IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Special Issue, 6, March 2019 E-ISSN: 2347-2693

Secure Data Storage Scheme using Blockchain for Federated cloud

Shaik.Munwar[1*], K.Ramani[2], K. Madhavi[3]

[1]Dept. of Computer Science, Govt. Degree College, Nagari. India
[2]Dept. of Information Technology Sree Vidyanikethan Engg. College Titupati, India
[3]Dept. of Computer Science &Engg. JNTUA college of Engineering. Ananthapuramu, India

Corresponding Author: Munwar.it@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7si6.5155 | Available online at: www.ijcseonline.org

Abstract—With the Internet technology development, the volume of data is growing immensely. To deal with huge volume of data, cloud storage has gained great attention from organizations and businesses because of its easy and efficient to adoption procedure. Traditional cloud storage has come to rely almost exclusively on large storage providers acting as trusted third parties to transfer and store data. Though, cloud Provider offers considerable security features, with increasing demands and usage, these centralized systems have become major targets for hacks and data breaches. This makes the data vulnerable and prone to tampering. In this paper, to address the above problems and proposed a distributed blockchain-based security scheme for storage in federated cloud, where users can divide their own files into fragments, encrypted those data fragments, and upload those encrypted fragments randomly into the federated clouds.

Keywords- Cloud storage, Security, Blockchain, Distributed Cloud computing, federated cloud.

I. INTRODUCTION

Cloud computing facilitates convenient, on-demand network access to a shared pool of computing resources like applications, storage, and infrastructure as services, which can be provisioned and discharged with trivial management effort or cloud service provider interaction [1]. Due to the notable benefits of lower admission cost, flexibility, device and location independency, scalability, easier maintenance and reliability, clouds have gain additional popularity accepted and ubiquitous [2].

Cloud storage is a type of data distribution system among servers and data centers which are able to work mutually for sharing and resource accessing by virtualization technology and provides a storage interface. Recently, cloud storage has achieved great interest from organizations and businesses because of its easy and efficient adoption. Users are moving their data to cloud, to access resources of application from anywhere, anytime and for getting the benefits such as flexibility, automatic installation of apps, disaster tolerance, software updates, and more. For key technologies, advantages, and challenges in different types of cloud storage are explained in [3]. The protection of data security [4] and users' privacy [5] is important, while storing the data in the cloud. Traditional cloud with large storage providers operated as trusted third parties to transfer and store data. Though, cloud Provider offers considerable security features, with increasing demands and usage, these centralized systems have become major targets for hacks and data breaches. This makes the data vulnerable and prone to tampering.

In traditional cloud, though the data is stored in several data centers, data is not fully distributed. The data is stored in several data centers with high density, and a massive amount which causes data leakage even if one of the data centers was broken down. For example, Verizon has collaborated with the Nice system to handle customer calls that use unsecured Amazon S3 data servers. Due to this, the bunch of six million records that keep a daily log from a subscriber called Verizon customer service, were able to be breached. Major leaks occurred when Deep Roots Analytics modified its AWS server, releasing sensitive information from 198 million Americans. Even large companies such as Home Depot, Target Corp and Anthem have been violating key data in recent years, affecting hundreds of millions of people.

These failures show that the cloud computing model for storage is not as secure as it may be because it has only one point of failure [6]. Even if encryption is used, the keys are stored with the cloud service provider. This reduces the security provided by encryption. Another problem is that the data is usually not encrypted during transmission. So the data can be captured during the transfer from a cloud to user's computer. Unfortunately, there are no effective solutions for the safety of storage in cloud.

The solution to make cloud storage faster and more secure is, using federated cloud and blockchain technology which proposed in 2008 and implemented in 2009 [7]. Federated cloud computing is the deployment and management of cloud computing services in heterogeneous external and internal clouds to meet the business needs. The customers of one cloud service can use the credentials from that service to make use of another cloud service without having the sign in separately. For decentralized storage, federated cloud is more suitable solution. In this scenario, A cluster of associated cloud providers trade their leftover resources between each other to gain a scale of market, expansion of their capabilities, and efficient use of their assets, for example, to conquer resource limitation for the duration of demands. In this model, the Service Providers (SPs) are getting their computing service utility by using the resources of either one Cloud Providers(CPs) or a combination of different cloud providers. In such a scenario, the Service Provider might be unaware of the federation and its contract is with a single cloud provider as shown in Figure 1.

