Data Security and Data Replication for Balancing the Server-side Memory Space

Swati Bhatt, Navin Patil, Mayur Borse, Amol Jadhav, Dr.G.M Bhandari

Department of Computer Engineering

JSPM's, ICOER, Wagholi, Pune.

Abstract: Now days we analysis the most important problem facing the any organization in cloud computing is data duplication, so this problem we analysis and avoid the reparative files/data on cloud storage based on different algorithm. The most important advantages of the file are minimum space required in cloud and avoid cloud size issue. We not only check the data duplication but also to protect the confidentiality of sensitive data while uploading the cloud server. To better protect data security proposed algorithm, on cloud storage. Secure data is can significantly reduce the communication and storage overheads on server side services provided by admin, and has potential applications in our big data-driven society. Existing data duplication technique are generally designed to the mobile application in the flash storage application ensure the efficiency and data availability, but not provide the better result in both conditions. In this proposed system we analysis the duplicate file on cloud storage also security apply using cryptography encryption technique. We use the data security using the advance encryption algorithm for encrypt the file that time we check the duplicate file using the hashing algorithm. Also we extend the system backup/recover option, server will i.e. local cloud provide the deleted file backup on requesting. This research paper gives the plain text (.txt file) as an input the system for checking the duplicate file using hash algorithm. We also use the key generation technique for user download the file for secure authorization.

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

I. INTRODUCTION

A today environment is all depend on the online service provide and number of customer are connected to the online social cloud network for use the data storage, access any services, where data is stored in the database of storage which are generally hosted by third parties. So in this section the cloud storage is increasing rapidly recent environment. The online storage provides users with benefits, ranging from cost saving and simplified study, to mobility opportunities and scalable service. All above properties used for data storage of user to use services, their personal data from the cloud storage. As per our research analysis the volume of data is the capacity of the cloud storage size is expected to increases 50 trillion gigabytes till 2020 year. The cloud storage system has mostly used in the all world for data storage; it fails to analysis the some emerging needs such as the abilities of data auditing integrity of user uploaded data on local cloud server and we detecting duplicated data/files by cloud servers. We analysis the problem and generate the system to overcome the above problem. The first we solve the problem is integrity auditing based on the data modification algorithm in the cloud computing. The local cloud is able to take action during the uploaded files action clients from the heavy burden of storage management and maintenance.

The secure cloud data storage used for further access and the data is transferred via Internet connection and stored in a selected domain, but network and domain not under control of the clients at all. These problems not originate which user take the unauthorized access from the cloud storage is susceptible to security threats from both outside and inside of the cloud [1], during these process some data will loss from the clients may be hidden by the uncontrolled cloud servers to maintain the reputation. The most important parameter is that for clients are the data security which is less accessed is deliberately deleted by the servers to maintain the cost and space. This system considering the large data size of the outsourced data files uploaded by registered user and the clients' constrained resource capabilities, the first problem is as how can the client efficiently analysis the verifications based on the proposed algorithm even without the local copy of data files. Then we solve this above all problem using the detecting is secure de-duplication file on cloud storage. The remove increased data on cloud server by the cloud server provider stored at remote cloud servers accompany the rapid adoption of cloud services.

According to the last survey of existing system, the most of the remotely data stored and that files are mostly deduplicated. [2], recently the 85% of the digital record of data is deduplicated. Due to this problem we introduce the proposed system explain I the fig. 1. The avoid data deduplication in which the cloud servers just only keep one file of same file to every user. Due to this technique it leads to a number of threats are affecting the cloud storage system [3][2], for example, a server search the client data and send or store the file which is same as other user but that files it can be dangerous sometimes to the cloud server. These problems can avoid based on the validation of each user data based on the proof that client owns a file that totally uses static or we can say a hash code.[3].

II. REVIEW OF LITERATURE

[1] A. Adya, W. J. Bolosky, D. Simon, J. R. Douceur, and M. Theimer., in this paper author studied the problem data replication of deduplication in multi environment. The authors proposed the use of the encryption technique to data analysis, i.e., secure keys from the hash of plaintext.

[2] K. Greenan, D. D. Long, M. W. Storer and E. L. Miller., in this research paper author explain some data security problems during data storage, and presented a new security model for secure data deduplication. However, these two techniques focus on server-side and client side deduplication and do not consider data loss (leakage) settings, against malicious users.

