International Research Journal of Computer Science (IRJCS)
Issue 05, Volume 4 (May 2017) SPECIAL ISSUE
ISSN: 2393-9842
www.irjcs.com

Survey on Dynamic Ownership Management for Secure Data De-duplication

[1]Ms Vijaya SC, [2]P Vasavi, [3]R Bhavitha Reddy, [4]K Swetha Kumari,

[1]Asst Professor,[2,3,4]*UG Students*
[1,2,3,4]*Department of CSE,*
Vemana Institute of Technology, Bangalore-34, India

Abstract: ***Data de-duplication is used in cloud storage to save bandwidth and reduce the storage space by keeping only one copy of same data. But it raises problems involving data ownership and security when multiple users upload the same data to cloud storage. Since encryption preserves privacy, yet its randomization property hampers de-duplication. Hence, there is a need of secure data deduplication scheme to prevent unauthorized access and data leakage. In recent times, a number of de-duplication schemes have been proposed to solve this problem. However, many systems suffer from security flaws because they do not reflect the dynamic changes in the ownership of outsourced data. In this paper, we review several deduplication techniques over encrypted data to achieve secure and efficient cloud storage service. Furthermore, proposed scheme uses RCE and group key management mechanism to ensure that only authorized access to the shared data is possible, which is considered to be the most important challenge for secure and efficient cloud storage service in the environment where ownership changes dynamically.***

Keywords: ***De-duplication, cloud storage, encryption, proof-of-ownership.***

I. INTRODUCTION

Cloud Computing is a widespread term used in today's world. It delivers infinite space for storage, readiness, user-friendliness from anywhere, anytime to entities. Now-a-day's number of users and their data in the cloud is continuously growing with higher memory space and upload bandwidth. Data de-duplication used in cloud storage providers to resolve these overheads. Deduplication is a process of removing multiple copies of same data, to reduce the storage space and save bandwidth. But when same data outsourced by users to cloud storage some challenges are arises on data ownership and security for sensitive data.

Today's cloud storage services like Dropbox and Google Drive etc. use a de-duplication scheme to save the network bandwidth and the storage cost. As data owners worried about their private data, they may encrypt their data before uploading in order to keep data privacy from illegal outside adversaries, as well as from the cloud service provider. As concern with authorized access and security, there are many encryption schemes proposed. De-duplication scheme takes benefit of data similarity to find the same data and scale down the storage space. In contrast, encryption algorithms randomized the encrypted files to make cipher-text same from theoretically random data. Encryption of the same data by dissimilar users with different encryption keys results in different ciphertexts, which makes it hard for the cloud server to decide whether the plain data are the same and de-duplicate them. Hence, traditional encryption makes de-duplication impossible for above reasons. The simplest implementation of traditional encryption can define as follows: Consider users A and B, encrypts the same file M under their secret keys SKA and SKB and stores corresponding cipher-text CA and CB. Then, further problems arise: First, how can the cloud server sense that the underlying file M is similar, and second is even if it can notice this, how can it allow both users to recover the stored data, based on their distinct secret keys? One simple way out is to let on each client to encrypt the file with the public key of the cloud storage server. Then, the server is capable to de-duplicate the identified data by decrypting it with its private key pair. Still, this solution grants access to the cloud storage server to get the outsourced plain data, which may break up the privacy of the data if the cloud server cannot be fully trusted. Convergent encryption plays the vital role in data deduplication and overcomes the drawback which discussed above. A convergent encryption algorithm works as follows: Firstly, it takes an input file and encrypts them with its hash value as an encryption key. Then, the ciphertext is given to the cloud server and user keeps the encryption key. As convergent encryption is deterministic, every time similar files encrypted into similar cipher-text irrespective of who encrypts them Hence, the cloud server can do de-duplication over the generated ciphertext. Then all data owners can download the ciphertext and decrypt it later as they have the same encryption key for the file. But convergent encryption has security weakness concern with tag consistency and ownership revocation. This paper formalizes a scheme to solve the challenge of ownership changes dynamically in the cloud system.

© 2014- 17, IRJCS- All Rights Reserved

The proposed scheme guarantees that only authorized access to shared data is possible. It is achieved by using a group key management mechanism in each ownership group.

II. LITERATURE SURVEY

In cloud computing, there have been many of the schemes, proposed for data deduplication over encrypted and unencrypted data of cloud storage. We are going to discuss about the data deduplication schemes over encrypted data and how it has been developed and improved further into Convergent Encryption (CE), Leakage-Resilient (LR) Deduplication scheme, Randomized Convergent Encryption (RCE) and Dynamic Ownership Management Scheme.

