Blockchain based Secure Data Storage and Access Control System using Cloud

Shubham Desai
Department of Computer Engineering
PCCOE, Pune, 411044
shubhamdesai085@gmail.com

Onkar Deshmukh
Department of Computer Engineering
PCCOE, Pune, 411044
onkardeshmukh51@gmail.com

Rahul Shelke
Department of Computer Engineering
PCCOE, Pune, 411044
rahulshelke3099@gmail.com

Harish Choudhary
Department of Computer Engineering
PCCOE, Pune, 411044
choudharyharish68@gmail.com

Prof. S. S. Sambhare
Department of Computer Engineering
PCCOE, Pune, 411044
ssambare69@gmail.com

Arjunsingh Yadav
Platform Engineer
Fynd, Mumbai
arjun.yadav015@gmail.com

Abstract—
Cloud storage today depends entirely on large storage providers. Such storage providers function as untrusted third parties that process data for storing, sending and receiving data from an entity. This style of system has many problems, such as high operating costs, software quality and data security. In this paper, we present a model of a multi-user access control system for databases that use blockchain technology to provide stable, distributed data processing. The system allows the data owner to upload the data via a web portal. So, the user who has the secret key to the particular data that has been uploaded to Cloud in encrypted form can only access the folder. Eventually, the system promotes data privacy by maintaining the immutability of the blockchain by processing it in the cloud. We have proposed a secure, blockchain-based data storage and access control system to increase the security of cloud storage.

Keywords— ***Blockchain, Cloud Storage, Smart Contract, Encryption, Decryption***

I. INTRODUCTION

There has been a challenge in the current era of maintaining vast amounts of data for large organizations that work globally. As a consequence, many organizations have switched to cloud storage, which has outstanding storing, distribution and upload services. The main problems faced by cloud computing are preserving the confidentiality and integrity of data in aiding data security. Most of the users prefer cloud to store their personal data. However, there are few security and copyright issues related with the data. The main problem when transmitting data to an external environment is that anyone other than the owner can have access to the information. Cloud providers do not ensure the high level of protection required for appropriate data security and privacy. At present, there are only few resources and methods available to secure data stored on cloud.

To overcome this problem, this paper proposes a system that provides data storage with the help of the Blockchain-based Secure Data Storage and Access System. Therefore, we propose to use Blockchain as a trustworthy environment to make cloud storage more secure and to protect exploitative attacks.

Blockchain is a tamper-resistant, decentralized electronic ledger that tracks transactions on a P2P network. Shared to every participant nodes in the network, the ledger forever registers the transactional data between the nodes in the network of an ordered chain of cryptographic hash-linked *blocks*. All authenticated and approved transaction blocks are linked and enchained from the end of the chain to the main current node, thus the term blockchain. The blockchain, thus, serves as a single source of truth, and all users of the blockchain network can see only such transactions that are allowable to them. [1]

Blockchain is a technology that enables all participants to create a Ledger comprising all transaction data and to modify their Ledgers to preserve validity when a new transaction occurs. Since the evolution of Web and cryptographic technologies, it has become possible for all participants to check the integrity of the transaction and a single point of error resulting from reliance on an official third party has been eliminated. It is a form of distributed ledger used to create a permanent and tamper-proof database of transactional information.

978-1-7281-4042-1/19/$31.00 ©2019 IEEE

The blockchain is an organized list which stores data in a form similar to the shared ledger and is intended to make it impossible to exploit unilaterally since the network members save and verify the blockchain. Every block is made up of a body and a header. The header contains the hash values of the current and previous blocks and nonce. The block information is retrieved from the server using the index process. Since the hash values stored in each neighbor in the chain are influenced by the values of the previous blocks, it is very difficult to falsify and change the recorded data. [2]

II. ADVANTAGES OF BLOCKCHAIN

Blockchain's immutability feature is accomplished by transactions that are settled upon and exchanged across the Blockchain. When the transaction is linked to the Blockchain, it will not be possible to change or remove the transaction. It also depends on the type of system– if the system is centralized, it can be modified and removed, because the decision is made by one individual. But if the system is distributed, such as the Blockchain, the transaction that is linked to the Blockchain is replicated to each device in the Blockchain network. This advantage makes Blockchain technology unalterable and invulnerable.

