IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue, 6, March 2019 E-ISSN: 2347-2693

Secure Data Storage Scheme Using Blockchain in Federated Cloud

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

[1,2]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/v7si6.139143 | Available online at: www.ijcseonline.org

***Abstract*—**With the development of Internet technology, the volume of data is growing immensely. To deal with large-scale data, cloud storage has gained great attention from organizations and businesses because of its easy and efficient to adoption. 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 we proposed a blockchain-based security scheme for distributed cloud storage, where users can divide their own files into encrypted data chunks, and upload those data chunks randomly into the federated clouds.

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

I. INTRODUCTION

Cloud computing facilitates convenient, on-demand network access to a shared pool of configurable computing resources such as applications ,services, servers, and networks, which can be provisioned and released with negligible management effort or cloud service provider interaction [1]. Due to the noteworthy 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 of servers and data centers able to work together for sharing and resource accessing by virtualization technology and provides a storage interface. Recently, cloud storage has gained great attention from organizations and businesses because of its easy and efficient to adoption. To access application resources from anywhere, anytime the users are moving their data to cloud for getting the benefits such as flexibility, automatic installation of apps, disaster tolerance, spending cuts, software updates, and more. For advantages, challenges and key technologies in different types of cloud storage one can refer to [3]. It's important to protect data security [4] and users' privacy [5] when users store their data in the cloud. 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.

For the current distributed cloud storage, the data stored in several data centers are not fully distributed. The data are still stored in several data centers at high density, and a massive amount of data will be leaked even if one of the data centers was broken down. For example, Verizon partnered with Nice Systems to handle customer service calls who utilized an unprotected Amazon S3 storage server. Because of this, 6 million records that held logs from customers who called Verizon customer service were able to be accessed. Another much more massive leak occurred when Deep Roots Analytics misconfigured their AWS server, releasing sensitive information of 198 million Americans. Even giant companies like Anthem, Target Corp, and Home Depot have had major data breaches over the last few years affecting hundreds of millions of people.

These failures aren't a one-time occurrence and they show that cloud computing model of centralized storage isn't as secure as it could be because it has a single 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. The data can hence be intercepted during transmission from the user's computer to the cloud. Unfortunately, there are still no effective solutions for the security of cloud storage.

The solution to make cloud storage faster and more secure is using federated cloud computing and blockchain which proposed in 2008 and implemented in 2009 [7].

Federated cloud computing is the deployment and management of multiple external and internal cloud computing services to match the business needs. The customers of one cloud service can use the credentials from that service to make use of the another cloud service without having the to sign in separately. For decentralized storage, federated cloud is more suitable solution. In this scenario, A group of cloud providers are federated and trade their surplus resources among each other to gain economies of scale, efficient use of their assets, and expansion of their capabilities [14], for example, to overcome resource limitation during spike in demands. In this model, the computing utility service is delivered to Service Providers(SPs) using resources of either one Cloud Providers(CPs) or a combination of different cloud providers. In such a scenario, the SP might be unaware of the federation and its contract is with a single cloud provider (Figure 1).

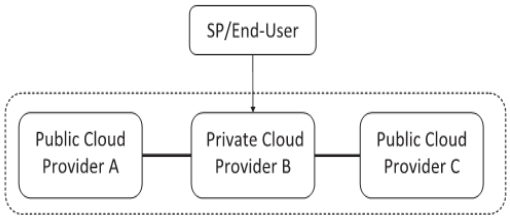

Figure1: Federated Cloud Scenario[12]

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.

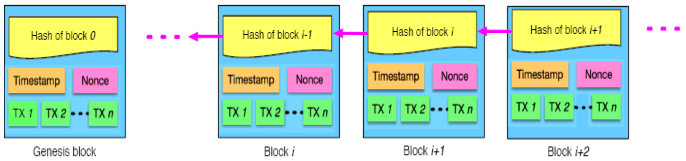

Figure 2: An example of blockchain which consists of a continuous sequence of blocks[11].

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.

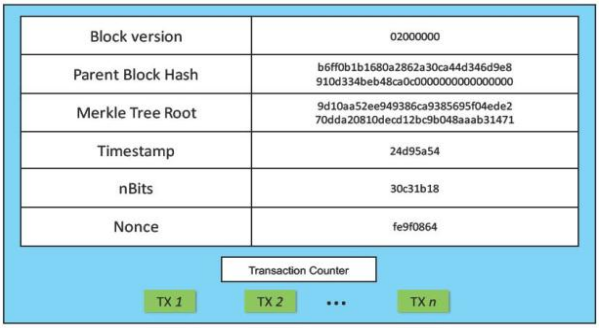

Figure 3: Block Structure[11].

