Indian Journal of Science and Technology, *Vol 9(25), 10.17485/ijst/2016/v9i25/87990, July 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Effectuation of Secure Authorized Deduplication in Hybrid Cloud

B. Mahalakshmi* and G. Suseendran

Department of Information Technology, School of Computing Sciences, Vels University, Chennai – 600117, Tamil Nadu, India;
maha.karthik921@gmail.com, suseendar_1234@yahoo.co.in

Abstract

Objectives: Different user can access the same data repeatedly and trying to store it in the memory of the cloud server. Due to this there is a problem of maintaining the storage space and bandwidth. The main purpose of this study is how the data is secured, whether the authorized person is accessing the data or not and finally to check whether same data is repeatedly stored in the memory to avoid duplication of the data. **Methods**: In deduplication, to guard the confidentiality of sensitive information, it's encrypted/decrypted by the planned convergent coding technique before outsourcing for higher protection of knowledge security. **Findings**: The convergent encryption method and open authorization protocol and deduplication are combined together and check the data for deduplication in a secured way. The possibilities of using other algorithms are also considered for further implementation.

Keywords: Authorized Duplicate Check, Confidentiality, Deduplication, Hybrid Cloud, Open Authorization

1. Introduction

One of the crucial challenges of cloud storage services are that the management of the growing volume of information. Recently deduplication is a data compression technique which is used in information management system to maintain the data securely and to avoid repetition of data. Deduplication concept is used to get better storage to cut back the amount of data while transferring it through internet. It reduces the redundant copy of same data to avoid replication by referring the physical copy of the information. In the file level deduplication, it reduces duplicate copies of a similar file. Deduplication also can occur at block level, in which it eliminates the duplicate chunk of information that occurs in non-identical files.

Although there are many benefits of information deduplication, security and privacy issues arise as users' sensitive data are at risk of insider and outsider attacks. Encryption techniques that were traditionally incompatible with information deduplication whereas providing information confidentiality. In the traditional encryption method, the encryption is done by different users who are using their own keys of other user data resulting various cipher text for the same information in which deduplication is not possible. To avoid the risk of confidentiality a convergent encryption technique has been used. It creates a convergent key which is computed by cryptographic hash value while encrypting/decrypting the data. When the user receives the convergent key he/she sends the cipher text to the cloud storage. In order to maintain the security a proof of ownership protocol is necessary to find whether the user is an authorized or not. So that the user can make some corrections in the file once the replication is found. The convergent encryption is used traditionally but the confidentiality is at some level only and the duplicate check with the authorization is also not possible. The

**Author for correspondence*

challenge is to do duplicate check with differential privilege at the same time in a confidential way.

Using the Fuzzy Clustering model the duplicate data are clustered into a group and the elimination is done easily so the level of deduplication is improved. A semantic Deduplication of Temporal Dynamic Records from Multiple Web Databases[1].

In cloud computing the data is encrypted in a way the user is having certain attributes and then the privilege rights are used for accessing the data. Hence the data is stored securely from unknown user[2].

The artificial intelligence technique is used to detect if any intrusions happens in private cloud. So that the real time data is secured using this technique. This model is proposed to use in the banking sector as it is a high end technique[3].

Investigation about data deduplication its techniques and changes introduced in deduplication due to virtualized data center and evolution of current cloud computing era, An investigation on Data De-duplication Methods And its Recent Advancement[4].

In a Hybrid cloud architecture a new deduplication system with differential duplicate check is proposed where the S-CSP resides in the public cloud. The duplicate check is done for files marked with the corresponding privileges are allowed by the user only, A hybrid Cloud Approach for Secure Authorized Deduplication[5].

2. Proposed Work

In hybrid cloud architecture the user can store the data in both public and private cloud. It is a combination of both public and private cloud. Non important activities are performed using public cloud where as critical activities are in private for security purposes. Even though we are using the private cloud for data storage there are many security issues. To avoid these issues we are using certain techniques to encrypt a data and decrypt to the privileged user. In this paper we discussed some of the key terms related to the authentication security issues. The open authentication algorithm is used for giving access permission to the user and the convergent encryption technique is used to encrypt the data with the key provided by the technique. Hence the privileged user can access the data in a secured way.