The transparency feature of the Blockchain is obtained through the copying phase of the transaction. As stated above, each transaction is reflected on either of the machines in the Blockchain network. Every member can look at all transactions which also means that each activity is shown to the Blockchain members i.e. it is transparent. [3]

SMART CONTRACTS:

Smart contract is an implementable code that executes on the blockchain to create, implement and impose the provisions of a contract between untrusted parties. Basically, it is self-operating in nature. The main objective of a smart contract is to dynamically enforce the terms of the contract once the defined requirements have been met. Thus, it offers small transaction fees with respect to conventional services that need a trustworthy third party to implement the contract. Various blockchain technologies can be used to build smart contracts, but Ethereum is the most popular one. This is because the Ethereum platform embraces the Turing-completeness feature that provides more complex and customizable contracts to be developed. [4]

III. RELATED WORK

Mr. Anup R. Nimje, Prof. V. T. Gaikwad and Prof. H. N. Datir [5] has compared seven different attribute-based encryption techniques namely ABE, KP-ABE, EKP-ABE, CP-ABE, CP-ASBE, HIBE, HABE, HASBE and concluded that Hierarchical Attribute Set-Based encryption (HASBE) provides the best access control mechanism. It is flexible and most efficient with less computation overhead than other techniques. The HASBE scheme consists of a hierarchical structure for application users by using a CP-ASBE delegation method. The HASBE system favors composite attributes due to dynamic and reliable attribute set combinations and enables effective client revocation due to multiple value attributes.

Pooja More [6] suggested a protection for cloud data using the Key Aggregate Cryptosystem centered on the attributes. To query the file stored in the cloud, the system uses trapdoor key and searchable keywords. On fetching of the file, it requires an aggregate key to encrypt and access a particular file from the data repository. For a particular community, the trapdoor key is publicly available, but access to the aggregate key is dependent on the characteristics of the information owner. The system utilizes the CP-ABE technique with a fixed length of ciphertext and key which enhances the flexibility by having two different folders, one document for key, and another document for a group of attributes instead of binding user attributes to a key.

Ilya Sukhodolskiy, Sergey Zapechnikov[7] suggested a blockchain-based user access platform for cloud storage. This offers a framework for accessing data that is stored in untrusted environments, i.e. cloud storage. For example, the data will be stored securely on the cloud, of instance, multimedia files, records etc. where the information recognizing the file is accessible on the blockchain. The data stored in a blockchain is anonymous, so it will be encrypted before it is sent for processing and access control. The client who wants to read a file must fit the access policy and have the key to decrypt it and open it. The information owner supplies the keys for decryption. The main advantages of the access control system are: the ability of flexible access policies, the ability of other stakeholders to modify access policies do not require additional measures to keep user keys unchanged, including granting and modifying access, the security and privacy of all transaction data, facts gain access to file, rejection of fact and the inability to edit and modify these data is guaranteed by using blockchain and smart contracts.

Maximilian Wöhrer and Uwe Zdun [8] have outlined six design patterns namely rate limit pattern, speed bump pattern, mutex pattern, balance pattern, check-effects-interaction pattern and emergency stop pattern that resolve security issues while writing code of smart contracts in Solidity. These patterns resolve the issue of loss of execution control after deployment of the contract, resulting from Ethereum's sandbox execution environment. This one-of - a-kind feature of Ethereum that enables autonomous execution of programs on the blockchain, but it also has disadvantages. Such disadvantages arise in various forms, either as negative bounce backs, or as adverse conditions as to how and when things are conducted, or as uncontrollable financial risks. By implementing the trends identified, we may fix these security issues and reduce conventional attack scenarios.

IV. PROPOSED SYSTEM

Cloud Technology is one of the most emphasizing and useful technology in the current era. Cloud storage has provided a digital storage solution that utilizes multiple servers across different locations to safely store the data. In the past few years, cloud storage has grown in different sectors and has become a direct challenger to local storage. Today cloud technology has become a promising paradigm of computing for different third-party service providers. The cloud provides the service of data storage which enables data owner to store their data on the cloud and give access to the organizations who needs that data.

