ISSN 2395-1621

Analysis of Data DE duplication on Cloud using the Privacy-Preserving Encrypted Files

#1Sarika Bhosale, #2Manaswi Butala, #3Avinash Kumar, #4Shubham Akole,
#5Prof. Rohini Patil

1sarikabhosale21@gmail.com,
2butalamanaswi@gmail.com,
3avinashandilya@gmail.com,
4shubhamakole1@gmail.com

#12345Department of Computers
SKNSITS, Lonavala.

ABSTRACT

Today is the most important issue in cloud computing is duplication for any organization, so we analysis this issue an avoid the reparative files on cloud storage. Avoidance of the file is advantages the cloud size issue. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, on cloud storage. In this system we check the duplicate file on cloud storage also security apply using encryption. We use the AES encryption algorithm for encrypt the file simultaneously we check the duplicate file using the hashing algorithm. Also enhanced this system using recover option, cloud provide the deleted file backup on requesting. This paper study on the plain text as a input the system for checking the duplicate file, we next stage we analysis the encrypted file as a input to the system and find the duplicate file on the cloud storage.

Keywords:- Duplication, Authorized Duplicate Check, Confidentiality, Cloud computing.

ARTICLE INFO

Article History

Received: 5th June 2018

Received in revised form :

5th June 2018

Accepted: 7th June 2018

Published online :

8th June 2018

I. INTRODUCTION

The Today work is all depend on the online and number of customer are connected to the online cloud network for use the storage, access any resources, where data is stored in pools of storage which are generally hosted by third parties. So in this scenario the cloud storage is increasing rapidly. The online storage provides users with benefits, ranging from cost saving and simplified convenience, to mobility opportunities and scalable service. Above all properties used for customers to use and storage their personal data to the cloud storage. As per our analysis the volume of data menace size of the cloud storage capacity is expected to achieve 50 trillion gigabytes in 2020. The cloud storage system has been widely used in the world; it fails to accommodate some main emerging needs such as the abilities of auditing integrity of uploaded data cloud files by clients and we detecting duplicated files by cloud servers. We generate the system and analysis both problems below.

The first solve the problem is integrity auditing in the cloud computing. The local cloud is able to remove unwanted action clients from the heavy burden of storage management and maintenance.

The online cloud data storage used for further access and the data is transferred via Internet and stored in an uncertain domain, not under control of the clients at all. These concerns originate from the fact that the cloud storage is susceptible to security threats from both outside and inside of the cloud [1], and some data loss from the clients may be hidden by the uncontrolled cloud servers to maintain the reputation. The most important thing is that for an ordinary clients the data which is rarely accessed is deliberately deleted by the servers to maintain the cost and space. We considering the large size of the outsourced data files uploaded by user and the clients' constrained resource capabilities, the first problem is as how can the client efficiently perform periodical in verifications even without the local copy of data files. So we solve this problem using the detecting is secure de-duplication file on cloud storage. The remove increased volumes of data stored at remote cloud servers accompany the rapid adoption of cloud services.

According to the last survey of EMC the most of the remotely stored files are deduplicated. [2], Recently the 75% of the digital data is deduplicated. Due to this the term came that is deduplication in which the cloud servers just keep only one file and keeps the link of that file for the user's who wants the same file to store. Due to this it leads to a number of threats affecting the storage system [3][2], for example, a server telling the client that it does not need to send or store the file which is same as other user and it can be dangerous sometimes.. These attacks originate from the proof that client owns a file that totally use static or we can say a hash code. [3]. Thus, the second problem is generalized as how can the cloud servers efficiently confirm that the client owns the uploaded file before creating a link to this file for him/her.

II. SYSTEM ARCHITECTURE

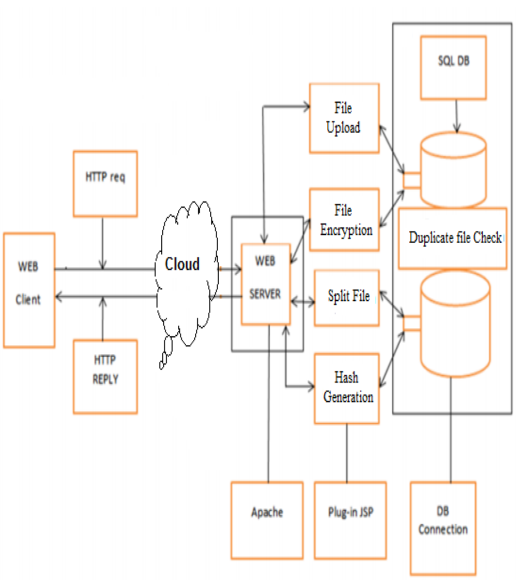


Fig 1. System architecture

Model:
Data Owner:
Data owner is the person to share the personals own files to the users. Here he give the secure access permission to access that given file.

Encryption:
Here encryption is apply to the given file using the advance encryption standard.
It uses a common secret key k to encrypt and decrypt information.

Convergent encryption:
Data owner gets key from each original data copy and encrypts data copy with the convergent key.

Proof of Ownership (AP):
Enable the users to provide their ownership of data copies to the storage server we choose proof of ownership.

Recovery of file:
It uses a common recovery option if some data will loss.

III. MATHEMATICAL MODEL

System Description:
Input:
Upload file ()
U : Upload file on cloud.
E : Encryption File.
S : Splitting file for security.
H : Hash value for each file.
Output:
Check Duplicate file on cloud storage

Input
Function Recovery (id, request, file)
ID : unique id for each file.
Request : User request for recovery of file.
File : Check file on cloud.
Output:
File will recover to data owner.

