A Survey on Hybrid Cloud Approach for Secure Authorized Deduplication

Dhokne Vaishali[1], Dr. Varsha Patil[2]

[1] *ME Student, Department of Computer Engineering, Mcoerc Nasik, Maharashtra, India*
2 HOD, Department of Computer Technology, Mcoerc Nasik, Maharashtra, India

ABSTRACT

Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data,and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself.We also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

Keyword : *Deduplication, authorized duplicate check, confidentiality, hybrid cloud*

1. INTRODUCTION

In Cloud computing large number groups of remote servers are connected through networked to allow centralized data storage and provide online access to computer services or resources . With cloud computing, number of resources can be connected through private or public network. In public cloud, services (i.e. applications and storage) are available for general use over the internet. A private cloud is a virtualized data center provide data storage services .Maximum amount of data stored in the cloud and shared by multiple user with their specific rights ,which define an access rights on stored data.

One of the major problem of cloud storage services is the management of increasing large volume of data. To reduce storage space in cloud computing, deduplication has been a well known technique which is being used by most of the users. Data Deduplication is specialized data compression techniques which is used to eliminate duplicate copies of repeating data. Another issue in cloud storage is security and privacy. That is any one can access data that are store on to the cloud . For security and privacy concern the encryption technique is used .Before uploading data on to the cloud user must encrypt data file with their own private key. But the same data copies of different user will produced different cipher texts, making deduplication impossible. To provide security for sensitive data with deduplication technique, the convergent encryption technique has been used to encrypt the information before outsourcing the data .

2. LITERATURE SURVEY

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller [1]developed models for secure deduplicated storage. These system model demonstrate that security can be combined with deduplication for removing duplicate copies of data and security is provided through the use of convergent encryption. In this technique number of user encrypt data with their convergent key that is encrypt with same ciphertext.

M. Bellare, S. Keelveedhi, and T. Ristenpartis[2] introduced a new cryptographic primitive, Message-Locked

Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure de-duplication usind deduplication technique. a goal currently targeted by numerous cloud-storage providers. They provide tag consistency for privacy and data integrity

J. Yuan and S. Yu. [3] Provide deduplication system in the cloud storage to reduce the storage space of the tags for integrity check. To increase the security of data deduplication and provide data confidentiality

M. Bellare, S. Keelveedhi, and T. Ristenpart.[4] It protect the data confidentiality by transforming the predicatable message into unpredicatable message. In this system, there is third party called key server is introduced to generate the file tag for duplicate check

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg.[5] Provide "proofs of ownership" (PoW) for deduplication Systems. Before uploading or downloading file client can efficiently prove to the cloud storage server that he/she has owns a file.

J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer[6] Convergent encryption provides data security and confidentiality in deduplication. A user derives a convergent key from original data copy and encrypts the data copy with the convergent key. User generate file tag for removing duplicate copies of data.

3. System Architecture

As shown in figure there are four modules are as follows

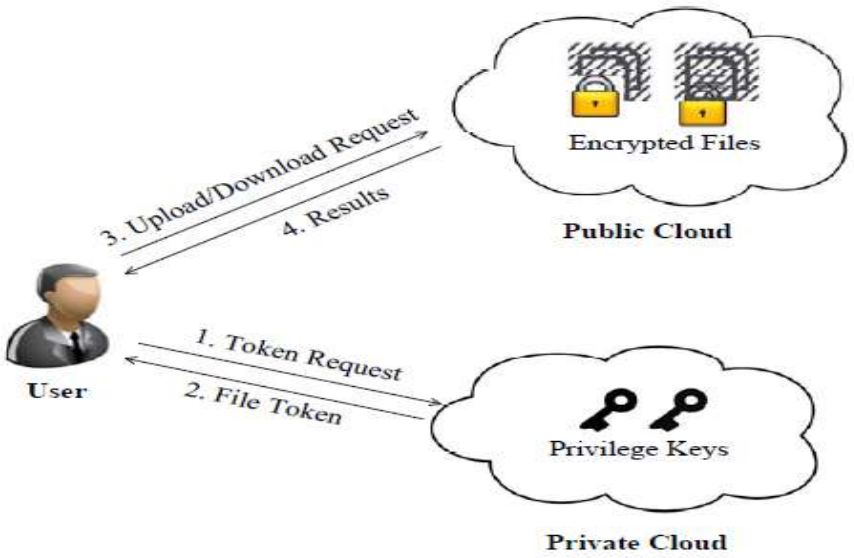

Fig -1:System Architecture

A) Cloud Service Provider(S-CSP)

The purpose of this entity to work as a data storage service in public cloud.On the half of the user S-CSP store the data.The S-CSP eliminate the duplicate data using deduplication and keep the unique data as it is.S-SCP entity is used to reduce the storage cost.S-CSP has abundant storage capacity and computational power.When user send respective token for accessing his file from public cloud S-CSP matches this token with internally if it matched then an then only he send the file or ciphertext with token, otherwise he send abort signal to user.After receiving file user use convergent key to decrypt the file.

B) Data Users

A user is an entity that want to access the data or files from S-SCP. User generate the key and store that key in private cloud.In storage system supporting deduplication, the user only upload unique data but do not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. Each file is

protected by convergent encryption key and can access by only authorized person. In our system user must need to register in private cloud for storing token with respective file which are store on public cloud. When he want to access that file he access respective token from private cloud and then access his files from public cloud.

C) Private Cloud

In general for providing more security user can use the private cloud instead of public cloud. User store the generated key in private cloud. At the time of downloading system ask the key to download the file. User can not store the secrete key internally .For providing proper protection to key we use private cloud. Private cloud only store the convergent key with respective file. When user want to access the key he first check authority of user then an then provide key.

D) Public Cloud

Public cloud entity is used for the storage purpose. User upload the files in public cloud. Public cloud is similar as S-CSP. When the user want to download the files from public cloud, it will be ask the key which is generated or stored in private cloud. When the users key is match with files key at that time user can download the file, without key user can not access the file. Only authorized user can access the file. In public cloud all files are stored in encrypted format. If any chance unauthorized person hack our file, but without the secrete or convergent key he doesnt access original file. On public cloud there are lots of files are store each user access its respective file if its token matches with S-CSP server token.

4. CONCLUSIONS

In this paper I conclude that authorized data deduplication technique is used for eliminating duplicate copies of data that are store onto the server .Private cloud used for storing different privileges related to file (secrete key, convergent key, proof of ownership), public cloud used for storing the files only. That means hybrid cloud used for storing data and avoid duplicate copies data that are store onto the server and reduce network bandwidth

5. REFERENCES

1] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In *Proc. of StorageSS*, 2008.

2] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296–312, 2013.

3] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. *IACR Cryptology ePrint Archive*, 2013:149, 2013.

4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.

5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500. ACM, 2011

6] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In *ICDCS*, pages 617–624, 2002.