SECURE DATA DEDUPLICATION WITH DYNAMIC OWNERSHIP MANAGEMENT IN CLOUD STORAGE

Mohammed Parveez[1], Srinidhi H R[2]

[1] *Student, Information Science, NIE Institute of Technology, Karnataka, India*
[2] *Assistant Professor, Information Science, NIE Institute of Technology, Karnataka, India*

ABSTRACT

In cloud storage services, deduplication technology is commonly used to reduce the space and bandwidth requirements of services by eliminating redundant data and storing only a single copy of them. Deduplication is most effective when multiple users outsource the same data to the cloud storage, but it raises issues relating to security and ownership. Proof of- ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a robust way. However, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers deduplication because of the randomization property of encryption. Recently, several deduplication schemes have been proposed to solve this problem by allowing each owner to share the same encryption key for the same data. However, most of the schemes suffer from security flaws, since they do not consider the dynamic changes in the ownership of outsourced data that occur frequently in a practical cloud storage service. A novel server-side deduplication scheme is proposed for encrypted data. It allows the cloud server to control access to outsourced data even when the ownership changes dynamically by exploiting randomized convergent encryption and secure ownership group key distribution. This prevents data leakage not only to revoked users even though they previously owned that data, but also to an honest-but-curious cloud storage server. In addition, the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme.

Keywords: - *Deduplication , cloud storage, proof of- ownership, encryption*

I. INTRODUCTION

Cloud computing provides scalable, low-cost, and location-independent online services ranging from simple backup services to cloud storage infrastructures. The fast growth of data volumes stored in the cloud storage has led to an increased demand for techniques for saving disk space and network bandwidth. To reduce resource consumption, many cloud storage services, such as Dropbox, Wuala, Mozy, and Google Drive, employ a deduplication technique, where the cloud server stores only a single copy of redundant data and provides links to the copy instead of storing other actual copies of that data, regardless of how many clients ask to store the data. The savings are significant, and reportedly, business applications can achieve disk and bandwidth savings of more than 90%. However, from a security perspective, the shared usage of users' data raises a new challenge. As customers are concerned about their private data, they may encrypt their data before outsourcing in order to protect data privacy from unauthorized outside adversaries, as well as from the cloud service provider. This is justified by current security trends and numerous industry regulations such as PCI DSS. However, conventional encryption makes deduplication impossible.

II. EXISTING SYSTEM

Most of the schemes have been proposed in order to perform a PoW process in an efficient and robust manner, since the hash of the file, which is treated as a "proof" for the entire file, is vulnerable to being leaked to outside adversaries because of its relatively small size. Under Deduplication on encrypted data scheme, data privacy is the primary security requirement to protect against not only outside adversaries but also inside the cloud server. Thus, most of the schemes have been proposed to provide data encryption, while still benefiting from a deduplication technique, by enabling data owners to share the encryption keys in the presence of the inside and outside adversaries. Since encrypted data are given to a user, data access control can be additionally implemented by selective key distribution after the PoW process. Not much work has yet been done to address dynamic ownership management and its related security problem.

III. PROPOSED SYSTEM

A deduplication scheme over encrypted data is proposed. First, dynamic ownership management guarantees the backward and forward secrecy of deduplicated data upon any ownership change. As opposed to the previous schemes, the data encryption key is updated and selectively distributed to valid owners upon any ownership change of the data through a stateless group key distribution mechanism using a binary tree. The ownership and key management for each user can be conducted by the semi-trusted cloud server deployed in the system. Thus, the proposed scheme delegates the most laborious tasks of ownership management to the cloud server without leaking any confidential information to it, rather than to the users. Second, the proposed scheme ensures security in the setting of PoW by introducing a re-encryption mechanism that uses an additional group key for dynamic ownership group. Thus, although the encryption key is revealed in the setting of PoW, the privacy of the outsourced data is still preserved against outside adversaries, while deduplication over encrypted data is still enabled and data integrity against poison attacks is guaranteed. The proposed scheme ensures that only authorized access to the shared data is possible, which is considered to be the most important challenge for efficient and secure cloud storage services in the environment where ownership changes dynamically. It is achieved by exploiting a group key management mechanism in each ownership group. As compared to the previous deduplication schemes over encrypted data, the proposed scheme has the following advantages in terms of security and efficiency.

