DOCUMENT SECURITY AND STORAGE ON BLOCKCHAIN

[1] ASHISH SHARMA [2] SIMRANJEET SINGH RANDHAWA [3] ADITYA KUMAR [4] KAPIL TYAGI
[1]Assistant Professor (CSE) MAIT, Delhi
[2]Student (CSE) MAIT, Delhi
[3] Student (CSE) MAIT, Delhi
[4] Student (CSE) MAIT, Delhi

Abstract : The blockchain is a distributed network that records digital transactions on a publicly accessible ledger. This paper explores whether blockchain technology provides a suitable platform for the preservation of digital documents. This paper suggests that the hash functions provide a better technique for authentication and storage of documents. Compared to digital certificates, hashing provides better privacy and security. It does not involve the third party for authenticating and the problem of single point storage is eliminated due to distributed nature of blockchain network.

Keyword - Blockchain, Ethereum, Smart contract, Hashing

I. INTRODUCTION

The blockchain has been with us since 2009. In the seven years of its life, it has been able to successfully thwart attempts to hack into it, manipulate it or co-opt it. While the technology is at a crossroads in its development, there are many use cases for its adoption across the industry, including the field of records management. The notarization of electronic records presents novel challenges for records managers. In a paper records condition, the maker of a record states responsibility for archive, or consents to an assertion explained in the report, by marking or countersigning it. From the records director's point of view, the report is the property of the gathering that marked it. The mark is synonymous with the report.

Thus using the concept of decentralization, proof of work and time stamping features of blockchain the authorization and accountability which in addition with hashing provides a way to preserve the content of documents, make author accountable and store the time of document preservation.

The project based on the above concept reads the content of the document which is uploaded instead of scanning the document as an image. Content once read is then hashed and that information is uploaded on blockchain network instead of the original content which makes the content safe. This process thus ensures the safety of content even if it is being uploaded on a public distributed ledger.

II. PROPOSED WORK

In this project, we present a way to store and authenticate the content and authorship of documents respectively. The process involves uploading the document which needs to be secured. Once uploaded the document's content is hashed using the concept of hashing. Similarly, the email of the author is also hashed. Both these hashes are joined together and hashed again. This hashing is then stored on blockchain along with the time stamp. Once the blockchain approves the upload then it becomes the part of blockchain forever.

All the steps described above are implemented using the Ethereum framework. Since ethereum framework is used for the project the concept of Smart Contract is also implemented. In this project thus all the middleman work is done by smart contracts. All the transactions are implemented by smart contract. This reduces the human interaction between any transactions and thus also reduces significantly the chance of any malicious interference in the process.

Therefore if implemented in a correct way the proposed project will be able to provide a way to secure a document and make its author accountable and help in avoidance and detection of corporate fraud and espionage.

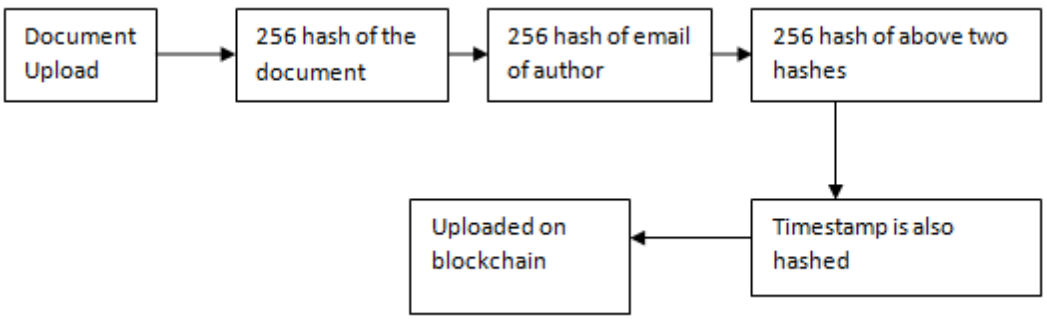

Figure 1: Flow Chart diagram of working of project

III. CONCEPTS USED

A. HASHING

Hashing is a form of document authentication in which documents are not signed directly. Instead, a hash function is used which calculates a hash value that confirms that the authentication of the document has taken place. There are two components to hashing:

- The hash function is a hexadecimal algorithm, such as SHA-256, that converts an input value into another compressed value. The input to the hash function is of arbitrary length but output is always of fixed length. In digital preservation, hash functions confirm that no changes have been made to a digital document.

