

Blockchain based Cloud Data Storage System

Aishwarya Patil^{#1}, Swapnajit Patil^{#2}, Sachin Rokade^{#3}, Vijay Sharma^{#4}

#Dept. of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

Abstract— In today's world, the simplest way to share data is through the internet. Cloud computing is a technology provided by the internet, which is dependent on large storage providers. These storage providers act as untrusted third parties who manage enormous data stored on the cloud. This data may contain sensitive information that belongs to multiple individuals or organizations. Such types of models may involve security issues like privacy and integrity. In this paper, we introduce a prototype of a multi-user system for access control to documents that use the blockchain technology for securing shared data storage. The data owner is allowed to upload the documents on the cloud using Web Portal and the user will request an access link of the document from the owner. Whenever the user tries to access the document using the provided link, a smart contract will be triggered which will send a notification to the owner. The owner will receive the notification to grant permission to the user. The user who has the permission to access a particular document stored on the cloud can only access it. The above operation on the document will be recorded on the blockchain. Owner can always see the logs to find any suspicious operation on the documents. Therefore, the privacy of data is ensured using the smart contracts, immutability property and ledger of blockchain.

Keywords— Blockchain, Cloud Storage, Cloud Triggers, Smart Contract, Logs, Ethereum, Ledger, Data Privacy, Data Security, Solidity, Cloud Provider.

1. INTRODUCTION

Nowadays, the problem faced by various organizations is storing an enormous amount of data. To address this issue, organizations have adopted cloud storage as an option to store data. Due to this in the last few years, cloud-based services have increased. These services facilitate remote storage of user data on the cloud as well as properties like sharing and transferring of data. Organizations need not maintain in-house storage because services are available irrespective of time and location across multiple platforms. Despite the mentioned benefits, there are various problems associated with cloud storage. They are maintaining the confidentiality and integrity of data. Data stored on the cloud may contain sensitive information. However, here copyright issues come into the picture. As we are uploading data on the external environment, anyone other than the owner can access data. Security is the most crucial parameter that should be taken into consideration while storing data on the cloud. But, cloud service providers don't ensure a high level of security. Currently, there are very few options available to guarantee the security of data on cloud servers.

The system presented in this paper will help to overcome all the issues mentioned with the help of Blockchain-based Secure Data Storage and Access System. In this paper, Blockchain enhances the security of the data stored on the cloud by maintaining logs of operations performed by the user. Additionally, Blockchain will secure the data from various attacks.

A. Blockchain

Blockchain is a technology that will serve for a bright future. It can help us develop systems that will be more safe and reliable. These systems will apply to any type of data irrespective of its nature. That is, we can use it for multimedia information, electronic documents, etc. To store this enormous amount of data directly on the blockchain is not good practice because it will lead to an increase in block size and number of blocks in the chain. So the records will be stored on the cloud and information that will help us identify, tampering in the document will be stored on the blockchain.[13]

Blockchain is a tamper-proof, distributed ledger that maintains all transactions on a public or private peer to peer network. All the member nodes in the network hold a copy of the ledger. Ledger keeps the transactional data of the nodes that are present in the system of an ordered chain of cryptographic hash linked blocks and is updated when any new transaction is executed in the network. Hence the integrity of the network is maintained. Transaction Blocks that are validated and confirmed are connected. This forms the chain of blocks, Thus the name blockchain. Therefore, blockchain acts as a single source of truth and all the members in a blockchain network can inspect only those transactions that apply to them.[17]

Blockchain is a technology that uses encryption to maintain the reliability of a network. Since it is a peer to peer network, it also solves the issue of a single point of failure due to interference of a third party. Each block in the blockchain has a header and body. Block header is a combination of previous and current block hash and nonce. Each block contains a previous block hash, which results in the formation of the chain. If any particular transaction is altered or deleted from a block, then this will modify the hash value for that block. Since the hash value of a block is affected by previous block hash, so the change in the hash value of any particular block will have to change the hash values of succeeding blocks. This

process of altering hash values of blocks is very difficult. Thus, it is also very challenging to forge or modify recorded data on the blockchain.[13]

B. Cloud System

A cloud system refers to the computing components like hardware, software and, infrastructure that enable the delivery of cloud computing services via a network[18] i.e. Internet. These services include SaaS (software as a service), PaaS (platform as a service) and IaaS (infrastructure as service). Users can access these computing services using browsers. This represents a computing model that shifts the computing workload to a distant location. Internet-based email applications are a good example of a cloud system that has a platform for the delivery of messaging services. Cloud computing is also called utility computing since consumer usage of cloud systems is metered and billed in a manner just like water or electric services.