Convergent Encryption (CE): LI[1] In order to keep data privacy against inside cloud server as well as outside challengers, users may want their data encrypted. However, conventional encryption under different users' keys makes cross-user de-duplication impossible, since the cloud server would always see different ciphertexts, even if the data are the same, regardless of whether the encryption algorithm is deterministic. Douceur introduces Convergent Encryption, which is the promising solution to this problem. In CE, a data owner derives an encryption key over data by using cryptographic hash function. Then computes the ciphertext using block cipher over data along with their encryption key. CE deletes data and keeps only encryption key after uploading ciphertext to the cloud storage. Since encryption is deterministic, on receipt of same file CE generates same ciphertext for it and the server does not store the file but instead updates meta-data to indicate it has an additional owner.

Advantages: Provides promising solution over conventional encryption and preserves data privacy.
Disadvantages: Convergent Encryption suffers from some security issues i.e. tag consistency problem. It means that integrity and security of data has been compromised due to the lack of proof of ownership process and dynamic ownership management.

Ramp Secret Sharing Scheme (RSSS) LI[2] formalizes a convergent key management scheme i.e. Dekey which is efficient and reliable for secure deduplication. Dekey set de-duplication between convergent keys and distributes those keys across multiple key servers while preserving the semantic security of convergent keys and privacy of outsourced data. Dekey is implemented using the Ramp secret sharing scheme. Dekey uses RSSS to collect convergent keys. Its idea is to permit deduplication in convergent keys and distribute the convergent keys over various KM-CSPs. Instead of encrypting the convergent keys on a per-user basis, Dekey builds secret shares on the original convergent keys (that are in plain) and assigns the shares over various KMCSPs. If many users share the identical block, they can access the same corresponding convergent key. This significantly decrease the storage overhead for convergent keys. In addition, this method provides fault tolerance and allows the convergent keys to remains accessible even if any subset of KM-CSPs fails.

Advantages: Provides reliable, efficient and fault tolerance convergent key mechanism for secure de-duplication.
Disadvantages: This scheme does not support dynamic ownership management issue in secure de-duplication.

Authorized De-duplication Hybrid Cloud LI[3] proposes an authorized de-duplication scheme where differential privileges of users, as well as the data, are considered in the de-duplication procedure in a hybrid cloud environment. He presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate check tokens of files are generated by the private cloud server with private keys. The figure shows the architecture of authorized de-duplication.

Advantages: This scheme provides authorized de-duplication over hybrid cloud for users who have different privileges.
Disadvantages: Data leakage.

III. PROBLEM DEFINITION

Storage efficiency functions such as compression and deduplication afford storage providers better utilization of their storage back ends and the ability to serve more customers with the same infrastructure. Data deduplication is the process by which a storage provider only stores a single copy of a file owned by several of its users. There are four different deduplication strategies, depending on whether deduplication happens at the client side (i.e. before the upload) or at the server side, and whether deduplication happens at a block level or at a file level. Deduplication is most rewarding when it is triggered at the client side, as it also saves upload bandwidth. For these reasons, deduplication is a critical enabler for a number of popular and successful storage services (e.g. Dropbox, Memopal) that offer cheap, remote storage to the broad public by performing client-side deduplication, thus saving both the network bandwidth and storage costs. Indeed, data deduplication is arguably one of the main reasons why the prices for cloud storage and cloud backup services have dropped so sharply.

IV. EXISTING SYSTEM

- *Security Concerns does not consider in the existing system.*
- *In existing system, user uploaded the file into cloud and each file can be uploaded 'n' number of times into the server.*
- *Makes the server with duplicate copy of file and it will attack by anyone and confidential level is decreased.*

© 2014- 17, IRJCS- All Rights Reserved

DISADVANTAGES

- *The key management is very complicated when there are a large number of data owners and users in the system.*
- *The key distribution is not convenient in the situation of user uses the system dynamically.*
- *Deduplication is most effective when multiple users outsource the same data to the cloud storage, but it raises issues relating to security and ownership.*

V PROPOSED SYSTEM

- *The proposed system uses hash function to avoid the duplication in cloud.*
- *Uses Elliptic Curve Cryptographic (ECC) algorithm for encryption and decryption process.*
- *Proposes an efficient group key management protocol in distributed group communication.*
- *High efficient.*
- *ECC algorithm provide high end security. Avoid duplication in cloud.*