2.1 Symmetric Encryption

Symmetric encryption[5] makes use of a standard secret key *K* to encode and decode data. A symmetric encryption method consists of three basic functions:

- *KeyGenSE(1_)!* using security parameter *1_* the key generation algorithm creates a key value *K*;
- *EncSE(K,M)! C* is the cipher text i.e. encrypted output which is generated by symmetric encryption algorithm using the key value *K* and *M* Message
- *DecSE(K,C)! M*is the Message i.e. decrypted output which is generated by symmetric decryption algorithm using the key value *K* and cipher text *C*.

2.2 Convergent Encryption Algorithm

The Convergent encryption technique[5-8] provides the deduplication in a confidential way. Data deduplication is one of the focused data compression technique for eliminating replication of duplicate copies in cloud storage. It is mainly used to get better storage utilization and also to cut back the amount of bytes while transferring data in the group. Instead of having multiple redundant copies of data in cloud storage, deduplication maintains a single physical copy of the data and referring the data to others where the information is needed. The user encrypts the data with the convergent key which is derived from each copy of data. A tag is created for each data by the user to find the duplicate copies. The user sends the tag value to the server for duplicate check whether the tags are same and If the tags are same then the data also same so the repetition of duplicate copies are eliminated here. Hence the data is secured and the tag cannot be presumed to compromise the convergent key. Each the encrypted information replica and its subsequent tag are going to be holding on the server aspect. Formally, a convergent encoding theme is often outlined with four primitive functions:

- *KeyGenCE(M)!K* is convergent key which is created for the message m for representing the data that the key generation rule that maps an information copy *M* to a convergent key *K*;

- *EncCE(K,M) ! C* is the cipher text i.e. encrypted output which is created using the convergent key value *K* and the message *M*

- *DecCE(K,C)! M* is message i.e. the decrypted output which is created using the convergent key *K* and the cipher text *C*

- *TagGen(M)! T(M)* is the output where *T* is the tag value for the original message *M*.

2.3 Proof of Ownership

Proof of ownership[9,10] is a concept for proving the ownership of data in the cloud storage. While storing the data initially, the server receives a value $\phi(M)$ where *M* is the message which is send by the user and the communication is done by the PoW an interactive algorithm. Before accessing the data the user have to prove the ownership by sending ϕ' to the server so that the value $\phi' = \phi(M)$. During the distribution of content in the network the user finds some attacker. The attacker knows a bit of details about the file and accomplices who is having the file. The formal security definition for PoW roughly follows the threat model during a content distribution network, wherever an offender doesn't recognize the whole file; however have accomplices who have the file. The constraint is based on the bounded retrieval model in which the attacker can get the fewer bits of a file and not the entire file.

2.4 Identification Protocol

There are many Identification protocols[11,12] are available and it is used to identify a person's identity. The two phases of the identification protocol is user and a server. The user will send his identity to the server, and the server performs some identification proof relates to the user's identity. The server confirms the user by sending the confirmation key whether he can have the identity to access the data. Hence the information is accessed by the original user and not shared by some others. These are done by some identification protocol.

2.5 Open Authorization Technique

Open Authorization[13] is an open protocol for token based authentication and authorization on the web. Open Authorization protocol is used to confirm the users if they're firmly authorized or not. This protocol is used on top of the design for approved deduplication the method for getting the token is named a flow.

Open Authorization protocol in different words a group of rules that permit a third party websites or application to access a user's information without the user eager to share login credentials. Its open supply protocol permits users to share their information and resources hold on one website with another site below a secure authorization theme supported a token-based authentication. The token generation using open authorization is explained in the Figure 1.

