A Secure Authorized Deduplication Scheme in Cloud Storage Environment

Arpitha M.S

Department of Computer Science & Engineering

Abstract— The traditional system has witnessed the trend of influencing cloud-based services for a large amount of data storage, processing, and distribution. Security and privacy are among top concerns for the public cloud environments. For these security challenges, it introduces a new client-side deduplication scheme for securely storing and sharing uploaded data via the public cloud. The main purpose of this proposed system has two-part. First, it provides data confidentiality towards unauthorized users. That is, every member computes a key per data to encrypt the original data that should be stored in a cloud. As such, the file data access is managed by the owner of the data. Second, by indicating access rights in the private server, an authorized user can decrypt an encrypted file data only with his private key.

Key words: Cloud Storage, Data Security, Deduplication, Confidentiality, Proof of Ownership

I. Introduction

The various growths of digital contents raise the demand for more cost-effective storage and network bandwidth for data transfer, along with new storage and network capacities.

In a multi-tenant environment, this system supports the transmission, storage, and an intensive computation of data transmitted in makes use of a different model. To prevent resource consumption in storage capacities and network, apply client-side deduplication for many cloud services such as Drop box and Memopal. It avoids the storage of duplicate data in the cloud and decrease more network bandwidth usage while transmitting the same data contents several times.

According to these advantages in saving resources, client-side data deduplication provides many security issues, due to the multi-owner data performs the confidentiality or the bandwidth consumption and the privacy of cloud storage. Consider one example, a user checks whether a file already uploaded by another user when trying to upload the same data file to the cloud server.

For this reason, different security policies have been proposed by many efforts. These policies are called Proof of Ownership protocol (PoW). It allows the storage server check an ownership of the user data, based on a static value. Several requirements are provided by these protocols, namely lightweight of verification and computation efficiency.

The new cryptography method has been introduced for secure Proof of Ownership scheme(PoW), based on the combination of convergent encryption and the hash code generation, it improves data security in cloud storage systems, between two users, provides dynamic sharing and efficient data deduplication is ensured. They have the idea in using the hash code generation, is used to derive a unique identifier of uploaded data. Using this identifier checks the availability of the same data present in public cloud servers and also obtains efficient access control in multiple sharing scenarios.

II. Methodology

The data storage in cloud server is very important concept because cloud storage was widely used by many organizations. Most of the IT companies are using cloud storage to store the large amount of data files. But many organizations suffering from lack of data confidentiality, duplicate data stored in cloud server and using the large amount of cloud storage space, consume more bandwidth while uploading large data copies.

For this reasons client-side data deduplication has been introduced, it is an important data compression technique for removing redundant data copies of repeating data. It works in client side instead of checking duplication in server side.

In the proposed storage system with client side deduplication, the first member sends hash code with respect to data content to the server and it checks if that hash value already exists in its database. If the hash value not presents in the database server ask for an entire data file to store. Otherwise, since the file already exists at the server, it tells the data owner about no need to send the data file it.

Deduplication takes place at two levels one is file level and another one is block level. For the file-level deduplication, it removes duplicate copies of a same data file. For the block-level deduplication, it removes a duplicate block of data that occur in non-identical files. This application takes place at a file level.

This system consisting of a group of registered clients (for example, employees of a company), who will use the public cloud and store data with deduplication technique. In this system, while greatly reducing storage space, it uses deduplication that can be regularly used for data backup and failure recovery applications. Such systems are well-known and are often more appropriate to user file backup and organization applications than richer storage space abstractions. There are three entities defined in this system, that is, users/member, private server and CSP in the public cloud as shown in Fig.1. The CSP in the public cloud is used to check if the contents of two data are the same and stores only one data of them.

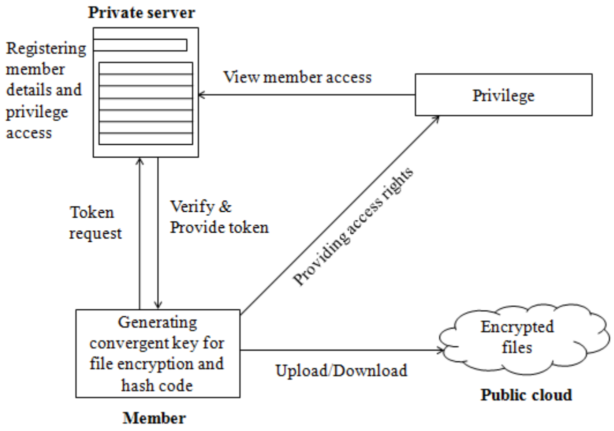

Fig.1: Architecture of Authorized Deduplication

- Private Cloud. In the cloud computing it compared with the traditional deduplication architecture, the private server is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the public cloud is not fully trusted in practice and the computing resources at data user/owner side are restricted, data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud is provided by private server. The private cloud manages the private keys for the privileges, who answer the file token requests from the users. The interface offered by the private server allows the user to submit files and queries to be securely stored and computed respectively.
- CSP. The cloud service provider is an entity in the public cloud that provides data storage service, the data outsourcing service and stores data with respect to the users is provided by the CSP in the public cloud. The CSP eliminates the storage of redundant data via deduplication and keeps only unique data to reduce storage cost. In this paper, data owner assumes that CSP in the public cloud is always online and has huge storage capacity and computation power.

