A Secure Data Deduplication for Multi-User Cloud Storage

Minal Zaryekar **Prof. Shweta Ashtekar** **Prof . Amruta Chintawar**

Abstract: Development in networking technology and growth in the requirement for computing resources have moved many organizations to outsource their computing needs and storage. Therefore the requirement of the reliable deduplication technique needed in a multi-user cloud storage system, in which user can acquire the ownership of files instantly by skipping the uploading process if the cloud server already exists the same file which is uploaded by other owners. Therefore to enables the user to successfully update the files on a cloud server and to verify that a server faithfully stores a file i.e. observe the integrity of outsourced files researchers suggested Dynamic Proof of Storage (PoS) in the single user domain. This proposes work suggest the concept of deduplicatable dynamic proof of storage for multi-user environments. It simultaneously achieves secure multi-user deduplication and dynamic Proof of Storage.

Index terms: Cloud storage, dynamic proof of storage, deduplication

I. INTRODUCTION

As in recent years, outsourcing of the data storage is getting more interesting to both academia and business due to the benefit of powerful accessibility, low cost and easy sharing. The cloud storage gets wide attention, as one of the outsourcing of data storage forms [2]. The cloud computing can administrate storage, applications and organize vast resource of computing. Due to these attractive characteristics both the business organizations and individuals are moved to outsource their data to the cloud storage. Thus users can administer their data themselves without the need of buying hardware and software.

The organizations such as Apple, Google, Microsoft and Amazon allocate their own cloud storage services, on which files can be easily uploaded by users on the servers, access them from various location and devices of the users choice, also can share them with the others.

Although due to the various advantages of the cloud storage services it is widely accepted in recent days. But outsourcing sensitive information such as personal health records, government documents, e-mails, company financial data etc. to remote servers causes privacy, security issues to become a critical concern[3].

When a data outsourced by users on the cloud storage, the user should be assured that the server possesses the original data faithfully. An important part in storing data on untrusted servers is verifying the reliability of data. Therefore when user outsources its data to cloud the data integrity is a most essential factor.

II. DEDUPLICATION

The deduplication simply means the elimination of duplicate data, especially in computer data. The deduplication in the cloud storage system is a process in which the amount of physical data stored on the cloud storage reduces by removing duplicated data from a data stream. The data deduplication is beneficial for organizations which require constant storing and copying of data for later reference or recovery purpose ie. highly redundant operations.

The Deduplication eliminates the sets of duplicate data and keeps only which is essential and unique, hence clearing storage space significantly as shown in fig 1. In deduplication using hash values, redundant data identified and only one copy is stored. The logical pointers created to other copies rather storing other actual copies of the redundant data.

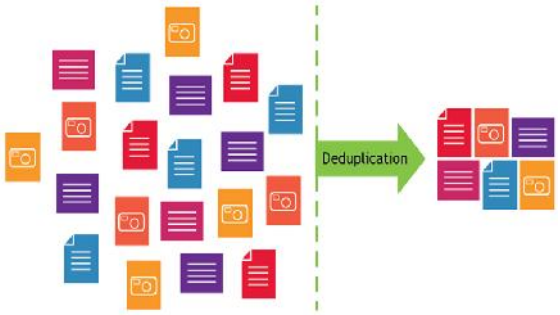

Figure 1: File level Deduplication

III. LITERATURE SURVEY

The Deduplication is beneficial when similar information by the various clients outsourced to the

distributed storage, But it raises problem identifying with possession and security. The objective is to store more information in less space. The deduplication commonly used by Cloud storage services. Deduplication reduces bandwidth and the space requirements of data storage services. It eliminates data which is not unique by storing only one copy of each block or file. The deduplication is more advantageous when applied to multiple users.

A huge number of online file storage services have been introduced in recent years. The basic functionality services provided to a user such as uploading and downloading files by a specific user. In addition to this more advanced features services offered such as real-time association, shared folders and minimization of data transfer. In Deduplication, balancing security raises new concerns with space savings. To maintain security during deduplication, the idea of using an encryption key has been explored.

In most of the existing dynamic PoSs, trust confirmation is produced by the generating secret key of the uploader. Thus, different entities who have the authority for files cannot be transferred it because of the client side cross-user deduplication, In this situation, the dynamic PoSs would not succeed.

