ISSN 2395-1621

Secure Authorized Deduplication using Hybrid Cloud

[#1]Prof. P. D. Kale, [#2]Farheen Shaikh, [#3]Jyoti Giri, [#4]Pooja Shinde

[1]pranotikale2@gmail.com
[2]farheen.shaikh07@gmail.com
[3]jyotigiri81295@gmail.com
[4]pspujashinde@gmail.com

[#1234]Bharti Vidhyapeeth College of Engg For Womens, Pune.

ABSTRACT

Data de-duplication is one of the important technique which is used for eliminating duplicate copies of same data. This technique used in many organization where enormous amount of data is stored on cloud. It is used in cloud storage to reduce the amount of storage space as well as save bandwidth used while transferring data. For example same file stored in different places by many users, so it generates multiple copies of same file on server and this files contain much of same data. That's why we needed De-duplication technique which eliminates these multiple copies of same file.To solve this problem in our propose system ,we are going to use concept of Hybrid cloud. Hybrid cloud is nothing but the combination of private cloud and public cloud. private cloud provides the security for that data that means only the authorized person can upload and download the files from the public cloud for this purpose user generates the key and stored that key onto the private cloud. At public cloud data is first encrypted and then stored.

ARTICLE INFO

Article History

Received 21st March 2016

Received in revised form :

23rd March 2016

Accepted : 25th March 2016

Published online :

28th March 2016

***Keywords*:** De-duplication, Authorized duplicate check, Confidentiality ,Security, Hybrid cloud - private cloud, public cloud.

I. INTRODUCTION

A cloud refers to a set of hardware, networks and storage devices with combined capability of providing any useful service. A Cloud is essentially a connected universe of machines, a massive pool-up of resources. Cloud computing delivers infrastructure, software, platform for user. Cloud computing is based on pay-
per-use means we have to pay only what we used. It provides services such as
1.Platform As A Service,
2.Software As A service,
3.Infrastructure As A service.
Cloud computing is internet-based. It provides a network of remote servers connected over the Internet to store, share, manipulate, retrieve and processing of data, instead of a local server or personal computer. It is combination of hardware, network and storage devices to fulfil services.
The benefit of cloud computing is to access data from anywhere at any time, user can work from anywhere. The most important thing is that customer doesn't need to buy the resource for data storage. when we discuss about security then authentication is very important which ensures that only authorized person can access cloud services.One critical challenge of cloud storage services is handling large amount of data, in that contains repeated data. To overcome this challenge we used Data de-duplication technique. It is a special type of data compression technique for eliminating duplicate copies of repeating data in cloud storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy.

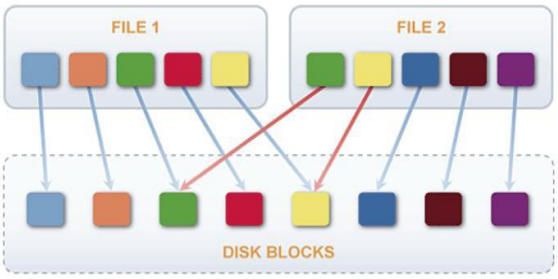


figure1: Example of De-duplication

Here same block from two files will be store only once in public cloud.

Hybrid Cloud

Hybrid cloud is nothing but the combination of private cloud and public cloud. Using concept of hybrid cloud to avoid duplicate copies of same data and to provides security to users data on public cloud which is not secure. As cloud computing becomes famous, an increasing amount of data is being stored in the cloud and used by users with specified privileges, which define the access rights of the stored data.

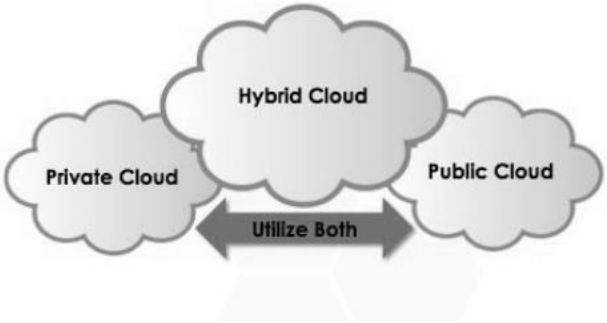


figure2:Cloud model

Inside a hybrid cloud, using the data de-duplication the encryption will become simpler. As we all know that the network is consist of Redundant amount of data, which is being shared by users and nodes in the network. Many large scale networks use the data cloud to store and share their data on the network. The node or user, which is present in the network have full rights to upload or download data over the network.