2. RELATED WORK

Maximilian Wöhrer and Uwe Zdun[1] have printed six style patterns particularly rate limit pattern, preventative pattern, mutex pattern, balance pattern, check-effects-interaction pattern and emergency stop pattern that address security problems once secret writing good contracts in Solidity. These patterns solve the matter of lack of execution management once the contract is deployed, ensuing from the distributed execution surroundings provided by Ethereum. This one-of-a-kind characteristic of Ethereum permits programs on the blockchain to be dead autonomously however conjointly has drawbacks. These drawbacks seem in several forms, either as harmful callbacks, adverse circumstances on however and once functions square measure dead, or uncontrollably high money risks at stake. By applying the bestowed patterns, we will address these security issues and mitigate typical attack eventualities. No such structured and informative style pattern language for Solidity, which will be used as steerage for developers or notice its application in automatic code generating frameworks, is obtainable at the instant.

Qiwu Zou, Yuzhe Tang, Ju Chen, Kai Li, Charles A Kamhoua, Kevin Kwiat, Laurent Njilla[2], have proposed a system called "ChainFS: Blockchain-Secured Cloud Storage", which is a multi-client file system in the cloud whose security is hardened by leveraging Blockchain. The root cause of clients/industries not moving to the cloud is security trust issues, that clients lack trust in the cloud as a third party. So, involvement of blockchain as a trusted third-party (TTP) hardens the security of cloud storage and also prevents the forking attacks in cloud file systems.

The prevention of forking attacks is due to the fact that forking file-system can be viewed as hard as double-spending transactions in Blockchain. In this blockchain-secured cloud storage, blockchain is involved in securing: a public-key directory which helps to manage user identities and is a foundation to establish trust among the users and an operational log server that records client-server interactions when accessing a remote file system. The system proposed "ChainFS" mainly consists of three parties, i.e. client, server and blockchain. The Filesystem in User Space (FUSE) client holds the client filesystem, client applications and special system calls like Remote Procedure Call (RPC) that are used to communicate with Secured Untrusted Data Repository (SUNDR) server. The SUNDR server stores the actual data in form of block storage and auditable logs of RPC operations. The blockchain works as a trusted third party that stores links to public-key directory from Certificate Authority (CA) and links to files from block storage. From implementation point of view, this system is an integration of Linux file system at client-end, Amazon S3 storage at cloud-end and Ethereum blockchain.

Fran Casino, Thomas K. Dasaklis, Constantinos Patsakis[3], provided associate degree insight into blockchain-based applications across multiple domains. They conjointly investigated this state of blockchain technology and its applications. They highlighted however specific characteristics of this riotous technology will revolutionise "business-as-usual" practices. Blockchain-enabled applications include supply chain, business, healthcare, IoT, privacy, and knowledge management any many more. They conjointly established key themes, trends and rising areas for analysis. They conjointly pointed to limitations the blockchain technology presents and the way these limitations spawn across completely different sectors and industries. Their analysis methodology enclosed following steps: Locating studies, Study choice and analysis, Analysis and synthesis. While blockchain applications square measure being widely deployed, several problems have however to be self-addressed. By doing this, blockchains can become not solely a lot of ascendable and economical however a lot of sturdy further. The options they provide don't seem to be distinctive if judged singly, and the bulk of the mechanisms supported are well-known for years. However, the mix of these options makes them ideal for several applications justifying the extreme interest by several industries. As blockchains become a lot more mature, their applications' square measure is expected to penetrate a lot of industries/domains. Main characteristics highlighted in their paper square measure accord mechanism, identity obscurity, protocol potency and consumption, immutability, ownership and management and group action approval with relevance to

public, private and united blockchain networks. We will infer that whereas several attempt to propose blockchains as a nostrum and an alternate to databases, this is often faraway from true, as steered by their paper. Since its origination, blockchain technology has shown promising application prospects. From the initial cryptocurrency to this good contract, blockchain has been applied to several fields. Their paper highlights few limitations like quality, Latency and measurability, Quantum resilience, Blockchain adoption and ability, Data management and privacy & security solutions, Big knowledge and computing.

