Proof of ownership in deduplicated cloud storage

[1] Rupali Bhimrao Sirsat [2]Nitin Talhar
[1]Student of computer department, [2]Assistant Professor
[1]AISSMS, COE, Pune-411001, Maharashtra, India.

Abstract – **Now-a-days, cloud computing provide a lot of storage space and magnificent parallel computing at effective cost. It provide many kind of services to the users. The main service is tendered by the cloud computing is storage amenity. Cloud computing becomes more popular and large amount of data being stored in the cloud. Due to this large amount of data one critical problem must be faced by users i.e. the management of ever increasing volume of data. In this large amount of data containing the replica of data is increases. Replication take place additional space in the cloud as a storage. To make data management scalable and mitigate the increased amount of data in the cloud data deduplication can be used. Data deduplication is a data compression technique which is used to improve storage utilization and efficiency of cloud storage. In this paper, we are going to use two cloud server, public cloud server and private cloud server. Public cloud server store data (hash, encrypted data) and compare hash which is generated by private cloud. After comparison hash is stored in the data base by private cloud. This proposed system used to mitigate the problem of data deduplication. Based on this we can improve storage utilization and efficiency of cloud storage.**

***Index Terms*-Deduplication, encryption, proof of ownership, cloud storage**

I. INTRODUCTION

An important requirement now-a-days for cloud storage service is to improve storage utilization. Especially in today's era where the redundancy of data increases rapidly, it is a challenge to mitigate useless and repeated data generated by many users. Data deduplication is a technique used for elimination of duplicate copies of data. It has been widely used in cloud storage to mitigate storage space and upload bandwidth. Today's commercial cloud storage services, such as Drop box, Mozy, Bitcasa and Memopal have been applying deduplication to user data to save maintenance cost[4].

There is some remarkable feature of cloud storage can be identified. It has high availability, high flexibility, and infinite virtual storage space. High availability which means user's data duplicated over cloud and whenever user want to access that data what they need it take guaranteed to provide data to the user. High flexibility it means it has high flexibility in as-per-pay. It means that the user can obtain additional storage immediately whenever user is willing to make an extra payment. Virtual infinite storage is most important feature of cloud storage, which means that users can backup whatever he/she wants to be uploaded to the cloud. A renowned example of Bitcasa which tender "unlimited storage" that enables the user to upload virtually everything.

Tendering a infinite storage space might increase economic burden on the cloud storage provider. The data deduplication technique can help to mitigate the cost storage. Data deduplication is gained by avoiding storing the same file multiple times.

There are two types of data deduplication depending on where the deduplication occurs.

1. Server side data deduplication.

After receiving the file server check first whether it already present in the storage. If file is present then server discard the file and if it is not then it create new file in the storage. We can see that server perform deduplication after receiving the file because it does not bandwidth saving.

2. Client-side data deduplication:

This side of deduplication adopt more combative method. It takes a calculation make hash of file and send hash of file before it uploading. After receiving hash it check in storage whether hash is already stored. User asked to send nothing and user associates the user with the existing file if it exist in storage. Otherwise user asked to upload the file.

3. Proof of ownership:

The intellection of proof of ownership (PoW) is to solve the problem of using a small hash value as a proxy for the entire file in client-side deduplication [5], where the antagonist could use the storage services as a content distributed network. This proof of mechanism in PoW provides a solution to protect the security in client-side deduplication. In this way, client can prove to the server that it actually has a file.

II. RELATED WORK

Cloud computing provide a lot of storage space and magnificent parallel computing at effective cost. It provides much kind of services to the users. The main service is tendered by the cloud computing is storage amenity. Nowadays cloud computing becomes more popular and large amount of data being stored in the cloud.

In existing system there is only one server and client. Simple communication is happened in between them. When client wants to upload a file it send request to cloud and it send acknowledge if file is present or not. If file is present then it send acknowledgment but if file is not present then it save the file. When second user want to save file on cloud which contain same data as user first but it contain quite additional data as compared to user one, in this case it save file on cloud as it is because of this it required more storage.

