Secure Deduplication and Data Security with Efficient Manner

Punitha. R*, Abila Rani. A*, Nagalakshmi. J, R. Ram Priya+
*UG Scholar, Department of Computer Science and Engineering,
Dhanalakshmi Srinivasan College of Engineering,
Perambalur, TamilNadu, India.
+ Assistant Professor, Department of Computer Science and Engineering,
Dhanalakshmi Srinivasan College of Engineering,
Perambalur, TamilNadu, India.

***Abstract*- Data deduplication is a compression technique for eliminating duplicate copies of repeating data, and has been widely in cloud storage space and uploads the bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication critical issues of making convergent encryption practical is to efficiency and reliably manage huge number of convergent key. In addition, the dare of privacy for sensitive data also take place when they are outsourced by users to cloud. Planning to address the above security test, this paper constructs the first effort to celebrate the idea of scattered reliable deduplication system. This paper recommends a new distributed deduplication system with upper dependability in which the data chunks are distributed from corner to cornering multiple cloud servers. The safety needs of data privacy and tag stability are also accomplishing by introducing a deterministic secret sharing scheme in distributed storage system.**

Keywords - Deduplication, reliability, secret sharing, distributed storage system.

I. INTRODUCTION

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. One critical challenge of today's cloud storage services is the management of the ever-increasing volume of data. To make data management scalable deduplication we are use convergent Encryption for secure deduplication services. Businesses, especially start-ups, small and medium businesses (SMBs), are increasingly opting for outsourcing data and computation to the Cloud. Today's commercial cloud storage services, such as Drop box, Mazy, and Memo pal, have been applying deduplication to user data to save maintenance cost [12]. From a user's point of view, data outsourcing raises security and privacy concerns. We must trust third-party cloud providers to properly enforce confidentiality, integrity checking, and access control mechanisms against any insider and outsider attacks. However, deduplication, while improving storage and bandwidth efficiency, is compatible with Convergent key management. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Many proposals have been made to secure remote data in the Cloud using encryption and standard access controls. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including insider attacks, mis-configured services, faulty implementations, buggy code, and the creative construction of effective and sophisticated attacks not envisioned by the implementers of security procedures. Building a trustworthy cloud computing environment is not enough, because accidents continue to happen, and when they do, and information gets lost, there is no way to get it back. One needs to prepare for such accidents.

1.1 MYSQL SERVER OVERVIEW:

The MySQL server provides a database management system with querying and connectivity capabilities, as well as the ability to have excellent data structure and integration with many different platforms. It can handle large databases reliably and quickly in high-demanding production environments. The MySQL server also provides rich function such as its connectivity, speed, and security that make it suitable for accessing databases.

1.2 Algorithm

S-CSP:

The S-CSP is an entity that provides the outsourcing data storage service for the users. In the deduplication system, when users own and store the same content, the S-CSP will only store a single copy of these files and retain only unique data. A deduplication technique, on the other hand, can reduce the storage cost at the server and save the upload bandwidth at the user side. For fault tolerance and confidentiality of data storage, we consider a quorum of S-CSPs, each being an independent entity. The user data is distributed across multiple SCSPs.

II. RELATED WORK

A private cloud is set up within an organization's internal enterprise datacenter. It is easier to align with security, compliance, and regulatory requirements, and provides more enterprise control over deployment and use. In

the private cloud, scalable resources and virtual applications provided by the cloud vendor are pooled together and available for cloud users to share and use. It differs from the public cloud in that all the cloud resources and applications are managed by the organization itself, similar to Intranet functionality. Utilization on the private cloud can be much more secure than that of the public cloud because of its specified internal exposure. Only the organization and designated stakeholders may have access to operate on a specific Private cloud.

