Seguro Digital storage of documents using Blockchain

Abhishek Jain[1], Aman Jain[2], Nihal Chauhan[3], Vikrant Singh[4], Narina Thakur[5]

[1,2,3,4,5] *Bharati Vidyapeeth's College of Engineering*

---***---

Abstract - *Disclosure of user's personal documents is a serious breach of privacy. Traditional database storage options have proven to be quite inefficient in protecting such sensitive data. Records can be stolen and tampered with on cloud based services. This paper proposes a decentralized system of storing documents using a novel technology, Blockchain. Blockchain technology has thus far been able to prevent unauthorized access with its secured cryptographic algorithms and its immutability makes the data tamper-proof.*

Keywords - Blockchain, Smart Contracts, File Storage

I. INTRODUCTION

Digital documents are instrumental in the modernization and revamping of our existing system of storing and managing hard copies of our essential records. While the benefits are numerous, digital documents are also open to potential abuses and threats. Large amount of sensitive information held in data centers is vulnerable to loss, leakage, or theft.

Current solution includes creation, storage and management of one's records from a single web portal. This has radically improved the storage, retrieval, sharing of the documents. Users have complete control over their data and can choose to share the same with a wide range of people, including friends, family and various organisation that requires them for their verification process. Many such services require a lot of capital due to the high cost of building and maintaining specialized data centers. This level of infrastructure can only be sponsored by large third party organisations.

This existing system is a matter of concern for several reasons. Cloud storage system has always been prone to data theft and loss while also relying too much on a central authority placing full trust on it for storage and maintenance of data.

In such a scenario of a trustless environment, blockchain emerges as the ideal solution. Before blockchain, banks had the monopoly over maintaining transaction records. The invention of blockchain democratised this process by creating a public decentralised ledger of the same.

Blockchain solves the most pressing issue of trust by eliminating the middle-man and using a decentralised mechanism to store data transactions. It uses cryptographic algorithms to prevent unauthorised access to all records and establishes an access control environment where only the person with the appropriate permissions can access the data. Its immutability ensures that records are tamper-proof and cannot be altered once entered further cementing the trust on the system.

II. RELATED WORKS

Recently rapid developments have been taking place in adapting the Blockchain technology in various greenfield areas. These fields canvas a large spectrum from banking[1], healthcare[2], edge computing[3] to governance[4].

Since its inception with Bitcoin[5], blockchain has been primarily used for distributed storage. The public trust mechanism has enabled the use of blockchains for decreasing privacy concerns over personal data stored on the cloud.

Zyskind et al[6], have proposed a model for using blockchains for transactions other than financial ones. Their system works for personal data and uses Blockchain as an automated access control manager for maintaining privacy.

Sia is another blockchain based system developed for storage by Vorick and Champine[7]. This system works on the basis of contracts which demand that the storage provider regularly prove to the clients that he/she is storing the client's data. It is a bitcoin derivative.

III. PRELIMINARIES

This section will throw light on the various tools that will be used in designing Seguro. The two major components of this system are the Blockchain network and Ethereum Smart Contracts.

A. The BlockChain Network

The highly secure cryptographic algorithms of Blockchains protect against unauthorised access and their immutable structure prevents tampering of the data.

It acts as a trustless proof mechanism for all transactions on the network. The global spread of the network is utilised for storing the ledger. Miner-accountants are responsible for maintaining these nodes. This stands in stark contrast against maintaining trust with the transaction counterparty or a separate third-party intermediary. This is the architecture of a new system of decentralized trustless transactions. All transactions between all parties on the network are disintermediated and decentralized on a global basis. This serves in ensuring the trust of the users.

Blockchain can be thought of as an extra application layer running over the existing Internet protocols. This creates a new tier in the Internet which enables a set of economic transactions. One is these is the immediate digital currency payment (universally usable cryptocurrency). The other is the longer-term, more complicated financial contracts. Blockchain has the capability to make a variety of transactions including financial contracts, hard/soft assets, or currency. It has many alternative uses as well. It can act as a register and inventory system for recording, tracking, monitoring and transacting assets. It is akin to a spreadsheet for registry and accounting of assets for transacting them on a global scale which can include any type of asset held by all parties worldwide. In a nutshell, blockchain can be and is used for asset registry, inventory, and exchange of hard assets (physical property) and intangible assets (votes, ideas, reputation, health data etc).

