Authenticated Deduplication System with Differential Privilege user Checks in Cloud

Shubha B P M tech CSE
M tech, CSE student
T.John Institute Of Technology
Bangalore, India

Mrs. Shimi Jeyaseelan
Asst. professor, Dept. of CSE
T.John Institute Of Technology
Bangalore, India

***Abstract*—Data deduplication is one of useful data compression scheme for eliminating duplicate copies of repeating data in the storage, and has been very popular and cheap in cloud system to reduce the amount storage space and bandwidth. For Security of sensitive data with supporting deduplication, the convergent encryption technique has been introduced to encrypt the data before uploading. This proposed system address Authenticated Deduplication with Security measures. In this system all users are having their own privilege and the user who is uploading the file first time can able to fix the Access Privilege and Deduplication Privilege. Based on these setting other user can able to access the file as well as they can deduplicate the file. For better security, this system uses Hybrid cloud approach. All the files uploaded into the cloud have to undergo two processes. First one is Encryption process for data Security and another one is Tag Generation Process to identify the file presence in the cloud. In this system all the keys with privilege are stored in Private Cloud server and encrypted data is stored in Public Cloud. This system avoids unnecessary duplication files in cloud storage, as well as it will authorized users to duplicate the file for fault tolerance.**

Index Terms*—Deduplication, authorized duplicate check, confidentiality, hybrid cloud

I. INTRODUCTION

One critical challenge of cloud storage services is the management of the ever-increasing volume of data. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent.

Deduplication can take place at either the file level or the block level. For file-level deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both insider and outsider attacks. Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible. It en-crypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys.

II. RELATED WORK

Traditional encryption, while providing data confidentiality, is incompatible with data deduplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making deduplication impossible. Convergent encryption[1] has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file.A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption allows the cloud to perform deduplication on the cipher texts and the proof of ownership prevents the unauthorized user to access a file.

However, previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization. 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. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (SCSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent, do not support the duplicate check with differential privileges.In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

A. Disadvantages of Existing System

- Various loose points in the existing system.
- There is no concept of hybrid cloud.
- Data can be stolen due to various bucks and errors.
- Authorized deduplication is not done.
- Deduplication cannot protect the security of predictable files

III. PROBLEM STATEMENT AND ASSUMPTIONS

A. 3.1 Problem Statement

Previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. There is only a single cloud in the existing system which makes very hard deduplication process. Existing deduplication system cannot prevent the privilege private key sharing among users.

B. 3.2 System Model

In the system model, our setting of interest is an enterprise network, consisting of a group of affiliated clients (forexample, employees of a company) who will use the S-CSP and store data with deduplication technique. In this setting, deduplication can be frequently used in these settings for data backup and disaster recovery applications while greatly reducing storage space. Here we considering the hybrid cloud which consists user, private cloud,and S-CSP in puplic cloud.

S-CSP: 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.

Data Users: A user is an entity that wants to store the data to the S-CSP and user and access the data any time.

Private Cloud: Compared with the traditional deduplication architecture in cloud computing, this is a new entity private cloud is 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.

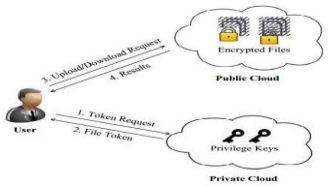

Fig 1 Architecture for Authorized Deduplication

Notice that this system is a novel architecture for data deduplication in cloud computing, which consists of a twin clouds that is.,the public cloud and the private cloud.Actually, this hybrid cloud setting has attracted more and more attention recently

C. 3.3 AdversaryModel

In this paper, we assume both public cloud and private cloud, suppose that all the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under the assumption,two kinds of adversaries are considered, that is,1)external adversaries which aim to extract secret information as much as possible from both public cloud and private cloud; 2) internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. Such adversaries may include S-CSP, private cloud server and authorized users.

D. 3.4 Design goals

In this paper, we address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for

• *Differential Authorization.* Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server.

• *Authorized Duplicate Check.* Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs duplicate check directly and tells the user if there is any duplicate.

IV. PROPOSED SYSTEM

Here we address the problem of deduplication with differential privileges in cloud computing, we consider a

hybrid cloud architecture consisting of a public cloud and a private cloud. Unlike existing data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely per-form 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.
A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP resides in the public cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
Proposed deduplication system provides,

System Setup: A symmetric key K_{pi} for each user will be selected and the set of keys will be sent to the private cloud. Each user is assumed to have a secrete key to perform the identification with servers. Each user has the privilege set. The private cloud server will maintain a table which stores each user's public information and its corresponding privilege set.

File Uploading: Suppose that a data owner wants to upload and share a file with user whose privilege belongs to the privilege set. The data owner needs interact with the private cloud before performing duplicate check with public cloud, the data owner performs an identification to prove its identity with private key. If it is passed, the private cloud server will find the corresponding privileges of the user from its table list. The user computes and sends the file to private cloud server, who will return back to the user for satisfying privilege. Then the user will interact and send the file to the public cloud

- if a file duplicate is found by the public cloud, the user proceeds the proof of ownership of this file with the public cloud. If the proof is passed, the user will be assigned a pointer, which allows him to access the file.
- If no duplicate is found, the user computes the encrypted file and uploads to the cloud server.

File Retrieving; Suppose a user wants to download a file. It first sends a request and file name to the S-CSP. Upon receiving the request and file name, the S-CSP will check whether the user is eligible to download file. If failed, S-CSP sends back an abort signal to the user to indicate the download failure.Otherwise, the S-CSP returns the corresponding ciphertext, user needs to decrypt to recover original file.

Advantages of Proposed System:

- A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP resides in the public cloud.
- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- The users without corresponding privileges cannot perform the duplicate check.

V. Security Analysis

Our System is designed to solve the differential privilege problem in secure deduplicaion. The security will be analyzed in terms of two aspects, that is, the authorization of duplicate chech and confidentiality of data.

VI Expected Outcome

To efficiently solving the problem of deduplication with differential privileges in cloud computing, we consider a hybrid cloud architecture consisting of a public cloud and a private cloud. Unlike existing 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 an 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. A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP resides in the public cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.

VII CONCLUSION

This proposed system address Authenticated Deduplication with Security measures. In this system all users are having their own privilege and the user who is uploading the file first time can able to fix the Access Privilege and Deduplication Privilege. Based on these setting other user can able to access the file as well as they can deduplicate the file. For better security, this system uses Hybrid cloud approach, in which the duplicate check tokens of the files are generated by the private cloud server with private keys, Furthermore, 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.

References

[1] 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.
[2] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In *Proc. of StorageSS*, 2008.
[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] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296– 312, 2013.
[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEE Transactions on Parallel and Distributed Systems*, 2013.
[6] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.
[7] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*, 22(1):1–61, 2009
[8] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012.
[9] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. *IACR Cryptology ePrint Archive*, 2013:149, 2013.
[10] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In *Technical Report*, 2013.
[11] R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, *ACM Symposium on Information, Computer and Communications Security*, pages 81–82. ACM, 2012
[12] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In *ASIACCS*, pages 195–206, 2013.
[13] 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.