Most important thing is that the Multi-user environment is not supported by existing POSs. The users' sensitive data susceptible to the outsider and insider malicious attack.

IV. DESIGN OBJECTIVES OF CLOUD COMPUTING

- **Improved availability** : As TPA store only a single copy of redundant data, hence availability improves.
- **Public auditability** : The Third Party Auditor can audit the user's data, without retrieving the data copy.
- **Storage capacity** : It improves storage capacity as TPA eliminates duplicated copies.
- **Privacy preserving** : It gives security and increases performance by checking the authenticity of the user and encrypting the data.

ALGORITHM USED FOR ENCRYPTION

Rijndael Algorithm (Advanced Encryption Standard)

Rijndael encryption algorithm is an encryption technique. This algorithm is used for encryption and decryption process, to encrypt sensitive data. It is an advanced AES algorithm. It is a symmetric key encryption algorithm. The advantages of the AES algorithm are high speed, performance, efficiency, security, simple to implement, flexibility, no royalty and versatility across a variety of platforms. Ability to run successfully on a large desktop, large computers and small devices such as smart cards. The key and block can be chosen independently from 128, 160, 192, 224, 256 bits. Rijndael algorithm uses very less system memory and it is simple to implement.

ALGORITHM USED FOR HASHING TASK

Secure Hash Algorithm

SHA stands for "Secure Hash Algorithm".

The hashing algorithm is designed by the United States National Security Agency. It is used for implementing a hashing task. The SHA algorithm gives better security as compared to other algorithms, as in SHA algorithm numbers of rounds per byte are more as compared to other algorithms. Although, SHA bit hash storing is expensive, but the time required by the SHA algorithm to generate a hash value and the number of cycles per bytes are efficient compared to the other algorithms. However, the number of cycles per byte in SHA is more as compared to other hashing functions, but the time taken by the SHA algorithm for generating the hashing value is much smaller than other algorithms. Thus, the SHA hash function is secure and also efficient hashing algorithm.

V. DESIGN STEPS OF THE IMPLEMENTATION

The following steps are involved to make data deduplication more secure.

- **Design Entry**: Users initially do the registration on the cloud. Then with the help of OTP data owner login into the system. A user can upload, download and claim for the integrity of the file to the cloud.
- **Perform an Encryption and Hashing**: When the file is uploaded system calculates the hash value of the file using the SHA algorithm. Then the file is encrypted by using hash value and that hash value is further encrypted by the encryption key.
- **Design implementation**: After verification of project successfully we need to execute graphs of different parameters to check the performance of the project. It reduces the storage size of duplicated data also provides stronger security to the data.

Fig 2 shows the design flow of the system.

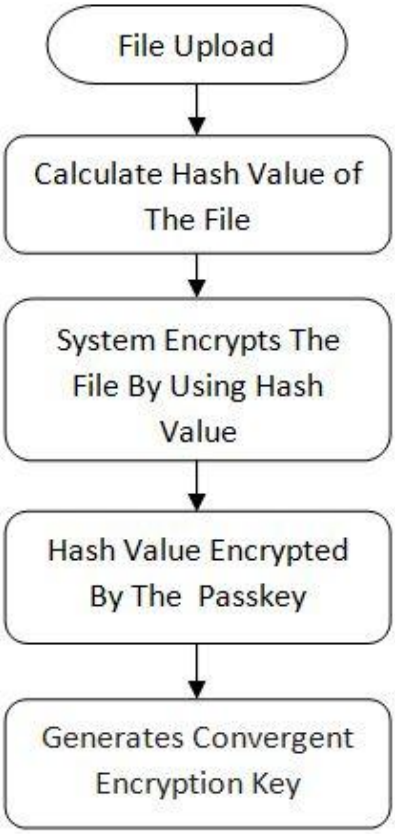

Figure 2 : Design flow of the system

IV PERFORMANCE ANALYSIS

Fig 3. Shows Graph of the storage space saved in cloud server due to deduplication. If the same file is uploaded by different users. For example, 10MB file uploaded by two different users, the existing system saves both files ,hence occupy 20MB storage space. But the current system store only a single copy and refer other redundant data to that copy.