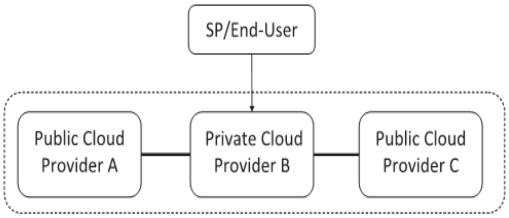


Figure1: Federated Cloud Scenario[12]

Blockchain technology can improve on current, centralized data security solutions, and help keep us safe and in control. Blockchain is a database or ledger that is shared across a network. This distributed ledger is encrypted such that only authorized parties can access the data. Since the data is shared, the records cannot be tampered [8]. Thus, the data will not be held by a single entity. The blockchain technology has the key characteristics, such as decentralization, persistency, anonymity and auditability. Blockchain can work in a decentralized environment, which is enabled by integrating several core technologies such as cryptographic hash, digital signature (based on asymmetric cryptography) and distributed consensus mechanism. With blockchain technology, a transaction can take place in a decentralized fashion. As a result, blockchain can greatly save the cost and improve the efficiency.

Blockchain architecture:

The blockchain is a sequence of blocks, which holds a complete list of transaction records like conventional public ledger [9]. Figure 2 illustrates an example of a blockchain. Each block points to the immediately previous block via a reference that is essentially a hash value of the previous block called *parent* block. It is worth noting that *uncle blocks* (children of the block's ancestors) hashes would also be stored in ethereum blockchain [10]. The first block of a blockchain is called *genesis block* which has no parent block.

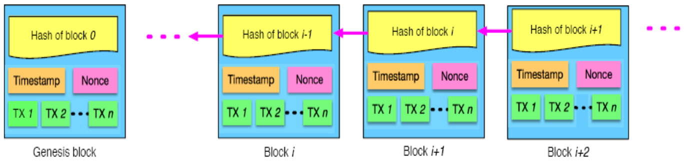


Figure 2: An example of blockchain which consists of a continuous sequence of blocks[11].

Block structure:

A block consists of the *block header* and the *block body* as shown in Figure 3. In particular, the block header includes:

Block version: indicates which set of block validation rules to follow.

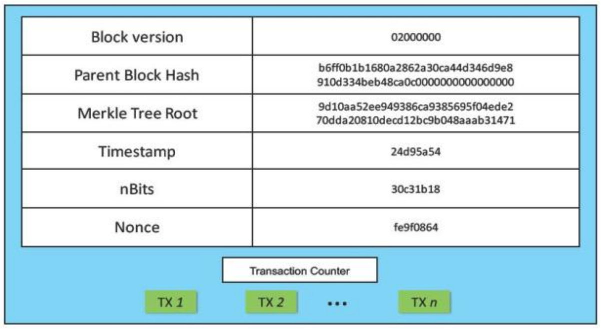
Parent block hash: a 256-bit hash value that points to the previous block.

Merkle tree root hash: the hash value of all the transactions in the block.

Timestamp: current timestamp as seconds since 1970-01-01T00:00 UTC.

nBits: current hashing target in a compact format.

Nonce: a 4-byte field, which usually starts with 0 and increases for every hash calculation.



Block version	02000000
Parent Block Hash	b6ff0b1b1680a2862a30ca44d346d9e8 910d334beb48ca0c0000000000000000
Merkle Tree Root	9d10aa52ee949386ca9385695f04ede2 70dda20810decd12bc9b048aaab31471
Timestamp	24d95a54
nBits	30c31b18
Nonce	fe9f0864

Transaction Counter
TX 1
TX 2
TX n

Figure 3: Block Structure[11].

By decentralizing data storage, we greatly improve the security of the data. Any attack or outage at a single point will not have a devastating effect because other nodes in other locations will continue to function. Instead of uploading your data on a centralized cloud, you distribute across a Federated cloud over the world. The cloud is shared, making it impossible to tamper and encrypted in a manner that only the owner can view the file. This is useful to make important records safe and decentralized.

To deal with the challenges and disadvantages as specified above, we combine distributed cloud (federated cloud) storage and blockchain technology to propose a blockchain-based distributed cloud storage architecture which can provide secure and reliable cloud storage services for individual or enterprises users.

The rest of this paper is organized as follows. In Section 2, we discuss related work of cloud storage and blockchain technology. Then we propose a novel blockchain based distributed cloud storage architecture in Section 3. We conclude this paper in Section 4.

II. RELATED WORKS

Cloud storage is a kind of Internet technology for sharing resources with IT-related capabilities. Traditional security policies mainly focus on data duplication, data encryption, access control, privacy preserving keyword search, network performance improvement and etc. in recent times, application data are increasing drastically hense a separate cloud cannot meet the storage demands of users. To deal with the situation mentioned above,

Zyskind et al. propose architecture with blockchain technology to protect information of enterprise and personal data through distributed storing file access permissions in the blockchain, but still its data storage uses a centralized cloud and to support a trusted third-party is required [13].

