Access Control System Using Randomized Tags In Data Clouds

C.Vishnupriya
Presidency college

Abstract-*Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.*

Keywords-Deduplication, Data cloud, cloud security

I. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers.

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive computer games.

The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

Cloud computing era have lots of research issues. Deduplication is one of them. It is a compression technique which identifies and locates the duplicate data. It then eliminates duplicate copies of repeating data and saves the space for data that needs to be physically store. Hence, the two main advantages of data deduplication are Reduction in Storage Allocation and Efficient Volume of Replication.

1.1 Types of deduplication strategies

According to the operational area, data de-duplication strategies can be classified into two approaches.

1.1.1 File-level De-duplication

File level de-duplication, as the name suggests, is always performed over a single file. Identification of same hash value of two or more files determines that the files are similar

1.1.2 Block-level Deduplication

Block level deduplication is performed over blocks. Firstly it divides the files into blocks and stores just a single copy of each block. Fixed-sized blocks or variable-sized chunks can be used with block-level deduplication.

II. RELATED WORKS

Q. He, Z. Li, and X. Zhang talks about various deduplication techniques. The basic principle of deduplication is to maintain only one copy of the duplicate data provided with a pointer to point to the the duplicate blocks. This can be

done at file level, block level or byte level. The new data are compared with old data at byte level and if they match, they are marked as duplicate. Data pointers are updated and the redundant copy is deleted.

Z. Li, X. Zhang, and Q. He, discusses various cloud storage techniques. With respect to data deduplication, they suggest to retain only the unique instance of the data, thus, reducing data storage volumes. An index of the digital signature is created by the data deduplication engine for the data segment along with the signature of a given repository to identify data blocks. To check whether the data block is already present, a pointer is provided by the index. In the copy operation, the data deduplication software found in a block of data inserts a link to the original data block index location instead of storing the data block again. Appearance of the same block more than once would generate more pointers to the indexing table. Moving data from one storage system to another which are at different geographical locations is referred to as data migration in cloud storage. It aims at cooperating and keeping load balance in cloud storage system. Migration of data into other cloud storage units should occur and the pointers to be kept in the old stored positions intact, or modify and update the index as changes occur. But this may bring overhead to network bandwidth and access bottleneck to concurrent clients.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg proposed the notion of "proofs of ownership" (PoW) for deduplication systems in which a client can prove to a server based on Merkle trees and the error -control coding that it indeed has a copy of a file without actually uploading it. However, their scheme cannot guarantee the freshness of the proof in every challenge. Furthermore, their scheme has to build Merkle Tree on the encoded data, which is inherently inefficient. This scheme do not consider about data privacy.

The Proof of Ownership (PoW) concept is introduced by Halevi, a challenge-response protocol enabling a storage server to check if a requesting entity is the data owner, which is based on a short value. In other words, while uploading a data file (D) to the cloud, user first computes and sends a hash value hash = H(D) to the storage server. This later maintains a database of hash values of all received files, and looks up hash. If a match is found, then data file D is already outsourced to cloud servers. With respect to these cloud tags, there is no need to upload the file to remote storage servers. If there is no match found, then the user has to send the file data (D) to the cloud.

Douceur et al study the problem of deduplication in a multitenant system in which deduplication has to be reconciled with confidentiality. The authors propose the use of convergent encryption, i.e., deriving keys from the hash of the plaintext, so that two users will produce the same ciphertext from the same plaintext block, and the ciphertext can then be deduced.

M. W. Storer, K. M. Greenan, D. D. E. Long, and E. L. Miller point out some security issues with convergent encryption, while, proposing a security model and two protocols for secure data deduplication. There are two approaches to secure deduplication, viz., authenticated and anonymous. While the two models are similar, they each slightly differ in security properties. These both can beapplied to single server storage as well as distributed storage. In the former, single server storage, clients interact with a single file server which stores both data and metadata. In the later, metadata is stored on an independent metadata server, and data is stored on a series of object-based storage devices (OSDs).

D. Harnik, B. Pinkas, and A. Shulman-Peleg discusses the shortcomings of client-side deduplication, and attacks to privacy and confidentiality that can be addressed without a full-edged POW scheme by triggering deduplication only after a small, but random, number of uploads.