[9] While cloud computing has many benefits, there are many inevitable security issues as well. Some security issues related to cloud storage are as follows-

a. Data privacy –

Nobody wants their information to be accessed until you permit them to do so. Privacy is the ability of the individual or group to separate themselves or to reveal information about themselves selectively. It also allows consumers to regulate their data when data is stored and handled in the cloud, and prevent theft, nefarious use, and unauthorized resale. This seems to be simple enough to manage if you store information upon-site, but what about the other cloud? As your own information is stored somewhere, this might not be possible to identify how open it is. How are you so sure that nobody can ever access it unless you keep the web servers on which it is stored? When you transfer personal data to the cloud, be mindful that you may be compromising important privacy controls.

b. Lack of control –

You take a lot of burden off your back when you depend on a third party to store data for you. But this is a two-edged sword. On the one side, you're not going to have to handle the data-on the other hand, someone else will. If something harms your data, such as power failures or ransomware attacks which will then directly affect access to your data. You're relying entirely on your provider to resolve these issues. The larger time your data spends unprotected, the riskier it becomes.

c. Data leakage –

A major part of the cloud data storage ensures that nobody outside the organization is trying to access documents. The purpose is to make sure the information is not sent to anyone outside the organization. The biz-critical or personal data is exposed to external sources which makes data leakage a serious issue.

d. Data Breaches –

These are the consequence of an assault or negligence and mistake by the worker. This is the main source of worry on cloud systems. Vulnerabilities in the implementation or inadequate safety procedures may also trigger information breaches. Employees can sign in to cloud systems from their machines or private laptops, thereby exposing the scheme to malicious attacks. It covers every kind of data not meant for public release, including personal medical records, financial records, confidential information and proprietary information.

e. System Vulnerabilities –

Cloud computing devices may involve system vulnerabilities, particularly in networks with complicated infrastructures and various third-party applications. Once a vulnerability has been identified with a common third-party system, it can be readily used against organizations.

V. OVERVIEW OF SYSTEM

To overcome these security issues mentioned above, we are proposing a "Blockchain-based secure data storage and access control system" model. In our model, we are using Blockchain technology to provide more security to the Cloud data. The links for accessing the documents will be stored on the blockchain. First, the data owner will upload the documents over the cloud in the encrypted form. The documents will be encrypted using the HABSE encryption technique. These documents can be decrypted by the aggregate key provided by the data owner.

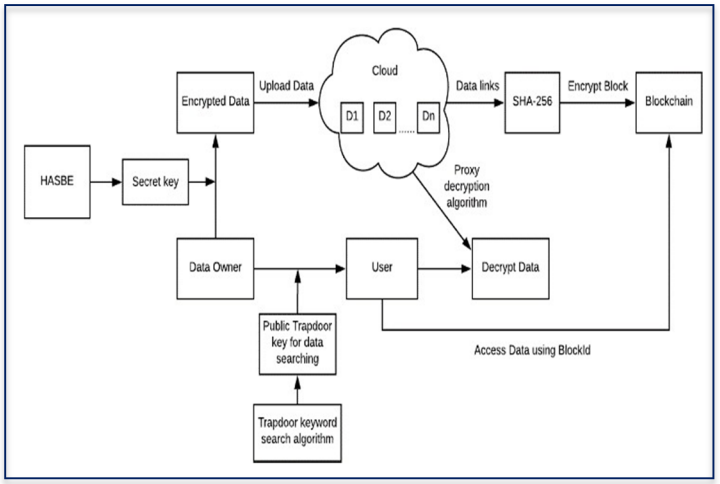

Fig.1. System Architecture

The user who wants to access files stored in the cloud must first verify their identification. Upon verification the user searches the required file using a keyword search for any number of documents. The information owner shall share the public trapdoor key for each file only with those users who are approved by the data owner. If the user locates the requested file over the cloud, he / she will ask the information owner to access the data stored in the cloud. When a information owner gets a request, he or she give the aggregate key or hashes keys to the user who demanded access to that data. The user can now access the link stored in the blockchain and using the aggregate key user can decrypt the encrypted data stored on the cloud. Hence, the requested file can be downloaded.