identifying the file will only be available on the blockchain. The data stored in a blockchain is public so it is encrypted before sending it to storage and control access to it. The user wishing to read a file must match the access policy and have the necessary keys to decrypt and download it. The keys for decryption are provided by the data owner. The important benefits of access control system are: the ability to customize the access policy for encrypted data without duplication, the ability of dynamic access policies, access policy change does not require additional actions from other members, which make the user keys remain unchanged, the integrity and confidentiality of information about all transactions, including granting and changing access, 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.

TABLE I

CLASSIFICATION AND MAIN CHARACTERISTICS OF BLOCKCHAIN NETWORKS

Property	Public	Private
Consensus mechanism	Costly PoW, All miners	Light PoW, Centralised Organisation
Identity and anonymity	Anonymous may be malicious	Identified users, Trusted
Protocol Efficiency and Consumption	Low efficiency High energy	High efficiency Low energy
Immutability	Almost impossible	Collusion attack
Ownership and management	Public permissionless	Centralised Permissioned whitelist
Transaction approval	Order of minutes	Order of milliseconds

R.Gowthami Saranya, A.Kousalya, in their paper[4] deals with the different algorithms or methods used for securing data in the public cloud. Their paper gives explanations about benefits, characteristics, security issues in cloud computing. It also gave comparative analysis of various cryptographic algorithms like DES, BLOWFISH, RC2, RC5, RC6, 3DES, RSA, AES, DSA, Diffie-hellman key exchange, TWOFISH, IDEA, ELGAMAL, Homomorphic encryption with respect to following parameters year of development, blocksize, keylength, security level provided and speed of execution. We infer from the paper that multilevel security architecture is required because Single security algorithms can't be trusted. Also algorithms used are situation dependent.

Shubham Desai, Rahul Shelke, Omkar Deshmukh, Harish Choudhary, Prof. S. S. Sambare proposed "Blockchain-Based Secure Data Storage and Access Control System using Cloud[6]" It provides a model to access data stored in untrusted cloud storage. The data will be stored in cloud storage and the information identifying the file will only be available on the blockchain. The data stored in a blockchain is public so it is encrypted before sending it to storage and control access to it. Here the HASBE technique is used for encrypting data stored on the cloud. The user wishing to read a file must match the access policy and have the necessary keys to decrypt and download it. The keys for decryption are provided hierarchically to users. The important benefits of the access control system [14] are fine-grained access control in cloud computing, dynamic access policies, the ability to customize the access policy for encrypted data without duplication, less computational overhead, flexible attribute set combinations and efficient user revocation.

Ilya Sukhodolskiy, Sergey Zapechnikov[5] proposed a blockchain-based access control system for cloud storage. It provides a model to access data stored in untrusted environments i.e. cloud storage. The data, for example, multimedia information, documents, etc. will be stored in cloud storage, wherein the information

Zibin zheng[19] provided an insight into Consensus mechanism namely- Proof of work[PoW], Proof of stake[PoS] and Practical byzantine fault tolerance[PBFT] that address Technical impediments:- limited storage, privacy concern, inefficient performance, and energy consumption exist, when adding a block in an existing blockchain network. They tend to analyze and compare these protocols in numerous respects. Moreover, they listed some challenges and issues that might hinder blockchain development and summarize some existing approaches for finding these issues. A good consensus algorithm means efficiency, safety and convenience and it is observed after comparing, a PoS algorithm is well suited for Public blockchain networks while talking about energy consumption. On the other hand, the PBFT algorithm is highly adopted in the private blockchain network because

of a lack of transparency in the network. Future directions given by the paper is conducting research and developing more efficient consensus mechanisms will make a significant contribution to the development of blockchain.

