ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Data Management with De-Duplication System in Cloud Computing

Battu Rakesh Raju
M.Tech Student,
Department of CSE,
Malla Reddy College of
Engineering and Technology.

Dr.G.Sharada
Professor,
Department of CSE,
Malla Reddy College of
Engineering and Technology.

Dr. Sujatha Dandu
Professor,
Department of CSE,
Malla Reddy College of
Engineering and Technology.

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.

INTRODUCTION

With the explosive growth of digital data, deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication has received much attention from both academia and industry because it can greatly improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as archival storage systems.

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. A brief review is given in Section 6. Especially, with the advent of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which motivates enterprises and organizations to outsource data storage to third-party cloud providers, as evidenced by many real-life case studies. According to the analysis report of IDC, the volume of data in the world is expected to reach 40 trillion gigabytes in 2020. Today’s commercial cloud storage services, such as Dropbox, Google Drive and Mozy, have been applying deduplication to save the network bandwidth and the storage cost with client-side deduplication.

There are two types of deduplication in terms of the size: (i) file-level deduplication, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and (ii) block-level deduplication , which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixed size blocks simplifies the computations of block boundaries, while using variable-size blocks (e.g., based on Rabin fingerprinting) provides better deduplication efficiency.

Though deduplication technique can save the storage space for the cloud storage service providers, it reduces the reliability of the 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. If such a shared file/chunk was lost, a disproportionately large amount of data becomes inaccessible because of the unavailability of all the files that share this file/chunk. If the value of a chunk were measured in terms of the amount of file data that would be lost in case of losing a single chunk, then the amount of user data lost when a chunk in the storage system is corrupted grows with the number of the commonality of the chunk.

Thus, how to guarantee high data reliability in deduplication system is a critical problem. Most of the previous deduplication systems have only been considered in a single-server setting.

However, as lots of deduplication systems and cloud storage systems are intended by users and applications for higher reliability, especially in archival storage systems where data are critical and should be preserved over long time periods. This requires that the deduplication storage systems provide reliability comparable to other high-available systems.

EXISTING SYSTEM:

- Data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges.
- Such architecture is practical and has attracted much attention from researchers.
- The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

DISADVANTAGES OF EXISTING SYSTEM:

- Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.
- Identical data copies of different users will lead to different ciphertexts, making deduplication impossible.

PROPOSED SYSTEM:

In this paper, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model.

ADVANTAGES OF PROPOSED SYSTEM:

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.

Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality.

IMPLEMENTATION MODULES:-

- Cloud Service Provider
- Data Users Module
- sec Cloud Module
- Sec+ Deduplication System

MODULES DESCRIPTON:-

Cloud Service Provider

- In this module, we develop Cloud Service Provider module. This is an entity that provides a data storage service in public cloud.

- The S-CSP provides the data outsourcing service and stores data on behalf of the users.
- To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data.
- In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

Data Users Module

- A user is an entity that wants to outsource data storage to the S-CSP and access the data later.
- In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users.
- In the authorized deduplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

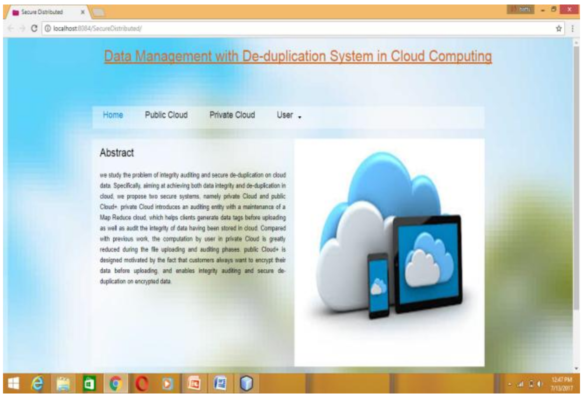
Sec Cloud Module

- Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service.
- Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud.
- The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

Sec+ De-duplication System

- We consider several types of privacy we need protect, that is, i) unforgeability of duplicate-check token: There are two types of adversaries, that is, external adversary and internal adversary.
- As shown below, the external adversary can be viewed as an internal adversary without any privilege.
- If a user has privilege p, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege p' on any file F, where p does not match p'. Furthermore, it also requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forge and output a valid duplicate token with p on any F that has been queried.

