International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 6 203 - 207

# Secure Authorized Deduplication for Hybrid Cloud Storage

Bhavanashri S Raut
Department Of Computer Engineering,
Jayawantrao Sawant College Of Engineering,Hadapsar
Savitribai Phule Pune University,
Maharashtra.
e-mail:bhavanashriraut6@gmail.com

Prof. H. A. Hingoliwala
Department Of Computer Engineering,
Jayawantrao Sawant College Of Engineering, Hadapsar
Savitribai Phule Pune University,
Maharashtra
e-mail:ali_hyderi_@yahoo.com

***Abstract*—** Cloud computing provides number of applications, as utilities in the internet .This applications create, configure and customize accessing referring the cloud computing as online utility. Cloud computing offers online data storage, infrastructure, services over networks and applications. Cloud storage is an widely popular offering of cloud computing. Cloud storage is used for increasing the number of users, access the users data from anywhere and also space for data storage in computing. Data deduplication means a type of data compression. This data compression to reduce its storage requirement using encoding of data. Data deduplication method used for replacement of multiple copies of data or eliminating duplicate copies of data. It also reduce storage space and save bandwidth. Deduplication is having one of the advantage for new security and privacy challenges with high cost. Basically, data deduplication means of reducing storage space in cloud. In this paper certain improves the speed of data deduplication with encrypted data reduces the cloud storage capacity of data. This paper first to show that addressing the problem of secure authorized data deduplication. Data deduplication works by eliminating data and ensuring that only one unique instance of data. Hence,Data deduplication is also called as single instance storage, because of the differential privilege of users considered in duplicate check. In this paper we implement that deduplication with encrypted data using SHA and MD5 algorithm for hybrid cloud storage.

***Keywords-*** *Deduplication, authorized duplicate check, confidentiality, hybrid cloud, security, private cloud, public cloud.*

*****

## 1. Introduction

Hybrid cloud computing with Deduplication gives a cost effective, measureable , robust location-independent infrastructure for data storage. This model of deduplication is used for data management and storage. Hence, lots of people pay attention to economize the capacity of cloud storage. Therefore how to utilize the cloud storage capacity becomes important issue compared to others .

## 2. Background

Deduplication is a specialized data compression model. This model is used for eliminating duplicate copies of repeating data. A Hybrid cloud is a made up of private cloud and public cloud. Some important data is present only in the enterprises private cloud . The other data is stored and accessible from a public cloud.

Public cloud describes cloud computing in the most common form. In public cloud ,resources are relatively accurate, worked on by the subject using it in the cloud over internet via web application or web services. The third party provider who shares resources and bill on relatively accurate utility computing basis in the public cloud.

Private cloud and internal cloud are synonyms. Private cloud describes cloud computing on private network. Hybrid clouds characterized by fast, most reliable and scalable as compared to other types of clouds. It is also potential cost savings of public clouds with the security. Hybrid cloud increases control and management of private clouds. Dedu-plication are of two types as follow. This types differentiated by the type of basic data units.

1) File-level deduplication: A file is a data unit. While examining the data of duplication, it typically uses the hash value of the file used as its identifier. If two or more files have the same hash value then they are assumed to have the same contents and only one with hash files will be stored.

2) Block-level deduplication: A file is stored into several fixed-sized blocks or variable-sized blocks of data. It computes hash value for each block for examining the duplication blocks. SO that we can easily make out that data stored is absolutely unique.

## 3. Literautre Survey

1) The paper written by author Paul Anderson about encrypted deduplication , he wrote about the security of data and users to increase the speed of backup and reduce the storage require-ment and support client-end confidentiality. It also supports a unique feature which allows immediate detection of common subtrees. It avoids the need to query the backup system for every file.

2) As author Mihir Bellere ,suggest that it pro-vides secure deduplicated storage residing Brute-Force attack and realize it in a system such as DupLESS. DupLESS uses PRF proto-col to encrypt client under message-based keys obtained from a key-server. With an existing service, It enables clients to store encrypted data. Deduplication perform service and still achieves strong confidentiality guarantees of storage.

3) The author Pasqualo Puzio, suggest that cost-wise deduplication with new security and pri-vacy challenges in deduplication technique is quite effective . ClouDedup, a secure and ef-ficient storage service. This storage service assures block-level deduplication and data con-fidentiality at the same time. It is based on convergent encryption. ClouDedup remains se-cure because of definition of a component that implements an additional encryption operation and an access control mechanism.