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 disadvantages and challenges as mentioned above, we integrate distributed cloud (federated cloud) storage and blockchain technology to propose a blockchain-based distributed cloud storage architecture that can provide secure and reliable cloud storage services for enterprises or individual 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 and it is important to either enterprises or individual users. Traditional security strategies mainly focus on information encryption, data deduplication , access control, privacy preserving keyword search, network performance improvement and etc. Recently, application data are becoming more and more intensive and a separate cloud cannot meet the storage demands of users. To deal with the situation mentioned above, Zyskind et al. propose architecture and their architecture uses blockchain to protect personal data through distributed storing file access permissions in the blockchain, but its data storage still uses a centralized cloud and requires a trusted third-party to support [13].

The Software Defined Storage (SDS) integrates a number of distributed cloud storage services [15]. When a cloud cannot meet the demands of users, their requests can be transferred to other cloud platforms. Compared with traditional cloud storage, the heterogeneity among cloud service providers such as different device types, hardware composition and etc. can be properly handled by SDS. This difference can be shielded by software-defined hardware [16] and software decoupling methods [17] to provide technical support for the aggregation of the upper storage resources and the unified scheduling platform. Inspired by SDS, the storage strategy in our architecture is a random storage strategy which takes users' fixed vacant storage space as the cloud storage space and then rents it to other users who need storage space. From the cloud service provider's perspective, the marginal cost of cloud resources is also increasingly prominent because of the demands to maintain a large number of servers and services. From another point of view, if we put the users' vacant storage space as cloud storage space, the cloud storage infrastructure costs will be greatly reduced. However, there are still many critical security issues in cloud storage.

In the aspect of storage security, authors of [18] have proposed a blockchain based solution for cloud storage. However, it only considers individual secure cloud storage rather than the security of the whole system. Before deploying blockchain technology in cloud storage, it must satisfy a condition that honest nodes constitute of at least half the computational power in the network. Meanwhile, the authors of [19] have also proposed a blockchain based P2P cloud storage network named Storj which implements end-to-end encryption allowing users to transfer and share data without a reliance on a third party data provider. Unlike [19], this paper focuses on the design of blockchain based architecture for distributed cloud storage, where both the security of users' data and the security of the architecture are considered.

Optimizing distributed cloud storage service transmission time resembles the resource scheduling optimization problem, and authors in [2] use a genetic algorithm to optimize the data resource scheduling between the scientific application and task of users' requirements. Besides, in cloud storage architecture, replication is one of the significant data reliability techniques. Furthermore, in order to improve cloud storage security in architecture, we combine the distributed cloud storage architecture with blockchain technology.

As mentioned above, the works on cloud storage security cannot be directly extended to solve the blockchain-based secure storage problem without a third party. However, some earlier works are also rarely considering on an architecture level, for examples, authors of [20] propose a blockchain-based system with private keyword search for secure data storage, authors of [21] respectively propose a blockchain-based data integrity checking framework and remote checking scheme for cloud storage, and authors of [22] propose a blockchain-based publicly verifiable data deletion scheme for cloud storage. Moreover, the security of cloud storage architecture is not the simple overlay of multiple specific security technologies. Thus, this paper studies a new blockchain-based security architecture design for distributed cloud storage to improve the security of the distributed cloud storage system.

III. ARCHITECTURE DESIGN

In this section, we present blockchain-based security architecture for distributed cloud storage. In this architecture, we first divide users' files into several blocks with the same size, encrypt these file blocks, sign them through a Digital Signature Algorithm (DSA) and upload them to a P2P federated cloud. Then we utilize blockchain technology as a trading mechanism between users who need cloud storage service and users who supply their vacant 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 blocks of same size.
Step 2:Encrypt these blocks 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.

