ISSN:0975-9646

A Review of Secure Authorized Deduplication with Encrypted Data for Hybrid Cloud Storage

Bhavanashri Shivaji Raut, Prof. H. A. Hingoliwala

Department of Computer Engineering, JSCOE PUNE
Pune, India

***Abstract*— Cloud Storage System are becoming increasingly popular with the continuous and exponential Increase of the number of users and the size of data. Data deduplication becomes more and more a necessity for cloud storage providers. Data deduplication is one of the important data compression technique for eliminating duplicate copies of repeating data . It has been widely used in the cloud storage to reduce the amount of storage space and save bandwidth. The advantage of deduplication unfortunately come with high cost in terms of new security and privacy challenges .The proposed scheme in this paper not only the reduces the cloud storage capacity but also improves the speed of data deduplication.**

To protect confidentiality of sensitive data while supporting deduplication the convergent encryption technique has been proposed to encrypt the data before outsourcing .This paper makes the first attempt to address the problem of authorized data deduplication .

Deduplication system is different from the traditional system ,because the differential privileges of users are further considered in duplicate check besides the data itself. Security analysis demonstrate that our scheme is secure in terms of the definitions specified in the proposed security model.

We show that our proposed system is authorized duplicate check scheme incurs minimal overhead ,compared to normal operations and also show that encryption for deduplicated storage can achieve performance and space saving close to that using the storage service.

Keywords: Deduplication, authorized duplicate check, confidentiality, hybrid cloud

I. INTRODUTION

Cloud computing provides a low-cost, scalable, location-independent infrastructure for data management and storage. Owing to the population of cloud service and the increasing of data volume ,more and more people pay attention to economize the capacity of cloud storage than before .Therefore how to utilize the cloud storage capacity well becomes important issue nowadays.

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data.

A Hybrid Cloud is a combined form of private clouds and public clouds in which some critical data resides in the enterprise's private cloud while other data is stored in and accessible from a public cloud.

Public cloud or eternal cloud describes cloud computing in the traditional main stream sense ,whereby resources are dynamically provisioned on a fine-grained ,self –services basics over the internet via web application /web services from an out-site third party provider who shares resources and bill on a fine –grained utility computing basis.

Private cloud and internal cloud are neologisms that some vendors have recently used to describe offerings the emulate cloud computing on private network. Hybrid clouds seek to deliver the advantages of scalability, reliability, rapid deployment and potential cost savings of public clouds with the security and increased control and management of private clouds.

Deduplication strategy can be categorized into two main strategies as follow, differentiated by the type of basic data units.

1).File-level deduplication: A file is a data unit when examining the data of duplication, and it typically uses the hash value of the file as its identifier. If two or more files have the same hash value, they are assumed to have the same contents and only one of these files will be stored.

2).Block-level deduplication: This strategy segments a file into several fixed-sized blocks or variable-sized blocks, and computes hash value for each block for examining the duplication blocks.

II. LITERATURE SURVEY

1.Fast and secure laptop backups with encrypted de-duplication

1.1 Back up algorithm(2010)

1.The data which is common between users to increase the speed of backup and reduce the storage requirement.

2.Supports client-end per user encryption is necessary for confidential personal data.

Disadvantages:

1. Network bandwidth can be a bottle-neck
2. Backing up directly to a cloud can be very costly

Conclusion

This provides the potential to significantly decrease backup times and storage requirement

2.Server Aided Encryption for De-duplicated Storage(2013)

1.That Provides secure deduplicated storage residting Brute-Force attack and realize it in a system –such as DupLESS

2.DupLESS that combines a LE-type base MLE Scheme

Disadvantages

1.That Provides secure deduplicated storage residting Brute-Force attack and realize it in a system –such as DupLESS

2.It is crucial to slow down brute force attacks.

Conclusion
DupLESS uses a low number of interactions with SS. It can work transparently on top of any Storage interface implementing a simple storage interface.

3. Secure Dedup: Server Deduplication with Encrypted Data for cloud Storage
1.Deduplication unfortunately come with a high cost in terms of new security and privacy challenges
Disadvantages
1.Does not impactthe overall storage and computational cost
Conclusion
A system achieves confidentiality and enables block level deduplication at the same time

4.Data Deduplication Scheme for cloud computing
4.1 Zhang fault Tolerant digital signature Scheme
1.Improves the speed of data deduplication 2.The Signature is computed for uploaded file for verifying the integrity of files.
Disadvantages
1.There is a problem of the worst case in that cloud storage server will regard all blocks as a new blocks and store all of these blocks, resulting in storing duplicate blocks the probability of the worst case is low and wont affect most.
Conclusion
To be concluded it improves the speed of data deduplication phase not only enhances the efficiency of data duplication.

5.Proofs Of Ownership in Remote storage System
5.1.Present solution based on Merkle Trees and Specific encoding
We identify attacks that exploit client side deduplication attempts to identify reduplication
Disadvantages
It is impossible to verify experimentally the assumption about the input distribution
Conclusion
Implemented a prototype of the new protocol and ran it to evaluate performance and asses the Pow scheme benefits.