SCREEN SHOTS:

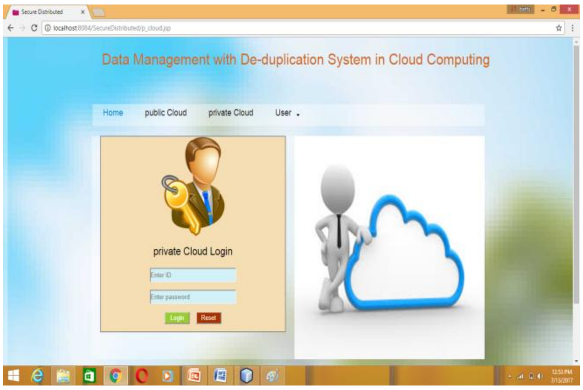


Data Management with De-duplication System in Cloud Computing

Abstract

we study the problem of integrity auditing and secure de-duplication on cloud data. Specifically, aiming at achieving both data integrity and de-duplication in cloud, we propose two secure systems, namely private Cloud and public Cloud+. private Cloud introduces an auditing entity with a maintenance of a Map Reduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. Compared with previous work, the computation by user in private Cloud is greatly reduced during the file uploading and auditing phases. public Cloud+ is designed motivated by the fact that customers always want to encrypt their data before uploading, and enables integrity auditing and secure de-duplication on encrypted data.

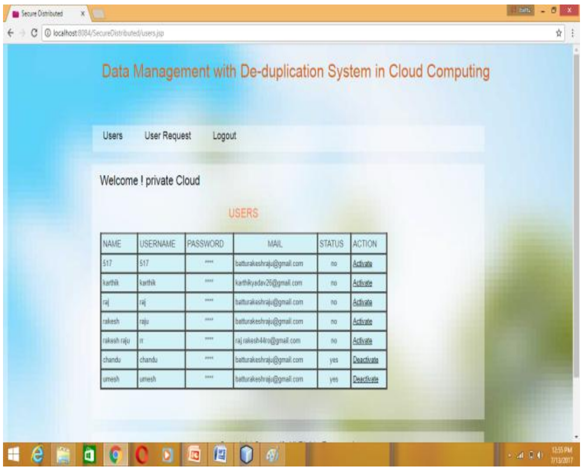
Fig: Home Page



Data Management with De-duplication System in Cloud Computing

private Cloud Login

Fig: Private Cloud Login Page


localhost:8084/SecureDistributed/users.jsp

Data Management with De-duplication System in Cloud Computing

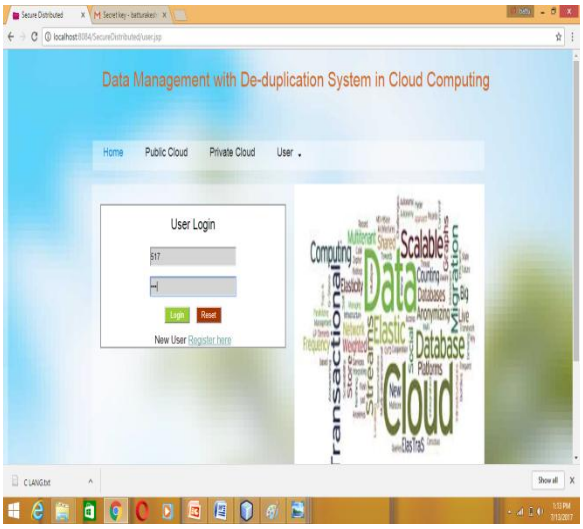
Users User Request Logout

Welcome ! private Cloud

USERS

NAME	USERNAME	PASSWORD	MAIL	STATUS	ACTION
517	517	****	batturakeshraju@gmail.com	no	Activate
karthik	karthik	****	karthikyadav26@gmail.com	no	Activate
raj	raj	****	batturakeshraju@gmail.com	no	Activate
rakesh	raju	****	batturakeshraju@gmail.com	no	Activate
rakesh raju	rr	****	raj.rakesh44ro@gmail.com	no	Activate
chandu	chandu	****	batturakeshraju@gmail.com	yes	Deactivate
umesh	umesh	****	batturakeshraju@gmail.com	yes	Deactivate

