Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 10, October -2017

# Review on Distributed Cloud Storage through Peer-To-Peer Network System

Mr. Tushar Gonawala, Mr. Rudra Patel, Prof. Vatsal Shah

*Information Technology, Birla Vishvakarma Mahavidyalaya, V. V. Nagar*

*Information Technology, Birla Vishvakarma Mahavidyalaya, V. V. Nagar*

*Information Technology, Birla Vishvakarma Mahavidyalaya, V. V. Nagar*

**Abstract-***A Distributed Cloud Storage would provide peer-to-peer storage system and would be more secure and robust compared to the normal centralized cloud storage. The data is stored in form of shards on the peers that can be later on retrieved by the client. The network takes care of security and privacy from malicious users through the use of hash challenge. Redundany methods are used for data retrival in cases when the storage miner is absent.*

***Keywords-*** *Bitcoin Blockchain; Distributed Cloud; Peer-to-Peer Network; Decentralized Storage; File Sharding; Encryption; Merkle Tree; Hashing*

## I. WHAT IS DISTRIBUTED CLOUD STORAGE

Distributed Cloud Storage is a term coined for a cumilative storage offered to clients by multiple individual peers who are independent storage providers for data storage and retrival. It is a decentralized storage system. For such a system to exist, no dedicated server or trusted party is required. Protocols and some verified coordinate operations are responsible for the functioning of such kind of storage network.

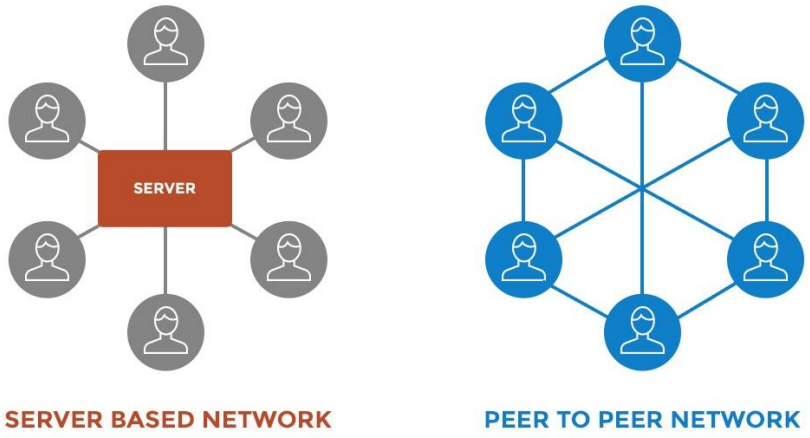


***Fig.1 Vizualization of the Server Based Network and the Peer To Peer Network[4]***

### 1.1. Participants

There are mainly 3 users in this type of storage netwok, viz. Client, Storage Miner and Retrival Miner.

**Client:** is the one who pays to store and retrival data to the Distributed Cloud Storage Network.

**Storage Miner:** is the one who provides storge space on the network. Storage Miner has to take responsibility that their system would be committed to store client's data for a specific period of time and would be available for retrival as and when required by the client. The Storage Miner is penalized for invalid or missing data set.

**Retrival Miner:** is the one who helps retrive data from the network. Unlike Storage Miner, Retrival Miner are not required to have committed data storage, but can obtain pieces of data from the client or the Retival Market.

### 1.2. Network

Network is a single entity that acts as an intermediate to run and manage protocols which are required for runing the cloud storage. Manages the information related to available storage, block size, validate the participants and audits based on their participation.

 

### 1.3. Ledger

The protocol runs on a Ledger based currency, which in general is refered as ledger-L. This ledger along with time is used to sequence the transaction for the network. The Ledger is an append only system that manages the verification of participants.

### 1.4. Market

There are total two markets for decentralized exchange of data i.e. for data storage and data retrival. Clients and miners set the prices for the services they request and provide with respect to the market.

## II. FILE AS ENCRYPTED SHARDS

Block is defined as an encrypted portion of file that would be stored on the shared network defined as distributed cloud. Splitting of the file that is also known as sharding, help for better security and performance in the cloud system. No masquerader would be able to get a complete copy of the file but would rather get a shard of a file in an encrypted form. The shard size is fixed make it an ambiguity on what is being stored in these shardes. Encryption and sharding of files makes large files more managable through distributed cloud storage.