[3] C. Namprempre, M. Bellare and G. Neven, in this paper provides either data security proofs or network attacks for a large number of identity-based identification and signature based schemes defined either explicitly or implicitly in existing literature.

[4] J. Xu, E.-C.Chang, and J. Zhou, in this proposed solution avoid the problem of data security based on a cryptographic (encryption) usage of symmetric key encryption used for enciphering the data file and asymmetric encryption for upcoming data files, due to the highest sensibility of this data file information towards several intrusions.

[5] S. Keelveedhi, M. Bellare and T. Ristenpart., in this paper proposed solution here the data which is common between multiple users to increase the backup speed and minimized the storage requirement namely backup algorithm. Here provide the better supports at client-end per user encryption are necessary for confidential personal data.

III. PROPOSED METHODOLOGY

A. *Architecture*

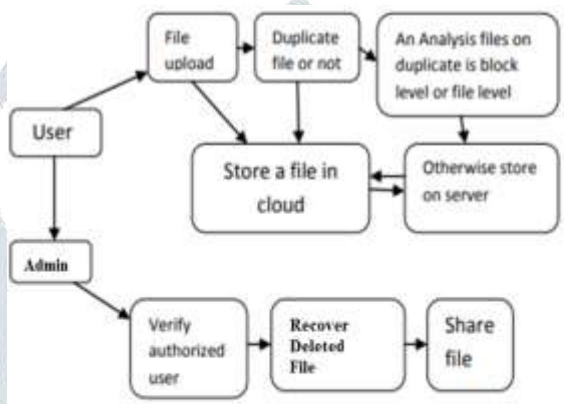

Fig 1. System architecture

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.

B. Algorithm.

AES Algorithm:
For encryption of data:-

START
Step 1- U=Upload (file), he input is consider as Text, is being converted to 128 bit plain text.
Step 2- R= Read (input file),
Step 3- K=Key generation (file)
  e.g= key=123456;
Step 4- E=Encrypt(file, key), encode the upcoming file
Step 5- C=Convert (file),
If(encrypt), then file convert plain to cipher text
  Split (file1, file2);

Stored (file)
Else, file not encrypted
Step 6- D=Decrypt (file), decode the file
if (decode), then file convert cipher text to plain
Combine (file1, file2);
Else, file not decoded
Step 7- Download file
END

Hash Algorithm:
START
Step 1- U=Upload (file)
Step –2 R=Read () file
Step –3 C(U,R)=Cloud server read uploaded file
Step- 4 Check calculate hash value
Generate SHA (file)
key="da39a3ee5e6b4b0d3255bfef95601890afd80709"
Step 5- Compare(R, key) hash value from existing hash value (k).
Step –6 S=Send (duplicate) duplication response whether the file already exists or not
Step – 7 If () the file does not exist
{
Script("Display file does not exist");
Upload(file,key);
}
Else{
Script("Display file exist");
}
Step 8- Store(file,key);
END

IV. SYSTEM ANALYSIS

In this proposed system we have created system logic based on java programing language. User data will stored in database for that we used MySQL tool. We have created a web application using java server pages with request and response system on local server. This system is totally depending on the uploaded text document on cloud server. Once system design then proposed algorithm evaluated minimum time for data duplication analysis and checking the recovery search time for tag generation and file deduplication checking for different file sizes.

