ISSN 2395-1621

Secure Reliability and Confidentiality to achieve Distributed Deduplication Systems

[#1]Chaitanya Ghule, [#2]Namrata Kulkarni, [#3]Ankur Vyas, [#4]Snehal Sakolkar

[1]chaitanyaghule3100@gmail.com,
[2]nakul.kulkarni97@gmail.com,
[3]ankurvyas1992@gmail.com,
[4]snehalsakolkar9@gmail.com

[#1234]Department of Computer Engineering
Zeal College of Engineering and Research, Narhe, Pune.

ABSTRACT

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. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. The security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. Security analysis demonstrates that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.

Keywords: **Distributed Deduplication Server, Ramp Secrete Sharing Scheme, Tag Generation, Message Access Control.**

ARTICLE INFO

Article History

Received: 26th May 2016

Received in revised form :

26th May 2016

Accepted: 28th May 2016

Published online :

29th May 2016

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. We implement our deduplication systems using the Ramp secret sharing scheme that enables high reliability and confidentiality levels. We have also use Tag Generation Algorithm which maps original data copy and generates a tag. It will be generated by user and will perform duplicate check with server. We have also used Message Authentication Code which is a small piece of information which authenticate message and provides the integrity and Authenticity assurance on message. Here we are also going to use database which stores the duplicate copy of data and the privileges to access these duplicate copy of data are left with the admin.

II. RELATED WORK

1. A Survey on: Secure Data Deduplication on Hybrid Cloud Storage Architecture

In this survey article proposed the secure deduplication with the Help of token generation and Secure upload download it can assure the user about high data security and also avoid data deduplication. Security analysis determine that given schemes are secure in terms of insider as well as outsider attacks specified in the proposed security model. As a proof of concept, it executed a prototype of proposed authorized duplicate check scheme and conduct testbed experiments on given prototype. In this paper we have to provide the different techniques to reduce the deduplication in cloud storage and maintain the security In future by using Cloud Service Provider (CSP) have significant resources to govern distributed cloud storage servers and to manage its database servers. It also provides virtual infrastructure to host application services. These services can be used by the client to manage his data stored in the cloud servers. The CSP provides a web interface for the client to store data into a set

of cloud servers, which are running in a cooperated and distributed manner. In addition, the web interface is used by the users to retrieve, modify and restore data from the cloud, depending on their access rights. Moreover, the CSP relies on database servers to map client identities to their stored data identifiers and group identifiers

2. Secure Deduplication And Data Security With Efficient And Reliable CEKM

The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We posit that secure deduplication services can be implement given additional security features insider attacker on Deduplication and outsider attacker by using the detection of masquerade activity. The confusion of the attacker and the additional costs incurred to distinguish real from bogus information, and the deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication.

3. Secure Authorized Deduplication with Encrypted Data for Hybrid Cloud Storage

Cloud computing has reached a maturity that leads it into a productive phase. This means that most of the main issues with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. This however does not mean that all the problems listed above have actually been solved, only that the according risks can be tolerated to a certain degree Cloud computing is therefore still as much a research topic, as it is a market offering. For better confidentiality and security in cloud computing we have proposed new deduplication 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. Proposed system includes proof of data owner so it will help to implement better security issues in cloud computing.

III. EXISTING SYSTEM

- A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications.
- Bellare et al. formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption variants for secure deduplication.
- Li addressed the key-management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files.
- Bellare et al. showed how to protect data confidentiality by transforming the predictable message into a unpredictable message.

Disadvantages of Existing System:

- Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners.
- Most of the previous deduplication systems have only been considered in a single-server setting.
- The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

IV. PROPOSED SYSTEM

In this paper, we show how to design secure deduplication systems with higher reliability in cloud computing. We 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. In other words, the secret shares of data will only be accessible by the authorized users who own the corresponding data copy.

