCHA, which is designed for a human to solve quickly, rather than a computer.

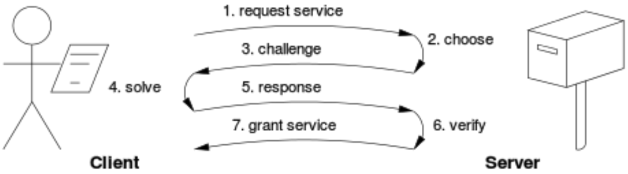


Fig. 2 Proof-of-Work Flow

Practically un-hackable

As the blockchain is permission less it is extremely hard as it will require to take the control over the most of nodes and possess CPU power equivalent to 51% of the total processing power of all nodes of the blockchain combined which is practically not feasible. New Transactions

Transactions

When the new transactions are made, they are broadcasted to all nodes into the network instantly and in order to mine a block, each node constantly keeps working on solving the proof-of-work function for its block. When a node finally finds a proof-of-work, it broadcasts the block over the network. Block is added to the network only if all the transactions done in it are valid. Hence new block is added into the blockchain and previous hash is incorporated into the new block.

5. IMPLEMENTATION

The major elements and attributes of the implementation are:

Block structure
These are details of the initial block which is also known to be the genesis block.

Mix hash

A 256-Bit hash which demonstrates, joined with the 'nonce', that an adequate measure of calculation has been completed on this block: Proof-of-Work (PoW). Block Header Validity, it permits to confirm that the block has truly been cryptographically mined [20], accordingly, from this viewpoint, it is legitimate.

Nonce

A 64-bit hash which is mixed with the hash code.

Index

It is just kept in order to keep the indexing of blocks in a block-chain.

Difficulty

It is a scalar value directly proportional to level of difficulty, higher the difficulty, more computation is required by the miner in order to prove the validity of the block. While testing the block-chain, its kept low in order to reduce the time delay in mining.

Time stamp

It is a scalar value which is equal to the output of time() function.

Storing the blocks

An in-memory JavaScript array is utilized to store the blockchain. The principal block of the blockchain is called a "genesis block", which is hardcoded in the framework.

Validation

At any given time, we should have the capacity to approve if a block or a chain of blocks are legitimate. This is needed particularly when we get new blocks from different nodes and must know on their validation with confidence.

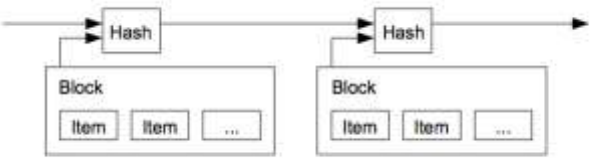


Fig. 3 Block and block's hash

Mining Algorithm

It is done in order to ensure digital computations and algorithms are solved by the nodes before the process of addition of blocks and we have used Proof of Work (PoW) [16] in order to do the same.

We introduce an extra value know as nonce in order to make sure that header file of the blocks is led by enough number of zeroes. Presently the mining capacity is modified to make the hash, yet in the event that the block's hash doesn't lead with enough zeros, we augment the nonce value, make the new header, ascertain the new hash and verify whether it leads with enough zeros or not.

Synchronization of Blocks. Reading the data signifies that when we have number of blocks in the block-chain, then it is required to sync the blocks as while there is only genesis block present, we can simply reach out to the block but synchronization is required in order to fetch the data from other blocks.

The process of synchronization of blocks require:

- When a node creates another block, it communicates it to the system
- When any node joins with another peer, it inquiries for the most recent block

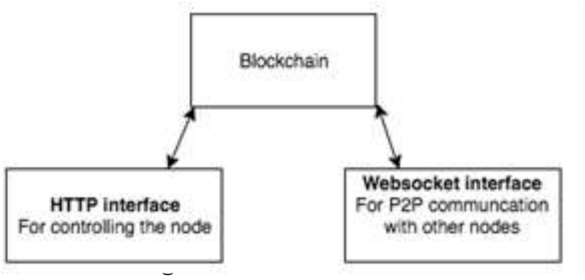
At the point when a node experiences a block that has an index bigger than the current known block, it either includes the block it's present chain or inquiry for the full blockchain.

Ethereum [7] is an open database that keeps a note of all the transactions. This database does not require any third party or a central authority to keep it up and secure it. It works as a system in which people can make peer-to-peer [5] transactions with no help from third parties.

The Ethereum blockchain is essentially a transaction-based state machine where the first state is a genesis block and final state is the current state in the blockchain.

Ethereum consists of smart contracts [13] that are used for running simple business logic, verifying signatures and cryptographic objects.

It is essentially a component including advanced resources and at least two parties, where some parties place resources in and resources are consequently redistributed among those parties as indicated by a formula in light of specific information that isn't known at the time the contract is started.