When one cloud provider is not in a position to meet the needs of a client, it can transfer the same requests to other cloud providers. In this case, the irregularity in devices and hardware composition among the cloud providers must be handled [14]. This heterogeneity and the distribution of cloud storage services can be managed by the Software Defined Storage (SDS)[15]. The SDS made this possible by software defined hardware [16] and software fragmentation methods [17] which help in aggregation of upper storage space and scheduling. Inspired by the SDS, in our scheme the storage strategy takes other cloud providers fixed storage space on rent basis and provide it to who needs storage space. From the cloud provider's point of view, the marginal cost of cloud resources is increased prominent because of the demands to preserve a large number of servers and services. From another perspective, if we put the another cloud providers vacant storage space as cloud storage space by forming cloud federation, the cloud storage infrastructure costs will be decreased in a great extent. However, there are some critical security issues in this federated cloud storage.

In [18] the authors proposed a solution for cloud storage with block chain. But, it considers security of data in the individual cloud but not the whole system. The authors of [19] also proposed a blockchain based solution with the name Storj for Peer-to-Peer cloud storage. Which implements end-to-end encryption which permits to share and transfer data without a depending on a third party data provider. Unlike [19], in this paper, block chain based solution for storage in federated cloud is proposed by aiming security of users data and security of the whole system.

To optimize the federated cloud storage, service transmission time scheduling optimization is similar to optimized resource scheduling. To optimize resource scheduling, authors of [20] use a genetic algorithm between user's requirements and applications. Along this in cloud storage replication is one of the important data reliability techniques.

As per the above survey, it is observed that security on cloud storage cannot be extended directly without using a third party though they use blockchain-based algorithms for secure storage. However, there are some efforts earlier which considers security on architecture level of federated cloud, for instance, in paper [21] the authors propose a blockchain-based system with private keyword search for secure data storage, in paper [22] proposed data integrity checking scheme based on a blockchain and remote checking scheme for cloud storage, and in paper [23], they proposed a publicly verifiable data deletion scheme for cloud storage using block chain. Thus, in this paper, new blockchain-based scheme for federated cloud storage is proposed to improve the security and integrity of data in cloud storage system.

III. ARCHITECTURE DESIGN

In this section, we present blockchain-based security scheme for federated cloud storage. In this scheme, we first divide users' files into several fragments of same size, encrypt these file fragments, then sign on them through a Digital Signature Algorithm (DSA) and upload them to a Peer-to-peer federated cloud. Then we use the blockchain technology between clients who need cloud storage service and cloud providers who provide their unused available storage space. Furthermore, we choose a random file replica placement strategy in this architecture so that users can retrieve their files quickly from the cloud and alleviate the burden of the P2P federated cloud. Finally, file integrity verification will be ensured by using the Merkle Hash Tree as a validation method.

```
Algorithm of operation()
{
    Step1: Divide the user file into several fragments of
           same size.
    Step 2:Encrypt these fragments using public key
           algorithm.
    Step 3:Sign them using hash algorithms.
    Step 4:Upload them to federated cloud and apply the
           blockchain algorithm.
    Step 5: Use the file replica strategy to retrieve the files
           quickly.
    Step 6:Verify the file integrity using Markle hash table.
}
```

A. Architecture Overview

a. Files are fragmented, encrypted and uploaded to Peer-to-Peer federated cloud

Considering the maximum transfer unit of a network, clients' files need to be fragmented and encrypted before uploading them to federated cloud. Actually, almost all clients' files are split into blocks of the same size, as per the convenience of network protocol for transmitting data fragments. We consider a particular client file CF consists of n data fragments:

$$CF = \{CF_1, CF_2, \dots , CF_n\}$$

For security, files should be encrypted before they are uploaded to the cloud to provide security so that client's data will not be retrieved by man in middle attacks . In the proposed scheme, cloud providers use RSA algorithm to generate a public-private key pair (K_{opu},K_{opr}) to encrypt and decrypt their file fragments without any third party key generation center.

$E_{opr}(CF)=\{E_{Kopr}(CF_1), E_{Kopr}(CF_2), {}_{Kopr}(CF_3) \dots E_{Kopr}(CF_n).\}$

Besides, to generate signature, a key pair (K_{opu},K_{opr}) using SHA 256 will also be generated by a digital signature algorithm. *Signature*=$H_{SHA}(E_{Kopr}(F_1))$

b. Using blockchain as a mechanism to provide security

A blockchain is a system in which records of time-ordered collection of transactions which are linked using cryptographic algorithms where information is appear as y distributed database. In this technology, files cannot be stored in the blockchain directly.