To solve this problem we can use to server in proposed system. Then second is security problem, To solve the new security problems in client-side deduplication, we propose a cryptographically secure and efficient scheme, called a provable ownership of the file (POF), in which a client proves to the server that it actually possesses the entire file without uploading the file. We provide rigorous security proof and extensive performance analysis.

Convergent key: Convergent encryption is a widely used technique to combine the storage saving of de-duplication to enforce confidentiality. In convergent encryption, the data copy is encrypted under a key derived by hashing the data itself. This convergent key is used for encrypt and decrypt a data copy [6]. We can make our data confidential through convergent encryption key.

Client side deduplication have some new security problems [10] [3]. When server tells to client that it does not have to send a file. It means that, some other client already have same file that already store on server cloud and containing secret information. To solve this problem chao yang, jian ren propose a cryptographically secured and efficient scheme call proof of ownership [10]. The proof of ownership is also introduced by halevi [7]. According to him it is challenge –response enabling protocol. Work assign to this protocol is to check whether requesting entity is data owner, based on short value. It means that when user want to upload a data file to cloud it first compute and send the hash value to storage server[8].

Recently, Ng et al. [9] propose a PoW scheme over encrypted data. That is, the file is divided into fixed-size blocks, where each block has a unique commitment. The hash-tree proof is then built, using the data commitments. Hence, the owner has to prove the ownership of a data chunk of a precise commitment, with no need to disclose any secret information. However, this scheme introduces a high computation cost, as requiring generation of all commitments, in every challenging proof request.

Existing System:

In existing system there is only one server and client. Simple communication is happened in between them. When client wants to upload a file it send request to cloud and it send acknowledge if file is present or not. if file is present then it send acknowledgment but if file is not present then it save the file. When second user want to save file on cloud which contain same data as user first but it contain quite additional data as compared to user one, in this case it save file on cloud as it is. Due to this it required more storage. And it is a main reason of facing data deduplication. In existing system it has only one cloud server for storing data. If any case data is corrupt or server has been crashed then there is possibility of losing data. In this case if you loss your confidential data then there is no backup or recovery option to get the data.

III. PROPOSED SYSTEM

- Problem Statement:
  To avoid deduplication of data and provide security to the data and make storage more efficient to the client.

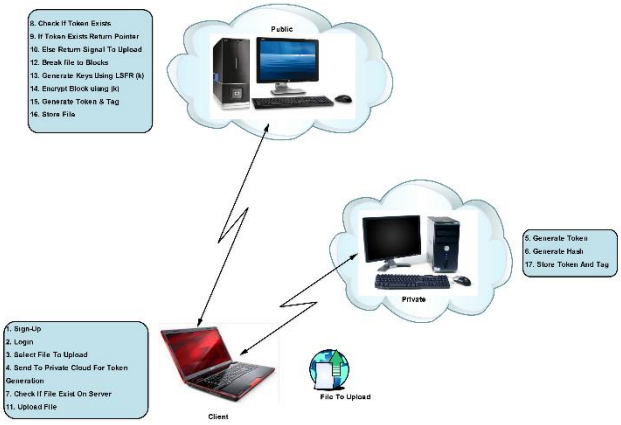

Figure 1: Proposed System Architecture

- Proposed Architecture:
  To overcome the disadvantage of existing system we will be using additional server for maintaining confidential data. It also provide backup in case of data corruption. The proposed architecture consists of three modules client, public cloud, and private cloud. In this paper we will be using basically 3 modules i.e. client, public cloud, server cloud.

  - Client :
    This module used for authorization that is performing operation of registration process. It is also used for file uploading. And then send to the private cloud for token generation.

  - Private cloud: user can used private cloud despite of public cloud for more security. Generated key is stored in private cloud by user. When user wants to download the file system ask key at the time of downloading.

To provide proper protection for the key we use private cloud. Its store only convergent key with respective file. When user want to access the file then first it check the authority of user and then provide key.

