# SECURITY FOR DE-DUPLICATION USING CLOUD COMPUTING

Arshin Pathan[1] Rahul Yadav[2] Anand Yadav[3] Amarja Adgoankar[4]
[1]Engineering Student, [2]Engineering Student, [3]Engineering Student, [4]Assistant Professor
[1]Computer Engineering Department,
[1]K. C. College of Engineering and Management Studies and Research, Mumbai, India

***Abstract:*** This paper represents that, many techniques are using for the elimination of duplicate copies of repeating data, from that techniques, one of the important data compression technique is data duplication. Many advantages with this data duplication, mainly it will reduce the amount of storage space and save the bandwidth when using in cloud storage. To protect confidentiality of the sensitive data while supporting de-duplication data is encrypted by the proposed convergent encryption technique before out sourcing. Problems authorized data duplication formally addressed by the first attempt of this paper for better protection of data security. This is different from the traditional duplication systems. The differential privileges of users are further considered in duplicate check besides the data itself. In hybrid cloud architecture authorized duplicate check supported by several new duplication constructions. Based on the definitions specified in the proposed security model, our scheme is secure.

***IndexTerms*** – **upload a file on cloud,check for duplication,avoid duplication on cloud server,download file from cloud.**

## INTRODUCTION

"Virtualized" resources to users as services across the whole internet providing by the cloud computing to hide platforms and implementation details. Highly available storage and massively parallel computing resources providing by the cloud services at low costs. Cloud computing widely spread in the world, maximum amount of data stored in the clouds and shred by the users with specified rights, which define as access rights of the stored data. One of the critical challenge of cloud storage services is the management of the duplication is one of the best technique to make the data management in the cloud computing. It has attracted more and more attention recently. In the data storage to reduce the data copies we go for duplication techniques. This duplication technique is a data compression technique The technique is used improve storage utilization and can also applied for network data transfer to reduce the number of byte that must be sent. Deduplication eliminates redundant data to reduce multiple data copies with the same content. Duplication only keeps one physical copy and referring other redundant data to that copy. Either the file level or block level, deduplication can take place. Same file duplicate

copies eliminated in file level de-duplication. In non-identical files, blocks of data that occur, this blocks of data eliminate with the block de-duplication.

## II. SYSTEM MODEL

By using the duplication technique, to store the data who will use S-CSP are consisted as group of affiliated client at high level. The main aim is enterprise all the network. To set the data back up and disaster recovery applications for reduce the storage space. We frequently go for de-duplication. Such systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions. There are three entities define in our system as shown in figure 1, those are, Users Private cloud S-CSP in public cloud

S-CSP. This is an entity that provides a dataϖ storage service in public cloud. 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 de-duplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power

Data Users. A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users.

Private Cloud. Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service.

# III. ALGORITHMS USED

In this section, we use two types of algorithms For file uploading and For file downloading.

**a.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 Display "file already exist"

END

**b.FOR DOWNLOADING A FILE BEGIN**

Step-1 Read fileStep

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

# IV. ACKNOWLEDGMENT

In this, we address the problem of privacy preserving de-duplication in cloud computing and propose a new deduplication system supporting for:

Differential Authorization: To perform duplicate check based on privilege of user is able to get his/her individual token. Without aid from the private cloud server and for the duplicate check outs token cannot generate by the user.

Authorized duplicate check: Authorized user isϖ 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 duplicate check the authority.

Unforgeability of file token/duplicate-check token:ϖ Unauthorized users without appropriate ectly and tells the user if there is any duplicate. privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the S-CSP.

# V. IMPLEMENTATION

Our implementation of the Client provides the following function calls to support token generation and de-duplication along the file upload process.

FileTag(File) – It computes SHA-1 hash of the Fileϖ as File Tag

TokenReq(Tag, UserID) – It requests the Privateϖ Server for File Token generation with the File Tag and User ID

DupCheckReq(Token) – It requests the Storageϖ Server for Duplicate Check of the File by sending the file token received from private server

ShareTokenReq(Tag, {Priv.}) – It requests theϖ Private Server to generate the Share File Token with the File Tag and Target Sharing Privilege Set

FileEncrypt(File) - It encrypts the File withϖ Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode, where the convergent key is from SHA-256 Hashing of the file

FileUploadReq(FileID, File, Token) – It uploads theϖ File Data to the Storage Server if the file is Unique and updates the File Token stored.

# VI. CONCLUSION

Notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test-bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

# VII. 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.