International Journal of Security and Its Applications
Vol. 10, No. 5 (2016) pp.17-26
http://dx.doi.org/10.14257/ijsia.2016.10.5.02

Achieving Secure Deduplication by Using Private Cloud and Public Cloud

Guljar P. Shaikh[1], S. D. Chaudhary[1], Priyanka Paygude[1], Debnath Bhattacharyya[2]

[1]*Information Technology Department,*
Bharati Vidyapeeth Deemed University College of Engineering,
Pune-411043, India
[2]*Department of Computer Science and Engineering,*
Vignan's Institute of Information Technology,
Visakhapatnam-530049, AP, India
guljarshaikh042@gmail.com, {sdchaudhary,pspaygude}@bvucoep.edu.in
debnathb@gmail.com

Abstract

At the present time, cloud computing furnish large volume of area for storage of data as well as huge equivalent computing at affordable rate. Due to advantages of cloud computing it turn out to be widespread; extreme quantity of data can be stored on cloud. But, rise in size of data has raised numerous new obstacles. Deduplication is one of significant compression method of data for reducing carbon copy of replicating data that being used in cloud to reduce the volume of storage area and it utilizes less network bandwidth. Data Deduplication is the convergent encryption technique that has been projected to encrypt the data before it's been sent out that preserve the confidentiality of responsive data. For better data security, it makes the initial effort that officially point out the dilemma of authorized data Deduplication. There are numerous new Deduplication structures that provides authorized duplicate check in hybrid cloud structural design that acquire negligible overhead over standard operation.

Keywords: *Deduplication, Security, data confidentiality, duplication authorized check, hybrid cloud*

1. Introduction

The Cloud computing offers apparently limitless resources that enables user as service crossways complete web, whereas masking platform and implementation particulars. Cloud storage SP offer highly available storage as well as computing resources at moderately short rates. Due to availability of cloud computing, a growing volume of data get saved on cloud & retrieve by users using unique stated privileges whenever needed. One of the major threat of cloud storage is to maintain the rising quantity of data. As a result how to make use of the cloud storage capacity well turn out to be important matter nowadays. To make data management more scalable in cloud computing, de duplication become famous scheme that diverting attention of more users. Data Deduplication is dedicated compression method for data that reduce replicas of duplicate data. This method for better storage space utilization and also can applied on network for transferring data to decrease the byte amount that is to be transmitted. Rather than maintaining numerous data copies with same content, Deduplication remove copied data keeping behind a unique copy. Using a Hybrid Cloud method we can achieve better results of Deduplication. It is nothing but the combination output of private clouds and public clouds. It offers consistency, scalability, potential cost savings & quick deployment with the improved control, superior security & managing of clouds that are private in nature. CE offers data

ISSN: 2005-4262 IJGDC

confidentiality at the same time produce Deduplication as realistic. Encryption and decryption performed by the key derived from file and that key is named as CK whereas it obtained by using hash function. After the key generated and data is encrypted, owners protect those keys then send cipher text to cloud. To keep away from not permitted access to the files, a safe or secured proof of ownership i.e. POW protocol is necessary for the confirmation that the user certainly owns the identical file when a replica/copy is found. Behind the confirmation, successive users with the similar data file will be given "pointer" that is available to user from the server without requiring uploading the equivalent file. With that pointer a user can download the file which is encrypted from server. And that file is decrypted by respective CKs only. Therefore, CE will permit the cloud to execute Deduplication on the cipher texts whereas PoW check for not permitted user to have access to file.

2. Related Area

Following are the list of papers we have referred for our proposed work.

2.1. A Hybrid Cloud Approach for Secure Authorized Deduplication

They explored new method of Deduplication which support authorized duplicate check in the Public and Private Cloud structure, to provide better data security in the cloud [1].

2.2. Secure Deduplication with Efficient and Reliable Convergent Key Management

It addresses the problem of achieving efficient and reliable key management in secure Deduplication. It implements Dekey using the Ramp secret sharing scheme that enables key management to adapt to different reliability and confidentiality levels [2].