}

3.1. Architecture Overview

3.1.1. Files are chunked, encrypted and uploaded to P2P federated cloud

Considering the network performance, users' files need to be chunked and encrypted before they are uploaded to the federated cloud. Actually, almost all users' files are split into blocks of the same size, which is limited by network protocol and is convenient for transmitting data packages. we consider a particular file F consists of n data segments:

$$F = \{F_1, F_2, \ldots, F_n\}$$

For security, files should be encrypted before they are uploaded to the cloud so that users' information will not be retrieved. In the proposed architecture, users run an elliptic curve cryptography (ECC) based key generation algorithm curve to generate a public-private key pair (K_{opu},K_{opr}) to encrypt and decrypt their files without any key generation center (KGC) or third party.

$E_{opr}(F)=\{ E_{Kopr}(F_1), E_{Kopr}(F_2), E_{Kopr}(F_3) \ldots E_{Kopr}(F_n).\}$

Besides, a signature key pair (K_{opu},K_{opr}) will also be generated by a digital signature algorithm named SHA256.

$Signature=H_{SHA}(E_{Kopr}(F_1))$

3.1.2. Use blockchain as a trading mechanism

A blockchain storing a time-ordered collection of widely accepted transactions, is an append-only distributed database. The blockchain technology enters peoples' sight after the success of Bitcoin proposed by Nakamoto. In Bitcoin, transaction size attracts many concerns and files cannot be stored in the blockchain directly.

The proposed architecture store file hashes, file location URLs (Uniform Resource Locator), file replicas location URLs and etc. instead of file blocks themselves in the blockchain to reduce the storage space. It's noticeable that each cloud has a copy of all transactions in the blockchain and the size of the transaction information is negligible to the clouds's hard disk so that the architecture can reduce a massive amount of memory space for users. In this architecture, an adversary cannot get anything about the raw users' file data from the blockchain, as only URLs and hash values are stored in it.

3.1.3. File storage strategy and file replicas replacement

Compared with traditional cloud storage architecture, our distributed cloud storage architecture stores file blocks to nodes in a P2P federated cloud randomly. Because a fault tolerance mechanism is necessary for every intelligent system, our architecture achieves the fault tolerance mechanism by using file replicas as data redundancy. Replication is the processes of sharing information and ensuring consistency among redundant resources. It improves the readability, fault tolerance or availability. A peer-to-peer file sharing system consists of peer nodes that want to share their resources and store files cooperatively to improve the services offered to the users. File replicas will be stored in the federated cloud randomly, and

their URLs will be stored in blockchain after being encrypted so that users can know and get their own file completely. The number of file replicas is determined by the network performance influencing by the file replicas placement strategy and the number of file blocks replicas.

3.1.4. File integrity verification

As shown in Fig. 4, Merkle Hash Tree (MHT) is constructed by calculation results based a one-way cryptographic hash operation like SHA256 [11]. Besides,

$SHA256^2$ is two times SHA256 encryptions.

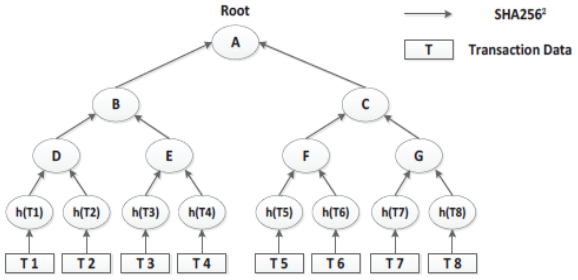

Figure 4:Structure of Markle Hash Tree

A MHT is first constructed by pairing data (e.g. in the Bitcoin system it usually refers to transactions), next hashing the pairs, then pairing and hashing the results until a single hash remains, the Merkle Root. In the tree, each leaf node containing information can be verified through its corresponding path. We can know whether the file data

blocks' information in the MHT's leaf nodes are tampered or not by comparing their Merkle Root.

3.15. Proof-of-retrievability

A basic Proof-of-irretrievability takes the form of a challenge –response protocol in which a node *P* demonstrates its possession of a file *F* and the fact that it can be correctly retrieved. To audit *P*'s possession of *F*, *P* receives a random challenge *c* at regular basis; it produces a response *r*, which it can be publicly verified without possessing *F*. A basic proof-of-retrievability scheme consists of three protocols:

Setup$(F) \rightarrow \{digest\}$. P computes a Merkle tree whose leaves are segments of the file *F* (with their indices) and whose root is *digest*. *P* outputs *digest* value.

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

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

IV. CONCLUSIONS

This paper has proposed blockchain-based security architecture for distributed cloud storage. 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] 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.

[21] 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.

[22] 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.