A. FIGURES AND TABLES

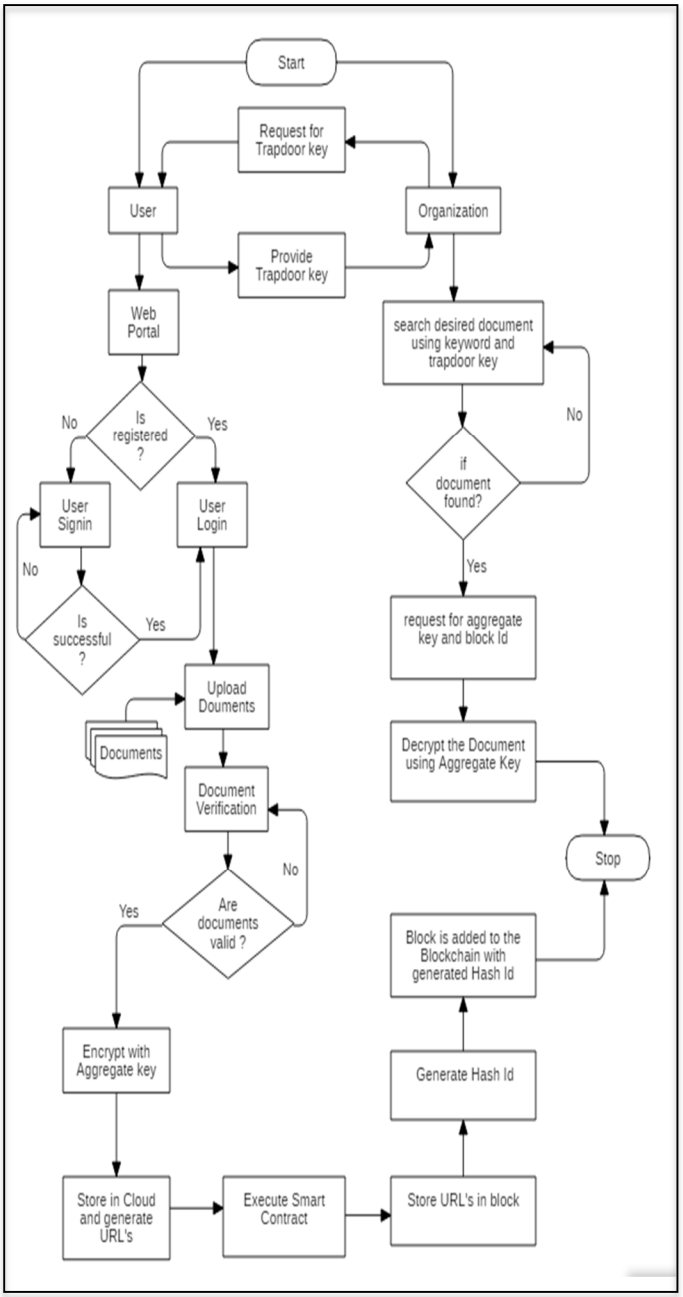

Fig.2. Flow Chart of Proposed System

VI. ANALYSIS OF ALGORITHM

Hashing algorithm is also termed as hash function and plays a vital role in modern cryptography [10]. The hash function is a public function, generally defined as H, we use it to map arbitrary brief message to a shorter, fixed-sized value, this value is an authentication code, we denote it as H(M). The H(M) is also called hash value, hash code, message digest or array fingerprint. The hash function can be viewed as a one-way cryptosystem in the password view, which means that it can only be encrypted, decryption is impossible. The hash value is a feature of the message bits, so it has error detection functionality because if one bit or a few bits are modified, the hash value will also be altered. In cryptography and data security technology, the hash function is an important tool that can help us achieve accurate, secure and reliable digital signature and certificate and is an important element in the safety authentication protocol. The most commonly used hash function can be divided into the following categories:

A] SHA-256:

The SHA256 algorithm is a cryptographic hash function and is used in both digital certificates and data integrity. As data, the SHA256 algorithm uses an arbitrary-length text that is less than 264 bits and produces a 256-bit reference digest as output. SHA256 is used to encrypt passwords, since the server side only needs to control the hash value of a particular user, rather than the hash key. In case an Intruder hacks the Server, this is useful as they will find only the hashed function and not the real password.