The proposed scheme, to reduce the storage space, instead of storing the entire files we store file fragment hashes, their location URLs, replicated fragmented file location URLs and etc. as per the blockchain technology It's obvious that each cloud in federation has a copy of all transactions and compared with size of the clouds hard disk the transaction information size is negligible so that the proposed scheme can reduce a enormous amount of memory space for clients. In this scheme, an opponent cannot get anything about the clients' information and data from the blockchain, as only URLs and hash values are stored in it.

c.. *File storage policy and file replicas replacement*

There is an additional problem of managing the data which is stored in various servers of federated cloud. Availability and optimized accessing are the key challenges. To cop up with these problems, in order to get better gain from file replication, an efficient method to store for file replicas is needed. We maintain the file replicas across different clouds. Unlike traditional cloud storage scheme, our federated cloud storage architecture stores file fragments in various servers of Peer-to-Peer federated cloud randomly. Our scheme achieves the fault tolerance mechanism by using file replicas as data redundancy. A peer-to-peer file sharing system consists of peer nodes of federated cloud share their unused resources cooperatively to improve the services offered to the clients. File replicas will be stored in the federated cloud randomly, and their URLs after being encrypted will be stored in blockchain so that clients can know and get their own file entirely. The number of file replicas is determined by the performance of network which is influenced by the file replicas placement strategy and the number of file fragment replicas.

d. File integrity verification

Verification of file integrity can be done by the Merkle Hash Tree. In which the hash of the data fragment is available at leaf nodes, and non leaf nodes are labeled with the hashed labels of their child nodes. As shown in Fig. 4, Merkle Hash Tree (MHT) is constructed to check the file integrity the calculation results are based a one-way cryptographic hash operation like SHA-256 [11]. Besides, SHA-256^2 represents two times encryption of SHA-256.

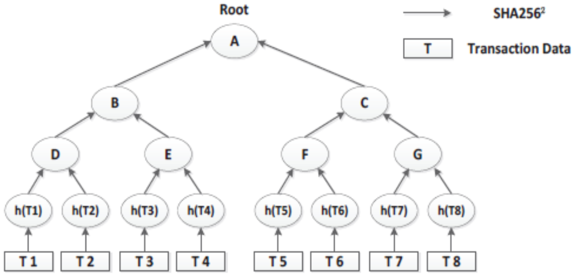


Figure 4: Structure of Markle Hash Tree

A MHT is first constructed by pairing the data, next hashing the pairs, then pairing and hashing the results until a single hash remains, the Merkle Root. In the tree, each leaf node having the information can be verified through its corresponding path. We can now identify whether the file fragment in the leaf nodes of MHT are tampered or not by comparing their Merkle Root.

e. Proof-of-retrievability

Suppose a node (server of a federated cloud) P possesses a file F, the Proof-of-retrievability assures that the node P retrieves the file correctly or not. It takes the form of a challenge-response protocol. To audit correctness of *P*'s possession of file *F*, a random challenge *c* is received by *P* at regular interval basis and produced a response *r*, which it can be publicly verified without having *F*. A basic proof-of-retrievability scheme consists of three protocols:

Setup(*F*) → *{digest}*: P calculate s a Merkle tree whose leaves are fragments of file *F* and whose root is a complete *digest*. *P* outputs *digest* value of whole file.

Prove(*R*) → $\{F_{ri\,,}\ \pi_i)\ r_i \in R\}$: $R = r_1, \dots, r_k \in [n]$ denotes a set of random challenge received by node *P*. *P* outputs a result as proof that for each challenge index *ri* in R, *F* contains F_{ri} and the accompanying path π_{ri} in the Merkle tree.

Verify (*digest*, *R*, F_{ri} , π_i) $r_i \in R$) → {0, 1}. The validation process verifies the Merkle path π_{ri} for each segment F_{ri} against the *digest*.

IV. CONCLUSIONS

This paper has proposed security scheme for data storage in federated cloud based on blockchain. Using of peer to peer federated cloud, download speeds can be boosted. As the data is distributed globally making it highly available. As the data is shared and encrypted which makes data as highly secured. The immutable nature of the block chain makes data accurate and unaltered. But we need to address the problems of network communication overhead, computational overhead, and consider the problems of federated cloud like, SLA and policy negotiation in using block chain.