3. PROPOSED SYSTEM

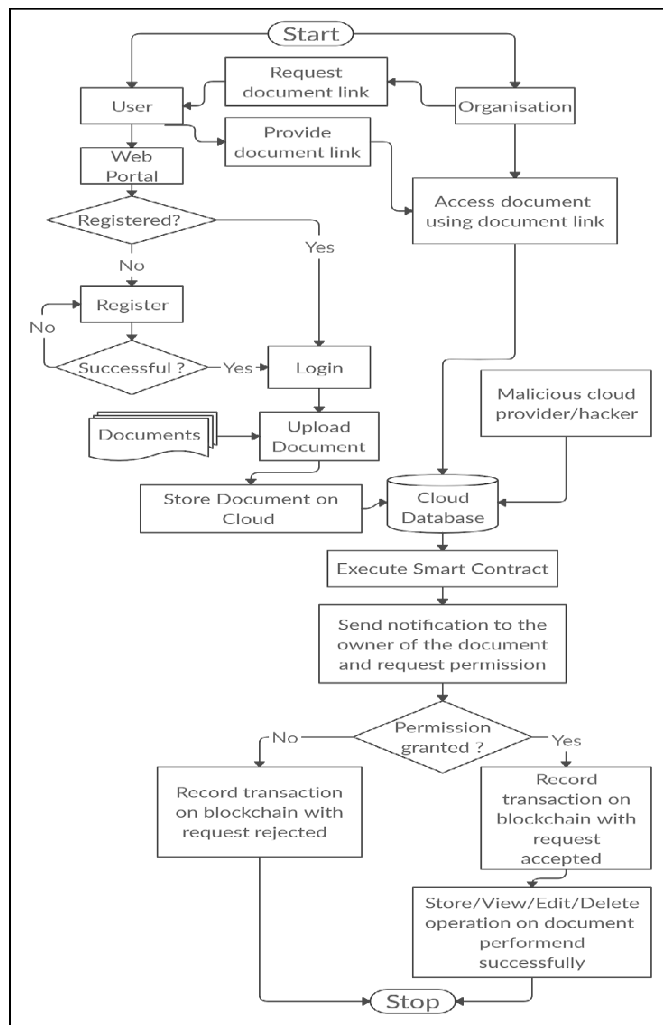


Fig. 1 Flow Chart of Proposed System

As we have discussed earlier that we can't trust the cloud provider in every case. To overcome this problem, we propose a blockchain based cloud storage and access control system. Basic idea is to store triggers in the database. Whenever any operation takes place on the document in cloud storage, a trigger will execute the smart contract. The smart contract will send the notification to the owner of the document. As soon as the owner gives the permission the further operation will be performed. At the same time, the transaction will also be recorded on the blockchain. Transaction will contain the following information: name of the requestor, operation he performed like read/ write/ update/ delete and whether permission was granted or not.

There are three stakeholders of a document : the owner himself, the third party organisation with whom he wants to share the document and the cloud provider. We consider our user as the owner of the document so we will use the words user and owner interchangeably.

First things first, we propose a decentralised app / web portal where users can register and start uploading their documents. As the document will be stored in the cloud database, it will trigger a smart contract and ask the user again to allow the create operation. Imagine a situation where a user's account is hacked and the user gives the permission through a mobile. In such a system the hacker will not be able to perform any operation without the user's permission. Even if by any chance a hacker is successful, his activity will be recorded on blockchain.

Third party organisations usually access documents through a shared link given by the owner with proper access rights. There seems no significant problem there. Still as a measure of precaution, the organisation will need permission granted from the owner. In the worst case, we could imagine a situation in which the organisation passes on the shared link to the next organisation. In that case this system is helpful because of the second step of permission grant.

Coming to our main motivation of keeping the document free from cloud provider's interference. Cloud provider can directly access the document in its database but whatever operation he tries to perform will have to pass through the second step of permission grant. Owner will get the notification. Even if somehow he is able to bypass the permission grant step. Or he is able to get the permission maliciously. It's malicious activity will be recorded on the blockchain which the user can always see through the logs. Further, the cloud provider can't modify the ledger of the blockchain.