II. PROPOSED SYSTEM

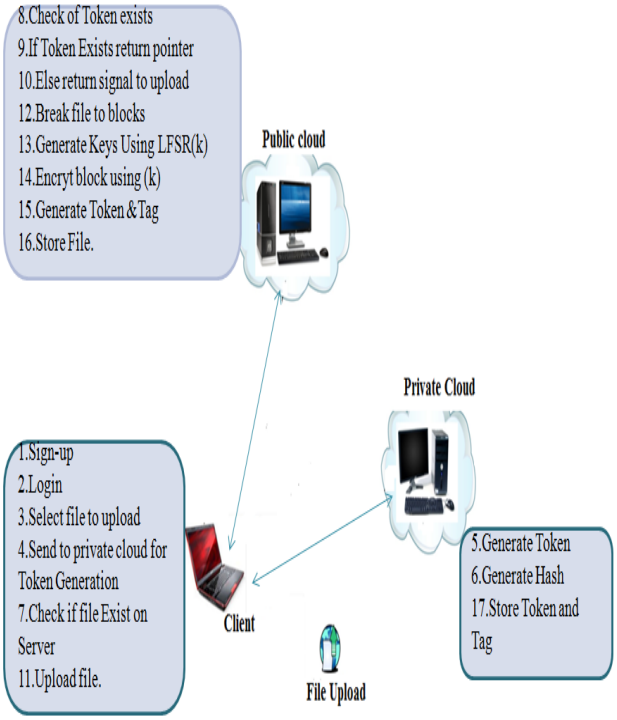


Figure3: Proposed system architecture

In our system we implement twin cloud approach that includes the public cloud and the private cloud.
In private cloud we used SHA-1 algorithm for calculating hash values of uploaded file by clients.
In the public cloud we use LFSR algorithm for encryption of data.When we upload file first time, file tokens are generated by private server for that file, after breaking files into tokens and calculate hash values using SHA-1 algorithm. This hash is send to public server through the client. On public server comparisons done between existing hash and current hash. If current hash already exist then that block does not stored else block will be store.

III.IMPLEMENTION

There are three entities in our proposed system.

1. User

First user need to register in private cloud for storing token of respective file .
If user want to access that file first he access respective token from private cloud and then use this token to access his files from public cloud. Token consist of file content f and convergent key Kf.
Implementation of the **Client** contains following function calls.

- File_Tag(file) - It calculates hash of the file as File Tag.

- Token_Req(Tag, UserID) - It requests the Private Server for File Token generation with the File Tag and User ID.

- Dup_Checker(Token) - It requests the StorageServer for Duplicate Check of the File by sending the file token received from private server.

• File_Upl_Req(FileID, File, Token) - It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.

2. Private Cloud
Most of the time user can use the private cloud instead of public cloud for more security. User store the generated key in private cloud. Private cloud only stores the convergent key with respective file.Implementation of the **Private Server** includes following function calls.

• Token_Gen(Tag, UserID) - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1 algorithm.

• Store_Hash(key) - It stores calculated hash on private cloud.

3. Public Cloud:
We use Public cloud for the storage purpose.
Implementation of public cloud includes following function calls.

• Duplicate_Check(Token) - It compare the file tokens for finding Duplicate.

• File_Store(FileID, File, Token) - It stores the File on Disk and update on public cloud.

IV. PRELIMINARIES

This section defines the notation used in the functions and algorithm.