B. Smart Contracts

A decentralized cryptocurrency, Ethereum[9], uses its built-in currency, Ether, for the execution of smart contracts. This Ether acts as the "fuel" which powers the programmable "smart contracts". Even though the basic technology of Ethereum is similar to that of blockchain, it is much more than that. A "contract" can be thought of as a program which provides services. These services range from voting systems and domain name registries to self-enforcing contracts and agreements, intellectual property, smart property, and distributed autonomous organizations.

In the big picture, Ethereum can be seen as a transaction-based state machine. It begins with a genesis state which evolves with subsequent transactions until it reaches some final state. This final state is accepted as the canonical "version" of Ethereum. This state generally contains some important information. This could be information about account balances, reputation, trust arrangements; any and all information which can be represented by a computer. Transactions thus represent a valid arc between two states; the ‘valid’ part is important—there exist far more invalid state changes than valid state changes.

IV. DESIGN

The design for the system will consist of a multi tier database system with a front end relational database that shall use primary user data consisting his identifier attributes like his Aadhar card number (National Unique Identification Number) along with his access keys saved in an encrypted format.

The second tier database will be implemented on a BlockChain fabric (eg. Ethereum), this data shall have chains of assets (individual data entries) accessible only via a public and private key pair.

On signup, each user shall receive a unique private key. The public key is also generated from private key by applying elliptical function on it. Each such key pair will link him to a unique chain on the BlockChain database.

Each chain shall then be incremented as and when new uploads are made by the user as new assets on the chain pertaining to the user.

Organisations will also be able to register and act as an issuing authority for official documents like land papers educational certificates, PAN card etc which will establish a greater level of trust between the user and any 3rd party with whom the documents needs to be shared. The organisations will be able to push the documents into the user's stack using his public key, which will then be added to the blockchain upon user's approval. Such documents will be marked as 'verified' as against the ones which the user uploads himself.

The user will have the ability to share the documents with any third party as and when required from within the app.

Each asset is not only protected by state of the art BlockChain encryption protocols but is also tamper proof as any such attempt will lead to a chain break.

The two tiers shall be implemented via a backend acting as the BlockChain backbone fully capable of transaction validation and a front end application layer that will have a controlled state access to the secure backend and will act as an interface for the users.

The design implementation using a BlockChain fabric which by its very nature is a distributed database speeds up access times and ensures safety of data via encryption protocols.

The multi tier system lets user access be paired with industry standard authentication and authorization systems on the frontend along with the use of encryption and hashing.

The proposed design pairs the efficiency, dependability and security of a BlockChain database with the user friendly node stack web application much like digital wallets to monetary BlockChain networks like BitCoin and Ethereum.

V. DISCUSSION

A comparison between Blockchain and Legacy Databases (See Appendix) reveals a lot of interesting information. The most important point of difference is the absence of a central authority which gives birth to the other differences. Ownership of the data is established via cryptographic key pairs in blockchain and by the central authority in the case of legacy databases.

In blockchain's case, data validity is ensured for the entire chain while the legacy databases only provide it for single instances in time. The consensus feature of blockchains maintains identical copies with all users while legacy databases require complex checking between central database and user database for ensuring agreement. All permissioned users inherently have identical access with blockchain.

Peer to peer networking for distributed data replication doesn't allow for a single point of failure. Since new blocks are dependent on the previous blocks, blockchains are protected against fraudulent changes while legacy databases suffer due to the inefficiency of keys and constraints.

VI. CONCLUSION

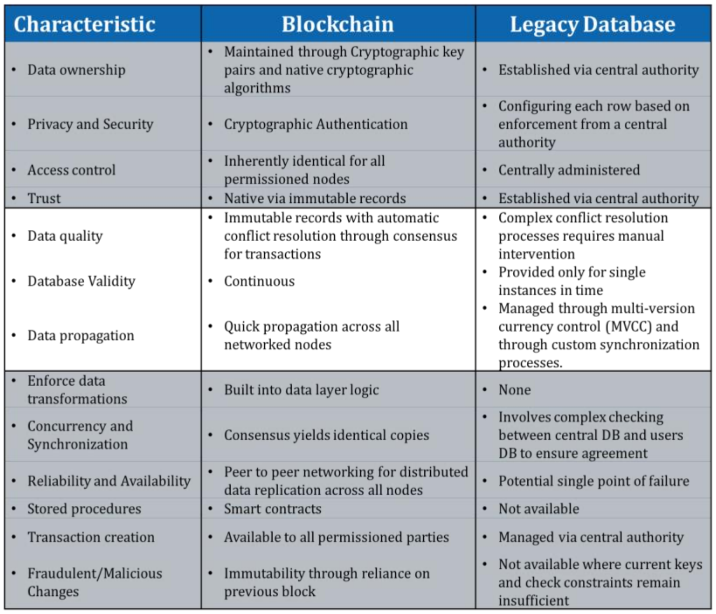
Blockchain has paved way for the next big transformation in the internet era. The benefits of blockchain becomes more apparent as we start taking privacy seriously which is the case with the digital documents.