If the user finds something malicious, it's better not to continue with such a cloud service provider.

4. SYSTEM IMPLEMENTATION TECHNOLOGIES

A. Cloud Storage

Various cloud storage services are provided by different cloud providers. These storage services provide different features.[18] The choice of choosing suitable storage service depends on various requirements of user like type of data to be stored (file, object, block, relation, etc.), pricing, storage type (persistent, temporary), caching, scalability, etc. For our system we require storage service that can support facilities like file storage, persistent storage, scalability, low pricing, etc.

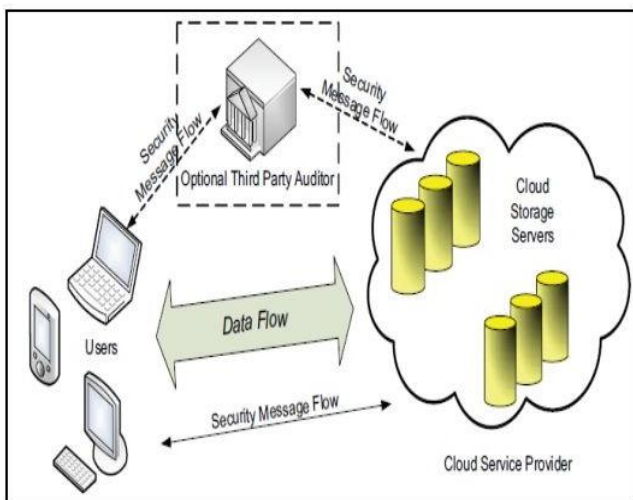


Fig. 2 Cloud Storage Architecture[7]

Google Cloud Filestore[8] is a managed file storage service for applications that need filesystem and use shared filesystem for data storage. Filestore gives users an easy, native experience for standing up managed Network Attached Storage (NAS) with their cloud virtual machine instances. Filestore is fast, simple, consistent, scalable. The availability of Filestore is 99.9%. It provides very fast file read/write throughput. According to pricing, the storage capacity can also be managed.

Amazon Simple Storage Service (Amazon S3)[9] is an object based storage service that provides features like scalability, data availability, security, and performance. Customers can use it to store and protect any amount of data for a range of use cases, such as websites, mobile applications, etc. It also provides 99.9% durability.

B. Cloud Storage Triggers

The triggers are used to respond to various events happening with data in storage. These events can be

created, updated, deleted, finalized, archived, etc. According to these events, triggers can perform various actions like sending notifications to the owner of data. The cloud providers provide services to manage these triggers and notification services.

Google Cloud Functions is a serverless, event-driven computing service within GCP. We can use it to create, implement and execute programmatic functions within the cloud, without caring about the underlying cloud infrastructure such as servers, storage and other resources. Google Cloud Functions provide features like simplified developer experience, increased developer velocity, pay only for what you use, avoid lock-in with open technology.

AWS Lambda is an event-based, serverless computing platform provided by Amazon as a part of AWS. This computing service executes code in response to events and automatically manages the computing resources according to the code. The function is invoked asynchronously by Amazon S3 along with an event that contains details about the object. Consistent performance, no need to manage server are key benefits of AWS Lambda.

C. Ethereum, Solidity and Smart Contracts

Ethereum is an open source, publicly available, blockchain-based distributed computing platform and environment for smart contract execution. Ethereum is a global, decentralized platform for money and transaction based applications. On Ethereum, we can write code that controls money, transaction and assets. It allows us to build applications accessible anywhere in the world. Ethereum was launched in 2015 and currently it is the world's leading programmable blockchain.