Fig: Private Cloud View User Requests


localhost:8084/SecureDistributed/user.jsp

Data Management with De-duplication System in Cloud Computing

Home Public Cloud Private Cloud User

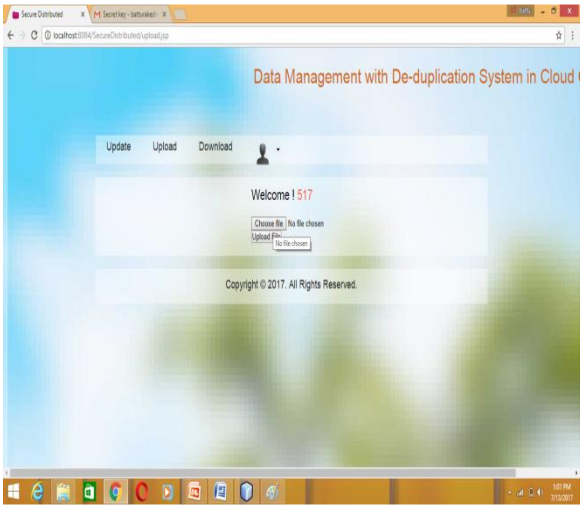
User Login

517

Login Reset

New User Register here

Fig: User Login Page


localhost:8084/SecureDistributed/upload.jsp

Data Management with De-duplication System in Cloud

Update Upload Download

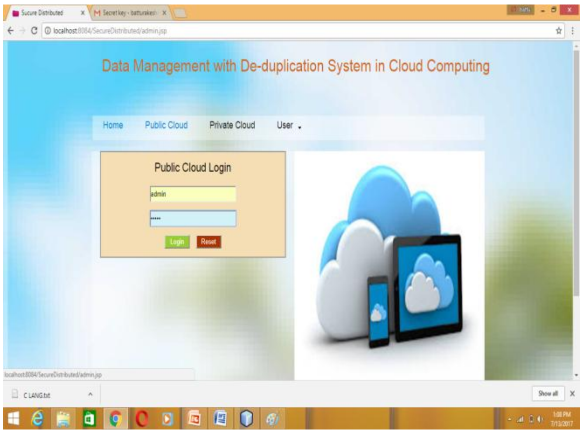
Welcome ! 517

Choose file No file chosen

Upload File

Copyright © 2017. All Rights Reserved.

Fig: File Upload Page


localhost:8084/SecureDistributed/admin.jsp

Data Management with De-duplication System in Cloud Computing

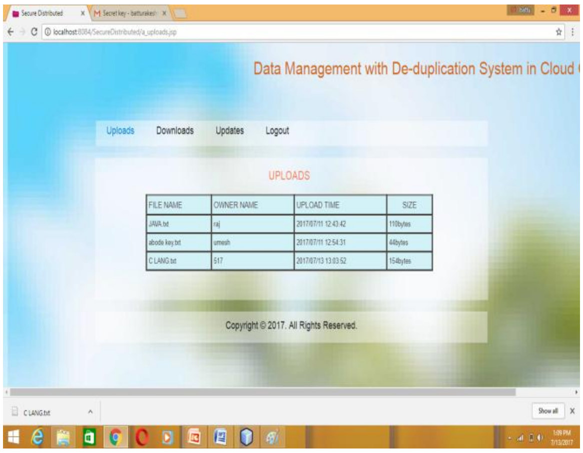
Home Public Cloud Private Cloud User

Public Cloud Login

admin

Login Reset

Fig: Public Cloud Login Page


localhost:8084/SecureDistributed/a_uploads.jsp

Data Management with De-duplication System in Cloud

Uploads Downloads Updates Logout

UPLOADS

FILE NAME	OWNER NAME	UPLOAD TIME	SIZE
JAVA.txt	raj	2017/07/11 12:43:42	110bytes
abode key.txt	umesh	2017/07/11 12:54:31	44bytes
C LANG.txt	517	2017/07/13 13:03:52	154bytes

Copyright © 2017. All Rights Reserved.

Fig: View Upload Files Page

CONCLUSION

In this paper, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several 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. Security analysis demonstrates that our schemes are 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 testbed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

REFERENCES

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

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

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.

[7] 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.

[8] 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 ICDCS, pages 617–624, 2002.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.