ISSN No: 2348-4845

International Journal & Magazine of Engineering, Technology, Management and Research

A Peer Reviewed Open Access International Journal

Authorized Data Deduplication Using Hybrid Cloud Architecture

S.Harisha
B.Tech (IT),
Dept of CSE,
MLR Institute of Technology,
Hyderabad.

D.Mahipal
B.Tech (IT),
Dept of CSE,
MLR Institute of Technology,
Hyderabad.

Mrs.Sujatha Rajamani
Assistant Professor,
Dept of CSE,
MLR Institute of Technology,
Hyderabad.

Abstract:

Data duplication is one of the major challenges faced by modern world, this paper is proposed in order to overcome such problem up to a certain extent. Data deduplication is used for compression and eliminating duplicate data. To protect data security this paper proposes authorized duplication check methods. A hybrid cloud is a combination of public and private clouds bound together by either standardized or proprietary technology that enables data and application portability. Proposed system aiming to efficiently solving the problem of deduplication with differential privileges in cloud computing.

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 hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model.

Keywords: Deduplication, authorized duplicate check, confidentiality, hybrid cloud

I. INTRODUCTION:

Cloud computing provides unlimited seamless virtualized amount of storage that can be accessed through internet throughout the world. Today’s CSP (cloud service providers) offer both highly available storage space and massive parallel computing resources at low cost. Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers.

To make data management scalable in cloud computing, deduplication has been a framed technique and has attracted more and more attention in recent times. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage.

However, previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files.

Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

II. CONVERGENT KEY TECHNIQUE:

Convergent encryption ensures data privacy in deduplication. A Convergent encryption technique has been proposed which enforces data confidentiality and provides compatibility for deduplication. Convergent key is used to encrypt and decrypt a data copy that is acquired by estimating and calculating the cryptographic hash value of the content of the data copy. A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the data copy, such that the tag will be used to detect duplicates. After key generation and data encryption, the users preserve the keys and cipher text is sent to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text is generated.

To prevent unauthorized access, a secure ownership proof protocol is required to provide the proof that user indeed owns the same file in case a duplicate is found. After the proof, consequent users with the same file will be allotted a pointer from the server without the necessity to upload the same file. The user is permitted to download the encrypted file from the server with a pointer, but can be decrypted only by the corresponding data owners with their convergent keys. Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts and decrypts a data copy with a convergent key and which is obtained by computing the cryptographic hash value of the content of the data copy.

Thus, convergent encryption permits the cloud to perform deduplication on the cipher texts and the ownership proof check prevents the unauthorized or illegal user to access the file. Previous deduplication systems could not support differential authorization duplicate check that is important in many applications. In differential authorization duplicate check, each user is assigned with set of privileges during system initialization. Each file which is uploaded to the cloud has set of privileges to specify which users are allowed to execute the duplicate check and file access. The user has to take the file and own privileges as inputs, before submitting the duplicate check request for some file.The user will be able to find a duplicate for the file if and only if there exists a copy of this same submitted file and a matched privilege stored in cloud. For example, in a company, different privileges will be assigned to employees. In order to save cost and manage the data efficiently, the data will be transferred to the storage server provider (SCSP) in the public cloud with specified privileges.

And to store one copy of the same file, the deduplication technique will be implemented. To provide privacy, few files will be encrypted and permitted for duplicate check by the employees with specified privileges to understand or grasp the access control. Traditional deduplication systems which are based on convergent encryption, provides confidentiality to certain extent but do not support the duplicate check with differential privileges. In other words, deduplication techniques that are based on convergent encryption technique, do not consider differential privileges. It leads to contravention, if one wants to realize both deduplication and differential authorization duplicate check simultaneously.

III. SECURITY ISSUES IN CLOUD:

The security will be analysed in terms of two aspects, that is, the confidentiality of data and the authorization of duplicate check. We suppose that all the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under this assumption, two kinds of adversaries are considered, that is, adversaries who aim to extract secret information as much as possible from both public cloud and private cloud, and internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. The data will be encrypted in our deduplication system before outsourcing to the storage cloud to maintain the confidentiality of data.

The data is encrypted with the traditional encryption scheme and the data encrypted with such encryption method which guarantees the security of data. System address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for Differential Authorization and Authorized Duplicate Check. Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any unauthorised user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server. Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs the duplicate check directly and tells the user if there is any duplicate.

The security requirements considered in two folds, including the security of data files and security of file token. Unauthorized users without appropriate privileges or file prevented from getting or generating the file tokens for duplicate check of any file stored at the Storage cloud. The users are not allowed to collude with the public cloud server. It requires that any user without querying the private cloud server for some file token, he cannot able to get any useful information from the token, which includes the privilege or the file information and to maintain the data confidentiality unauthorized users without appropriate privileges or files, prevented from access to the underlying plaintext stored at Storage cloud.