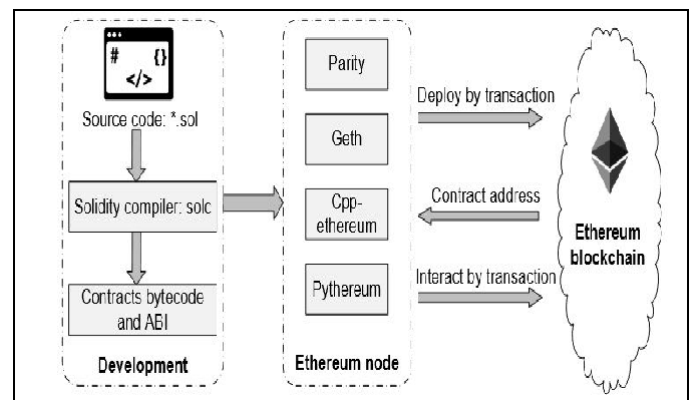


Fig. 3 The Process of smart contract's development, deployment, and interaction[10]

Solidity is a high-level, object-oriented language for implementation of smart contracts. Solidity was influenced by various features of C++, Python and

JavaScript and is designed to target the virtual machine in Ethereum platform that is EVM (Ethereum Virtual Machine). Statically typed, support for inheritance, support for libraries and support for complex user-defined types are some salient features of solidity. With Solidity we can create contracts which can be applied in various applications such as voting, crowdfunding, blind auctions, land registry, etc.

Smart contracts are programs that are stored on a blockchain and automatically execute when defined conditions are met. In simple manner, they are programs that run as they've been set up to run by the people who developed them. The smart contract is a protocol that digitally facilitates, verifies, or enforces a contract. Smart contracts allow the execution of transactions without third parties. Smart contracts can be used in various domains like financial services, healthcare, insurance, etc. Autonomy, trust, backup, safety, speed, savings, accuracy are some features that smart contracts provide.

D. DApps(Decentralized applications)



Fig. 4 DApp Stack

Decentralized applications or DApps, open up the options and services of the blockchain to the complete world. A DApp depends on the practicality of a blockchain for its infrastructure and operations. In its simplest type, a DApp contains a user interface as a front-end, and a back-end that has a blockchain and a smart contract. Consider, for example, a wallet application client, a Bitcoin blockchain decentralised infrastructure. This can be just like the design of a web browser and a web server, however with one important distinction. A DApp is usually created with a non-blockchain back-end. Design and develop end to end decentralised applications using Truffle IDE, smart contracts, a web client, and a meta mask client. Typically, a Dapp contains a web front-end, and a blockchain rear , and the code connecting the two. In such a design, the front-end of a Dapp channels any external input from the users to the blockchain infrastructure and returns any response back to them. It starts transactions

to invoke functions written in smart contracts. The front-end of a Dapp is often as easy as a command line interface. It may also be a classy web app or a simple to use mobile app. The front-end development might involve development of a web consumer with the HTML, JavaScript, and CSS's, and different internet assets.

5. COMPARATIVE ANALYSIS

TABLE II
COMPARISON OF PROPOSED SYSTEM WITH EXISTING SYSTEMS

Existing systems	Proposed system
Many theories propose the use of encrypted documents	No overhead of encryption and decryption of the large documents
Using Cryptography involves sharing various keys	No overhead of sharing keys
Faster in execution	It takes more time to execute due to extra permission step and transaction logging step
Don't use blockchain system	Extract the benefit of blockchain ledger and its tamper proof nature
Less expensive	Cost might increase due to addition of blockchain system
One step verification in most cloud systems	Two steps verification.
No logs for user	Users can see the logs on the blockchain.
Cloud providers have sole right on document storage. He might operate on it with bad intentions	Cloud provider need to take permission from user before doing any operation on document
Shared document links can be shared further. User is unaware of who is using the link	If a document link is shared with an organisation other than intended to use, it will require permission from the user.

If an account is hacked, the user loses full control.	If a user performs the verification step from another device, say, mobile phone, then even a hacker can't perform any operation on the document.
---	--

6. FUTURE WORK

Time stamping can be included i.e. links provided to users by the owner of the document will be valid for a particular time period only. As the proposed system can give permanent access to files once permission is granted, time bound access can be given a thought. Implementation of such a system is not yet present in the market. Hence this system can be implemented as a future work. Storing documents on multiple cloud providers at the backend can be given as an add on feature. Doing this can cut down cost for the user and increase competition among cloud providers. Distributing documents among multiple cloud providers prevents the concentration of power among any one. Efficiency of execution of smart contracts is something which can always be improved.