V. RESULT

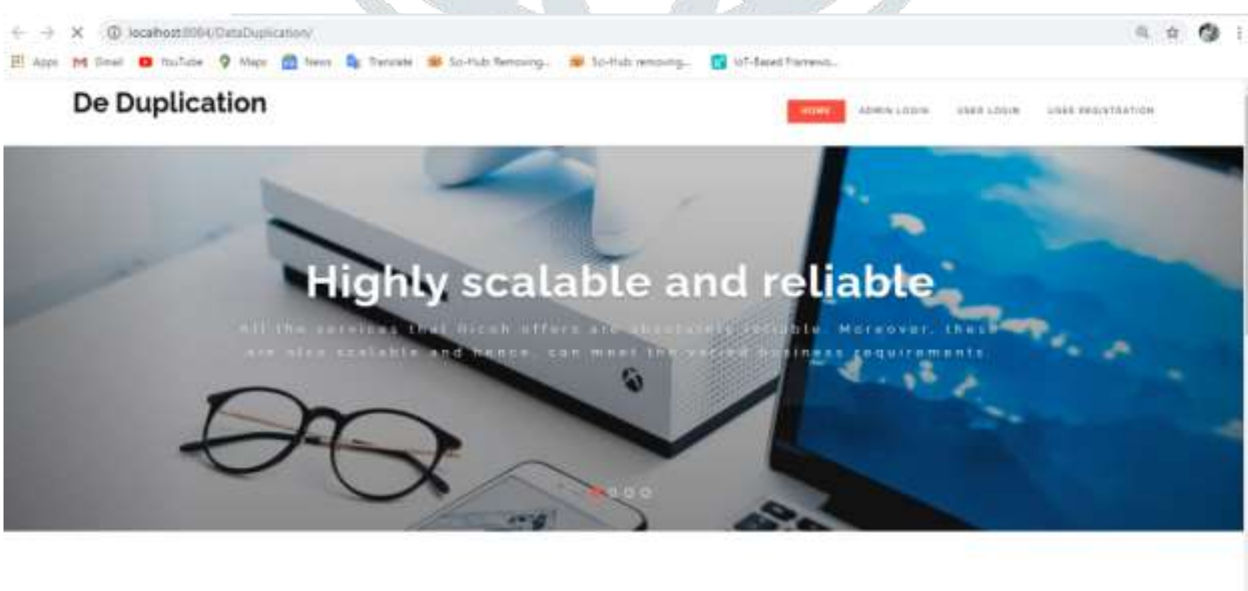

Fig 2. Home page

Fig 2. Login page

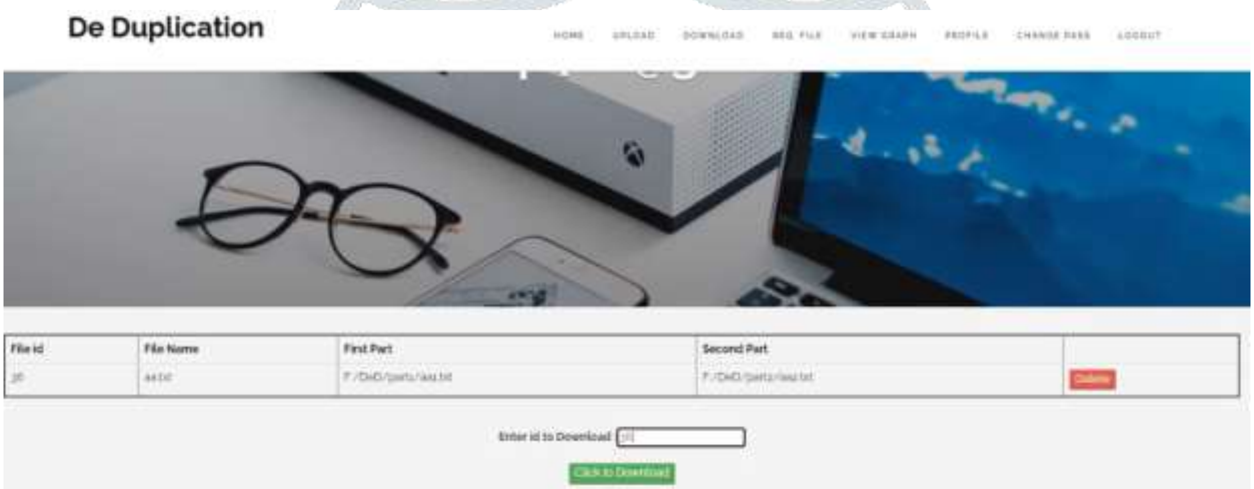

Fig 4. File uploaded information

VI. CONCLUSION

Finally we have reviewed the data deduplication techniques for better security and minimizing the cloud storage space in cloud computing environment. In this system we remove the redundant data of this redundant data is an important task for keeping the cloud storage clean and scalable. Proposed system duplicate data elimination has a great advantage for cloud storage. We have surveyed various techniques for deduplication.

REFERENCES

[1] W. J. Bolosky, D. Simon, J. R. Douceur, A. Adya, and M. Theimer. "Reclaiming space from duplicate files in a server less distributed file System". 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, Storage SS 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.