VI. DESIGN PHASE

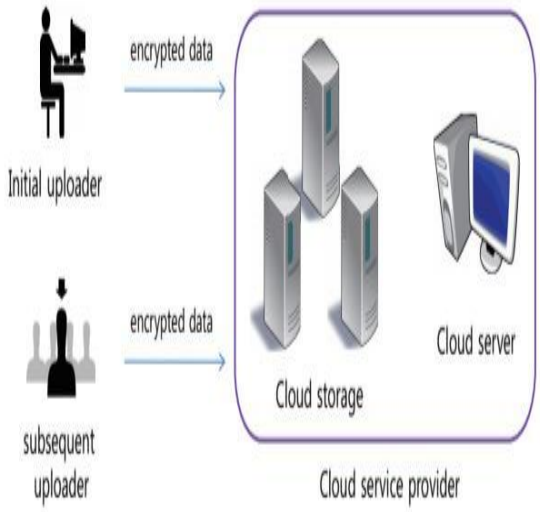

Fig.1 System Architecture

TABLE I: COMPARISON OF DATA DE-DUPLICATION TECHNIQUES

PARAMETERS V/S TECHNIQUES	EFFICIENCY	TAG CONSISTENCY	OWNERSHIP MANAGEMENT
CE	Less efficient than proposed scheme	No	No
RSSS	Efficient and Reliable	Yes	No
Authorized de- duplication	Efficient and secure in terms of insider and outsider attacks	-	No
Proposed Scheme	Proposed Scheme Highly Efficient and more secure than all	Yes	Yes

VII. METHODOLOGY

MODULES

1. USER REGISTRATION AND LOGIN

In this module a new user can be allowed to register by providing the user details like name,email,age etc.the user can allow to login before which the group must generate a team to make the user valid and process him towards the group. The validation of user details, inserting a new record in to the registration table will be taken care.

© 2014- 17, IRJCS- All Rights Reserved

2. USER JOINING THE GROUP AND FILE UPLOAD

For every user a key would be generated using which the user gets the authentication to join the group with a key. If file upload process user chooses a file from his system and generate hash key for each file. Hash key generation is provided to avoid duplication of the file to the cloud. If the file is already in the cloud the user cannot upload the file.

3. FILE ENCRYPTION AND STORAGE IN CLOUD

After the validation of the file from the user with cloud, we apply a cryptographic technique to improve the security level in the cloud. We implement ECC algorithm which converts a file in to a binary format and it gets encrypted and is stored on to the cloud. The data that is stored on to the cloud will be in encrypted format.

4. USER FILE REQUEST AND DOWNLOAD

Any user who has registered earlier and joined the group with a valid key can request the file to the cloud. The cloud service provider after authenticating the user can receive the file request, decrypt the file using ECC algorithm and send the requested file to the user. Then the file will be downloaded in the users location

ALGORITHMS

1. KEK<--------KEK Gen(U)

The KEK generation algorithm takes a set of Users (U) as inputs and outputs KEKs for each user in U.

2.C<----------Encrypt (M,I^x)

The encrytion algorithm is a randomized algorithm that takes an input data M and a security parameter x and outputs ciphertext (C) of the data.Ciphertext consists of the encrypted message and its tag information for indexing.

3. C^1<---------- Re-Encrypt (C,G)

The re-encryption algorithm takes ciphertext(C) and ownership group(G) as input and outputs C^l such that only the valid owners in G can decrypt the message.

4. C^1<----------Decrypt (C,K,Gk)

The decryption algorithm is a deterministic algorithm that takes C,encryption key (K) and owner group keyGk and outputs a message M.Gk talks abour user revocation.

CONCLUSION

In this paper, we have reviewed different data deduplication techniques over encrypted data that are used in the cloud computing for secure data storage. Traditional encryption makes deduplication impossible because of the randomization property of encryption. Recently, several deduplication schemes are proposed to solve this issue by allowing each owner to share the same encryption key for the same data. Convergent encryption has different encryption variants for secure deduplication. Though, CE suffers from security flaws with regard to tag consistency and ownership revocation. Furthermore, many schemes could not achieve secure access control under dynamic environment. Hence, not much work has yet been done to address dynamic ownership management and its related security problem. Thus 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.

REFERENCES

[1].P. Balasubramanyam Reddy, G. Nagappan, "A Survey on Secure Cloud Storage with Techniques Like Data Deduplication and Convergent Key management," International Journal of Scientific Research in Computer Science,2016 IJSRCSEIT volume1.

[2].Trupti Deore, Prof. J. V. Shinde , "A Review on Secure and Authorized Data De-Duplication in Hybrid Cloud," International Journal of Scientific Research in Computer Science,2016 IJSRCSEIT volume6.

[3].R.Thilagavathi, S. Ramasamy, R.K. Gnanamurthy, A Study of De Duplication using De Key with efficient and Reliable Convergent Key Management in Cloud Storage," International Journal of Scientific Research in Computer Science,2015 IJSRCSEIT volume5.

[4].Harnik, B. Pinkas, and A. Shulman-Peleg, "Side Channels in Cloud Services: De-duplication in Cloud Storage," IEEE Security Privacy, vol. 8, no. 6, pp. 40-47, Nov./Dec.2010.

© 2014- 17, IRJCS- All Rights Reserved