AWS offers a broad set of services that help us move faster, lower the costs, and scale the applications. AWS provides a variety of computing and networking services to meet the needs of applications.

A client using access key and a secret key which are provided by AWS console to a management console.

- Data Users. A user/member is an entity that wants to upload data storage to the CSP and access the data later. In the deduplication storage system, the user only uploads unique data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized deduplication system, in the setup of the system each user is issued a set of privileges. Each file is protected with the convergent encryption key and privilege keys to understand the authorized deduplication with differential privileges.

III. RESULT

In order to evaluate the performances at the client side, we first conduct data encryption and decryption tests locally. Second, we evaluated the time consumption of random data upload in the cloud, or data download from remote servers.

1) Client-Side Computation: To evaluate client side computation costs, we conduct data symmetric encryption and decryption, using a symmetric cryptographic algorithm. Fig. 2 shows the computation overhead of data encryption and decryption at client side, with different sizes of data contents.
2) Upload/Download: In this section, we evaluated the time consumption of random data upload in the cloud, or data download from remote servers. This approach consists at first, to generate a random data file of a fixed size. Then, we encrypt the data file based on the AES algorithm. The length of the used enciphering key is 256 bits (AES-256-CBC). Afterwards, we upload the encrypted file and we download it from the cloud.
3) Client Access Control Integration: In order to include the security procedures at the client side, we first choose an asymmetric algorithm based on the use of RSA algorithm. This application performs asymmetric encryption and decryption using the implementation of the RSA algorithm provided by the cryptographic service provider (CSP).

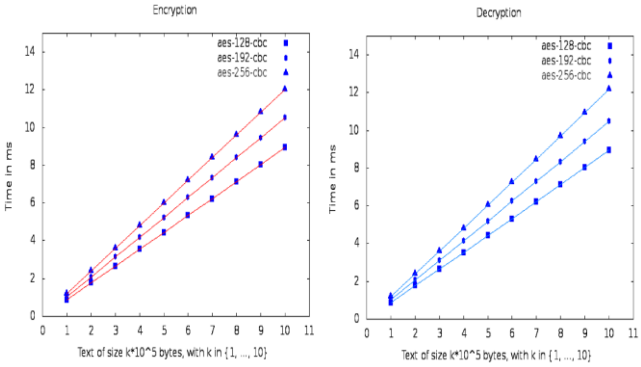

Fig. 2: Computation overhead of data encryption and decryption at the client side with different data size.

IV. CONCLUSION

The organization depends on cloud storage services for the huge amount of data storage space. It will provide secure data storage for secure authentication. Here using the properties of convergent cryptography to provide the innovative solution to the uploaded data security and privacy.

This application introduce cryptographic usage of symmetric encryption and asymmetric encryption, due to several attacks it destroys the highest sensibility of this information. These algorithms are used to generate and share the key with privileged user. In addition, thanks to the hash code generation properties, it will support data deduplication, as it employs a pre-verification of data reality, in cloud servers, which is helpful for saving bandwidth. Besides, this solution is also used to avoid the unauthorized access to data and the two levels of access control verification is provided during the sharing process, it avoids leakage of data.

Finally, many organizations facing the challenges that are in the cloud data storage security and giving first importance, and can identify many research problems are remaining.

REFERENCES

[1] NesrineKaaniche, Maryline Laurent Institut Mines-Telecom, Telecom SudParis, "A Secure Client Side Deduplication Scheme in Cloud Storage Environments".6th International Conference On New Technologies, Mobility And Security Year 2014

[2] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman Peleg."Proofs of ownership in remote storage systems". In Proceedings of the 18th ACM conference on Computer and Communications security, CCS'11, pages 491–500, New York, NY, USA, 2011. ACM.

[3] D. Harnik, B. Pinkas, and A. Shulman-Peleg. "Side channels in cloud services: Deduplication in cloud storage". IEEE Security And Privacy, 8(6):40–47, 2010.

[4] W. K. Ng, Y. Wen, and H. Zhu. "Private data deduplication protocols in cloud storage". In Proceedings of the 27th AnnualACM Symposium on

Applied Computing, SAC ’12, pages 441–446, New York, NY, USA, 2012. ACM.

[5] 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.