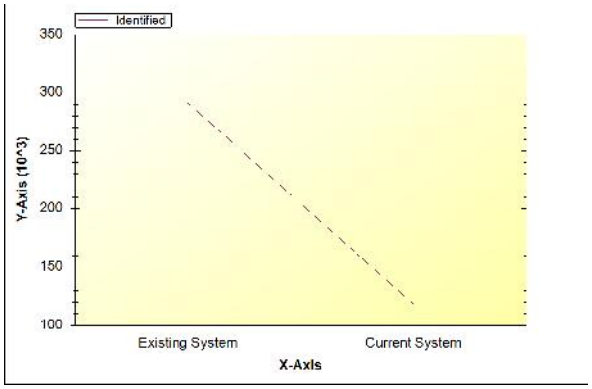

Figure 3 : Storage Graph

Fig 4. Shows Graph of the security. It compares the security level of the existing system and the current system. In the current system, security level increases due to the hash value are again encrypted by the passkey.

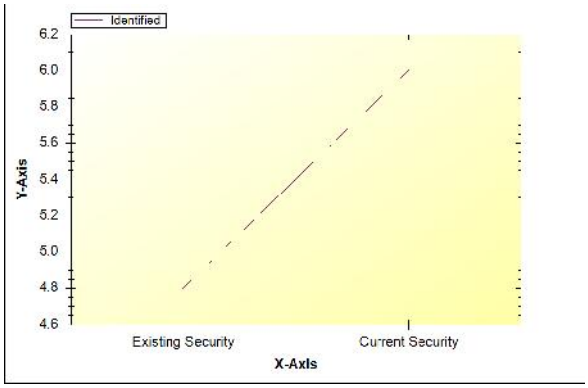

Figure 4: Security Graph

CONCLUSION

In this project, we introduced Deduplication in multi user cloud storage for saving storage capacity and bandwidth. In recent years data growing rapidly on cloud storage thus proposed system helps to reduce an expensive requirement of storage and also improves the performance. The advanced AES algorithm which is used to encrypt sensitive information. It is the best combination of security, high speed, performance, easy implementation, flexibility, efficiency and versatility across a variety of platforms. SHA algorithm is used for implementing the hashing task. It gives better security compared to other algorithms. The data privacy and confidentiality maintained by integrity checking if a file is tempered by other owner or admin. The graphical result shows the comparison between storage space requirement and security in an existing system and proposed system.

FUTURE WORK

Although cloud storage data deduplication has been widely used to reduce storage space by eliminating redundant data and storing only one copy. The data privacy preserving is also an important thing for secure data deduplication.The convergent encryption is proposed for secure data privacy. In this scheme clients confidential data preserved by applying a private-key encryption technology.

REFERENCES

[1] DeyPoS: Deduplicatable Dynamic Proof of Storage for Multi-User Environments Kun He, Jing Chen, Ruiying Du, Qianhong Wu, Guoliang Xue, and Xiang Zhang

[2] Z. Xia, X. Wang, X. Sun, and Q. Wang, A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data, IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340352, 2016.

[3] Z. Xiao and Y. Xiao, Security and privacy in cloud computing, IEEE Communications Surveys Tutorials, vol. 15, no. 2, pp. 843859, 2013.

[4] G. Ateniese, S. Kamara, and J. Katz, Proofs of storage from homomorphic identification protocols, in Proc. of ASIACRYPT, pp. 319333, 2009.

[5] C. Erway, A. Kupc u, C. Papamanthou, and R. Tamassia, Dynamic provable data possession, in Proc. of CCS, pp. 213222, 2009

[6] R. Tamassia, Authenticated Data Structures, in Proc. of ESA, pp. 25, 2003.

[7] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, Proofs of ownership in remote storage systems, in Proc. of CCS, pp. 491 500, 2011.

[8] "Summary on Deduplicatable Dynamic Proof of Storage for Multi-User Environments" Megha T P1, Poornima B G2

[9] "Boosting Effciency and Security in Proof of Ownership for Deduplication "Robert Di Pietro Universit di Roma Tre, Italy dipietro@mat.uniroma3.it,Alessandro SorniottiIBM Research Zurich, Switzerland aso@zurich.ibm.com

[10] "client side deduplication for encrypted data."Xu et al. School of Computing University of Leeds, Leeds, LS2 9JT United Kingdom scwl, p.m.townend, j.xu @leeds.ac.uk