2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) | 978-1-6654-8962-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAISS55157.2022.10011076

Document Management System using Blockchain

Shanmugaraja.P
Associate Professor,
Information Technology,
Sona College of Technology ,
Salem, Tamil Nadu, India.
shanmugaraja@sonatech.ac.in

Susmitha.K.S
Student,
Information Technology,
Sona College of Technology ,
Salem, Tamil Nadu, India.
susmitha.18it@sonatech.ac.in

Swadha.S
Student,
Information Technology,
Sona College of Technology ,
Salem, Tamil Nadu, India.
swadha.18it@sonatech.ac.in

Vijay.R
Student,
Information Technology,
Sona College of Technology ,
Salem, Tamil Nadu, India.
vijay.18it@sonatech.ac.in

Naveen.G
Student,
Information Technology,
Sona College of Technology ,
Salem, Tamil Nadu, India.
naveen.18it@sonatech.ac.in

***Abstract*— Creating fake documents and certificates by duplicating them from the real ones has become most popular in the recent days. Moreover, documents are stored in the respective offices in a centralized database, and to access them one should always be physically present in that office. Also, during the pandemic, most of the schools, colleges, and offices were subjected to online document storage. Therefore, this paper proposes a blockchain system as a solution for the taut storage of documents in the form of electronic files on a blockchain network. The immutable property will aid the document from being duplicated, and will result in the permanent entry of the document in the distributed databases. Also, the blockchain belongings of the decentralized ledger structure will help to keep the files safe and unaffected forever.**

***Keywords*— Blockchain, Centralized database, Secure storage.**

I. INTRODUCTION

Documents have become a paramount proof in the present way of living. Also, along with this rising needy for documents, the pilfering and duplicating of documents also increases day by day. So, there is a very great need for the shielding and authenticating of files to keep them from being demolished, operated, or duplicated into some other file. Also, some documents are so chief in a way that if they are lost, they cannot be rehabilitated or repaired [1]. Apart from all this, the person has to be in-person to visit the appropriate office for restoring their document and this process takes much more time. Because of the dominance of document storage, the pay-off in the society also has escalated with a great increase in duplication and creation of any educational or property-related files by giving a certain amount of money [10]. These problems can be overcome by using blockchain as a service for storing files. This will be a mechanism for easy recovery of documents, complete validation of data, and easy elevation of data as well. All these systems will help greatly in reducing trickery, bribes, and pilfering of documents.

II. BACKGROUND

Document management currently, in this epoch, is done by storing it in a centralized database (DB) system. In this centralized database system, if a single change is made to the database, then every other person/office connecting to that database for getting information will get erroneous results. Also, there is expunging and overhauling of data in the centralized DB system [19], thus making the changed information irretrievable. Thus, the duplicates of documents and thefts can be easily done in the current epoch [11]. These problems are dealt with in the proposed model by using a decentralized ledger system, instead of centralized databases.

The blockchain is a data structure that separates the data into containers - the blocks. The blocks are quite comparable to nodes in a linked list. Each block links to the previous block using a hash. A blockchain is made up of blocks that create a continuous chain. The end of the blockchain is always where fresh blocks are added [3][4]. Transactions, among other things, are the most important aspects of the blockchain, hyper ledger and smart contract etc.

Each component is explained below:

1. *Transaction:*

The transaction is recorded in blockchain as an asset or like an ownership and it is verified and accessed all over the nodes. These transactions are processed using timestamp as in Fig. 1.

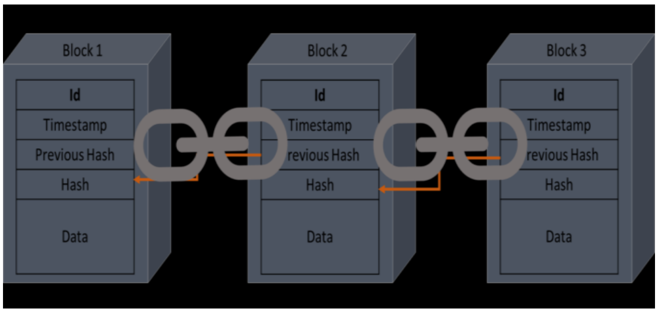


Fig.1: Timestamp

2. *Peer network:*

Blockchain is a ledger which runs on a p2p network and is managed by a central server over the internet. As a result, they aren't vulnerable to a single point of failure or malware [15].

Authorized licensed use limited to: PES University Bengaluru. Downloaded on January 07,2025 at 05:01:28 UTC from IEEE Xplore. Restrictions apply.

For the system to last longer, every member should share the storage and computational energy.

3. *Ledger system:*

Ledgers are independent nodes to record, store and synchronize transactions in their respective e-ledgers.

4. *Smart contract:*

Smart contract is an application program saved on a blockchain that runs on a certain user defined conditions and are written on solidity.

