International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

# Data Deduplication in Cloud Storage with and Ownership Management

**Nandhana Kuttappan**

*Student, Dept. of Computer Applications, Christ Knowledge City College, Kerala, India*

---

**Abstract** : *Secure Data de-duplication with Dynamic Ownership Management in cloud storage service. This technology is used to reduce the space for the cloud storage. The data de- duplication providing eliminating redundant data and storing only a single copy. De-duplication method is very effective when multiple users storing the same data in outsource. This time relating security and ownership issues. The 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. Then many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers de duplication because of the randomization property of encryption.*

***Key Words***: de-duplication, cloud storage, encryption, proof-of-ownership, revocation, Convergent encryption,

## 1. INTRODUCTION

CLOUD computing System provides scalable, low-cost, and location-independent online services and simple backup services to cloud storage infrastructures. The new world technology are day to day fast growth of data volumes stored in the cloud storage has increased demand .Then techniques for saving disk space and network bandwidth. This problem reduce resource consumption, many cloud storage services, such as Dropbox, Wuala, Mozy, and Google Drive. The deduplication provides technique, where the cloud server stores only a single copy of same data and provides links to the copy instead of storing other user actual copies of that data, then not considering of how many clients ask to store the data. The savings are expressive, revealing and purportedly, business applications can achieve disk and bandwidth savings of more than 90 percent. However, the security perspective, the shared usage of users' data raises a new challenge.

In cloud storage sevice, deduplication technology is commonly used to reduce space and bandwidth requirement of service by eliminting redundant data and storing only a single copy of them. De-duplication is most effective when multiple users outsource the same data to the cloud storage, but it rasises issues relating

To security and ownership. The 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 de-duplication techniques take provides advantage of data similarity to identify the same data and reduce the storage space .The de-duplication methods in using encryption algorithms randomize. Encryption algorithms encrypted files in order to make cipher text indistinguishable from theoretically random data. Encrypted same data by different users with different encryption keys results in different cipher texts, which makes it difficult for the cloud server to determine whether the plain data are the same and de-duplicate them.

The Convergent encryption effectively resolves this problem. Convergent encryption algorithm encrypts an input file with the hash value of the input file as an encryption key. The cipher text is given to the server and the user retains the encryption key. Since convergent encryption is deterministic, 1 identical files are always encrypted into identical ciphertext, regardless of who encrypts them. Cloud storage server can be perform deduplication over the ciphertext, and all owners of the file can download the ciphertext .Then after the proof-of-ownership (PoW) process optionally and decrypt it later since they have the same encryption key for the file. However convergent encryption suffers from security flaws with regard to tag consistency and ownership revocation.

The ownership Management in cloud storage .In the case of ownership dropping, suppose multiple users have ownership of a ciphertext outsourced in cloud storage. As time discontinued, some of these users may request the cloud server to delete or modify their data, and then, the server deletes the ownership information of the users from the ownership list for the corresponding data.

On the other hand, when a user uploads data. The user uploading that data already exist in the cloud storage,. The user should be deterred from accessing the data that were stored before he obtained the ownership by uploading it (backward secrecy).2 These dynamic ownership changes may occur very frequently in a practical cloud system, and it should be properly managed in order to avoid the security degradation of the cloud service.

The proposed scheme ensures that only authorized access to the shared data is possible. Which is considered to be the most important challenge are 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.

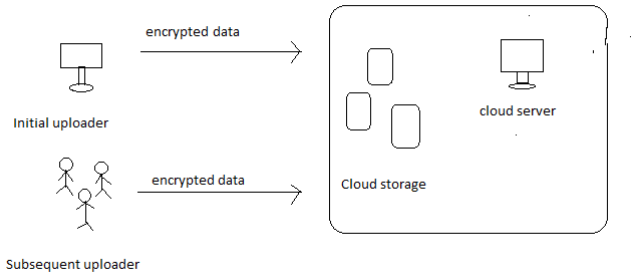
## 2. RELATED WORK

The main aim behind this project is to reduce the problems of portability of storage a space by making use of the concept of cloud computing storage. Cloud computing provides different services for actions like Storage, Operating Systems, Machine Learning, Security, etc. The proposed scheme features a re-encryption technique that enables dynamic updates on any ownership changes in the cloud storage. The ownership change occurs in the ownership group of outsourced data, and the data are re-encrypted with an immediately updated ownership group key, which is securely delivered only to the valid owners. Ownership management in cloud storage for secure and better data de-duplication.