- The hash value is the output of a specific length that permanently identifies the input data (Pedro, 2015, p.95).

The hash work is a restricted procedure. This implies the client can make the hash from input information, yet not utilize the hash to uncover the information. Should a records chief modify even one piece from the info information and afterwards attempt to apply a similar hash work, the administrator will create totally different hash esteem.

B. TIMESTAMPING

A timestamp demonstrates that a specific dataset existed at one point in time (Pedro, 2015, p. 99). The blockchain strategy makes time stamped hinders through distributed innovation, subsequently, disintermediating Time-Stamping Authorities (TSAs). Diggers on the Bitcoin blockchain timestamp each square which contains ten minutes of exchanges. The diggers are, viably, working as a dispersed TSA. This implies there is no requirement for intermittent re-timestamping of marks because of lapsing keys. In limited time materials for its new BLT cryptographic calculation, the product security organization Guard time expressed the time and trustworthiness of the mark can be demonstrated numerically, without dependence on the security of keys or of confided in parties (Guard time, 2016). Amanti (2016) expressed that the time it takes for a TSA to check a move is estimated in seconds, though the blockchain's confirmation takes minutes (para. 23). He additionally noted two different focal points of blockchain timestamping over TSA timestamping:

- Long-term preservation can be achieved without the maintenance costs that come with a TSA-issued certificate (Amati, 2016, para. 24). 10 ‖ See Also: Vol. 3 (Spring 2017)
- Archivists can exploit the convenience of verifying the signature with the document and public key without having to safeguard the digital signature on a central server (Amati,2016)

C. SMART CONTRACT

A smart contract is a PC convention planned to carefully encourage, check, or uphold the arrangement or execution of an agreement. Smart contracts permit the execution of tenable exchanges without outsiders. These exchanges are traceable and irreversible. Smart contracts were first proposed by Nick Szabo, who instituted the term, in 1994.

Defenders of smart contracts assert that numerous sorts of legally binding provisions might be made incompletely or completely self-executing, self-upholding, or both. The point of smart contracts is to give security that is better than customary contract law and to diminish other exchange costs related to contracting. Different cryptographic forms of money have actualized sorts of smart contracts.

Ethereum enables engineers to program their own particular shrewd contracts, or 'self-sufficient specialists', as the ethereum white paper calls them. The dialect is 'Turing-finished', which means it bolsters a more extensive arrangement of computational directions and along these lines in this task custom savvy contracts are made.

Smart contracts can:

- Function as 'multi-signature' accounts, so finances are spent just when a required level of individuals concur.
- Manage understandings between clients, say, on the off chance that one purchases protection from the other.
- Provide utility to different contracts (like how a product library works).
- Store data around an application, for example, area enrolment data or participation records.

IV. Flow of Transaction in a Blockchain Network

The diagram significantly explains the flow of transaction in a blockchain network.

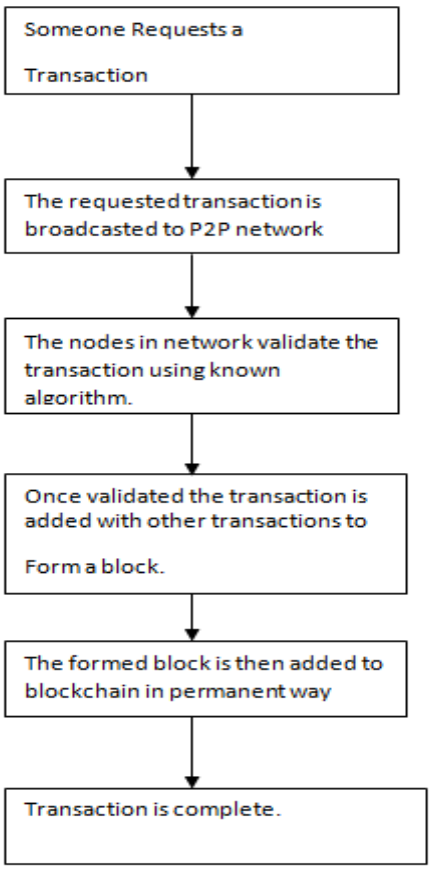

Figure 2: Block diagram of transaction on a blockchain

V. HOW IT PROVIDES SOLUTION

The project provides security and accountability to documents by using the concepts discussed above. However, the complete process working and the solutions provided by this project are explained as below. The security of data or the content of documents is secured by using the concept of hashing. Instead of uploading the complete document on blockchain only its hash is uploaded thus ensuring the safety of content. The author's accountability is maintained with the help of hashing the content of documents hash and authors email which in case of inspection would result in different hash if only one letter is changed or deleted. The timestamp of documents also helps in maintaining the time of uploading of document which ensures that no document can be copied and uploaded causing conflicts as the record also contains the timestamp thus clearly catching which document was original and which was copied. The blockchain records remain on blockchain network forever even if the user who actually uploaded the document has left the network. Thus once uploaded and approved the record of the document remains on the blockchain forever and thus it can be scrutinised whenever required making the document accountable and secure forever.