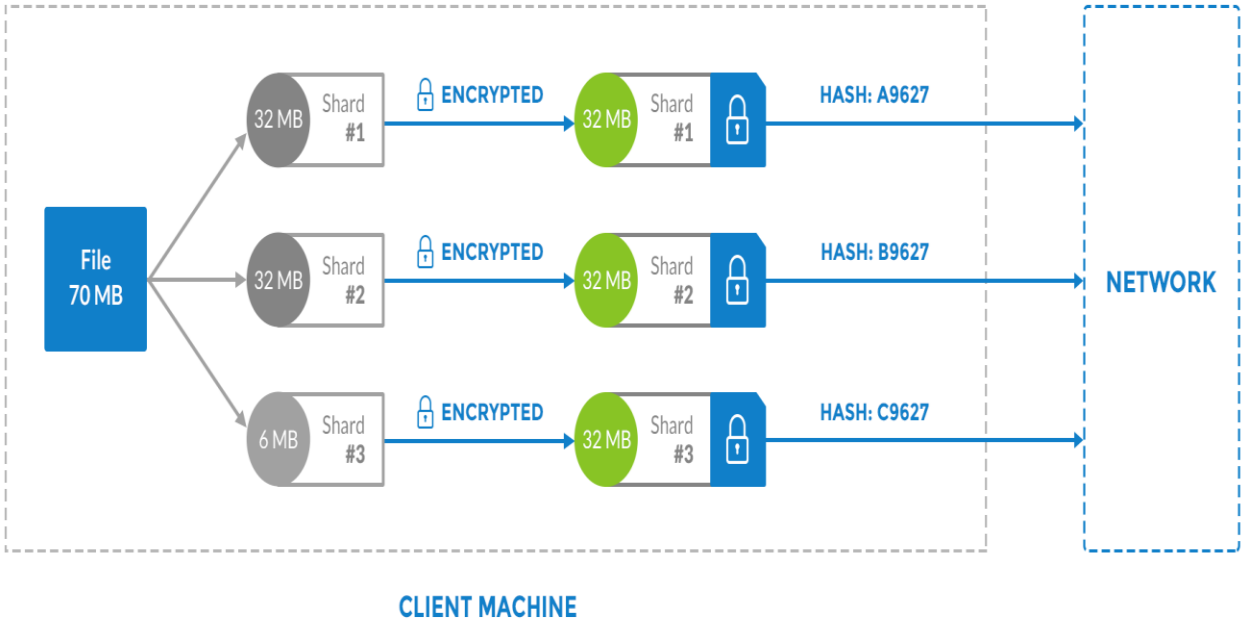


***Fig 2. Visualizing the shardingprocess[4]***

a. A file is split into shards based on the fixed file size. If at all extra space is left then zero is filled.
b. Each shard is encrypted with an external encryption key that serves as an input to the encryption algorithm defined by the user.
c. The shard once created would be directly transmitted to the network.

## III. PROOF OF STORAGE

The main concern for the client would be the integrity and the availability of data in form of shards that are stored on the network. It is mandatory client has to audit that the shards are not modified and are cryptographically correct. This is done through the MerkleTree[7] which is inserted in the blockchain that is passed into the network.

 