Here the author (SANDRO RAFAELI AND DAVID HUTCHISON) specifies Group communication can benefit from IP multicast to achieve scalable exchange of messages. There is a challenge of effectively controlling access to the transmitted data. The IP multicast by itself does not provide the any mechanisms for preventing non group members to have access to the group communication. Encryption can be used to protect messages exchanged among group members. Then distributing the cryptographic keys becomes an issue. The paper have proposed several different approaches to group key management. This approaches can be divided into three main classes: centralized group key management protocols, decentralized architectures and distributed key management protocols. There are three classes described here and an insight given to the features and goals. The area of group key management is then surveyed and proposed solutions are classified according to those characteristics.

Here in(Mihir Bellare1 and Sriram Keelveedhi) This paper considers the problem of secure storage of outsourced data in a way that permits deduplication. The first time able to provide the privacy for messages that are both correlated and dependent on the public system parameters. The new ingredient that makes this possible is interaction. We extend the message-locked encryption (MLE) primitive of prior work to interactive message locked encryption (iMLE) where upload and download are protocols. Our scheme, providing security for messages that are not only correlated but allowed to depend on the public system parameters, is in the standard model. We explain that interaction is not an extra assumption in practice because full, existing deduplication systems are already interactive.

Deduplication techniques can be categorized into two different approaches: deduplication over unencrypted data and deduplication over encrypted data.




*Architecture of a data deduplication system*

Data de-duplication architecture define the security model. According to the granularity of de-duplication, de-duplication schemes are categorized into file-level or block-level schemes. Since the block-level de duplication can easily be deduced from file - level de-duplication, we consider only file level de-duplication for simplicity's sake

Data de-duplication system, consists of the following entities:

1. Data owner
2. Cloud service provider

Data owner client who owns data, and wishes to upload it into the cloud storage to save costs. The owner data encrypts and outsources it to the cloud storage with its index information, that is, a tag. If a data uploads owner data that do not already exist in the cloud storage, this data upload is called an initial uploader; if the data already exist, it is called a subsequent uploader.

Cloud service provider provides cloud storage services. The cloud server deduplicates the outsourced data from users if fundamental and stores the deduplicated data in cloud. 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. The cloud server controls access to the stored data based on the ownership lists and manages group keys for each ownership group as a group key authority.

## 3. PROPOSED FRAMEWORK

Deduplication technology is used to reduce the space and bandwidth requirement of services by eliminating redundant data and storing only a single copy of them.

We propose a novel server-side deduplication scheme for encrypted data. The deduplication technique where the cloud server stores only a single copy of redundant data. Deduplication techniques take advantage of data similarity to identify the same data and reduce the storage space. The convergent encryption resolves this problem effectively. Convergent encryption algorithm encrypts an input file with the hash value of the input file as an encryption key.

Convergent encryption is deterministic, identical file are always encryption into identical ciphertext, regardless of who encrypts them. The cloud storage sever can perform deduplication over the ciphertext and all owners of the can download the cipryption and decrypt it later since they have the same encryption key the file. These dynamic ownership changes may occur very frequently in a practical cloud system. Its importance to secure dedpulication, because the encryption key is derived deterministically and rarely update after the initial key derivation. A long as revoked user keep the encryption key. Then they can access the corresponding data in the cloud storage at any time regardless of the validity of their ownership. The proposed scheme ensures that only authorized access to the shared data is possible.

The propose a secure deduplication scheme for encrypted data. That data has been provide dynamic ownership management capability. This scheme is developed based on a randomized convergent encryption scheme .The randomize the encrypted data, which renders the proposed scheme secure against the chosen-plaintext attack while still allowing deduplication over the data. To handle dynamic ownership management, the cloud server must obtain the ownership list for each data, since otherwise revocation cannot take effect. This setting where the cloud server knows the ownership list does not violate the security requirements, because it is allowed only to reencrypt the ciphertexts and can by no means obtain any information about the data encryption key of users.

**4. IMPLEMENTATION**

Cloud computing provides scalable, low-cost and location independent online services ranging from simple backup services to cloud storage infrastructures. In the cloud system fast growth of data volumes stored in the cloud storage has led to an increases demand for techniques for saving disk space and network bandwidth. The de-duplication technique where the cloud server stores only a single copy of redundant data 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 saving are significant and reportedly, business application can achieve disk and bandwidth saving of more than from a security perspective the shared usage of users data raises new challenge.

The de-duplication scheme implemented for encrypted data that has dynamic ownership management capability. Data de-duplication is one of the techniques to prevent the unauthorized user of data accessing and create duplicates data on cloud then the encryption techniques to encrypt data before stored on cloud server. Cloud storage usually contain business critical data and process hence high security is the only solution to retain strong trust relationship between the cloud user and cloud service providers.