4) The author Iuon-Chang ,Lin Pio-ching given as improves the speed of data deduplication . Up-loaded file is verified for integrity by specially computed signature. In this paper to implement Zhang Fault Tolerant digital Signature Scheme. The scheme proposed by Zhang can detect and correct errors efficiently in digital signature. It based on top of Zhangs scheme. This paper proposes a novel data deduplication method to improve not only the utilization of cloud stor-age capacity but also the speed of deduplication in cloud storage.

5) Shai Halevi, Danny Harnik, Benny Pinkas,"Proof of Ownership in Remote Storage System", introduce the notion of proofs-of-ownership (PoWs).Which lets a client efficiently prove to a server that the client holds a file, rather than just some short information about it. The concept of proof-of-ownership, under rigorous security definitions and efficiency requirements of Petabyte scale storage systems. We present solutions based on Merkle trees, specific encodings and analyze their security.

6) The author M.Shamala Devi and A V khanna to suggest that the to optimize the private cloud storage backup in order to provide high throughput to the users of the organization by increasing the deduplication efficiency. It is highly desirable to improve private cloud backup storage efficiency by reducing the deduplication time.

7) The author Ee-chien chang, proposed that a secure client-side deduplication scheme. In that addressed an important security concern in cross-user client-side deduplication. In his scheme convergent encryption and custom en-cryption methods are not semantically se-cure.This is disadvantage of the paper.

8) Author Kehuan Zhang, Xiaoyong Zhou, Yangyi Chen and XiaoFeng Wang a new, to preserve data privacy and to minimize computational overheads ,generic secure computing frame-work needs to be built to support automatic splitting of a data-intensive computing job and scheduling of it across the public and private clouds. Sedic includes a privacy aware execu-tion framework that automatically partitions a computing job according to the security levels of the data and distributes the computation between the public and private clouds. Sedic is based on MapReduce, which includes a map" step and a reduce" step: the map step divides input data into lists of key-value pairs and as-signs them to a group of concurrently-running mappers; the reduce step receives the outputs of these mappers, which are intermediate key-value pairs, and runs a reducer to transform them into the final outputs.

9) The author Qingji Zheng ,Shouhuai Xu show, somewhat surprisingly, that the two aspects can actually co- exist within the same framework. This is possible fundamentally because of the following insight: The public verifiability of-fered by Proof of Data Possession ( PDP ) and Proof of Retrievability ( POR) schemes can be naturally exploited to achieve POW.

## PROPOSED SYSTEM

In the proposed system we are providing proof of data for authorized data deduplication with the help of data owner. Same proof of data is used at the time of uploading of the file. Firstly each file is uploaded to the cloud. This uploaded file is also provided with the set of instructions . This specifies which kind of users is allowed to perform the duplicate check and access the files in storage. Before submitting his/her duplicate check request for some file, after the user needs to take this file and his/her own instructions as inputs. Hence, the user is able to find a duplicate copies for this file if and only if there is a copy of this file and a matched instructions stored in cloud storage for deduplication. Fig 3.1 Show that Proposed System Framework Communication between client , Private and Public Cloud.

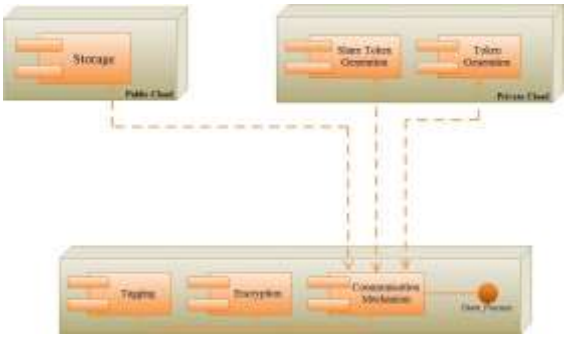


Fig. 3.1: Proposed System Framework Communication

In this work Public Cloud is used for storage data. Private Cloud is used for performance the operations like share token, token generation. Client has been performed operations like tagging of file , encryption of file and communication between private and public cloud.

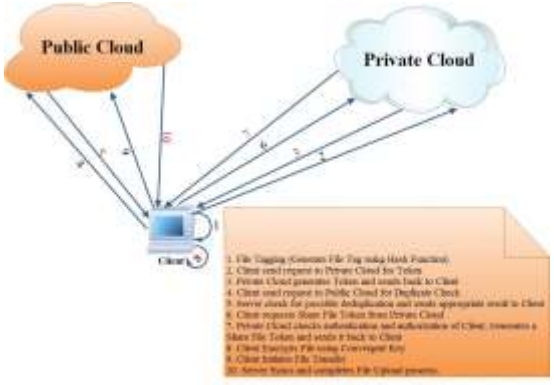


Fig. 3.2: Proposed System of Operations

### 3.1 Encryption of files

To avoid duplication of files , files are encrypted and decrypted using cipher text.