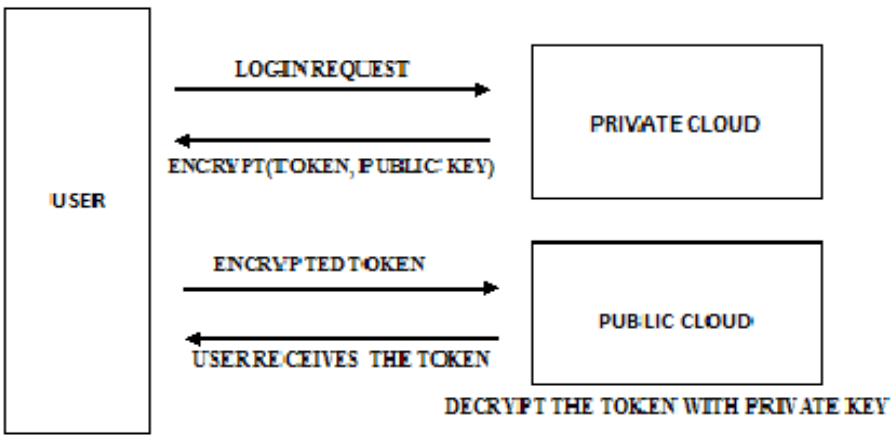

Figure 1. Token generation using Open Authorization.

It works like a client-server model in that the primary storage is a server and the application accessing the stored data is a client. The client creates a session to validate the user by establishing the Open Authorization interface and secret key for accessing the data. Once the session is started the user (i.e. the client) is forwarded to the login credentials for opening the data. In the Triple Crown method, initially a token is send to the requesting client as an acknowledgement for authentication permission to access the data.

3. Authorized Deduplication Architecture in Hybrid Cloud

Hybrid cloud is a combination of both public and private cloud. Depends upon the usage, computing needs and workload the hybrid cloud is giving a greater flexibility and more deployment needs between the public and private cloud. In Hybrid Cloud convergent encryption has been used to implement data confidentiality. Data copy is encrypted below a key which is derived by hashing the data itself. This convergent key is used for encrypt and decrypt a data copy. Moreover, such unauthorized users cannot decode the cipher text even join together with the S-CSP (storage cloud service provider). Security analysis shows that the system is protected in terms of the definitions specified in the planned security model[14,15].

The three entities defined in a Hybrid cloud model of authorized deduplication are explained Figure 2.

3.1 S-CSP (Storage Cloud Service Provider)

S-CSP is a data storage services in public cloud. On behalf of the user it stores the data and provides data outsourcing services. It keeps the unique data to cut back the cost of storage using data deduplication. S-CSP is available all the time, hence we are having large amount of storage capacity and the computing power