B] MD-5:

The MD5 cryptographic algorithm is a one-way hashing method which recognizes a token of any length as source and produces a fix-length digest value to be used to authenticate the original message as result. The algorithm takes an random length message as its source and produces a 128-bit ' fingerprint ' or ' text digest ' of the source as its result. It is presumption that producing two messages with similar message digesting or producing message of any kind with a given pre-defined target message digesting is computationally infeasible. It is specifically designed for digital signature implementations where a huge document has to be ' compressed ' securely before being authenticated with proprietary (secret) keys such as RSA under a public key cryptographic algorithm.

C] Comparison of Hashing Algorithms:[11]

a. Time required to Encrypt Data:

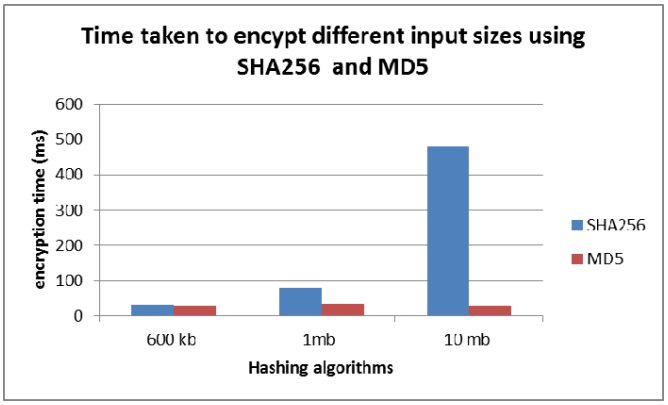

Fig.3. Time required to encrypt different input sizes using a hashing algorithm [11]

Fig. 3 Displays a contrast between SHA256 and MD5 for the different source size encryption time. From the chart it can be observed that there is minor difference between the 600 kb output size encryption times. Further, when the source size increases, the gap becomes bigger. SHA256 spent more time than MD5 when the source size grew to 1Mb. The time required to encrypt SHA256 is about 5 times greater than MD5 when the source size is 10MB.

b. Power Consumption:

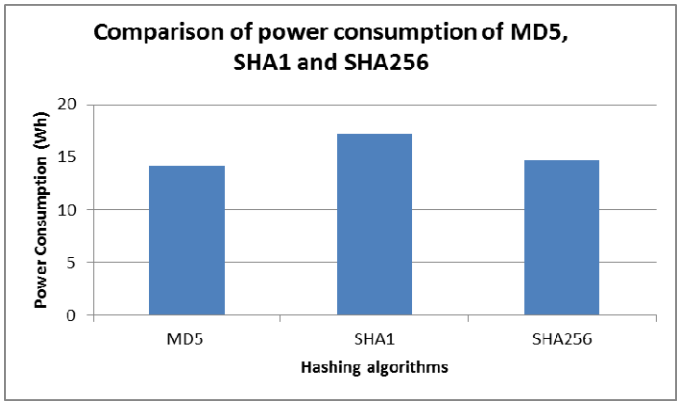

Fig.4. Comparison of MD5, SHA1 and SHA256 power consumption [11].

Fig. 4 displays MD5, SHA1 and SHA256 power consumption. It can be observed that MD5 uses less energy than SHA256 followed by SHA1.

c. Latency Comparison:

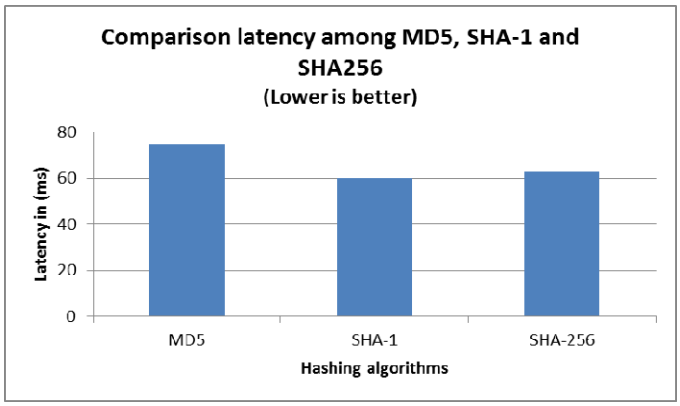

Fig .5. Latency comparison between MD5, SHA1 and SHA256 [11].