IV. PROPOSE AUTHORISED DUPLICATION CHECKER FOR DEDUPLICATION:

There are three entities defined in system, that is, users, private cloud and storage cloud service provider in public cloud as shown in Fig. 1. The Storage cloud performs deduplication by checking if the contents of two files are the same and stores only one of them and the access right to a file is defined based on a set of privileges. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denotes the tag with specified privileges. A user computes and sends duplicate-check tokens to the public cloud for authorized duplicate check.

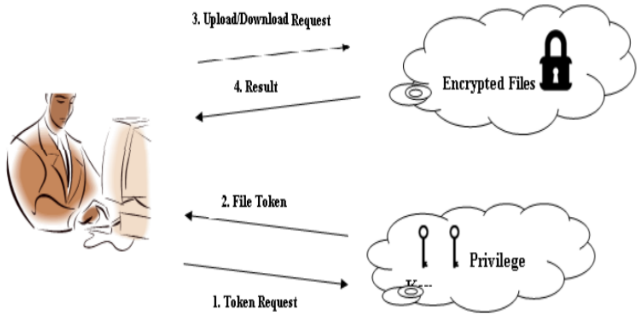


Fig 1: Architecture of Authorized deduplication

While Users have access to the private cloud server, a semi trusted third party which perform duplicable encryption by generating file tokens for the requesting users.

Storage Cloud

This is an entity that provides a data storage service in public cloud. The storage cloud service provider provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the storage cloud eliminates the storage of redundant data via deduplication and keeps only unique data.

Data User

A user is an entity that wants to outsource data storage to the S-CSP and access the data later when needed. In a storage system supporting deduplication, to save the upload bandwidth the user can only uploads unique data but does not upload any duplicate data, which may be owned by the same user or the different users. In authorized deduplication system, each user is issued a set of privileges in the setup of the system and each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

Private Cloud

The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users and this interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

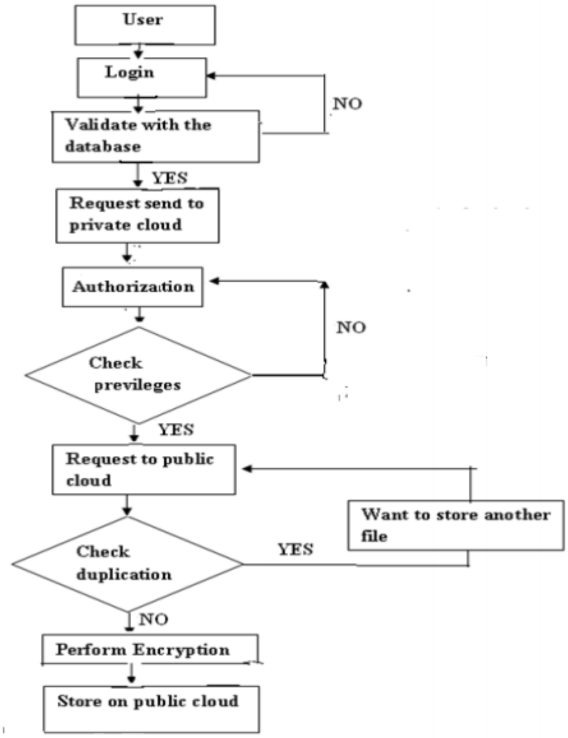


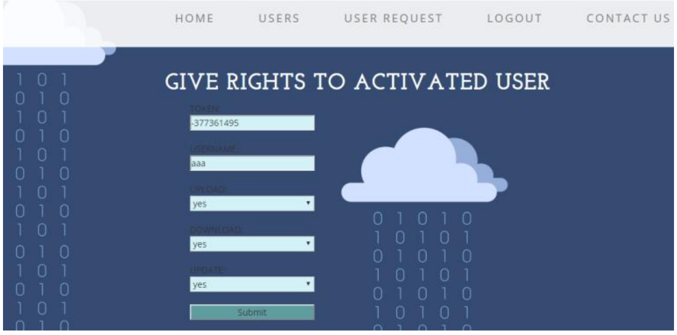
Fig 2: Flow Diagram of Proposed Method

In deduplication system, hybrid cloud architecture is introduced to solve the problem of unauthorized deduplication of file. The private keys for privileges will not be issued to users directly, which will be kept and managed by the private cloud server. The user needs to send a request to the private cloud server to get a file token. The user needs to get the file token from the private cloud server to perform the duplicate check for some file. The private cloud server also checks the user's identity before issuing the corresponding file token to the user.