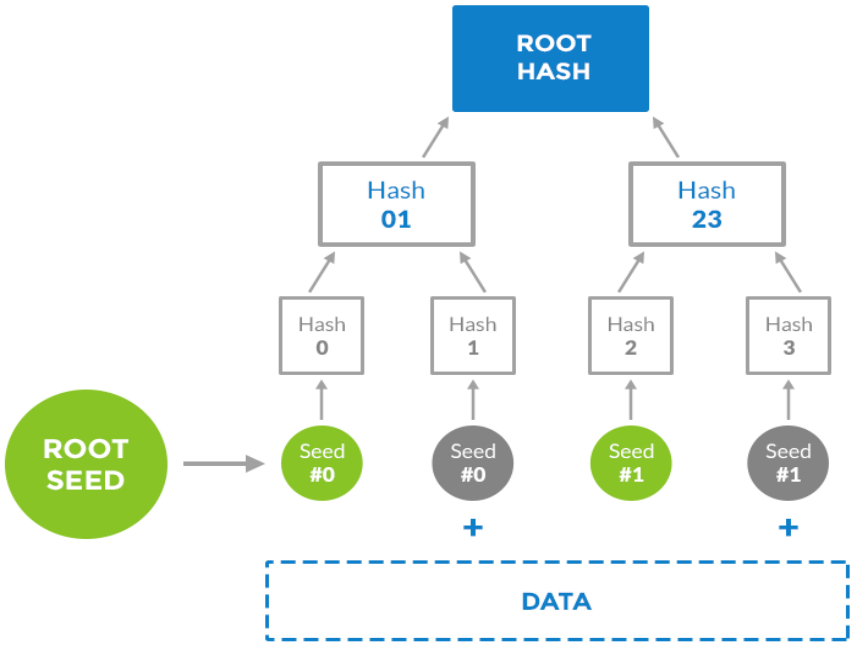


***Fig. 3 Making the Merkle Tree from the data[4]***

Method that is recommended is through hash challenge, where the client creates a series of seeds which is then added to the file and hashed to generate a unique hash answer. [4] This process is refered as heartbeat.

If at all the file is modified the resultant hash answer would be different from the one that client generated and henceforth the shard would fail the hash challenge. The hash response is verified by the MerkleTree[7]. The checking of hash answers or hash challenges done through brute force lays an overhead of calculating time ultimately reduces the efficiency of the system. As a result, 3 techniques of generating heartbeats are developed.

**3.1. Full Heartbeat**

Full shard file is used to generate hash response. Using this process of hash generation ensures integrity up to extreme level but is in-efficient and consumes more time.

**3.2. Cycle Heartbeat**

Herein the shard is divided into n discrete parts and turn by turn each and every part is verified in a cyclic manner. This system is more Input-Output efficient but the major disadvantage is that the system is open to potential risk from attackers with only an integrity of 1/n.

**3.3. Deterministic Heartbeat**

Encoding and selection of the n divided parts is done in such a way that even a minor change in the shard will be detected and would fail the hash challenge. This is most efficient as there is a balance between integrity and efficiency.

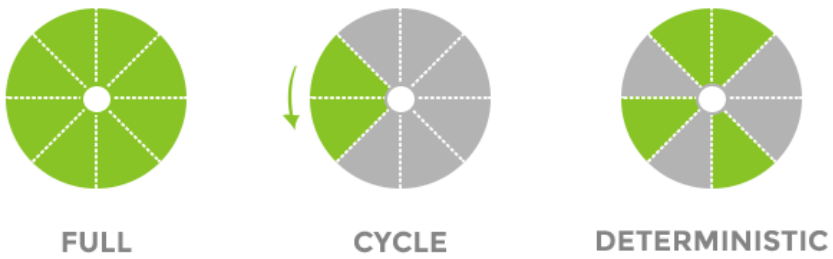


***Fig. 4 Different types of Heartbeats[4]***

 

## IV. PROOF OF REDUNDANCY

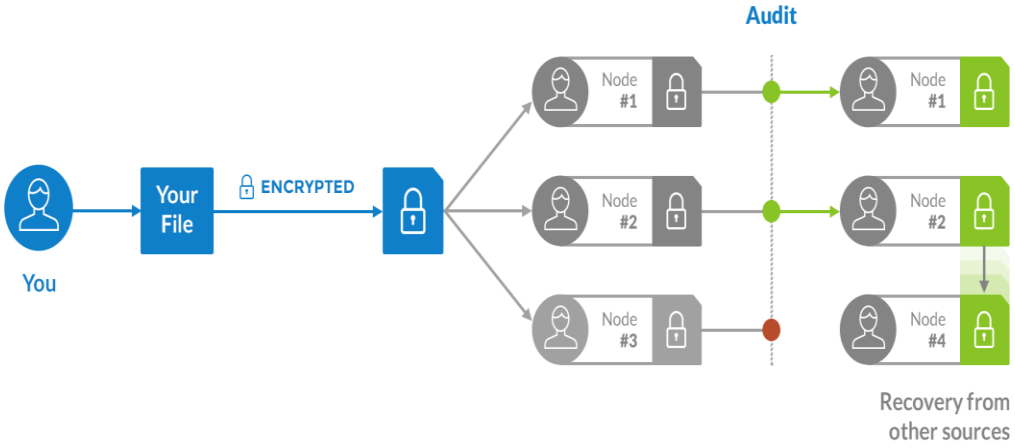


***Fig. 5 Auditing the redundancy of the shards on different peers[4]***

In a normal centralized cloud storage system, data redundancy is handled through RAID method or storing it at multiple datacenters. But in distributed cloud storage as there is no centralized dedicated server for handling the stuff, shards are stored using K-of-M erasure encoding scheme for multiple miners. The redundancy is maintained on the network layer rather than on the physical layer. The main aim for keeping redundancy of shards is to ensure non-disruptive data access even if the storage miner is out of reach or has turned off the system.