7. CONCLUSIONS

The proposed system secures the data which is stored in untrusted environments. The implementation of blockchain with cloud will be very efficient to solve various problems in cloud based data storage like data privacy, data breaches, data leakage, data loss, system vulnerabilities etc. It also gives the solution to problems like malicious cloud providers accessing the document from backend, hackers accessing the document, and uncontrolled sharing of document links. To implement the system, some of the algorithms have been selected of acceptable time complexity, functionality and efficiency. The system we propose adds value to existing cloud systems. It provides a better and efficient solution than keeping encrypted documents on the cloud. Proposed system can be incorporated in the existing systems by mere addition of triggers in the database that execute smart contracts. In addition, the proposed system also gives the user an advantage to use multiple cloud storages and watch logs of operations performed on the documents. The involvement of an untrusted third party that is a cloud storage provider will be easily controllable with the help of the proposed system.

ACKNOWLEDGEMENT

We express our sincere thanks to our research guide Prof. G.B. Sambare for his encouragement and support throughout our research, especially for the useful suggestions given and having laid down the foundation for the success of this work.

We also thank all the web communities for enriching us with their immense knowledge.

REFERENCES

- [1] Maximilian Wöhrer, Uwe Zdun, " Smart contracts: Security patterns in the ethereum ecosystem and solidity", International Workshop on Blockchain Oriented Software Engineering (IWBOSE) ,IEEE, 2018.
- [2] Qiwu Zou, Yuzhe Tang, Ju Chen, Kai Li, Charles A. Kamhoua, Kevin Kwiat, Laurent Njilla, "ChainFS: Blockchain-Secured Cloud Storage", IEEE 11th International Conference on Cloud Computing (CLOUD), 2018.
- [3] Fran Casino, Thomas K. Dasaklis, Constantinos Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues", Elsevier, 2018.
- [4] R. Gowthami Saranya, A. Kousalya, "A comparative analysis of security algorithms using cryptographic techniques in cloud computing", IEEE, 2017.
- [5] 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.
- [6] Shubham Desai, Rahul Shelke, Omkar Deshmukh, Harish Choudhary, Prof. S. S. Sambhare "Blockchain Based Secure Data Storage and Access Control System using Cloud" IEEE - ICCUBE 2019.
- [7] Gurudatt Kulkarni, Rani Waghmare, Rajnikant Palwe, Vidya Waykule, Hemant Bankar, Kundlik Koli, "Cloud Storage Architecture", IEEE 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA), 2012.
- [8] Google Cloud Platform Documentation: <https://cloud.google.com/docs>
- [9] AWS Documentation: <https://docs.aws.amazon.com/>
- [10] Xiaoqi Li, Peng Jiang, Ting Chen, Xiapu Luo, Qiaoyan Wen, "A Survey on the Security of Blockchain Systems", Beijing University China, 2018.
- [11] Rongzhi Wang, "Research on data security technology based on cloud storage", 13th Global Congress on Manufacturing and Management, GCMM, 2016.
- [12] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends", IEEE 6th International Congress on Big Data, 2017.
- [13] Julija Golosova et.al. "The Advantages and Disadvantages of Blockchain Technology", IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2018.
- [14] Mr Anup R. Nimje et.al. "Blockchain Attribute Based Encryption Techniques in Cloud Computing Security: An Overview" IJCTT Volume 4 , Issue 3-2013.

Sangsuree Vasupongayya -"Blockchain-Based Access Control Model to Preserve Privacy for Personal Health Record Systems", Research Article, 2019

[15] Naresh vurukonda, B.Thirumala Rao, -"A Study on Data Storage Security Issues in Cloud Computing", 2nd International Conference on Intelligent Computing, Communication & Convergence, (ICCC-2016), 2018.

[16] Guang Chen, Bing Xu¹, Manli Lu¹ and Nian-Shing Chen, -"Exploring blockchain technology and its potential applications", [Elsevier] 2018

[17] Jin Ho Park, -"Blockchain Security in Cloud Computing: Use Cases, Challenges.", Seoul National University of Science and Technology, SoongSil University, Korea; Research Article, 2017

[18] Zibin Zheng¹, -"An Overview of Blockchain Technology: Architecture, Consensus", National Engineering research center of China, [IEEE] 2017