1. Start
2. Get unencrypted file tag

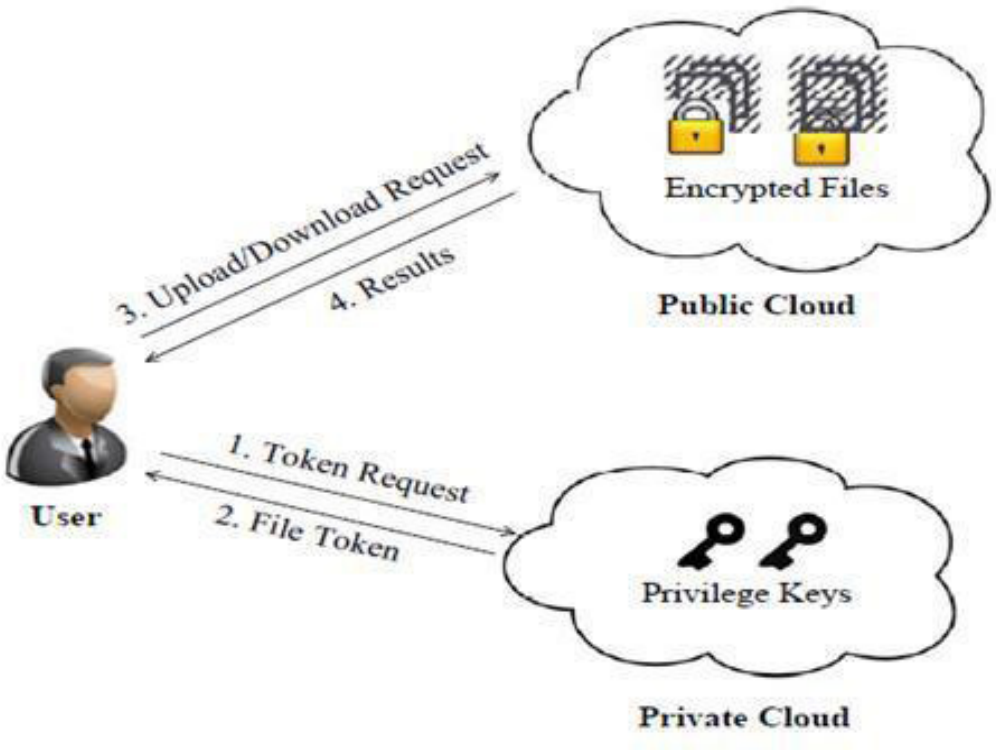

Figure 2. Authorized deduplication architecture.

3. Accept the privilege based token from user
4. Validate the token and assert privilege level
5. Run deduplication check only on privileged files
6. If the same tag is found along the privileged files, mark deduplication search as successful and grant access to the encrypted file
7. Stop.

3.2 Data users

A data user wants to access the information and outsource data storage to the S-CSP. The user only uploads the unique data and avoids uploading duplicate copies to maintain the bandwidth in which it is accessed by the same or different user. The authorization and the privileges are confined by having convergent key and privilege key on each files.

3.3 Private Cloud

This entity is used to secure the cloud services by providing a file token to the users in which the privilege is managed by the private cloud by private key. Due to insufficient security in public cloud while computing resources at the user side, the private cloud provides the user with some better infrastructure and environment between the user and the public cloud. The private cloud manages the private key by answering the file token request from the users.

1. Start
2. Request user authentication based on their credentials
3. Token key is generated based on unique privileges
4. Send the token
5. Stop.

The private cloud only has the privileges to access the user credentials whereas public cloud is not. Even though a compromised token key is changed regardless of the user credentials, the public cloud does not have the access rights to encrypt or decrypt a file. Deduplication is done only by the privileged users.

4. Convergent Encryption and Open Authorization Algorithm

Convergent encryption is a cryptographic method for elimination of duplicate file without having the access to the encryption keys for sensitive data. Open authorization is a secure delegated access for the owner to server resources. It provides a service from the server for the owner to authorize third party access without sharing their ID. By an authorized server the resource owner can allow by issuing an access token to the third party clients. Then the client can able to access the resource by the resource server. Here we use two types of algorithm one is for checking whether the user is a privileged one for uploading and downloading another one is encryption algorithm for eliminating the data securely.

File Uploading

START:

1. Reading the file
2. Cloud server checks whether or not the user is public or private
3. If user is public, then
   a. The Cloud server checks the file for duplication
   b. It Sends duplication response whether or not the file already available or not
   c. If file doesn't exist it displays → file doesn't exist, then
   d. It encrypts and uploads the file
   e. Else it displays → file previously exist

4. If the user is private, then
   a. The Cloud server checks the file for duplication
   b. It Sends duplication response whether or not the file available or not
   c. If file doesn't exist it displays → file doesn't exist then
   d. It encode and uploads the file
   e. Else it displays → file subsist

STOP

File Downloading

START:

1. Reading the file

2. Cloud server checks the user whether or not he is public or private

3. If user is public
   a. Cloud server test the file for duplication
   b. Sends duplication response whether or not the file already exists or not
   c. If file exist already it displays → file exist
   d. Then it decode and downloads the file
   e. Else, it displays → file doesn't subsist

4. If the user is private
   a. Cloud server checks the file for duplication
   b. Sends duplication response whether or not the file already exist
   c. If file exist already it displays → file exist
   d. Then it decodes and downloads the file
   e. Else, it displays → file doesn't subsist