Fig. 5 contrasts the MD5, SHA1 and SHA256 latency. It can be observed that the most delay is caused by MD5, followed by SHA-256 and SHA-1.

d. Hashing algorithms - Comparison

Table 1 compares SHA1, SHA256 and MD5 in terms of 10 MB input size encryption time, energy consumption, latency and protection level. MD5 is the best with respect to time for authentication and energy consumption. Among these three algorithms, SHA1 consumed the most power. MD5 has the highest latency with SHA256 followed by SHA1. SHA256 is eventually safer and more efficient than SHA1 and MD5.

Comparison among hashing algorithms			
Parameters	Algorithms		
	SHA1	SHA256	MD5
Encryption(10Mb)	-	480ms	27ms
Power Consumption (Wh)	17.5	14.8	13
Latency (ms)	60	62	74
Security	Prone to attack	Secure	Prone to attack

Table.1. SHA1 SHA256 and MD5 comparison.[11]

VII. CONCLUSION

The proposed approach suggests a prototype of a secure cloud storage system using blockchain. The proposed system secures the data which is stored in unreliable environments. Some of the security algorithms of acceptable time complexity, usability, and efficiency was selected to implement the system. The data will be stored in the cloud, where the file location information will only be visible on the blockchain. The data stored in the blockchain is publicized, so it is encrypted before sending it to cloud and control access to it. The users willing to read a file should satisfy the access policies and have the relevant trapdoor key and search keyword to search for a file stored in the cloud. The aggregate key issued by the data owner is used to decrypt a certain file from the document pool in order to download it.

The proposed system incorporates HASBE technique with a fixed length cipher text and key, which enhances the efficiency of the system. The system uses blockchain to store links of encrypted documents that are available on the cloud.

So, the proposed system provides an effectual solution for current cloud storage systems.

VIII. REFERENCES

[1] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends", IEEE International Conference on Big Data (Bigdata Congress), 2017.

[2] Jin Ho Park and Jong Hyuk Park, "Blockchain Security in Cloud Computing: Use Cases, Challenges, and Solutions", Department of Computer Science and Engineering, Seoul National University of Science and Technology, (SeoulTech) 01811, Korea, 18 August 2017.

[3] Julija Golosova, Andrejs Romanovs**, "**The Advantages and Disadvantages of the Blockchain Technology", IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2018.

[4] Maher Alharby and Aad van Moorsel, "Blockchain-Based Smart Contracts: A Systematic Mapping Study", Fourth International Conference on Computer Science and Information Technology (CSIT), 2017.
[5] Mr. Anup R. Nimje, Prof. V. T. Gaikwad, Prof. H. N. Datir, "Blockchain Attribute-Based Encryption Techniques in Cloud Computing Security: An Overview", International Journal of Computer Trends and Technology, Volume 4, Issue 3- 2013.
[6] Pooja More, **"**Cloud data security using attribute-based key-aggregate cryptosystem**",** International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016.
[7] Ilya Sukhodolskiy, Sergey Zapechnikov, "A Blockchain-Based Access Control System for Cloud Storage," IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 2018.
[8] Maximilian Wöhrer, Uwe Zdun, " Smart contracts: Security patterns in the ethereum ecosystem and solidity", International Workshop on Blockchain Oriented Software Engineering (IWBOSE), 2018.
[9] Naresh vurukonda, B.Thirumala Rao, "A Study on Data Storage Security Issues in Cloud Computing", 2nd International Conference on Intelligent Computing, Communication & Convergence, 2016.
[10] H. Khali, MIEEE, R. Mehdi, A. Araar, "A System-Level architecture For Hash Message Authentication Code", 12th IEEE International Conference on Electronics Circuits and Systems, 2005.
[11] Aquino Valentim Mota, Sami Azam, Bharanidharan Shanmugam, Kheng Cher Yeo, Krishnan Kannoorpatti, "Comparative Analysis of Different Techniques of Encryption for Secured Data Transmission", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), 2017.