Episteme: an online interdisciplinary, multidisciplinary & multi-cultural journal
Bharat College of Arts and Commerce, Badlapur, MMR, India
Volume 7, Issue 3 **December 2018**

AUTHORIZED DATA DEDUPLICATION FOR SECURE CLOUD

By Ankita Awasarmal

Abstract

Data deduplication is one among vital knowledge compression techniques for eliminating duplicate copies of repetition knowledge, and has been wide employed in cloud storage to cut back the amount of storage space and save bandwidth. The main advantage of using cloud storage from the customers' expectation view is that customers will scale back their expenditure in buying and maintaining storage infrastructure whereas solely paying for the quantity of storage requested, which may be scaled-up and down upon demand. To protect the confidentiality of sensitive knowledge whereas supporting deduplication, the focused secret writing technique has been planned to encrypt the information before outsourcing. For raised shield knowledge security, this paper makes the primary plan to formally address the matter of licensed knowledge deduplication. Completely different from ancient deduplication systems, the differential privileges of user's area unit more thought-about in duplicate check besides the info itself. Addition to this we present many new deduplication constructions supporting licensed duplicate check in a hybrid cloud design. Security analysis demonstrates that our theme is secure in terms of the definitions as per the planned security model. As a proof of construct, we have a goal to implement a paradigm of our planned licensed duplicate check theme and conduct test bed experiments using our paradigm.

Keywords

Deduplication, authorized duplicate check, confidentiality, hybrid cloud

Objective of study

To study the data deduplication & to understand the complete working of cloud using authorized deduplication.

Introduction

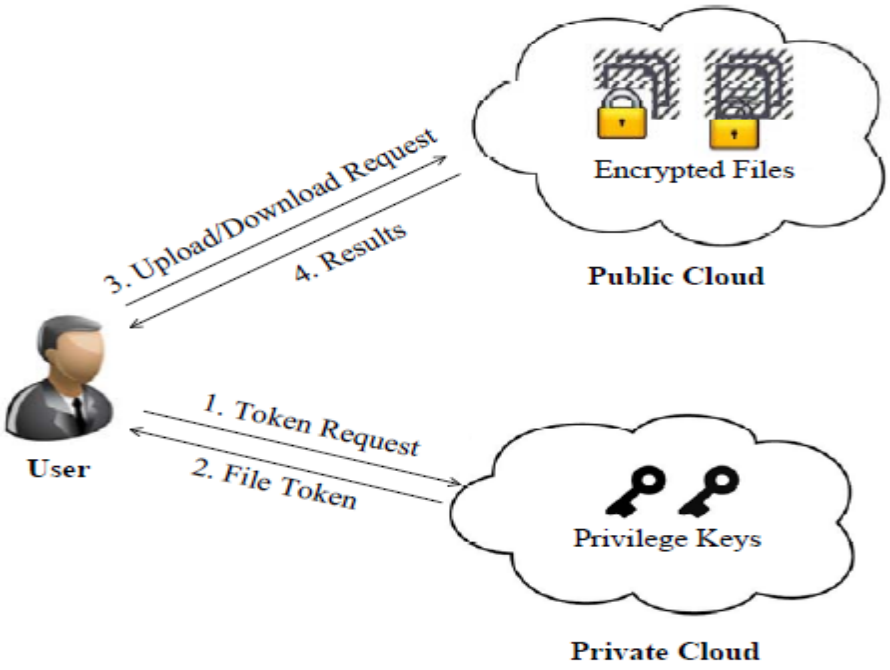
Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data.

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.

Need for deduplication

Deduplication can take place at either the file level or the block level. For file-level de duplication, it eliminates duplicate copies of the file. De duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.To prevent unauthorized access, a secure proof of ownership (POW) 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.

Architecture of Authorized Deduplication



Architecture of Authorized Deduplication

Modules

1. Cloud Service Provider
2. Data Users Module
3. Private Cloud Module
4. Secure Deduplication System

1. 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. S-CSP is always online and has abundant storage capacity and computation power.

2. Data User 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.

3. 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.

4. 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. the external adversary can be viewed as an internal adversary without any privilege.

5. Key Requesting

Once the key request was received, the sender can send the key or he can decline it. With this key and request id which was generated at the time of sending key request the receiver can decrypt the message.

Recommendation

- **Following Technique Used in Data Deduplication Process**

Symmetric Encryption

Symmetric encryption uses a common secret key κ to encrypt and decrypt information.

A symmetric encryption scheme consists of three primitive functions:

• KeyGenSE(1 λ) → κ is the key generation algorithm that generates κ using security parameter 1 λ ;

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

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

Convergent Encryption

Convergent encryption provides data confidentiality in de-duplication. A user (or data owner) derives aconvergent 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. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Note that both the convergent key and the tag are independently derived and the tag cannot be used to deduce the convergent key and compromise data confidentiality. Both the encrypted data copy and its corresponding tag will be stored on the server side.

A convergent encryption scheme can be defined with four primitive functions:

1. KeyGenCE(M)!K is the key generation algorithm that maps a data copy M to a convergent key K.

2. EncCE(K, M)!C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a ciphertextC;

3. DecCE(K, C)!M is the decryption algorithm that takes both the ciphertextC and the convergent key K as inputs and then outputs the original data copy M;

Proof of Ownership

The notion of proof of ownership(POW) enables users to prove their ownership of data copies to the storage server.Specifically, POW is implemented as an interactive algorithm (denoted by POW). The verifier derives a short value $\phi(M)$ from a data copy M. To prove the ownership of the data copy M, the properneeds to send $\phi$to the verifier such that $\phi = \phi(M)$.

PSEUDO CODE

Step1:Calculate the two convergent key values

Step2: Compare the two keys and files get accessed.

Step3: Apply de-duplication to eradicate the duplicate values.

Step4: If any other than the duplicates it will be checked once again and make the data unique.
Step5: That data will be unique and also more confidential the authorized can access and data is stored.

Identification Protocol

An identification protocol Π can be described with two phases: Proof and Verify. In the stage of Proof, a prover/user U can demonstrate his identity to a verifier by performing some identification proof related to his identity. The input of the prover/user is his private key skU that is sensitive information such as private key of a public key in his certificate or credit card number etc. that he would not like to share with the other users. The verifier performs the verification with input of public information pkU related to skU . At the conclusion of the protocol, the verifier outputs either accept or reject to denote whether the proof is passed or not. There are many efficient identification protocols in literature, including certificate-based, identity-based identification.

Conclusion

The notion of authorized data de duplication was suggested to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de duplication 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.

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.

References

- Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," Tech. Rep. IBM Research, Zurich, ZUR 1308-022, 2013.
- M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data de duplication," in Proc. 4th ACM Int. Workshop Storage Security Survivability, 2008, pp. 1–10.
- M. Bellare and A. Palacio, "Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks," in Proc. 22nd Annu. Int. Cryptol. Conf. Adv. Cryptol., 2002, pp. 162–177.
- S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twinclouds: An architecture for secure cloud computing," in Proc. Workshop Cryptography Security Clouds, 2011, pp. 32–44.

- Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou,"A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 26, NO. 5, MAY 2015.
- P. Anderson and L. Zhang. "Fast and secure laptop backups with encrypted de-duplication". In Proc. of USENIX LISA, 2010.
- M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Server aided encryption for deduplicated storage". In USENIX Security Symposium, 2013.
- Pasqualo Puzio, Refik Molva ,MelekOnen ,"CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage", SecludIT and EURECOM,France.

Bio

Ankita Awsarmal is working in Bharat College of Arts and Commerce, Badlapur. She can be contacted at **ankitabhalerao05@gmail.com**