Secure Document Vault using Blockchain and IPFS (InterPlanetary File System)

Suhasini N1,Nikhil C B2, Shamshtabrej M Y3, Karthik P R4, Venkatesh S

suhasini.n@reva.edu.in, nikku10227@gmail.com, tabrezy8228@gmail.com, rashinkarkarthik@gmail.com, venkateshstalian@gmail.com

Abstract.

In the past few years the Blockchain technology has gained tremendous amount of trust for its security because of its decentralized in nature. Data such as government or non-government document plays very important role in day-to-day life. The interference in the data change is irrelevant which causes the data to lose and data stolen which led to bigger problems. To overcome it, this project uses encryption methods to secure the data. The data could be any document e.g. Driving license, lease documents etc. The data store on the website cannot be accessed by any other party apart from the user itself. The technologies which are used are blockchain and IPFS. The IPFS(InterPlanetary File System) returns the hash (SHA-256) of the file which is present on the network in which this model is going to save or share the data in distributed manner. Which is Distributed file system. By using the hash code user can access the secure file. After getting the hash code from the IPFS this algorithm, will be going to encrypt the hash and store it into the blockchain which acts as the secure database.

Keywords— encryption, security, cryptography, algorithm, IPFS, blockchain

1. INTRODUCTION

This article will tell us about the project which is based on blockchain and IPFS. Which is helps the user to store the document in the IPFS and then store its hash code in the blockchain. The drawback is that by using the hash anyone can access the file which is not much secure to overcome it the model will further encrypt the hash code then store it in the blockchain. Islet leads to the which type of encryption algorithm should be used and how much time it will be consumed to encrypt and then store is it flexible? All the questions will be answers below. First of all, the blockchain technology used is the Ethereum based blockchain. Which is the digital money on the private or a public network.

The Blockchain technology, IPFS, and the distributed ledger are now getting massive attention in the IT Industries which triggering several projects in different fields, but the financial industry is one of the most big industry that is considered the main use of the blockchain concept. The interplanetary File System(IPFS) is protocol that provides a peer-to-peer network for storing and fetching the data as in the form of distributed file system. The files and the blocks it contains are given unique fingerprint known as a Cryptographic hash, IPFS will not provide duplicate hash for the same network. A cryptographic hash function is an algorithm that takes an arbitrary amount of input data from an identifier and produces an output of a fixed size of cipher text. It stored the cipher text in the blockchain.

Blockchain technology is a distrusted network that provides high security and stores the information in the form of digital ledger technology, blockchain technology provides secure transfer without an intermediary. Blockchain technology provides a permanent, record of each transaction, Blockchain is fully transparent and shows real-time transactions. Distrusted Ledger Technology (DLT) that stores entries in blocks of the transaction, grouped and hashed into the cryptographic hash algorithm. Each block relates to all the blocks before and after in thought a distinctive hash pointer which increases more security to the blocks, if anything changed in a block then the hash value will also change.

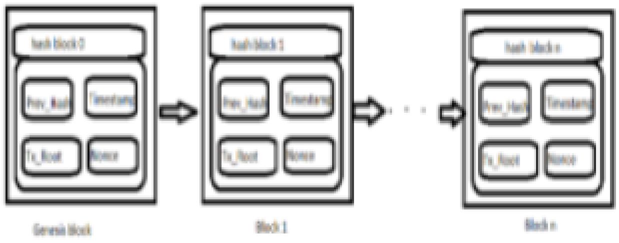

Fig. 1. Blocks in the blockchain.

In Fig. 1 show like how the block is interconnected with each other by using the previous hash which hold the previous block address in the network. Every node consists of timestamp, [9] a nonce, the hash value the nonce is the method is a random number to verify the hash, this will ensure the structural integrity of the block chain up to the first block which is called as “genesis block”. In the hash part which is unique from each other which is lead to less frauds since if the change of a block in the chain can cause the hash value to change. If the majority of nodes in the network agree through a consensus mechanism on the validity of each transaction in the block and on the validity of each block, the block is created and add to the chains. That way once the information add to the block chain cannot be deleted or modified. The best example is Bitcoin, in Bitcoin the block is created by those who got reward because of miners of the Bitcoins, miners are the validators of the block. Now, people all over world can trust each other and transfer different types of assets to each other through the internet.

Our main goal is that the user may able to trust a platform where he can store and manage his private documents in that platform form. The data which is store in the blockchain should not be visible as directly to the end users as encrypted for in encryption RSA algorithm might be the good solution to the problem. So, for In this project the algorithm is to encrypt the data and then store in the blockchain.