III. EXISTING SYSTEMS

The authors of [2] presented a method that used a limited shared data technique for dividing medical files. Data packets should be restricted as separation data inside the doctor-patient context, as they could provide an unsecured site for theft to gain access to private medical reports for which the patient has opted not to divulge specific information. If there are more partners inside the network, the writers agree that recognizing a patient using a variety of techniques is likewise not a viable alternative for provisioning confidential material. The key information of a patient can be determined by combining ages, religion, region, day of enrollment, and medical issues [5] utilizing best fit consensus techniques. The author of [14] suggested a "ceremonious major data exchange consensus" that make data shareable among disparate organizations.

IV. PROPOSED SYSTEM

In the proposed system shown in Fig. 2, Interplanetary File System (IPFS) is used for storing the documents in a secure manner which is based on content addressing. In the case of storing files here, no copy of the file was created; instead, only an encrypted hash is provided to the recipient through which he can access the file.

IPFS being content addressed helps in lowering the bandwidth expending and quickening the file ingress. User can upload their personal files, and they are only accessible to their certificates, important documents, etc. Using decentralized system to store files, provides the user with high reliability as the files are at risk from theft where anyone can post any type of document (pdf, image) to share with other users [16]. Users can also plea a specific document by circulating a request.

An API has been created in which the user can upload the documents using metamask gas price [12]. When the user uploads a document, transaction hash will be generated for the specific files and when it is shared, the transaction hash of the sender and the user will be stored in the decentralized database.

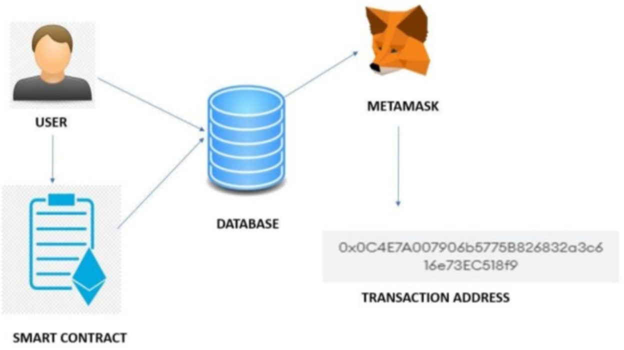


Fig.2: System Model

V. METHODOLOGY

File upload

A file in .pdf or .jpeg format can be uploaded by the end user. Regardless, such files include unstructured data, in which they all follow the same pattern, which has been authorized by each person's internal regulatory authorities.

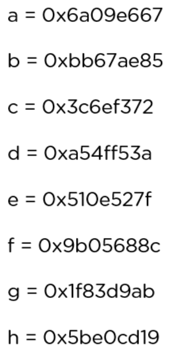
Smart contract

The uploaded file is then given to the smart contracts where the execution of the document is automated for the immediate outcome to reduce the time complexity [18]. This is then fed into data storage which contains plenty of versatile data which will generate in the form of hash.

Database

The database contains the document and the hash value of the respective file, and in that, BigchainDB is used which offers decentralization [7], immutable assets. The data stored in the database is transformed into a 256 long bit hash function using the SHA-256 hash algorithm.

A hashing with such a range of 256 bits is known as SHA-256 (Secure Hash Algorithm). It's a hash function for the power steering. A data is divided as 512 = 16 32-bit blocks, with each block requiring 64 successions. It will use this to set default values for the eight buffers.