2.1 Process of Deduplication

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. As a result, deduplication system improves storage utilization while reducing reliability. Furthermore, The challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of distributed reliable. deduplication system

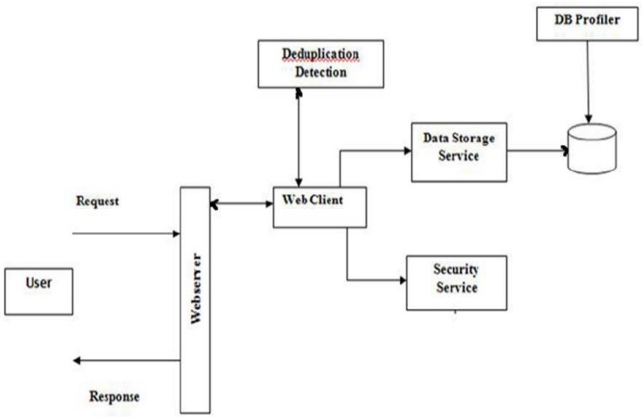


Fig.1.1 Prosess of deduplication

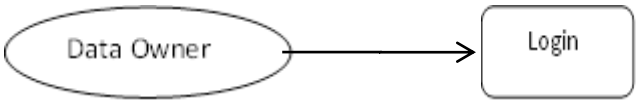


Fig.1.2 Login Data Flow Level 0

Fig.1.3 New User Registration Data Flow Level 1

Fig.1.3 Search Over Files Data Flow Level 2

III. PROBLEM ANALYSIS

3.1 OUR CONTRIBUTION

In this paper, show how to design secure deduplication systems with higher reliability in cloud computing. Introduce the distributed cloud storage Servers into deduplication systems to provide better fault tolerance. To further protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. In more details, a file is first split and encoded into fragments by using the technique of secret sharing, instead of encryption mechanisms. These shares will be distributed across multiple independent storage servers. Furthermore, to support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server. Only the data owner who first uploads the data is required to compute and distribute such secret shares, while all following users who own the same data copy do not need to compute and store these shares any more. To recover data copies, users must access a minimum number of storage servers through authentication and obtain the secret shares to reconstruct the data.

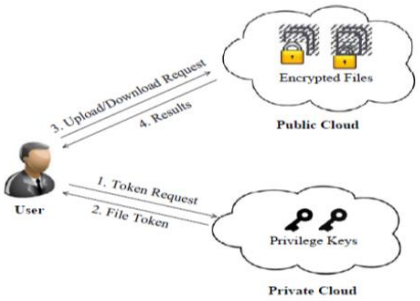


Fig.2. System Architecture

3.2 Analysis

The secret shares of data will only be accessible by the authorized users who own the corresponding data copy. We implement our deduplication systems using the Ramp secret sharing scheme that enables high reliability system supporting the same level of reliability.

Algorithm for Key Generation:

Rijndael Algorithm (AES)

AES is a new cryptographic algorithm that can be used to protect electronic data. Specifically, AES is an iterative, symmetric-key block cipher that can use keys of 128, 192, and 256 bits, and encrypts and decrypts data in blocks of 128 bits (16 bytes).

IV. EXPRIMENTAL RESULTS

New User Registration module:

This module allows the new users to register themselves. They need to register by giving their personal information. During registration they user has to set the type i.e. data owner or user. Once the user is registered he/she becomes an existing user. Then they will be redirected to the login module.

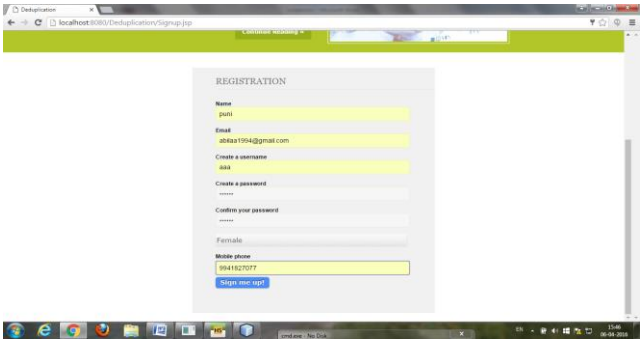


Fig.3.1 Module 1 (New User Registration)