To do encryption and decryption from plain text to cipher text and vice versa respectively, we used three basic function:

1. Key generation algorithm(k):

   Where k is generated using security parameter 1.
2. Symmetric encryption algorithm(k,M):
   Where it takes Secrete key k and message M as input and

output is cipher text C.

3. Symmetric decryption algorithm(k, C):

Where it takes secrete k and cipher text C as input and output is original message M. Encryption helps for data confidentiality in data deduplication.

Convergent key with original data copy helps the users to avoid duplication of data. For the same tag is used to detect duplicate data.

### 3.2 Design Goals

Proposed a new deduplication system follows:

#### 3.2.1 Differential Authorization:

Every authorized individual get a token for his /her respective file from cloud. As the basis of his /her instructions. Other unauthorized or common individual can not get a token for duplicate check based as in the instructions.

#### 3.2.2 Authorized Duplicate Check:

Authorized user has a private key. This private key generate query for certain file with help of instructions he owned in his own private cloud. In public cloud duplicate check directly happen to detect the duplicate copies of data.

3.2.3 Unforgeability of file token/duplicate-check token:
Unauthorized user cant proper instructions will not generate token for respective file. This file taken is used for duplicate check of any file in cloud. Du-plicate check token are generated in private cloud.

Prototype of proposed authorized deduplication system is as follows:

In this model, three entities are used as seperate java socket programs.
A Client Program: It is used by data users. Here file is uploaded in cloud storage.
Private Server: It is used as private cloud. Two things are done and those are not only managing private keys but also computing file token.

Storage Server: It is used to store and deduplicate file in storage.
In this prototype ,client provides support for token generation and deduplication of the file upload process.
FileTag(File) Generate file tag using hash function. It computes SHA-1 hash of the File.

TokenReq(Tag, UserID) In this function Client send requests to the Private Server for File Token generation with the User ID and also File tag;
DupCheckReq(Token) In this function client send requests to public cloud for duplicate check of the file. Its by sending the file token received from private server.
ShareTokenReq(Tag, fPriv.g) In this func-tion Client send requests the Private Server to generate the Share File Token. With their File Tag and Target Sharing Privilege Set.
FileEncrypt(File) This function client en-crypts the file using Convergent Encryption key.
FileUploadReq(File ID, File, Token) It up-loads the File Data to the Storage Server. Then if the file is Unique and updates the File Token stored in cloud storage.
TokenGen(Tag, UserID) - It loads the assoc-ated privilege keys of the user. To generate the token with HMAC-SHA-1 algorithm.

### 3.3 SHA-1 Algorithms:

The SHA Algorithm is a cryptography hash func-tion. It is used in digital certificate and also data integrity. It is used for computing a compressed representation of a message or a data file. SHA is a fingerprint for use with digital signature applica-tions. The message which is less than 264 bit in length. Here, secure hash algorithm works with that type of messages. Message digest is the output of SHA and length of these type of messages is 160 bits (32 bits extra than MD5).

Message digest is the output of SHA and length of these type of messages is 160 bits (32 bits extra than MD5) .MD5 processes a variable-length message into a fixed-length output of 128bits.

### 3.4 Message Digest Algorithm Steps:

This algorithm is based on message length. It re-quires 8 bit of message length and too fast but also take long message. Message digest algorithm used for hash generating with the help of SHA Algorithm.

Step 1:- Padding bits and Append Length Step 2:- Divide the input into 512-bit blocks Step 3:- Initialize Channing variables

Step 4:- Process blocks
Step 5:- Hashed Output

**Uploading And Downloading Algorithm**

**FOR UPLOADING A FILE**

BEGIN
Step –1 Read file
Step –2 Cloud server checks for duplication
Step –3 Sends duplication response whether the file already exists or not
Step – 4 If the file does not exist
4.1 Display “file does not exist”
Step – 5 Then it uploads the file
Step – 6 If the file already exist
6.1 Display “file already exist”
END