Four new secure deduplication systems are proposed to provide efficient deduplication with high reliability for file-level and block-level deduplication, respectively. The secret splitting technique, instead of traditional encryption methods, is utilized to protect data confidentiality. Specifically, data are split into fragments by using secure secret sharing schemes and stored at different servers.

As we know that data Availablity is an important factor so, here we introduce a database at the backend which stores duplicate copy of data , In case of data lost, the database admin is the only person who has the access to that duplicate copy.

Advantages of Proposed System:

Distinguishing feature of our proposal is that data integrity, including tag consistency, can be achieved.

To our knowledge, no existing work on secure deduplication can properly address the reliability and tag consistency problem in distributed storage systems.

Our proposed constructions support both file-level and block-level deduplications.

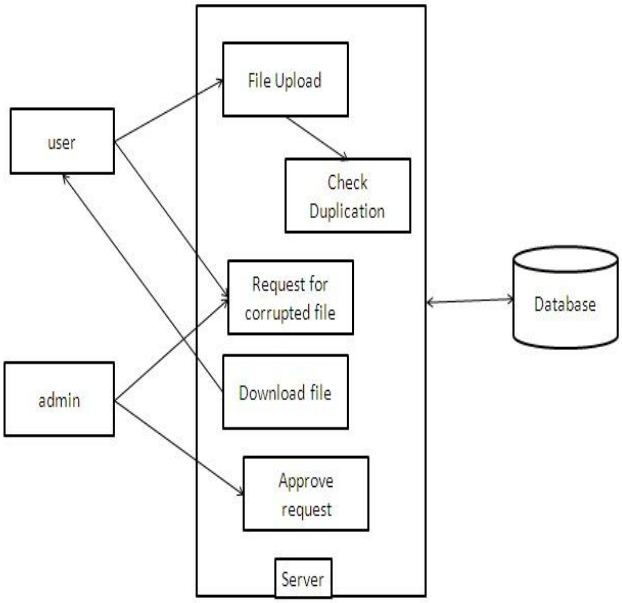
Security analysis demonstrates that the proposed deduplication systems are secure in terms of the definitions specified in the proposed security model. In more details, confidentiality, reliability and integrity can be achieved in our proposed system. Two kinds of collusion attacks are considered in our solutions. These are the collusion attack on the data and the collusion attack against servers. In particular,

the data remains secure even if the adversary controls a limited number of storage servers.

We implement our deduplication systems using the Ramp secret sharing scheme that enables high reliability and confidentiality levels. Our evaluation results demonstrate that the new proposed constructions are efficient and the redundancies are optimized and comparable with the other storage system supporting the same level of reliability.

- As we have introduce the database at the backend so, In case of data lost admin has the privillege to access the duplicate copies of data, due to which data availability can be achieved.