Our project name “Secure Document Vault” is the application where the end user can store his private documents which is very secure and unerasable our application is based on blockchain and IPFS once the end user stores his document in our application, he/she will not able to modified or changed. IPFS have the option to edit the uploaded data only to authenticated end users that create a various resiliency network enable continuous availability with or without backbone internet connection. It means a better connection to develop the world this application will help in during natural disasters, or just when some uncertainty some thig happened with the wi-fi or systems. Today’s most of the modern website are inefficient and expensive. The IPFS work on the Peer-to-Peer network which saves big on bandwidth making it possible to efficiently distribute high volumes of data without duplication.

2. LITERATURE REVIEW

The blockchain technology can used in the banking system. To prevent the fraud transfer of the money or assets. In this paper author talk about the [1] The data stores in chronological manner for document in blockchain technology. for that they are using the IPFS for storing the original soft copy of the data documents author says block chain-based structures is less eco-friendly in compare to general data base approach. So, they go with the IPFS to store massive amount of statistical data as in the form of remotely. Since the block chain is every lasting data which cannot be modified. In this each of the block is unique fingerprint known as a cryptographic hash. Which is also used for verification and validation.

According to the author [2] the block chain is the revolutionary that has been under laying for many years because of the issue with scalability the decentralized has been down by the centralized which is run in current system. The distributed storage system which this model, can say the IPFS is used to bypass storage obligations and increase throughput. They talk about dual blockchain by adding the master block reference in the ledger instead of the original blockchain. Their analysis shows that this method can achieve up to 25.8x higher throughput and nearly 1685-time lesser lodger register size than Bitcoin Core.

[3] In this paper author says Blockchain technology is a distributed technology that supports distributed infrastructure and it is a computing pattern. After that coming version, introduced the super account book and later introduced blockchain 3. This paper is a combination of the core of blockchain technology and kernel. "Blockchain Technology + Accounting Services", this technology used in the accounting industry by combing blockchain technology and IoT (Internet of things) based on the amount of data

There are several research papers. But the paper which is write the [4] Jammulspti Ravi Teja where he talks about the maintains of the documents which is important like bonds, government related documents, property documents and many more. And, authority where the documents In this project, re get altered which is very insecure to overcome this drawback, he provides the solution by implementing the Merkle tree algorithm for data storing.

[5] Blockchain technology has received great attention in overall field in recent years. And has Reached great height in the past 10 years However at some extent, blockchain data volume is continuously increasing due to non-deletable, immutable and non-additional features. This theoretical method is proposed in this paper to increase throughput and reduce the storage. In this paper they focused on the IPFS-based blockchain data storage to solve the problem. By using the characteristics of the IPFS network and the features of the IPFS hash function, blockchain data is significantly reduced. Compression ratio can reach 0.00818. Based on analysis, it also has good performance in security and sync speed of a new node

Since, The IPFS work on peer- to- peer network [6] In this paper author talk about the peer-to-peer version-controlled file system that synthesize learning. based on the IPFS and blockchain they utilize the use of the IPFS and block chain in way that the individual user can access the file in the secure way. They address the high- throughput problem for users. IPFS by introduce the role of the content service providers and the block chain to combined IPFS with the storage model. Based on their analysis this provide scheme can effectively solve the problem of security.

3. METHODOLOGY

For understanding our project first one must know how the blockchain work for the best example is the bitcoin which is now day more popular the bitcoin is based on the blockchain is just like a database for [10] storing the transaction. In general, the blockchain acts as the database which stores the information, unlike most database which in the centralized system but the blockchain is a decentralized system which means that the data does not store in the particular system it stores the data in the network which is called the Ethereum blockchain network. The blockchain is like the block which is connected in the manner of a linked list. Where each block holds the address of the previous block. Because of the decentralized manner, there is no authority that private the user to access the data it is highly secure, and transparency of the transaction which is made our project include the high use of the blockchain.

For our project we used two technology 1. Ethereum Blockchain which is used to store the data about user file as the hash file. 2. IPFS (Inter Planetary File System) it is basically used for the sharing data in a distributed file system.

Secure Document Vault Algorithm

Step 1: - get the data or file from the user and store in the object variable

Step 2: - convert the file into file Buffer

Step 3: - send the file document to the IPFS

Step 4: - then IPFS create a node for the file revert back the hash of the node where the file is store.

Step 5: - check whether the data existed in the block chain if existed print the data is already available.

Step 6: - if not store the hash into the block chain with some additional data like (document name, categories, existence etc)

Step 7: - Display the file from the block chain.

The full elaboration of the algorithm is below

A. get data file from the user

In this project for getting the user data, we have to call the document object function which help to fetch the data file from the user personal system and store it in the object variable. For this model used the react for the user experience where the react will handle the getting file function.