2.3. Message-Locked Encryption and Secure Deduplication

MLE contributes a technique to acquire protected Deduplication i.e. capacity-well organized for secure storage, an objective currently intended by various providers of cloud-storage [3]. They give the definitions for privacy as well as for integrity that they called as a tag consistency.

2.4. DupLESS: Server-Aided Encryption for Deduplicated Storage

Dupless is an architecture that provides protected storage defies brute-force attacks for duplication. Clients encrypt under message-based keys by an oblivious PRF protocol. It offers clients to save that data with an obtainable service, however accomplish strong confidentiality [4].

2.5. Proofs of Ownership in Remote Storage Systems

They have offered solution based on specific encoding and Merkle Trees [5]. Those are used for identifying attacks that exploit client side deduplication attempts to recognize deduplication. In POWs term Client proves to server that it in fact holds the whole data of the file rather than just short information about it [6].

3. Proposed Work

Here in this new deduplication technique we have used Hybrid Cloud structure to eliminate previously encountered issues. Authorized Duplicate Check is performed to indentify replicas from stored Files.

3.1. Deduplication

The Deduplication is an expert compression method for removing duplicate copies of similar data/chunks contents from file in the cloud. Deduplication eliminates unnecessary data by maintain the single original unique copy and use reference pointer of this copy from other redundant data rather than keeping many copies which contains the alike content type of data.

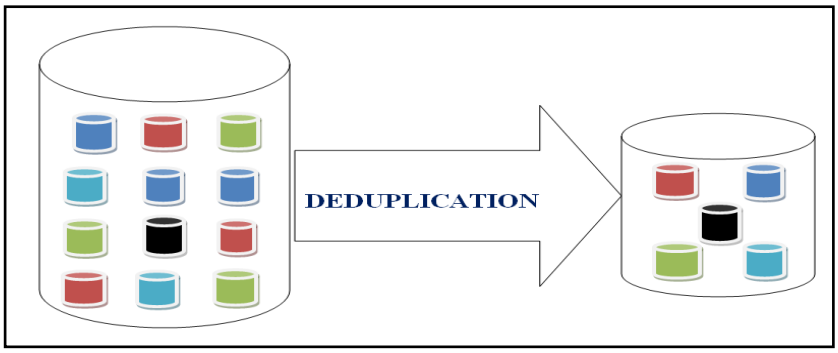

Figure1. Deduplication Process

It has two main approaches can be classified as follow:

3.1.1. Deduplication at File-level: A file is nothing but a data unit. While investigating duplication in file, it characteristically uses hash function and generates hash value for file which treated as file identifier. In case that more than one file has the same hash value, they are considered to have the similar contents and only one unique file of these files will be stored or saved in Storage.

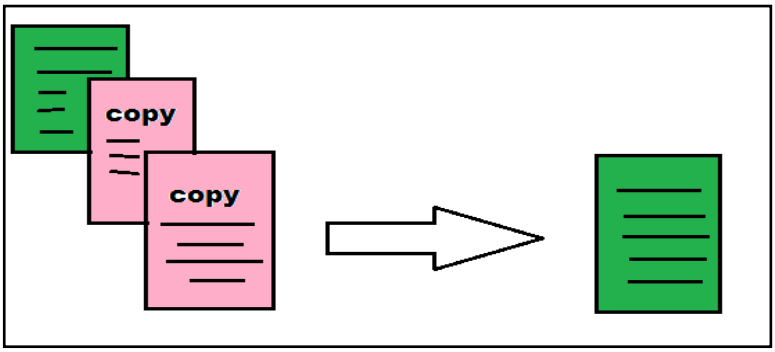

Figure 2. File Level Deduplication

3.2.2. Deduplication at Block-level: In this kind of Deduplication, it divides entire file into numerous fix size chunks or variable size chunks then it calculates a hash value for each of the chunk for investigating duplication blocks. Block-level Deduplication reduces duplicate chunks of files that may occur in non- alike files.