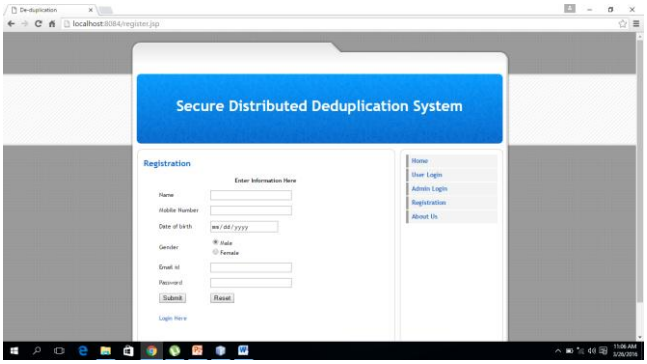
System Architecture

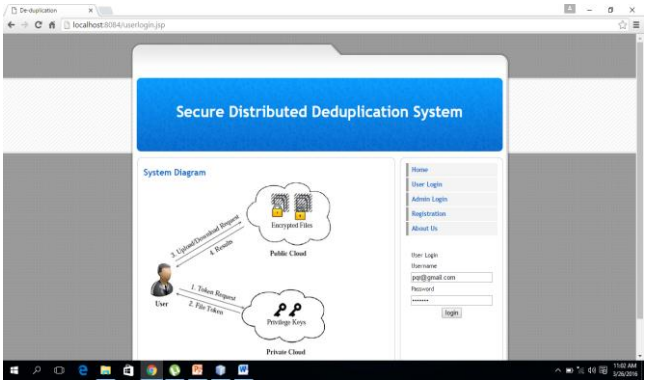


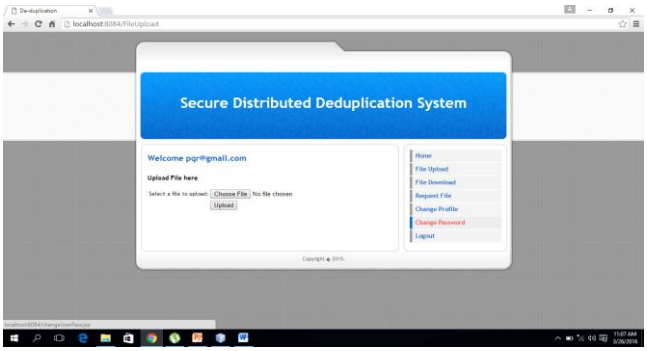
Implementation

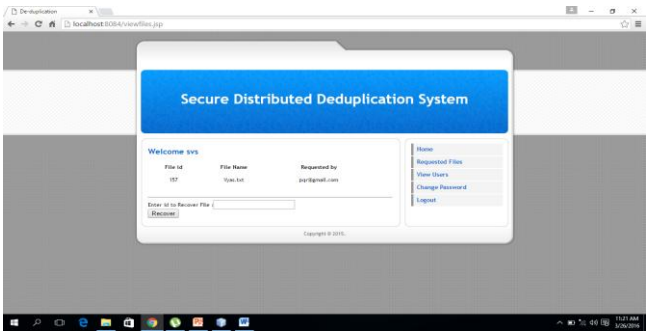
We describe the implementation details of the proposed distributed deduplication systems in this section. The main tool for our new deduplication systems is the Ramp secret sharing scheme (RSSS). The shares of a file are shared across multiple cloud storage servers in a secure way. The efficiency of the proposed distributed systems are mainly determined by the following three parameters of n, k, and r in RSSS. To download a file F user has to give the unique id of file so that file can be downloaded from the storage. To upload a file F, the user first performs the file-level deduplication. Different from the above constructions, the user needs to compute the secret shares $\{Fj\}1 \leq j \leq n$ of the file by using the Share algorithm. Then, ϕFj = TagGen(Fj) is computed and sent to S-CSP. For Block level deduplication, the file F is firstly divided into a set of fragments. For each block, the duplicate check operation is the same as the file-level check except file F is replaced with block Bi. we also consider how to achieve the integrity of the data stored in each S-CSP by using the message authentication code.

V. RESULT









We proposed the distributed de-duplication systems to improve the reliability of data while achieving the confidentiality of the users' outsourced data without an encryption mechanism. Four constructions were proposed to support file-level and fine-grained block-level data de-duplication. The security of tag consistency and integrity were achieved. We implement our de-duplication systems using the Ramp secret sharing scheme and demonstrated that it ensure small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations. we have also used Tag generation algorithm to map original data copied and perform duplicate check with server. And also Message Authentication code to authenticate mesage and provide integrity and authenticity assurance on message. Here we are going to store duplicate

copy of data which can be only accessed by admin as he is the only person who hasprivillege to accessed the data at the time of data lost.

REFERENCES

[1] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang Senior Member, IEEE and Mohammad Mehedi Hassan Member, IEEE and Abdulhameed Alelaiwi Member, IEEE, "Secure Distributed Deduplication Systems with Improved Reliability", IEEE Transactions on Computers, 2015.

[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows,"http://www.emc.com/collateral/analyst-reports/idcthe-digital-universe-in- 2020.pdf, Dec 2012.