IV. SYSTEM ARCHITECTURE

In this section, we describe the data deduplication architecture and define the security model. According to the granularity of deduplication, deduplication schemes are categorized into (coarse-grained) file-level or (fine-grained) block-level schemes. Since block-level deduplication can easily be deduced from file-level deduplication, we consider only file-level deduplication for simplicity’s sake. Thus, a data copy refers to a whole file in this paper.

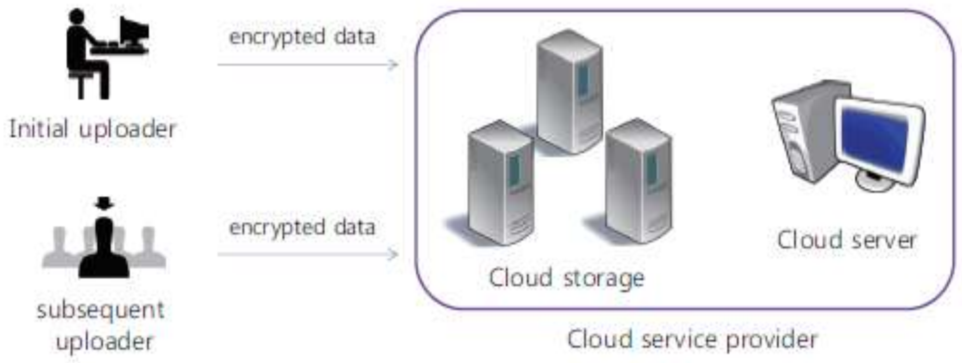

Fig 1: ARCHITECTURE OF A DATA DEDUPLICATION SYSTEM

Fig. 1 shows the architecture of the data deduplication system, which consists of the following entities.

1) Data owner: This is a client who owns data, and wishes to upload it into the cloud storage to save costs. A data owner encrypts the data and outsources it to the cloud storage with its index information, that is, a tag. If a data owner uploads data that do not already exist in the cloud storage, he is called an initial uploader; if the data already exist, called a subsequent uploader since this implies that other owners may have uploaded the same data previously, he is called a subsequent uploader. Hereafter, we refer to a set of data owners who share the same data in the cloud storage as an ownership group.

2) Cloud service provider: This is an entity that provides cloud storage services. It consists of a cloud server and cloud storage.

3) Cloud server: The cloud server deduplicates the outsourced data from users if necessary and stores the deduplicated data in the cloud storage. The cloud server maintains ownership lists for stored data, which are composed of a tag for the stored data and the identities of its owners.

V. CONCLUSION

Dynamic ownership management is an important and challenging issue in secure deduplication over encrypted data in cloud storage. In this study, we proposed a novel secure data deduplication scheme to enhance a fine-grained ownership management by exploiting the characteristic of the cloud data management system. The proposed scheme features a reencryption technique that enables dynamic updates upon any ownership changes in the cloud storage. Whenever an ownership change occurs in the ownership group of outsourced data, the data are reencrypted with an immediately updated ownership group key, which is securely delivered only to the valid owners. Thus, the proposed scheme enhances data privacy and confidentiality in cloud storage against any users who do not have valid ownership of the data, as well as against an honest-but-curious cloud server. Tag consistency is also guaranteed, while the scheme allows full advantage to be taken of efficient data deduplication over encrypted data. In terms of the communication cost, the proposed scheme is more efficient than the previous schemes, while in terms of the computation cost, taking additional $0{:}1$ to $0{:}2$ ms compared to the RCE scheme, which is negligible in practice. Therefore, the proposed schemeachieves more secure and fine-grained ownership management in cloud storage for secure and efficient data deduplication.

VI. ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible so I would like to acknowledge all those whose guidance and encouragement served as a beacon light and crowned the efforts with success. I would like to sincerely thank the project guide, Mr. Srinidhi H.R, Assistant Professor, Dept. of Information Science and Engineering, NIEIT for providing relevant information, valuable guidance and encouragement to complete this project.

VII. REFERENCES

[1]. M. Dutch, “Understanding data deduplication ratios,” SNIA Data Management Forum, 2008.
[2]. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” Proc. StorageSS’08, 2008.
[3]. D. T. Meyer, and W. J. Bolosky, “A study of practical deduplication,” Proc. USENIX Conference on File and Storage Technologies 2011, 2011.
[4]. W. K. Ng, W. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage,” Proc. ACM SAC’12, 2012.