a = 0x6a09e667

b = 0xbb67ae85

c = 0x3c6ef372

d = 0xa54ff53a

e = 0x510e527f

f = 0x9b05688c

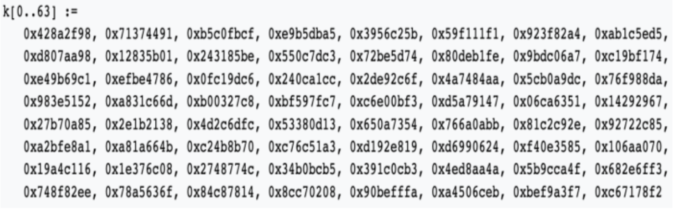
g = 0x1f83d9ab

h = 0x5be0cd19

Fig.3: Buffers

It then stores the key values in the form of array, from k[0] to k[63] as in Fig.4.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on January 07,2025 at 05:01:28 UTC from IEEE Xplore. Restrictions apply.



```
k[0..63] :=
   0x428a2f98, 0x71374491, 0xb5c0fbcf, 0xe9b5dba5, 0x3956c25b, 0x59f111f1, 0x923f82a4, 0xab1c5ed5,
   0xd807aa98, 0x12835b01, 0x243185be, 0x550c7dc3, 0x72be5d74, 0x80deb1fe, 0x9bdc06a7, 0xc19bf174,
   0xe49b69c1, 0xefbe4786, 0x0fc19dc6, 0x240ca1cc, 0x2de92c6f, 0x4a7484aa, 0x5cb0a9dc, 0x76f988da,
   0x983e5152, 0xa831c66d, 0xb00327c8, 0xbf597fc7, 0xc6e00bf3, 0xd5a79147, 0x06ca6351, 0x14292967,
   0x27b70a85, 0x2e1b2138, 0x4d2c6dfc, 0x53380d13, 0x650a7354, 0x766a0abb, 0x81c2c92e, 0x92722c85,
   0xa2bfe8a1, 0xa81a664b, 0xc24b8b70, 0xc76c51a3, 0xd192e819, 0xd6990624, 0xf40e3585, 0x106aa070,
   0x19a4c116, 0x1e376c08, 0x2748774c, 0x34b0bcb5, 0x391c0cb3, 0x4ed8aa4a, 0x5b9cca4f, 0x682e6ff3,
   0x748f82ee, 0x78a5636f, 0x84c87814, 0x8cc70208, 0x90befffa, 0xa4506ceb, 0xbef9a3f7, 0xc67178f2
```

Fig.4: Array of key values

The whole hash is divided up into 512-bit pieces. It performs 64 rounds of operation on every block, with the result of each block plastering as the input for another block [17]. For the transaction, the resulting 256-bit hash function is passed into metamask.

```
While i do
s=expand(block)
    k=HM0
    l=HM1
    m=HM2
    n=HM3
    0=HM4
    p=HM5
    q=HM6
    r=HM7
```

Calculation of the hash

First, there are eight variables, each with its own prime value and the first 32 bits of the fragmented result of the quadratic formula of the first eight numbers are,

$C1(0) = 0x6a045857$

$C2(0) = 0xbb89ae90$

$C3(0) = 0x3c9ef346$

$C4(0)= 0xa54ff53a$

$C5(0) = 0x510e647f$

$C6(0) = 0x9b05489c$

$C7(0) = 0x1f83d89ab$

$C8(0) = 0x0be0cd467$

Blocks are then executed one at a time:

The 64 blocks are fabricated Wi from M(t) for l = 1 to n, as indicated above.

$(k0,k1,k2,k3,k4,k5,k6,k7) = (H (l-1) 1 , H(l-1) 2, H(l-1) 3, H(l-1) 4 , H(l-1) 5 , H(l-1) 6 , H(l-1) 7, H(l-1) 8)$

```
    $T1 = h + \Sigma1(k4) + Ch(k4, k5, k6) + Kt + Wt$
    $T2 = \Sigma0(k0) + M aj(k0, k1, k2)$
        $k8 = k7$
        $k7 = k6$
        $k5 = k4$
        $k4 = k3 + T1$
        $k3 = k2$
        $k2 = k1$
        $k1 = k0$
        $k0 = T1 + T2$
    End while
        HM0=K+HM0
        HM1=L+HM1
        HM2=M+HM2
        HM3=N+HM3
        HM4=O+HM4
        HM5=P+HM5
        HM6=Q+HM6
        HM7=R+HM7
    End WHILE
```

The value of Hj(t) is calculated.

```
        $C1(t) = C1(l-1) + k$
        $C2(t) = C2(l-1) + l$
        $C3(t) = C3(l-1) + m$
        $C4(t) = C4(l-1)+ n$
        $C5(t) = C5(l-1)+ 0$
        $C6(t) = C6(l-1)+ p$
        $C7(t) = C7(l-1)+ q$
        $C8(t) = C8(l-1)+ r$
    End while
```

After the last block has been processed, the hash of the message is the addition of the variables CN i.

$C=C1(N) \|C2(N) \|C3(N) \|C4(N) \|C5(N) \|C6\|C7(N) \|C8(N)$

VI. RESULTS

A situation has been created where the documents can be stored securely and the execution time for the existing system (Traditional algorithm) and the proposed system (Improved algorithm) varies, as SHA256 algorithm is used. Therefore, the scalability of the system increases by reducing the time taken which is shown in Fig.5.

Authorized licensed use limited to: PES University Bengaluru. Downloaded on January 07,2025 at 05:01:28 UTC from IEEE Xplore. Restrictions apply.