- Public cloud: it is used for storing data. First client upload files on public cloud. When client want to download that file then it will be ask the key which is stored in private cloud. When client key is match with files key then client can download data from public cloud. Without key client cannot access the file from public cloud. In public cloud all data is stored in encrypted format. In case unauthorized person hack file, but without convergent key hacker doesn't access original file.

IV. PROPOSED ARCHITECTURE IMPLEMENTATION

When client upload their data on cloud at that time he thinks that his data will be safe. But when he knows that their data is not secure from hacker i.e. unauthorised person also can access their data very easily then it's a big problem in front of client. This proposed system will provide security to the confidential data of client. In the implementation of this system we are going to used three modules client, private cloud, and public cloud. Normally, when client want to upload file on cloud it will be directly uploaded on cloud. By implementing this proposed this when client want to upload the file on cloud it must go through registration process. In registration process client get some unique id for login. After that client select file for uploading and then send to the private cloud for token generation. After generating token at private cloud it also generate hash of the file. Main use of this cloud is to store encrypted key of data that will used at the time of accessing data.

Public cloud is used for storage purpose. Client upload file on this cloud. It first checks whether the file on server. If it does then it return pointer otherwise is send signal to upload file. For avoid replicas of data it break file into block and assign block id. After breaking file into block it generate key by using LFSR technique for avoid duplication and provide protection from unauthorized user. It generate token and tag and then store file on public cloud in the form of block or chunk. And generated token and tag is stored on private cloud. When client need to access the data public cloud then it ask for key which is generated or stored in private cloud. When client key is matched with private key then client will be able to access data from public cloud. If client key and key in private cloud is not match then client cannot access the data from public cloud. Here, there is no need to remember key to the client. It is already stored in private cloud in encrypted format. So by implementing this system we can provide security to the data and also provide efficiency to the storage by performing block operation and avoiding deduplication in the file storage.

V. CONCLUSION

In this proposed system, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in deduplicate check. Our data are securely store in public cloud in encrypted format, and our key is store in private cloud with respected file. There is no need to user remember the key. It is already store in private cloud so without key anyone cannot access data from public cloud. This proposed system provide more efficiency and security by using authorized deduplicate check. Result is it is helpful to improve storage efficiency and performance of storage cloud.

REFERENCES

[1] Chia-Mu Yu, Chi-Yuan Chen, and Han-Chieh Chao, "proof of ownership in deduplicate cloud storage with mobile device efficience" in 2015 IEEE Network • March/April 2015 0890-8044/15/$25.00 © 2015 IEEE

[2] R. Di Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ASIACCS '12, pages 81–82, New York, NY, USA, 2012. ACM.

[3] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ASIA CCS '13, pages 195–206, New York, NY, USA, 2013. ACM.

[4] J. Mirkovic, H. Bryhmi and C.M. Ruland, "A framework for the development of Ubiquitous patient support system", in pervasive health 2012, May 21-24, San Diego, United States, Pg. 81-88.

[5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, ''Proofs of Ownership in Remote Storage Systems,'' in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds., 2011, pp. 491-500.

[6] Anu George, Mr. sandeep hegade "authenticated deduplication system with access control and security measures" in international journal of computer science and information technology research, vol. 3, issue 2, pp: (301-305), month: april-june 2015.

[7] S. Halevi et al., "Proofs of Ownership in Remote Storage Systems," Proc. ACM Conf. Computer and Communications (CCS), 2011.

[8] Nesrine Kaaniche, Maryline Laurent" A Secure Client Side Deduplication Scheme in Cloud Storage Environments", IEEE Transactions on Mobility and Security (NTMS) in Cloud Computing,Issue Date:April.2.2014

[9] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12, pages 441–446, New York, NY, USA, 2012. ACM.

[10] Chao Yang, Jian Ren and Jianfeng Ma "Provable Ownership of File in De-duplication Cloud Storage", Globecom 2013 - Communication and Information System Security Symposium.