**FOR DOWNLOADING A FILE**

BEGIN
Step –1 Read file
Step –2 Cloud server checks for duplication
Step –3 Sends duplication response whether the file already exists or not
Step –4 If the file exist
4.1 Display “file exist”
Step –5 then it downloads the file
Step –6 If the file does not exist
6.1 Display “file does not exist”

END

## 4. EXPERIMENTAL RESULTS

### A. *Server Started and user registration*

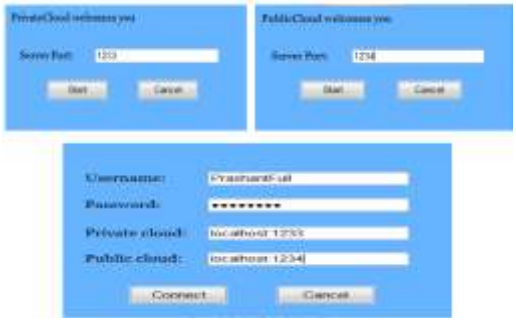


Figure1: Connect All Port and Start Connectivity

### B. *Registration Sucessful and server start*


Figure2: Private Cloud (Server) started


Figure: Public Cloud(Server) Started


Figure:File Uploading

### C. *User Choose Access Specifier*


Figure: Choose the Access Specifier

### D. *Types Of Aceess Specifier*


Figure :User Accessing Level


Figure: Completed Upload File

### E. *UploadinfgFile for Block Level Deduplication*


Figure: File Upload for Block level Deduplication

*F. Chechout the Block level deduplication*


Figure:Duplicate contents check


Figure: Duplicate Contents check

*G. Graphical representation of file levl deduplication*

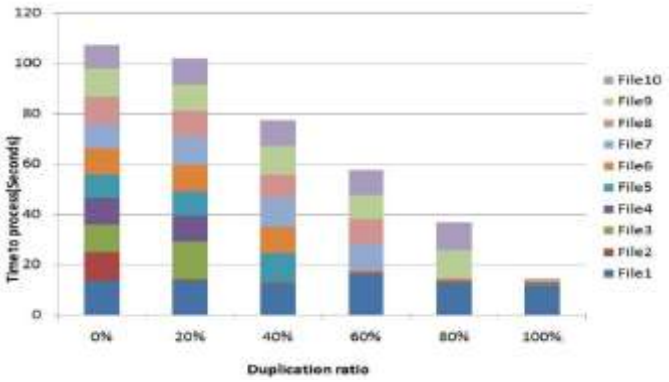


Figure: Statistics Graph on File Level De duplication

## 5. CONCLUSION

Cloud computing has reached a maturity that leads it into a productive phase. It has been addressed to a
degree that clouds have become interesting for full commercial exploitation of data. This however does
not mean that all the problems listed above have actually been solved, only that the according risks can be tolerated at certain limit. For better confidentiality and security in cloud computing we have proposed new deduplication constructions supporting authorized duplicate check in hybrid cloud system architecture. In which the duplicate-check tokens of files are generated by the private cloud server with their private keys. Proposed system includes proof of data owner. So it will help to implement better security issues in cloud computing.

## REFERENCES

[1] P. Anderson and L. Zhang. “Fast and secure laptop backups with encrypted de-duplication”. In *Proc. of USENIX LISA*, 2010.
[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. “Dupless: Server aided encryption for deduplicated storage”. In *USENIX Security Symposium*, 2013.
[3] Pasqualo Puzio, Refik Molva ,MelekOnen ,”CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage”, SecludIT and EURECOM,France.
[4] Iuon –Chang Lin, Po-ching Chien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)
[5] Shai Halevi, Danny Harnik, Benny Pinkas,”Proof of Ownership in Remote Storage System”, IBM T.J.Watson Research Center, IBM Haifa Research Lab, Bar IIan University,2011.
[6] M. Shyamala Devi, V.Vimal Khanna,Naveen Balaji ”Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup”,IACSIT, August,2014.
[7] J. Xu,E-C. Chang and J.Zhou.“Weak Leakage-Resilient Client –Side deduplication of Encrypted Data in Cloud Storage” Institute for Info Comm Research,Singapore,2013
[8] Tanupriya Chaudhari , Himanshu shrivastav, Vasudha Vashisht, ”A Secure Decentralized Cloud Computing Environment over Peer to Peer”,IJCSMC,April,2013
[9] Mihir Bellare, Sriram keelveedhi,Thomas Ristenart ,”DupLESS: Server Aided Encryption for Deduplicated storage” University of California, San Diego2013