Stop

The uploading and downloading of a file is done using an open authorization algorithm and convergent encryption algorithm. The open authorization technique is used to check whether the person is an authorized or not and then the convergent encryption algorithm is for encrypting/decrypting a file for secured reasons. The open authorization algorithm creates a key and a tag for a file. The tag is for checking the data i.e. The file is already available or not. And the key is for checking that the authorized person only accessing the file. Cloud server initially creates a tag value while a user wants to upload a file in the cloud storage. The tag value represents that the data is the original copy. The user then uploads the file in the cloud server. The key is used to encrypt/ decrypt a file after the tag processing is over. Finally the uploading and downloading of file is over using these tag and key value created by the server with the help of these algorithms. The general model of the algorithm is explained above. Cloud storage contains lots and lots of data. The deduplication compression method is for reducing the storage space and the data is secured using many algorithms. Here we tend to use the open authorization algorithm for verifying whether the person is an authorized and duplicate check[16] also done.

5. Conclusion

Hybrid cloud design provides plenty of advantages for both public and private cloud. Now a day's most of the users use hybrid cloud to store information. In cloud storage increasing volume of information and security problems may be a major concern. So as to cut back the storage complexity of space and to efficiently utilizing the information, data deduplication idea is employed. For secured deduplication the Open Authorization and also the convergent encryption techniques are used for duplicate check in a protected way and keep the information in cloud with user privileges.

6. References

1. Devi RP, Thigarasu V. A semantic Deduplication of Temporal Dynamic Records from Multiple Web Databases. Indian Journal of Science and Technology 2015 Dec; 8(34):1–7. Doi:10.17485/ijst/2015/v8i34/75103.
2. Manjusha R, Ramachandran R. Secure Authentication and Access System for Cloud Computing Auditing Services Using Associated Digital Certificate. Indian Journal of Science and Technology. 2015 Apr; 8(S7):220–27. Doi: 10.17485/ijst/2015/v8iS7/71223.
3. Rajendran PK. Hybrid Intrusion Detection Algorithm for Private Cloud. Indian Journal of Science and Technology. 2015 Dec; 8(35):1–10. Doi: 10.17485/ijst/2015/v8i35/80167.

4. Kaurav N. An investigation on Data De-duplication Methods And its Recent Advancement 2014.
5. Li J, Li YK, Chen X, Patrick P, Lee C, Lou W. A hybrid Cloud Approach for Secure Authorized Deduplication IEEE, 2015 May 1; 26(5)1206 –16.
6. Bellare M, Keelveedhi S, Ristenpart T. Message-Locked encryption and secure deduplication, 2013.
7. Douceur JR, Adya A, Bolosky WJ, Simon D, Theimer M. Reclaiming space from duplicate files in a serverless distributed file system. 2002 Jul; 1–14.
8. Bellare M, Keelveedhi S, Ristenpart T. Dupless: Server aided encryption for Deduplication storage. 2013; 179–94.
9. Halevi S, Harnil D, Pinkas B, Peleg AS. Proofs of ownership in remote storage systems. 2011 Aug; 1–13.
10. Pietro R D, Sorniott A, Boosting efficiency and security in proof of ownership for deduplication, 2012 May 2. p. 81–2.
11. Bellare M, Namprempre C, Neven G. Security proofs for identity-based identification and signature schemes, 2009, pp. 1–18.
12. Bellare M, Palacio A. Proofs of security against impersonation under active and concurrent attacks. 2002 Sep 13; 162–77.
13. George C. Efficient Secure Authorized Deduplication in Hybrid Cloud using OAuth. 2015 Mar; 4(3):200–05.
14. Nandini J, Reddy RN. Implementation of Hybrid Cloud Approach for Secure Authorized Deduplication, 2015.
15. Li J, Chen X, Li M, Li J, Lee P, Lou W. Secure deduplication with efficient and reliable convergent key management. 2013; 25(6):1615–25.
16. Srinivas V, Vardhan MG. Authorized Duplicate Check Scheme in Cloud using Hybrid Cloud. 2015 Apr 18; 26(5):1206–16.