J. Yuan and S. Yu. Here, the data owner outsources the erasure-coded file to the cloud server along with its corresponding authentication tags. The integrity of the outsourced file is audited by a user (who may not be the owner) who challenges the cloud with a challenging message. On receiving this message, the cloud generates the information proof based on the public key and sends it to the user. The user verifies the data integrity with the proof information, using our verification algorithm. In order to deduplicate the data, when a user wants to upload a data file that already exists in the cloud, the cloud server executes a checking algorithm to check if this user actually possesses the whole file. If the user passes the checking, he/she can directly use the file existed on the server without reuploading it.

Zhu et al try to solve the problem of disk bottlenecks while deduplicating. A significant challenge is to identify and eliminate duplicate data segments at a high rate on a low-cost system that cannot afford enough RAM to store and access a large metadata of the stored blocks.

Camble et al. proposes the "Sparse Indexing" deduplication system which uses a different approach to avoid

the chunk lookup disk bottleneck. Here, the chunks are sequentially grouped into segments. These segments are then used to search similar existing segments using a RAM based index, which stores only a small fraction of the already stored chunks. In contrast to other approaches, Sparse Indexing allows to store a chunk multiple times if the similarity based system is not able to detect the segments, which already have stored the chunk. Therefore, Sparse Indexing is a member of the class of approximate data deduplication systems.

Stanek et al. presented a novel encryption scheme that provides differential security for both popular data and unpopular data. For popular data that are not particularly sensitive, the traditional conventional encryption is performed. For unpopular data, another two-layered encryption scheme with stronger security, while supporting deduplication is proposed. In this way, they achieved better tradeoff between the efficiency and security of the outsourced data.

M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless gives the Server aided encryption for deduplicated storage proposes different security mechanisms. Confidentiality can be preserved by transforming predictable message into unpredictable form. One new concept introduced as Key server (Third party auditor) to generate the file tag for duplicate check.

III. METHODOLOGY

our system model which includes main components such as key request handler and tag generation, deduplication check, convergent encryption and decryption. A new deduplication system obtained for differential duplicate check is proposed under this hybrid cloud architecture where in the public cloud resides the Secure Client Service Provider (S-CSP). The user is only allowed to perform the duplicate check for files marked with the corresponding privileges. To support authorized deduplication, the tag of a file F will be recognized by the file F along with the privilege. To show the difference with

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.

Private 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.

Secure Deduplication 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.

IV. RESULTS

Cloud storage and data deduplication techniques have pulled the attention of many researchers recently. He et al. [4] have researched over various cloud storage techniques and have recommended some techniques for reducing the storage volumes. The authors have proposed data deduplication engine that generated an index of digital signatures. This index also provides pointers for knowing the presence of data blocks.

new concept of data migration is emerging now-a-days. It is nothing but the relocation of data from one storage to other storage. Both the storages are geographically separated. New concept known as Proof of Ownership is implemented for deduplication process by Halevi et al. [5]. Here a client can manifest based on Merkle-Hash Tree [5]. But the proposed system doesn't take into attention providing the data security. The client-side deduplication has many limitations which are discussed by Harnik et al. [6]. An excellent survey on various deduplication techniques is done in [7]. The concept of deduplication is very simple, it says only a single copy of the duplicate data must be maintained and there should be a pointer for pointing the duplicate blocks. And this process can be attained in three levels: file level, byte level and block level.

V. 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.

FUTURE SCOPE

The traditional approach of convergent encryption cannot be suited in this secure deduplication as it is susceptible to brute-force attack. To overcome this drawback, modified version of convergent encryption can be used by introducing two approaches - domain separation and cryptographic tuning This gives a better authorized deduplication approach.

REFERENCES

[1] K. Jin and E. Miller, "The effectiveness of deduplication on virtual machine disk images" In Proc. SYSTOR 2009: The Israeli Experimental Systems Conference..

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.Fröhlich, B. and Plate, J.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[4] Z. Li, X. Zhang, and Q. He, Analysis of the key technology on cloud storage, in International Conference on Future Information Technology and Management Engineering, 2010, pp. 427428.

[5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491500. ACM, 2011.

[6] D. Harnik, B. Pinkas, and A. Shulman-Peleg. Side channels in cloud services: Deduplication in cloud storage. IEEE Security & Privacy, 8(6), 2010.

[7] Q. He, Z. Li, and X. Zhang, Data deduplication techniques, in International Conference on FutureInformation Technology and Management Engineering,pp.431-432,2010.

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

[9] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[10] 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.

[11] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.