B. Convert the file into file Buffer

In this project, will convert the file into the file buffer which is act as the temporary file object. Without effecting the original file once it saved it will automatically update the file this will allow user to elaborate its documents before updating

C. send the documents to the IPFS

To actual work this first this project, one must send the document to the IPFS node. Where the IPFS will evaluate and the check whether there was duplicate or not and store the file in the node and send back the IPFS hash. The hash is work in the principle of DAGs (directed acyclic graphs) specifically, they use Merkle DAGs this will help uniquely identifier that is a hash of the node’s contents. In this project, we are using the ‘infra’ which is the public IPFS API’s for storing the document for more information visit the infra.io this website provides the user to store the documents. And after receiving the hash code from the IPFS In this project, will store the data in the block chain. The main drawback is the by using the hash code any one can access the file which is not efficient so overcome this project, encrypt furthermore for privacy.

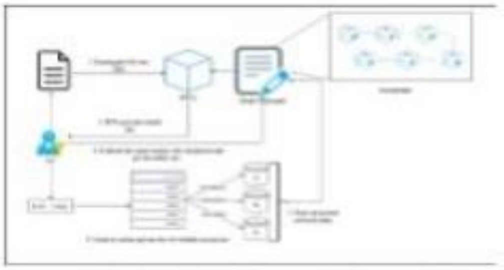

Fig. 2. Working techology of the secure document vault

D. checking for duplicate

Checking for the duplicate is important if the same document object is available at the block chain is will lead to wastage of the block chain as In this project the data redundancy which leads to further complication in the block chain [8]

E. store in the block chain

In this project, after getting the IPFS hash further encrypt the data from the IPFS by using the RSA Encryption algorithm with user details and categories which that document belongs and store into the block chain. After it will ask the gas to pay for transaction to be completed.

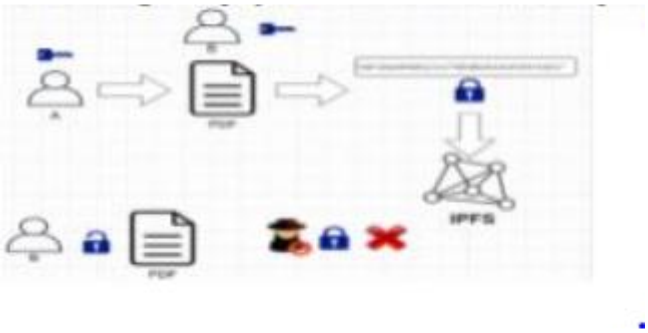

Fig. 3. Key authentication

Let’s say the file random.jpg file will store in the block chain first it will send the doc to IPFS then IPFS will going to process it and send the hash code of the file back to program. Where the program will encrypt the hash and add random private key which look like.

"0xF42ba36FdaFe215c574D1f6a18A61d419044A915" +private Key only the user who has the same key can only access. If other than user who has a key try to access it will be going to show the error and it will not show anything. After words it will store back into the block chain with some additional information like user id categories from which it is belongs and some more information like (document name, categories, existence etc).

4. RESULT

In this project the result got is from the prototype which the data is created in project, will be further modifying it the below screen capture will provide the demo of our project. In this project we see that sending the file to IPFS in fig.5 will created the buffer of the file which is been uploaded and in the same figure how much Gas we have to pay for the updating in the block chain. In figure 6 In this project, successfully get the file and displaying on the screen.

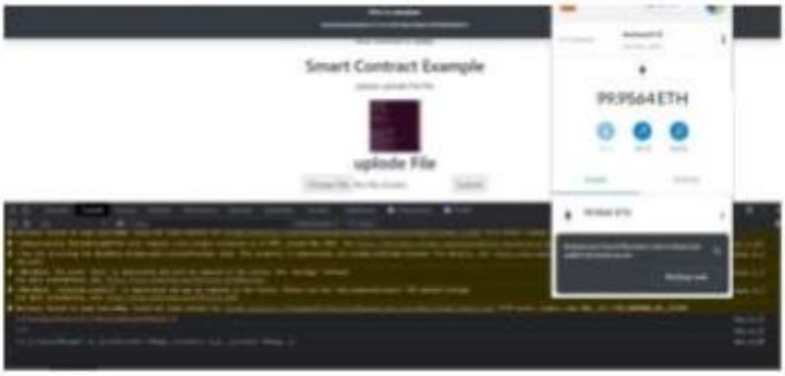

Fig. 4. Screen shoot (1) before updating

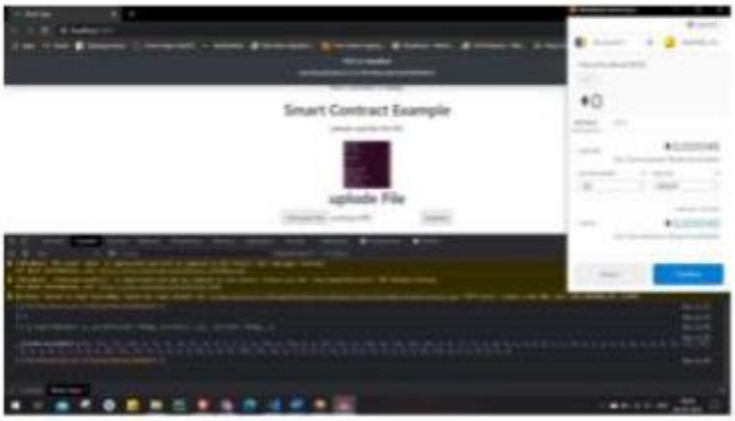

Fig. 6. Screen shoot (2) updating document.