Using Blockchain for storage of digital records solves the problem at hand, i.e. data leakage. It provides a secure space for storing all the records. Since the data is decentralised, the smooth functioning of the system is not dependent on any particular cloud service provider. It protects the sensitive data from reaching any third person without the owner's permission. Since the records are immutable, the records once created cannot be altered even by the owner or the issuing authority itself for any kind of personal gains.

Our proposed system, Seguro, utilises the security capabilities of Blockchain technology to ensure the person's privacy with respect to his/her documents while also implementing an efficient, easy-to-use sharing mechanism that facilitates document verification at any third party during the registration or any other process. By implementing the proposed model, users will be able to store all their documents on an online medium without any risk of data leakage.

VII. APPENDIX

Table 1 Comparison between Blockchain and Legacy Database[15]



Characteristic	Blockchain	Legacy Database
• Data ownership	• Maintained through Cryptographic key pairs and native cryptographic algorithms	• Established via central authority
• Privacy and Security	• Cryptographic Authentication	• Configuring each row based on enforcement from a central authority
• Access control	• Inherently identical for all permissioned nodes	• Centrally administered
• Trust	• Native via immutable records	• Established via central authority
• Data quality	• Immutable records with automatic conflict resolution through consensus for transactions	• Complex conflict resolution processes requires manual intervention
• Database Validity	• Continuous	• Provided only for single instances in time
• Data propagation	• Quick propagation across all networked nodes	• Managed through multi-version currency control (MVCC) and through custom synchronization processes.
• Enforce data transformations	• Built into data layer logic	• None
• Concurrency and Synchronization	• Consensus yields identical copies	• Involves complex checking between central DB and users DB to ensure agreement
• Reliability and Availability	• Peer to peer networking for distributed data replication across all nodes	• Potential single point of failure
• Stored procedures	• Smart contracts	• Not available
• Transaction creation	• Available to all permissioned parties	• Managed via central authority
• Fraudulent/Malicious Changes	• Immutability through reliance on previous block	• Not available where current keys and check constraints remain insufficient

VIII. REFERENCES

[1] Peters, Gareth W., and Efstathios Panayi. "Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money." Banking Beyond Banks and Money. Springer, Cham, 2016. 239-278.

[2] "Ekblaw, Ariel, et al. "A Case Study for Blockchain in Healthcare:"MedRec" prototype for electronic health records and medical research data." Proceedings of IEEE Open & Big Data Conference. 2016."

[3] Samaniego, Mayra, and Ralph Deters. "Using blockchain to push software-defined IoT components onto edge hosts." Proceedings of the International Conference on Big Data and Advanced Wireless Technologies. ACM, 2016.

[4] Atzori, Marcella. "Blockchain technology and decentralized governance: Is the state still necessary?." (2015).

[5] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." (2008).

[6] Zyskind, Guy, and Oz Nathan. "Decentralizing privacy: Using blockchain to protect personal data." Security and Privacy Workshops (SPW), 2015 IEEE. IEEE, 2015.

[7] Vorick, David, and Luke Champine. "Sia: simple decentralized storage." (2014).

[8] Wilkinson, Shawn, Jim Lowry, and Tome Boshevski. "Metadisk a blockchain-based decentralized file storage application." Technical Report. Technical Report (2014).

[9] “Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum Project Yellow Paper 151 (2014).”

[10] “Wright, Aaron, and Primavera De Filippi. "Decentralized blockchain technology and the rise of lex cryptographia." (2015).”

[11] Ali, Muneeb, et al. "Blockstack: A Global Naming and Storage System Secured by Blockchains." USENIX Annual Technical Conference. 2016.

[12] Crosby, Michael, et al. "Blockchain technology: Beyond bitcoin." Applied Innovation 2 (2016): 6-10.

[13] “Delmolino, Kevin, et al. "A programmer’s guide to ethereum and serpent." URL: https://mc2-umd. github. io/ethereumlab/docs/serpent_tutorial.pdf.(2015).(Accessed May 06, 2016) (2015).”

[14] “Swan, Melanie. Blockchain: Blueprint for a new economy. " O'Reilly Media, Inc.", 2015.”