6.Enhanced Dynamic whole file De-duplication(DWFD) for space optimization in private cloud storage backup
6.1.DWFD scheme is designed
To optimize the private cloud storage backup in order to provide high throughput to the users of the organization by increasing the de-duplication efficiency
Disadvantages
It is not sufficient used to development of chunk level deduplication and block level reduplication
Conclusion
It is highly desirable to improve the private cloud backup storage efficiency by reducing the de-duplication time

7.Weak leakage –resilient client side Deduplication of encrypted data in cloud storage
We propose a secure client –side deduplication scheme
Disadvantages
Convergent encryption and custom encryption methods are not semantically secure
Conclusion
Addressed an important security concern in cross-user client –side deduplication

III. THE PROPOSED SYSTEM

In the proposed system we are achieving the data deduplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

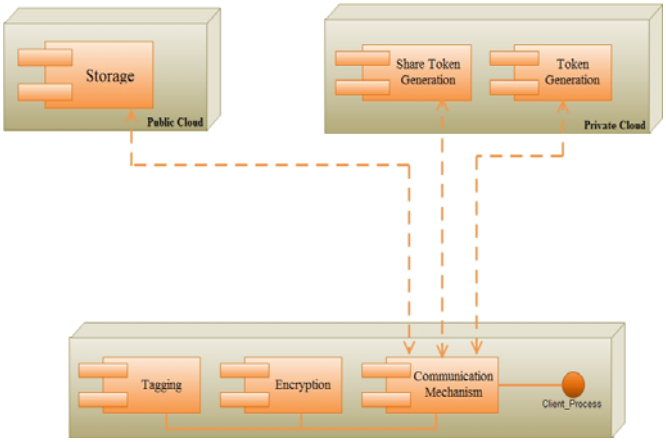

Fig1. Proposed System Framework Communication

Fig1. Shows the Proposed System Framework Communication between client ,Private Cloud ,Public Cloud.
In this work Public Cloud is used for storage data ,Private Cloud is used for performance the operations like share token generation ,token generation, also Client is performed operations like tagging of file ,encryption of file, communication between private cloud and public cloud.

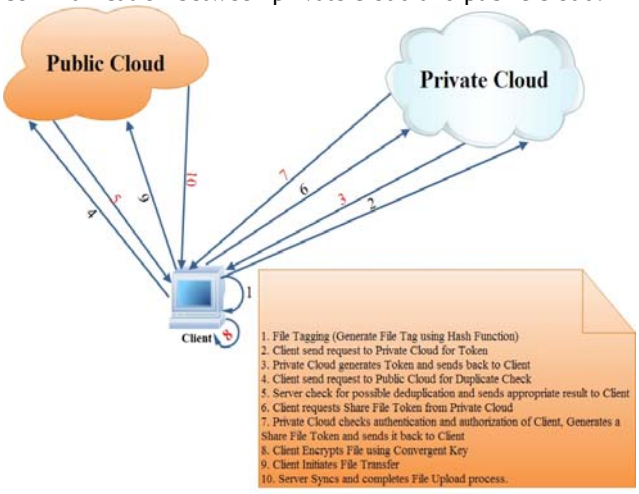

Fig2.Proposed System of Operations

ENCRYPTION OF FILES

Here we are using the common secret key *k* to encrypt as well as decrypt data. This will use to convert the plain text to cipher text and again cipher text to plain text. Here we have used three basic functions:

KeyGenSE: *k* is the key generation algorithm that generates κ using security parameter 1.

EncSE (k, *M*): *C* is the symmetric encryption algorithm that takes the secret κ and message *M* and then outputs the ciphertext *C*;

DecSE (k, *C*): *M* is the symmetric decryption algorithm that takes the secret κ and ciphertext *C* and then outputs the original message *M*.

CONFIDENTIAL ENCRYPTION

It provides data confidentiality in deduplication. A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a *tag* for the data copy, such that the tag will be used to detect duplicates.

PROOF OF DATA

The user have to prove that the data which he want to upload or download is its own data. That means he have to provide the convergent key and verifying data to prove his ownership at server.

CONCLUSION

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.

REFERENCES

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

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. “Dupless: Server aided encryption for deduplicated storage”. In *USENIX Security Symposium*, 2013.

[3] Pasqualo Puzio, Refik Molva ,MelekOnen ,”CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage”, SecludIT and EURECOM,France.

[4] Iuon –Chang Lin, Po-ching Chien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)

[5] Shai Halevi, Danny Harnik, Benny Pinkas,”Proof of Ownership in Remote Storage System”, IBM T.J.Watson Research Center, IBM Haifa Research Lab, Bar IIan University,2011.

[6] M. Shyamala Devi, V.Vimal Khanna,Naveen Balaji ”Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup”,IACSIT, August,2014.

[7] Weak Leakage-Resilient Client –Side deduplication of Encrypted Data in Cloud Storage” Institute for Info Comm Research,Singapore,2013

[8] Tanupriya Chaudhari , Himanshu shrivastav, Vasudha Vashisht, ”A Secure Decentralized Cloud Computing Environment over Peer to Peer”,IJCSMC,April,2013

[9] Mihir Bellare, Sriram keelveedhi,Thomas Ristenart ,”DupLESS: Server Aided Encryption for Deduplicated storage” University of California, San Diego2013