In K-of-M erasure encoding, the client has to select the values of K and M in such a way that a right balance is created between robustness of shards and the cost of storing the redundant data. A normal data is distributed in to 3-4 miners whereas a highly important data is spread out amongst as many as 500 miners so that the data shards become robust but would ultimately increase the cost of the system.

To solve the problem of a malicious miner sending the wrong data to the client, the system encrypts the shards using a unique cryptographic key. Even the redundant shard would have a unique key. Erasure encoding makes sure that no malicious miner transfers data to the client. So, if at all a malicious miner gets access to the key, he won't be able to pass the hash challenge.

## V. BLOCKCHAIN

For a network to achieve integrity and robustness of data the best methodology is using the Satoshi-Style Blockchain[1]. Being a public ledger, it ensures accurate retrieval of data and would be more secure from attacker. The data of the file is not stored in the block chain, rather the metadata is stored. The metadata includes information regarding the file hash, location where the data is present and the Merkle Root. The data is inserted in to the blockchain via a standard transaction as extra metadata. Doing this in Bitcoin Blockchain is expensive and not feasible, Florinachain [2] would be the best mechanism for such type of transcation.

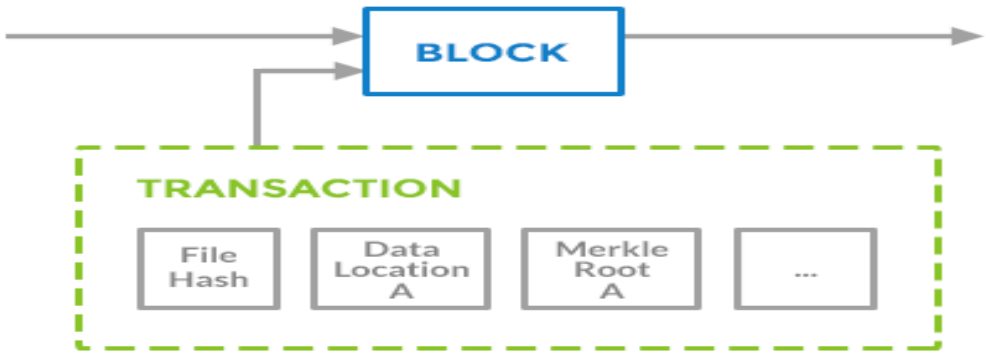


*Fig. 6 Transaction in a Blockchain[4]*

 

## VI. CONCLUSION

It is possible to eliminate a third party centralized storage for successful working of cloud storage. Peer-to-peer distributed decentralized cloud storage is the best example of network based storage. The problems occurring while storing and retrieving the shards from various peers are solved using various redundancy and storage methods. Moreover, the end-to-end encryption done on each and every shard (even the redundant shard) makes it nearly impossible for malicious miner or client to get access to it. Blockchain transaction with a change in the metadata and extra metadata makes it more secure as well as speeds up the process of verification.

## REFERENCES

[1] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, (2009):https://bitcoin.org/bitcoin.pdf

[2] Florincoin, 2014: http://florincoin.org/florincoin.pdf

[3] White Paper: https://www.sia.tech/whitepaper.pdf

[4] Storj, 2014: https://storj.io/storj2014.pdf

[5] Filecoin: https://filecoin.io/filecoin.pdf

[6] Enigma: https://www.enigma.co/enigma_full.pdf

[7] R.C. Merkle. Protocols for public key cryptosystems, (April 1980). lIn Proc. 1980 Symposium on Security and Privacy, IEEE Computer Society, pages 122-133.

[8] Gregory Maxwell, Proof of Storage to make distributed resource consumption costly: https://bitcointalk.org/index.php?topic=310323.0

[9] HovavShacham, Brent Waters, Compact Proofs of Retrievability, Proc. of Asiacrypt 2008, vol. 5350, Dec2008, pp. 90-107

 