In the de-duplication method we have to detect the duplication copy of the file. Same type of data cloud store in at a onetime then the same data another user store in cloud system, but that same data duplicated .The user saved data already stored in cloud server.

For each operation, are include a benchmark timing. Each cryptographic operation was implemented using the Crypto++ library ver. 5.6.2 [34] on a 3.4 GHZ processor PC. The key parameters were selected to provide a 128-bit security level. The implementation uses an MD5 as a cryptographic hash function to generate a 128-bit key and tag, and an AES with Electronic Code Book (ECB) mode as an encryption/decryption function. Data encryption and decryption time, denoted by Enc and Dec, respectively, are measured for different data sizes

The basis of the encryption and decryption time, we measured the total computation cost for the upload and download of each scheme,. For the upload procedure, the proposed scheme requires the same computations as the CE and RCE schemes. For the download procedure, the proposed scheme needs one more key decryption operation than does the basic RCE scheme. However, since the symmetric key size is much smaller than the typical data size in the cloud (e.g., document file, or multimedia data), the additional 128bit key decryption time (i.e., 0:129 ms) in the proposed scheme would be relatively negligible as compared to the data decryption time in a pragmatic cloud computing system.

The Data de-duplication providing security requirements.

1. Data Privacy :

   Data privacy defing, the cloud server is no longer fully trusted even if honest. Therefore, plain data should be kept secret from the cloud server as well as from unauthorized users who cannot prove ownership. In order to ensure security and data privacy, user may encrypt their data before uploading to cloud. In such cases same data or different user many upload same data in encrypted from and it results in the existence of duplicated data. Data deduplication is a technique designed it identify and eliminated redundant data.

2. Data Integrity:

   De-duplication is a technique where the server stores only a single copy of each file, regardless of how many clients asked to store that file such that the disk space of cloud servers as well as network bandwidth are saved. In the de-duplication scheme, data integrity may be threatened by a poison attack on tag consistency. Without loss of generality, we suppose an attacker and another user u have the same data M. In the proposed scheme, the poison attack on tag consistency is easily detected, as in the basic RCE scheme

3. Collusion Resistance:'

The collusion resistance, proved unauthorized users who have not valid ownerships of cloud data should not be able to decrypt them even if they collude. In the proposed method, in order to decrypt the cipher text and obtain the plain data, users should have knowledge of both the data encryption key L and the ownership group key GK. The some unauthorized users may be able to obtain the data encryption key, 9 it is impossible to have the ownership group key GK. This is because the KEK assignment for ownership group key distribution in the binary KEK tree is information theoretic, that is KEKs are assigned randomly and independently of each other. Whenever any ownership change occurs in an ownership group, the ownership group key is rekeyed immediately and the data are re-encrypted using the updated group key. Even if the unauthorized users collude with each other, they cannot obtain the current ownership group key, since none of the KEKs in their path keys in the KEK tree is used to encrypt and distribute the ownership group key. Therefore, the proposed scheme is secure against collusion attack of the unauthorized users.

## 5.CONCLUSIONS

Dynamic ownership management is an important and challenging issue.in secure de-duplication over encrypted data in cloud storage. In this study, we proposed a novel secure data de-duplication 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 enable updates upon any ownership changes in the cloud storage

## ACKNOWLEDGEMENT

I would like to extend my heartfelt thanks to Ms. GEETHU KRISHNA KARTHA, HOD, Assistant Professor, Department of Dual Degree Master Of Computer Applications for their valuable advices and guidance throughout my course of study. My most sincere thanks go to my mentor

I express my sincere thanks to Ms. ELSA SABU. Assistant Professor .my project guide for the valuable advices and guidance throughout the completion of project.

Also I like to express my gratitude to all Teachers and staffs of Christ knowledge city for their kind co –operation and help.

I extend my thanks to all my friends for their moral support and encouragement. Last but not least we thank sour parents and benefactors who inspired me always to do our best.

## REFERENCES

[1] J. Li, et al., "Secure distributed deduplication systems with improved reliability," IEEE Trans. Comput., vol. 64, no. 2, pp. 3569–3579, 2015.

[2] S. Rafaeli and D. Hutchison, "A survey of key management for secure group communication," ACM Comput. Surveys, vol. 35, no. 3, pp. 309–329, 2013..

[3] M. Bellare and S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," in Public-Key Cryptography. Berlin, Germany: Springer, 2015, pp. 516–538