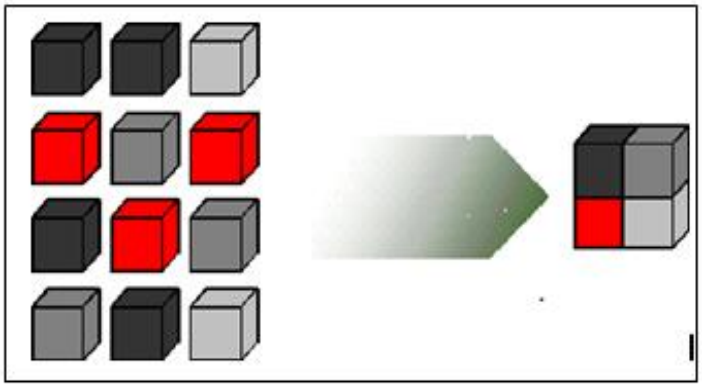

Figure 3. Block Level Deduplication

3.2. Hybrid Cloud Architecture for Secure Deduplication

Combined form of the public and private cloud is said to the Hybrid cloud which is launched for resolving the difficulties in security concern. To execute the checks on the Deduplication i.e. duplicate checks for the several files, it is necessary for users or the vendors that it should obtain the token from the particular file or data copy which is from the cloud server is a private one. And this private CS will furthermore make sure about the user's/owner's identity prior to issuing the equivalent file-token to that particular user. The duplicate check is authorized for that file which can be operated by the user by means of the public cloud earlier than uploading that file to the cloud. Based on result generation by performing the duplicate checks, the user either decides to upload this file or to run the PoW. The CE technique or the method has been introduced for encrypting the data before its been outsourced to the cloud.

At the top stage, a setting of attention is a venture network, having of a collection of united clients those will utilize the functions or the services provided by S-CSP and then it will store the data or the chunks by applying Deduplication techniques. Here In this system, we can say the Deduplication performed on the data or the files can be often used within this environment for backing up the data and also for the purpose of disaster recovery, at the same time as very much reducing space for storing the data. Such systems are extensive and are frequently more appropriate to the vendors for having the file backups as well as for organization applications in comparison of better-off storing concept.

In hybrid architecture of the cloud, as shown in Figure 4, three elements or entities are characterized and those are:

- A users or owners,
- Private cloud and
- S-CSP i.e. service provider which is located inside of the public cloud.

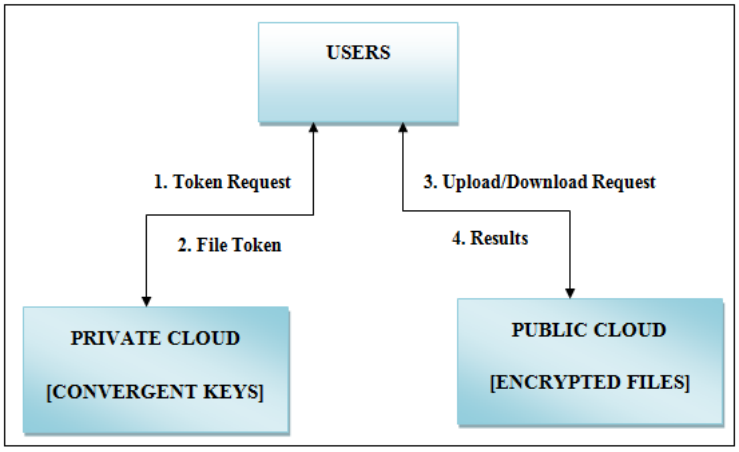

Figure 4. Architecture of Authorized Deduplication

3.2.1. S-CSP: Storage provider is nothing but kind of an entity which makes available storage services of data inside a public cloud. S-CSP offers outsourcing data services and then it stores that data in support of a users of the data. To trim down the cost of storage, the Service Providers eradicates the storing of an unneeded data by using Deduplication technique or methods whereas it also maintains only distinctive form of data rather keeping all the files having similar content.

3.2.2. Users: The owners or users are those who would like to outsource their data in public cloud to S-CSP then it access that stored data later whenever required. In this system of storage, in support of Deduplication method, a user uploads only unique single data copies/files although it is impossible for them to upload any duplicated data files.