Acronym	Description
S-CSP	Storage cloud service provider
POW	Proof of Ownership
F(Key)	Convergent encryption key for file F
H	Hash function

Table 1

1. Symmetric encryption
In Symmetric encryption a common secret key *κ is* use to encrypt and decrypt information. It consists of three primitive functions:
1.KeyGenSE() → κ
It is the key generation algorithm that generates κ using security parameter
2.EncSE(κ,M) → C
It is the symmetric encryption algorithm that takes the secret κ and message M and then outputs the ciphertext C
3.DecSE(κ,C) → M
It is the symmetric decryption algorithm that takes the secret κ and ciphertext C and then outputs the original message M.

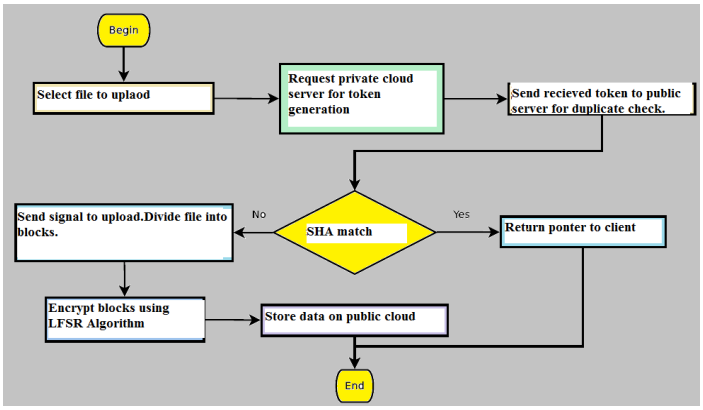
2. Convergent encryption
In traditional encryption different users encrypt data with their own key so for same data different cipher text is generated which makes de-duplication difficult.
Convergent encryption, also called as content hash keying, is used to produces identical cipher text from identical plaintext files.The simplest implementation of convergent encryption can be defined as: Bob derives the encryption key from his file F such that K = H (F), where H is a cryptographic hash function. Convergent encryption scheme can be defined with four primitive functions:

1. KeyGenCE (M) -> K is the key generation algorithm that maps a data copy M to a convergent key K;
2. EncCE (K, M) ->C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a ciphertext C;
3. DecCE (K, C) ->M is the decryption algorithm that takes both the ciphertext C and the convergent key K as inputs and then outputs the original data copy M
4. Tag Gen (M) -> T (M) is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

3. Proof of ownership
The notion of proof of ownership enables users to prove their ownership of data copies to the storage server. Specifically, PoW is implemented as an interactive algorithm run by a prover and a verifier. The verifier derives a short value ϕ(M) from a data copy M. To prove the ownership of the data copy M, the prover needs to send ϕ′ to the verifier such that ϕ & prime;= ϕ(M). The formal security definition for PoW roughly follows the threat model in a content distribution network, where an attacker does not know the entire file, but has accomplices who have the file. The accomplices follow the "Bounded retrieval model", such that they can help the attacker obtain the file, subject to the constraint that they must send fewer bits than the initial min-entropy of the file to the attacker

4. S-CSP
S-CSP used in the public cloud storage. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data.

Uploading mechanism:



V. Test Result

We have implementing our project on three machines having minimum i3 processer. One of them is client who uploads file, other is private cloud who calculates hash value of files, last one is public cloud for storing encrypted data.

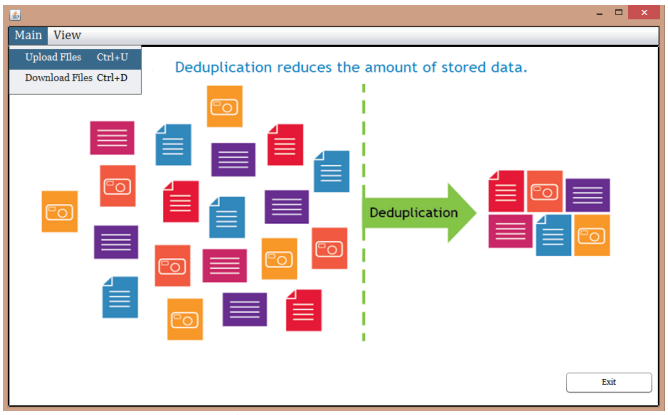


figure4:File uploading

1.When we upload file first time, file tokens are generated by private server for that file and it will stored on public sever.