REFERENCES

[1] National Institute of Standards and Technology special publication.no.800-145, “The NIST definition of cloud computing” Sept. 2011

[2] Shan, Chen, Chang Heng, and Zou Xianjun. “ Inter-cloud operations via NGSON” IEEE communications, 2012.

[3] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li. Enabling public auditability and data dynamics for storage security in cloud computing. IEEE transactions on parallel and distributed systems, 22(5):847–859, 2011.

[4] Yinghui Zhang, Dong Zheng, and Robert H Deng. Security and privacy in smart health: Efficient policy-hiding attribute-based access control. IEEE Internet of Things Journal, 2018. doi:10.1109/JIOT.2018.2825289.

[5] Yinghui Zhang, Xiaofeng Chen, Jin Li, Duncan S Wong, Hui Li, and Ilsun You. Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing. Information Sciences, 379:42–61, 2017.

[6] http://techgenix.com/blockchain-technology-for-cloud-storage/

[7] Nakamoto, S. “Bitcoin: A Peer-to-Peer Electronic Cash Systems https://bitcoin.org/bitcoin.pdf namecoin (2014).

[8] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https: //bitcoin.org/bitcoin.pdf, 2008.

[9] Kosba, A., Miller, A., Shi, E., Wen, Z. and Papamanthou, C. (2016) ‘Hawk: the blockchain model of cryptography and privacy-preserving smart contracts’, Proceedings of IEEE Symposium onSecurity and Privacy (SP), San Jose, CA, USA, pp.839–858.

[10] Buterin, V. (2014) A Next-Generation Smart Contract and Decentralized Application Platform, White Paper.

[11] Zibin Zheng and Shaoan Xie,”Blockchain challeng es and oppor tunities: a survey” Int. J. Web and Grid Services, Vol. 14, No. 4, 2018.

[12] Adel Nadjaran Toosi, Rodrigo N. Calheiros, and Rajkumar Buyya. 2014. Interconnected cl oud computing environments: Challenges, taxonomy, and s urvey. ACM Comput. S urv. 47, 1, Articl e 7 (April 2014), 47 pages. DOI: h ttp://dx.doi.org/10.1145/2593512.

[13] Guy Zyskind, Oz Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In Security and Privacy Workshops (SPW), 2015 IEEE, pages 180–184. IEEE, 2015.

[14] Antonio Celesti, Francesco Tusa, Massimo Villari, and Antonio Puliafito. 2010a. How to enhance cloud architectures to enable cross-federation. In Proceedings of the 3rd International Conference on Cloud Computing (Cloud’10). Miami, FL, 337–345.

[15] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick PC Lee, and Wenjing Lou. A hybrid cloud approach for secure authorized deduplication. IEEE Transactions on Parallel and Distributed Systems, 26(5):1206–1216, 2015.

[16] Xiaoming Zhu, Bingying Song, Yingzi Ni, Yifan Ren, and Rui Li. Software defined anything from software-defined hardware to software defined anything. In Business Trends in the Digital Era, pages 83–103. Springer, 2016.

[17] Qiang Fu, J¨org-Uwe Pott, Feng Shen, and Changhui Rao. Stochastic parallel gradient descent optimization based on decoupling of the software and hardware. Optics Communications, 310:138–149, 2014.

[18] Guy Zyskind, Oz Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In Security and Privacy Workshops (SPW), EEE, pages 180–184. IEEE, 2015

[19] Shawn Wilkinson, Tome Boshevski, Josh Brandoff, and Vitalik Buterin. Storj a peer-to-peer cloud storage network, 2014.

[20] Lizhen Cui, Junhua Zhang, Lingxi Yue, Yuliang Shi, Hui Li, and Dong Yuan. A genetic algorithm based data replica placement strategy for sci- entific applications in clouds. IEEE Transactions on Services Computing, pages 1–13, 2017. doi:10.1109/TSC.2015.2481421.

[21] Liu, Xiao Liang Yu, Shiping Chen, Xiwei Xu, and Liming Zhu. Blockchain based data integrity service framework for iot data. In Web Services (ICWS), 2017 IEEE International Conference on, pages 468–475. IEEE, 2017.

[22] Yong Yu, Man Ho Au, Giuseppe Ateniese, Xinyi Huang, Willy Susilo, Yuanshun Dai, and Geyong Min. Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage. IEEE Transactions on Information Forensics and Security, 12(4):767–778,2017.

[23] Changsong Yang, Xiaofeng Chen, and Yang Xiang. Blcokchain based publicly verifiable data deletion scheme for cloud storage.Journal of Network and Computer Applications, 103:185–193,2017.doi:10.1016/j.jnca.2017.11.011.