All the above solutions are provided by the blockchain technology and the features which are provided by the blockchain technology. The concept of decentralization and record being stored forever on blockchain provides the project an edge over the classic server-client model and digital signatures.

VI. CONCLUSION AND FUTURE WORK

This project proposes the idea of using blockchain technology and smart contracts to safely confirm and store the content and creation of documents along with time they were published so as to avoid and detect fraud and reduce the chances of government or corporate espionage.

The future work can be extended to using image processing to scan the documents instead of uploading them for their content and using features of image processing to completely remove the involvement of manual notary.

REFERENCES

[1] T. G. Peter Mell, "The NIST Definition of Cloud," *Recommendations of the National Institute*, September 2011. Allen, C. (2015, October 9). Schnorr signatures: An overview. *WebOfTrustInfo*. Retrieved from https://github.com/WebOfTrustInfo/rebooting-the-web-of-trust/blob/master/topics-and-advance-readings/Schnorr-Signatures--An-Overview.md.

[2] Allen, C., Brock, A., Buterin, V., Callas, J., Dorje, D., Lundkvist, C., Kravchenko, P., Nelson, J., Reed, D., Sabadello, M., Slepak, G., Thorp, N. & Wood, H. T. (2015). *Decentralized public key infrastructure. A White Paper from Rebooting the Web of Trust*. Retrieved from https://github.com/WebOfTrustInfo/rebooting-the-web-of-trust/blob/master/final-documents/dpki.pdf

[3] Amati, F. (2016, January). Using the blockchain as a digital signature scheme. Medium blog. Retrieved from https://blog.signatura.co/using-the-blockchain-as-a-digital-signature-scheme-f584278ae826

[4] Anon. (2015, October 31). The trust machine: The promise of the blockchain. *The Economist*. Retrieved from http://www.economist.com/news/leaders/21677198-technology-behind-bitcoin-could-transform-how-economy-works-trust-machine

[5] ANSI X9.95-2012. Trusted time stamp management and security. Retrieved from https://www.sec.gov/rules/proposed/s72703/iac120105.pdf.
Bentov, I., Lee, C., Rosenfeld, M. & Mizrahi, A. (2014).

[6] Proof of activity: Extending Bitcoin's proof-of-work via proof-of-stake. [Extended Abstract] *Performance Evaluation Review*, 42 93), 34-37.

[7] Blockchain.info. Retrieved from https://blockchain.info.

[8] BlockNotary. Retrieved from https://www.blocknotary.com/

[9] D. U. G. Radhika D.Bajaj, "Dr. U.M. Gokhale," *International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031* , vol. Volume 02 , no. Issue 05 , May 2016 .

[10] Cumming, K. & Findlay, C. (2016). Report on blockchain: Applications and implications. *Recordkeeping Roundtable*. Retrieved from https://rkroundtable.org/2016/04/03/report-on-blockchain-applications-and-implications/.

[11] Ethereum Project (2017). Retrieved from https://ethereum.org

[12] Findlay, C. (2015). Decentralised and inviolate: the blockchain and digital archives. Retrieved from https://rkroundtable.org/2015/01/23/decentralised-and-inviolate-the-blockchain-and-its-uses-for-digital-archives/.

[13] Van Garderen, P. (2016, May 17). Blockchain and digital preservation – Part2. Presentation at Simon Fraser University [Video file]. Retrieved from https://www.youtube.com/watch?v=S2N0m9YDgZw.

[14] Van Garderen, P. (2016, May 17). Blockchain and digital preservation – Part3. Presentation at Simon Fraser University [Video file]. Retrieved from https://www.youtube.com/watch?v=onx3f6xmEsI&t=276s.

[15] Lemieux, V. (2016). Blockchain technology for recordkeeping: Help or hype? Retrieved from http://www.blockchainubc.ca/main/dissemination.