3.2.3. Private Cloud: In contrast to previous Deduplication i.e. traditional Deduplication system/architecture in the cloud, this is *i.e.*, Private cloud concept is a fresh entity launched for smooth the progress of consumer's secure utilization of service provided by cloud. In particular, the fact that the computing assets at user's side which are controlled moreover a public cloud is not that totally trusted in practical practice, while a private cloud is capable to make available data/file owner with implementation surroundings and make an infrastructure work effectively as a boundary between the owner and a public cloud.

Table 1. Notions used in this Paper

Acronym	Description
S-CSP	Storage Cloud Service Provider.
pkU & skU	Public and Secret Key of User's.
PoW	Proof of Ownership.
CKs and PKs	Convergent and Privilege (Private) keys

3.3. Execution

Here we are representing Pseudo code for algorithm which we have preferred.

3.3.1. Deduplication Algorithm.

Input: File F

Output: Unique File uploaded

a. Begin DeDuplicatation

b. File Level Deduplication []

DupCheckReq($\emptyset' fp$)= DupCheck($\emptyset fp$, F)

c. For each (Contents in File[]) do

d. SHA-1← H(F)//calculate Hash value

e. End do

f. If (F{$\emptyset fp$,$\emptyset' fp$} = F{$\emptyset fp$, $\emptyset' fp$}) then

g. Uploading denied.

h. Else

i. (F{$\emptyset fp$,$\emptyset' fp$} ≠ F{$\emptyset fp$, $\emptyset' fp$}) then

j. F←Sks

k. SCSP←F

l. End if

m. End

Table 2. Terminologies used in Pseudo Code

Terminologies	Description
Sks	Secrete Keys.
Pl	Privilege Level.
H(F)	Hash Function for file.
$\emptyset fp$	$Tag\ of\ the\ File\ with\ its\ Privilege$
$\emptyset' fp$	$Token\ of\ the\ File\ with\ its\ Privilege$

3.3.2. File Downloading Algorithm.

Input: Select File

Output: File Downloaded

a. Begin FileDownloading

b. RequestFileDownload(Req, File Name)

c. LookupSCSP (File Name,$\emptyset' fp$)

d. File[] Contents ←Collect File Contents ()

e. F ≈ Sks

f. Sks send to Email Id

g. FileDownload(FileID, F, $\emptyset' fp$) // Downloaded

h. End

3.3.3. File Uploading Algorithm.

Input: Browse File

Output: File stored in encrypted format

a. Begin FileUploading
b. BrowseFile
c. Set F={Pl}
d. FileUploadReq(FileID, File, Token)
e. Private Cloud←Request H(F)
f. $\emptyset fp = TagGen(F, Kp)$
g. $\emptyset' fp = TokenGen(\emptyset fp, Uid)$
h. ShareTokenReq (($\emptyset fp$, {Priv.})
i. DupCheckReq ($\emptyset fp, \emptyset' fp$)

j. If (File(Contents) = File (Contents))
k. File Upload Request Denied.
l. Else
m. (File (Contents) ≠file (Contents)) then
n. AES← (F = EncCE {Sks, F})
o. Sks = KeyGenCE(F)
p. (F,$\{\emptyset' F, p\}$) //File uploaded
q. End if
r. End

4. Results and Evaluations

We have evaluated our system in consideration of different factors, that varying in different parameters. It includes:

4.1. Time Performance

The performance of this paper is analyzed under consideration of a variety of files and their sizes. Based on different file sizes time performance is evaluated.

Table 2. Time Performance

File Size	Upload File (seconds)	Download File (Seconds)
10bytes	15	0
1 kb	19	5
10kb	22	7
100kb	25	10
1mb	27	10

4.2. Storage

A set of sample files all along with their replica's are saved on the system. In below Figure, test runs are plotted on the x-axis and file size on the y-axis. Results incidental for storage space after deduplication process; used are:

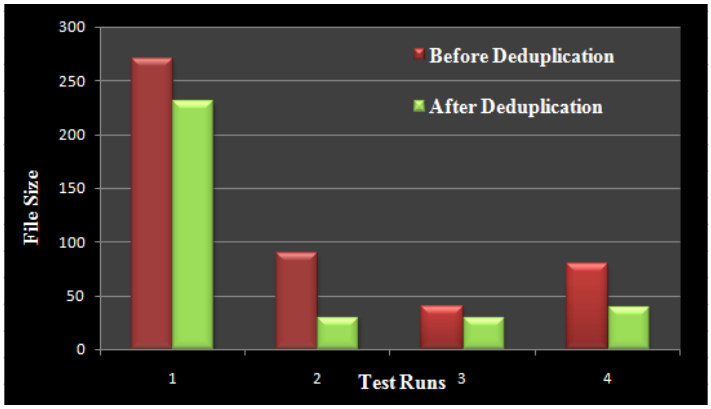

Figure 5. Storage Space Saved

4.3. Bandwidth

When a client requests for files from the cloud, bandwidth consumption is reduced by sending only unique files. In below Figure, where the x-axis is file size (in MB) and the y-axis is the bandwidth consumption. From the graph it can be seen that efficiency of bandwidth utilization is increased with the proposed approach.

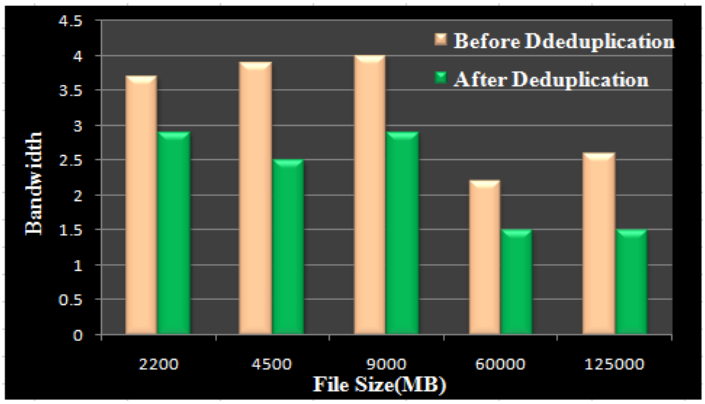

Figure.6. Bandwidth Consumption

5. Conclusion

This paper evaluates representations about the Deduplication proofs of the security. We have mentioned deduplication types, deduplication to keep away from the replicate copies of the files of data of the users and it also offers the better security. A hybrid cloud approach is used to provide the enhanced security and less utilization of network bandwidth. This paper identifies the Deduplication and the effectiveness with the security. We have presented tested and evaluated results of the proposed system for the users to understand them the security aspects.

6. Future Work

Currently we have designed system in consideration of single cloud; thereby in future, it may comprehensive used for sky computing and may provide stronger security aspects to systems of multi-users.

References

[1] J. Li, Y. Kit Li, X. Chen, P. P. C. Lee and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", In IEEE Transactions on Parallel and Distributed Systems, DOI:10.1109/TPDS.2014.2318320, **(2015)** April 7, pp. 1206-1216.
[2] J. Li, X. Chen, M. Li, J. Li and P. P. C. Lee, "Secure Deduplication with Efficient and Reliable Convergent Key Management", In IEEE Transactions on Parallel and Distributed Systems, DOI: 10.1109/TPDS.2013.284, **(2014)** May 12, pp. 1615-1625.
[3] M. Bellare, S. Keelveedhi and T. Ristenpart, "Message-locked encryption and secure deduplication", Proceedings of EUROCRYPT, Athens Greece, **(2013)** March 3, pp. 296-312.
[4] M. Bellare, S. Keelveedhi and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage", In USENIX Security Symposium, Washington DC, **(2013)** August 14-16, pp. 179-194.
[5] S. Halevi, D. Harnik, B. Pinkas and A. Shulman-Peleg, "Proofs of ownership in remote storage systems", Proceedings of the 18thACM Conference on Computer and Communications Security, Hangzhou, China, **(2011)**, pp. 491-500.
[6] V. Satish Radia and D. Kumar Singh, "Secure Deduplication Techniques: A Study", International Journal of Computer Applications, **(2016)**, pp. 41-43.