Ethereum smart contracts [18] are written in Solidity, which is a programming language that runs on ethereum virtual machine.
For the development and testing purpose we have used *testrpc* which is a *Node Js* based ethereum client. It utilizes ethereumjs to reproduce full client behavior and make creating Ethereum applications considerably fast. Also, we are using web3 for integration of our frontend with backend (smart contracts). We have used a collection of libraries – *web3.js,* which allows us to interact with a remote ethereum node using an *HTTP* connection.

We are deploying the application on a private test network. Other private networks such as *Rinkeby* can be used. Ethereum decentralized application requires that a special file called the smart contract be used to implement functions and classes usable by the peers connected into the network. This contract file contains the definitions that are used to implement the data structure of the storage that is being implemented in our decentralized application. It also provides an interface for utilizing the full functionality of the app.

6. FUTURE SCOPE

Proof-of-Work reasons that nodes which offer critical assets into the framework are basically trustable [21]. Utilizing comparable thinking we could characterize another dynamic measure of assuming that this depends on node behavior. Great actors that take part into mining are rewarded. We could set the trust of every node and reward in the future too keeping in mind the past behavior and so it does not have to keep computing forever. Proportionally, since we are managing a double arbitrary variable. A straightforward method to surmised this likelihood is by tallying the quantity of good and awful moves a node makes, at that point utilizing the capacity to end it.

With this measure, the system could give more weight to trusted nodes and compute blocks more productively. Since it requires investment to procure confide in the system, it ought to be impervious to possible breach. This instrument could conceivably draw in different kinds of breach, for example, nodes expanding their reputation just to act malignantly at a later time. This may be moderated by arbitrarily choosing a few hubs, weighted by their trust, to vote on each square, at that point taking the similarly weighted dominant part vote. This ought to keep single on-screen characters from having excessively impact, paying little attention to their trust level.

7. CONCLUSION

Personal data hence should be limited or restricted from the reach of third parties offering us free services in exchange of valuable information as it may lead to various possible attacks and misuse or rather could be used to manipulate.
Instead, users should have full rights over their data. Our solution provides a bridge between a user and third-party services where the user can control the access over data and restrict the use of the information being used by third-party services. And this is done by preparing a stack using the blockchain technology and decentralized storage.

Moreover, the blockchain perceives the users as the proprietors of their own information. Organizations hence can center around using information without being at all worried about appropriately securing and overseeing them. Besides, with a decentralized system like ours, settling on lawful and administrative choices and laws about gathering, putting away and sharing delicate information will be simpler.

8. REFERENCES

[1] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.
[2] K Schwab, A Marcus, JO Oyola. Personal data: The emergence of a new asset class. In an initiative of the world Economic forum, 2011.
[3] European Commission. Proposal of comprehensive reform of data protection rules to increase user’s control of their data to cut costs for business. 2012
[4] David Shrier, Weige Wu, Blockchain and Infrastucture, May 2016.
[5] Shwan Wilkinson, Tome Boshevski: Storj: A peer-to-peer cloud storage network.
[6] Jaun Perez. Facebook, google launch data portability programs to all, 2008.
[7] Jon Evans. Bitcoin 2.0: Sidechains and ethereum and zerocash, oh my! 2014
[8] Michal Resk. How much information is there in the world? 2014
[9] Muneeb Ali, Jude Nelson, Blockstack: A global naming and storage system secured by blockchains, June 2016
[10] Craig Gentry: Fully homomorphic encryption using ideal lattice. STOC vol 9.

[11] Adi Shamir. How to share a secret. Communications of ACM.

[12] Scaling the facebook data warehouse, 300pb 2014.

[13] Ahmed Kosba , Andrew Miller , Elaine Shi, Zikai Wen, Charalampos Papamanthou. Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts. 2016 IEEE Symposium on Security and Privacy (SP)

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

[15] Kyle Croman, Christian Decker, Ittay EyalAdem Efe, GencerAri Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, Emin Gün Sirer, Dawn Song, Roger Watten hofer. On Scaling Decentralized Blockchains. Springer (LNCS, volume 9604).

[16] Bentov, I., Lee, C., Mizrahi, A., Rosenfeld, M.: Proof of activity: extending bitcoin's proof of work via proof of stake.

[17] Shawn Wilkinson, Jim Lowry. Metadisk: Blockchain-Based Decentralized File Storage.

[18] Nicola Atzei, Massimo Bartoletti, Tiziana Cimoli. A Survey of Attacks on Ethereum Smart Contracts (SoK). International Conference on Principles of Security and Trust (2017).

[19] Ian Miers, Christina Garman, Matthew Green, and Aviel D. Rubin. Zerocoin: Anonymous distributed e-cash from bitcoin. In Proceedings of the 2013, IEEE Symposium on Security and Privacy, SP '13, pages 397–411, Washington, DC, USA, 2013. IEEE Computer Society.

[20] D. Chaum. Blind signatures for untraceable payments in CRYPTO '82. Plenum Press, 1982, pp. 199–203.

[21] S. Barber, X. Boyen, E. Shi, and E. Uzun, "Bitter to better – how to make bitcoin a better currency," in Financial Cryptography 2012, vol. 7397 of LNCS, 2012, pp. 399–414