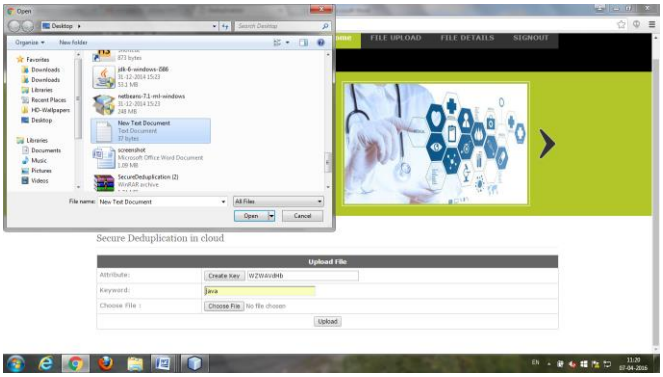


Fig.3.2 Module 2 (File Upload module)

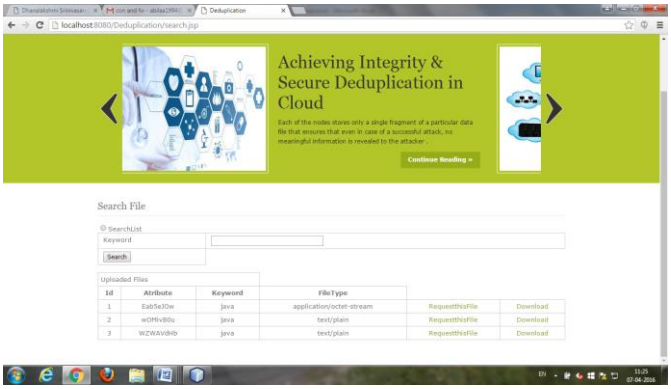
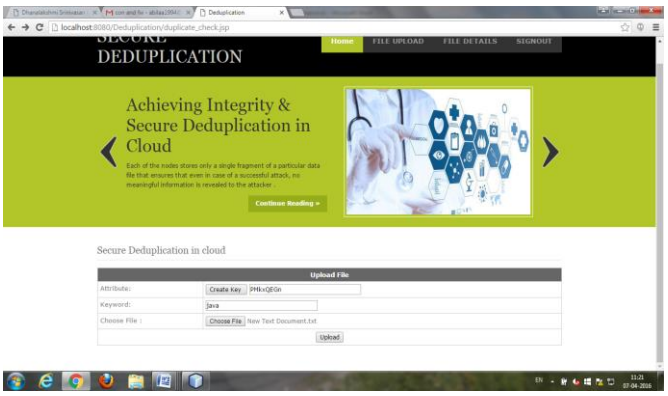


Fig.3.3 Module 3 (Search File Module)



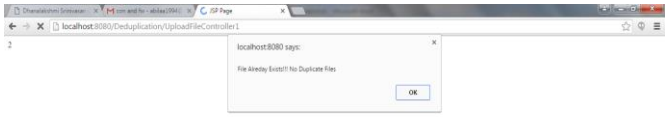


Fig.3.4 Module 4 (Duplication check Module)

File Upload module:

This module is available to the data owner category. In this module the data owner can upload the files. Each file has a unique id. The data owner should mention the name of the file while uploading. The file owner encrypts his files and outsources the cipher texts to the server. The server validates the outsourced cipher texts and stores them for the owner. The uploaded file is stored in the server. For each file a secret key will be generated.

Search File Module:

Here the user can search for a particular file based on the given query. The files related to the given query will be shown. If the user needs to access a Particular file that file key will be sent to his/her mail. Using that key he/she can access the file. Search File Enter the Key.

Duplication Check Module:

In this module, the uploaded file is encoded using Base64 Encoding, and the file content is stored in the server. Here the duplication check is processed, if the encoded file content and the content already stored in the server matches, then the file is marked as duplicate and it tells the owner that the content already exists in the cloud server.

Decryption of Files:

Decryption is the process of converting encrypted data back into its original form, so it can be understood. When the user inputs the correct key that is used at the decryption process, this will extract the original content that is encrypted.

V. CONCLUSION

The notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplication check. Also presented several new de-duplication constructions supporting authorized duplicate check in cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that the proposed scheme is 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.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in USENIX Security Symposium, 2013.
[2] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. of USENIX LISA, 2010.
[3] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proc. of StorageSS, 2008.
[4] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: the least-authority filesystem," in Proc. of ACM StorageSS, 2008.
[5] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in *Technical Report*,2013.