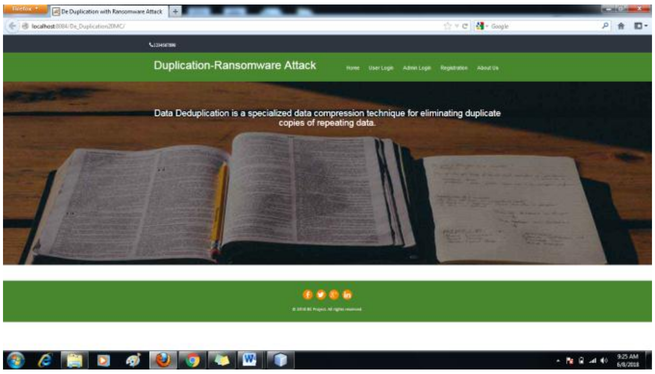
Algorithm Process for detection duplicate file:

```
START
Step –1 Read file
Step –2 Cloud server checks for duplication
Step- 3 Check calculate hash value
Step 4- Compare hash value from existing file value.
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
```

IV. SYSTEM ANALYSIS

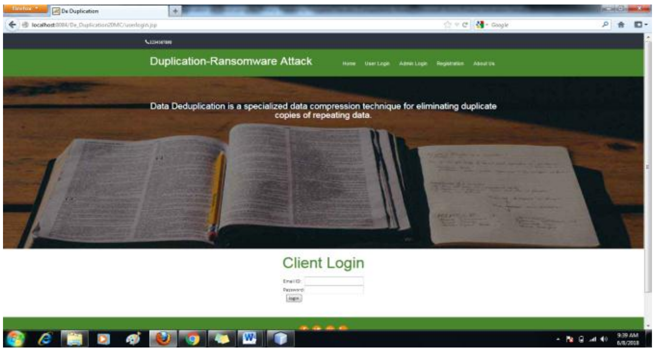
We have created system in java. Data is stored in mysql database. We have created a web application with local server. Web application that communicates with local server and Trustee Server using REST API. We have uploaded text document on cloud. We have evaluated time required for tag generation and file deduplication checking for different file sizes. Here we also calculate the file each file size for analysis purpose.

V. RESULT



Data Deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data.

Fig 2. Home page

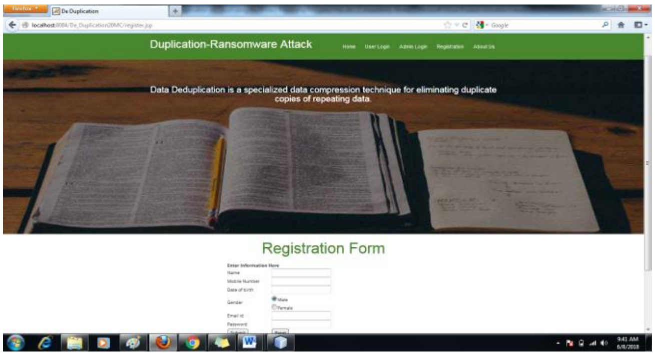


Data Deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data.

Client Login

Email ID
Password

Fig 3. User Login



Data Deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data.

Registration Form

Enter Information Here
Name
Mobile Number
Date of birth
Gender ● Male ○ Female
Email id
Password

Fig 3. User registration

VI. CONCLUSION

In this paper we reviewed the deduplication techniques for better confidentiality and security in cloud computing. The detection of redundant data and removal of this redundant data is an important task for keeping the cloud storage clean and scalable. This duplicate data elimination has a great advantage for cloud storage. We have surveyed various techniques for deduplication.

VII. ACKNOWLEDGMENT

I wish to express my profound thanks to all who helped us directly or indirectly in making this paper. Finally I wish to thank to all our friends and well-wishers who supported us in completing this paper successfully I am especially grateful to our guide Prof. Rohini Patil madam for him time to time, very much needed, valuable guidance. Without the full support and cheerful encouragement of my guide, the paper would not have been completed on time.

REFERENCES

[1] 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 Proceedings of 22nd International Conference on Distributed Computing Systems (ICDCS, 2002).

[2] M. W. Storer, K. Greenan, D. D. Long and E. L. Miller. "Secure data deduplication". In Proceedings of the 4th ACM International Workshop on Storage Security and Survivability, StorageSS 08, pages 1–10, 2008.

[3] M. Bellare, C. Namprempre, and G. Neven, "Security proofs for identity-based identification and signature scheme". 2009

[4] J. Xu, E.-C.Chang, and J. Zhou, "Weak leakage-resilient client side deduplication of encrypted data in cloud storage", 2013

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Server aided encryption for deduplicated storage". In USENIX Security Symposium, 2013.

[6] Z. Li, X. Zhang, and Q. He, Analysis of the key technology on cloud storage, in International Conference on Future Information Technology and Management Engineering, 2010, pp. 427428.

[7] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491500. ACM, 2011.

[8] D. Harnik, B. Pinkas, and A. Shulman-Peleg. Side channels in cloud services: Deduplication in cloud storage. IEEE Security & Privacy, 8(6), 2010.

[9] Q. He, Z. Li, and X. Zhang, Data deduplication techniques, in International Conference on Future Information Technology and Management Engineering,pp.431-432,2010.

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

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

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

[12]W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.