The users perform the authorized duplicate check for this file with the public cloud before uploading this file. The user either uploads this file. If a file duplicate is found, the user needs to run the Proof of ownership protocol with the cloud storage service provider to prove the file ownership. Otherwise, if no duplicate is found then the data owner performs an identification to prove its identity with private key. If it is passed, the private cloud server will find the corresponding privileges of the user from its stored table list and send to the user then user can upload his files. The same way user can download his file from storage cloud.

V. EXPERIMENTAL RESULTS:

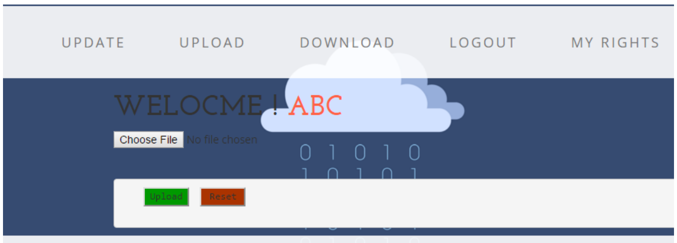
The final results of the designed system are given below. From those results we get the detailed information to Check de-duplication and upload the files, checking for Duplication, file uploading, file downloading. Detailed procedure of the proposed system is given. Based on this we confirm that securely authorized de-duplication is successfully achieved with hybrid cloud approach. The output images given as below,



GIVE RIGHTS TO ACTIVATED USER

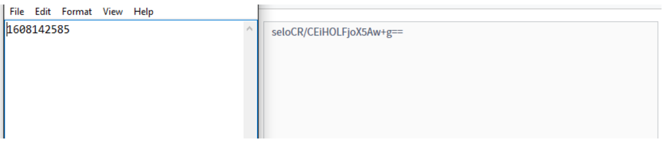
TOKEN: -377361495
USERNAME: aaa
UPLOAD: yes
DOWNLOAD: yes
UPDATE: yes

Fig 3: Token generation



WELOCME ! ABC

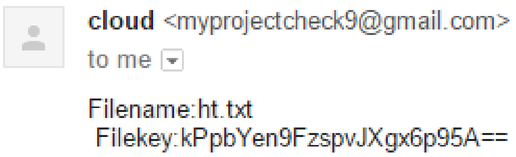
Fig 4: Data uploading to public cloud



```
1608142585
```

```
seloCR/CEiHOLFjoX5Aw+g==
```

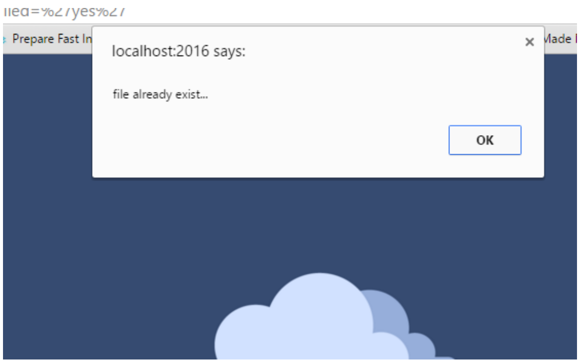
Fig 5: Data before and after uploading



cloud <myprojectcheck9@gmail.com>
to me

Filename:ht.txt
Filekey:kPpbYen9FzspvJXgx6p95A==

Fig 6: Key and file details sent to mail



localhost:2016 says:

file already exist...

OK

Fig 7: On uploading duplicate data

VI. CONCLUSION:

Hybrid clouds offer a greater flexibility to businesses while offering choice in terms of keeping control and security. Hybrid clouds are usually deployed by the organizations willing to push part of their workloads to public clouds either for cloud bursting purposes or for projects requiring faster implementation because hybrid clouds vary based on company needs and structure of implementation. In proposed system authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check system presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, the duplicate-check tokens of files are generated by the private cloud server with private keys. Proposed system is secure in terms of insider and outsider attacks specified in the proposed security model.

REFERENCES:

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou” A Hybrid Cloud Approach for Secure Authorized De-duplication” in vol: pp no-99, IEEE, 2014.

[2] OpenSSL Project. http://www.openssl.org/.

[3] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

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

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart. Messagelocked encryption and secure eduplication. In EUROCRYPT, pages 296– 312, 2013.

[6] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[7] M. Bellare and A. Palacio. Gq and schnorr dentification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.

[8] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. chneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[9] 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.

Author’s Details:

S.Harisha, B.Tech (IT), Dept of CSE,
MLR Institute of Technology, Hyderabad.

D.Mahipal, B.Tech (IT), Dept of CSE,
MLR Institute of Technology, Hyderabad.

Mrs.Sujatha Rajamani, Assistant Professor, Dept of CSE,
MLR Institute of Technology, Hyderabad.