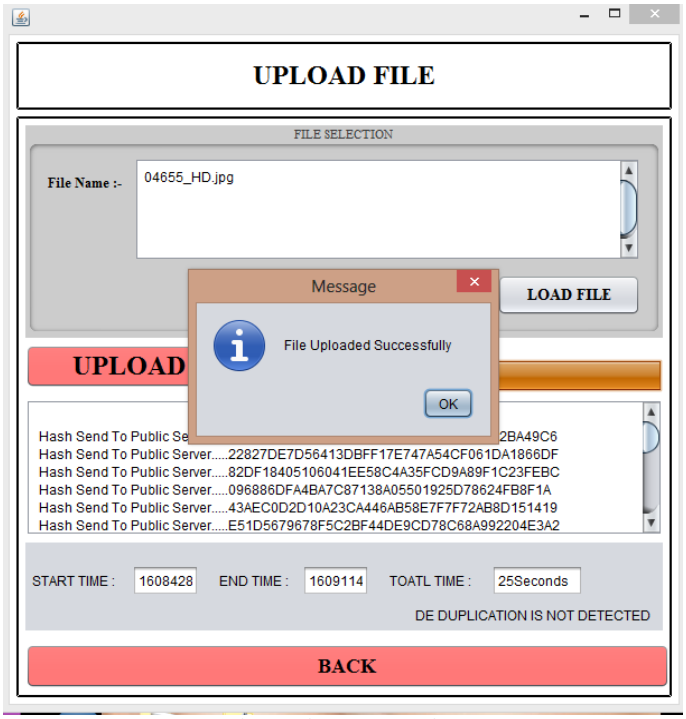


figure5:file uploading first time

2. Next time if we try to upload same file by renaming it then it will not store. it gives massage as "De-duplication is detected".

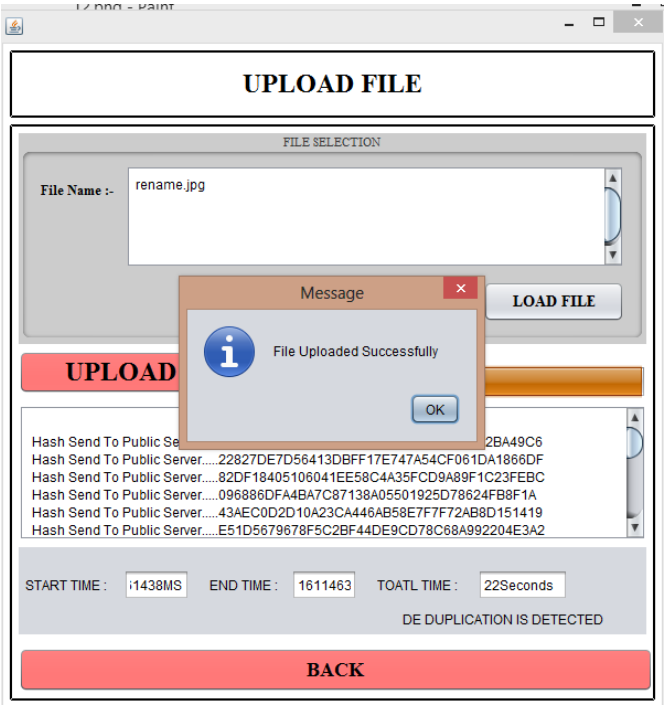


figure6:file uploading second time

V. Conclusion

In this paper we presented the idea of de-duplication which is very important in now a days. Users key is stored on private cloud which ensures the security of our system as well as data is encrypted before storing on public cloud.

REFRENCES

1. Jin Li et al "A Hybrid Cloud ApproaPh for Secure Authorized Deduplication" IEEE Transaction on Parellel and Distributed System VOL:PP NO:99 YEAR 2014.

2. Bhushan Choudhary, Amit Dravid ,"A Study on Authorized Deduplication Techniques in Cloud Computing ",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, April 2014.

3. M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Serveraided encryption for deduplicated storage", USENIX Security Symposium, 2013..

4.Halevi,et al, “Proofs of ownership in remote storage systems”. ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

5. Pasquale Puzio et al, “ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage”

6. J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