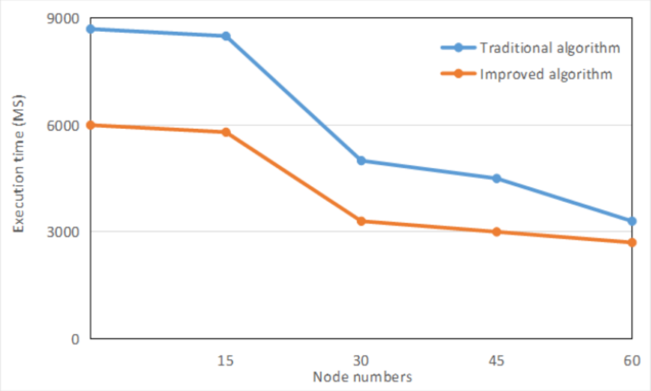


Fig.5: Comparison of time taken by Traditional and Improved algorithms

VII. CONCLUSION

This study has proposed a blockchain-based document storage system that stores the documents, images, etc., in a form of hash which is then processed by the SHA-256 algorithm. The proposed system does not have any analyzing mechanism to analyze what is written in the document. This information can be currently elucidated by a person. But in the future, the nodes can be made smart through Natural Language Processing and the ability to understand one document from the forged one can be provided, thus raising a security awareness whenever it finds any theft in the currently received document.

REFERENCES

[1] F. Benhamouda, S. Halevi and T. Halevi ,"Supporting the private data on Hyperledger with secure multiparty computation," in IBM Journal of Research and Development,(March-May 2019).

[2] Academy of Medical Sciences. Clinical trials data sharing: science, privacy and ethics,(14 July 2014) http://www.acmedsci.ac.uk/viewFile/535a3c3962a46.pdf.

[3] D. T. T. Anh, M. Zhang, B. C. Ooi, and G. Chen, “Untangling blockchain” IEEE Transactions on Knowledge and Data Engineering,(1–1, 2018)

[4] S. lnes, J. Ubacht, and M. Janssen, “Blockchain in government: Benefits and implications of distributed ledger technology for information sharing,” Available: http://www.sciencedirect.com/science/article/pii/S0740624X17303155

[5] G. Tripathi, "A blockchain-based approach for the smart healthcare using health care management system",(2020).

[6] Official Ethereum docs- https://ethereum.org/en/

[7] N. Z. Aitzhan and D. Svetinovic, “Security and privacy in decentralized system based energy trading through blockchain,” IEEE Transactions on Dependable and Secure Computing,(1–1, 2016)

[8] DOCSTRACK -DOCUMENT MANAGEMENT SYSTEM Volume: 05 Issue: 01 | Jan2018

[9] Official Hyperledger Docs - https://hyperledgerfabric.readthedocs.io

[10] B.Saravanan, V.Mohanraj, Dr.J.Senthilkumar “A fuzzy entropy technique for dimensionality reduction in recommender systems using deep learning” Soft Computing –Springer Vol. 23, No. 8, pp.2575-2583, April 2019

[11] G.Mohanraj,V Mohanraj, J Senthilkumar, Y Suresh,” A hybrid deep learning model for predicting and targeting the less immunized area to improve childrens vaccination rate ”, Intelligent Data Analysis 24(6):1385-1402,2020.

[12] Thiyaneswaran, B., Anguraj, K., Kumarganesh, S., Thangaraj, K.” Early detection of melanoma images using gray level co-occurrence matrix features and machine learning techniques for effective clinical diagnosis”, International Journal of Imaging Systems and Technology, 2021, 31(2), pp. 682–694.

[13] Jeba Emilyn Jeyaswamidoss , Kesavan Thangaraj, Kadarkarai Ramar, Muthusamy Chitra ,” A rough set based rational clustering framework for determining correlated genes:, ACTA MICROBIOLOGICA ET IMMUNOLOGICA HUNGARICA, Vol. 63,No.02,Pp:185-201,2016.

[14] Y. Jiang, X. Cheng, J. Zhu, Y. Xu, A consensus mechanism based on multi-round concession negotiation, Comput. Stand. Interfaces 74, (2021).

[15] R. Kumar, R. Tripathi, Blockchain-based framework for data storage in peer-to-peer scheme using interplanetary file system, 2020.

[16] X. Li, P. Jiang, T. Chen, X. Luo, Q. Wen, A survey on the security of blockchain systems, Future Gener. Comput. Syst. 107 ,(2020) ,841–853.

[17] H. Lin, M. Wang, Repudiable ring signature: stronger security and logarithmic-size, IACR Cryptol. ePrint Arch. 2019, (2019), 1269.

[18] Kim, Soohyeong, Kwon, Yongseok, Cho, Sunghyun, 2018. A Survey of scalability solutions on blockchain. In: 9th Int. Conf. Inf. Commun. Technol. Converg. ICT Converge. Powered by Smart Intell. ICTC 2018, pp,1204–1207.

[19] K. Chokkanathan, P. Shanmugaraja, , Shivashankar Ramasamy, Rujira Ouncharoen, Nopasit Chakpitak, “A Survey on Role of Block Chain in Smart Cities”, International Journal of Computer Science and Network Security, VOL.21 No.7, July 2021. Pages. 1-7

Authorized licensed use limited to: PES University Bengaluru. Downloaded on January 07,2025 at 05:01:28 UTC from IEEE Xplore. Restrictions apply.