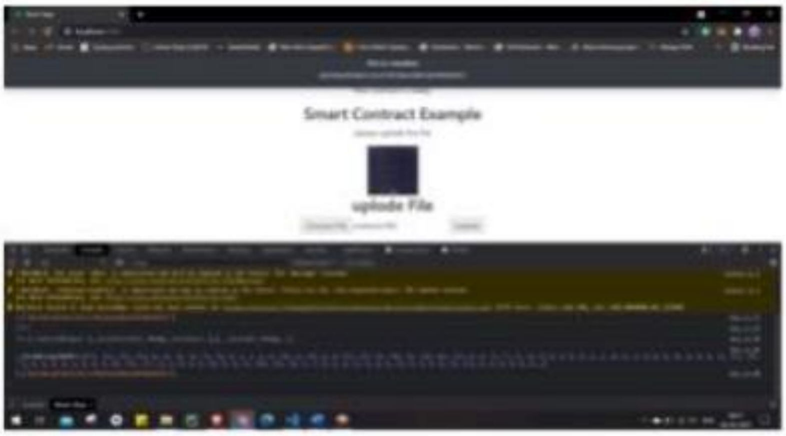

Fig. 7. Screen shoot (3) after adding document

The project can be further moficable by providing the addition features as user interface. Ex

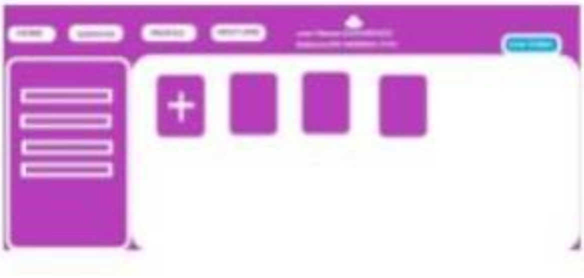

Fig. 8. Modification interface

5. CONCLUSION

Since nowadays these type of project come to know that security is very important for every individual. Now the world governs by the data which has all the information about the user like what he wants, what will be his preferences . So, the hacker might try to steal the information from the user and used it to harm them, the technology like blockchain and IPFS is very secure and reliable the user needs to get to know about this technology. Our main goal is that the user may able to trust a platform where one can store and manage his private documents in that platform form. The data which is store in the blockchain should not be visible as this project, as encrypted for in encryption RSA algorithm might be the good solution to the problem. And the node which is created by the IPFS to store the file. The user might not know where exactly the node is created so if the file got to change the hash code also change accordingly so the IPFS creates the copy of the node in the several systems. The blockchain will provide high security where each block is connected like a linked list. So, this will make the highly secure system.

6. REFERENCES

[1] Paras Pant; Ruchika Bathla; Sunil Kumar Khatri, “A Model to Implement and Secure Online Documentation using Blockchain”, 2019 4th International Conference on Information Systems and Computer Networks (ISCON).

[2] MARYLAND. Soharab Hossain Sohan;Minahz Mahumound;MA Baten Sikder”Increasing throughtput and reducing block problem using IPFS and the Dualblockchain Approach”

[3] Songyue Liu; Shangyang He, “Application of Block Chaining Technology in Finance and Accounting Field”.

[4] Jammulapati Ravi Teja, “Proposing method for Public record maintenance using Block chain”.

[5] Qiuhong Zheng; Yi Li; Ping Chen; Xinghua Dong, “An Innovative IPFS-Based Storage Model for Blockchain”

[6] Yongle Chen; Hui Li; Kejiao Li; Jiyang Zhang, “An improved P2P file system scheme based on IPFS and Blockchain”.

[7] Deepak K. Tosh, Sachin Shetty, Xueping Liang, Charles Kamhoua, Laurent Njilla “Consensus Protocols for Blockchain-based Data Provenance: Challenges and Opportunities”.

[8] Y. Zhang, S. Kasahara, Y. Shen, X. Jiang and J. Wan, "Smart contract-based access control for the internet of things"

[9] mafiadoc.com

[10] Subramaniam, Muthusamy, Ayyaswamy Kathirvel, E. Sabitha, and H. Anwar Basha. "Modified Firefly Algorithm and Fuzzy C-Mean Clustering Based Semantic Information Retrieval." *Journal of Web Engineering* (2021): 33-52.