Twin Cloud based Deduplication Scheme with user's Privileges

K. Jamuna[1] R. Meera[2] R. Satya[3]
Final Year B.E.
K. Ramakrishnan College of Engineering

N. Vijaya Raj
Assistant Professor, CSE
K. Ramakrishnan College of Engineering

***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 and width. 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 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 test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.**

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

I. 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.To make data management scalable in cloud computing, deduplication has been a well-known technique and has attracted more and more attention recently. 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. 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.

II RELATED WORK

Secure deduplication. With the advent of cloud computing, secure data deduplication has attracted much attention recently from research community. Yuan and Yu proposed a deduplication system in the cloud storage to reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality, This shows how to protect the data confidentiality by transforming the predictable.

A. A First Attempt

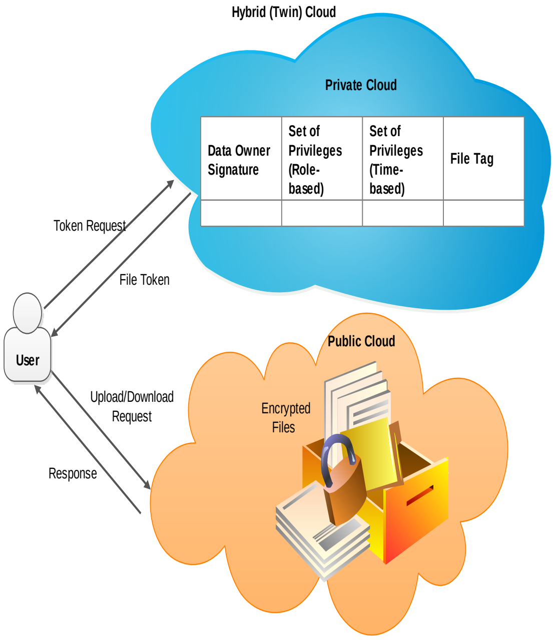
Before introducing our construction of differential deduplication, we present a straightforward attempt with the technique of token generation TagGen; k_pÞ above to design such a deduplication system. The main idea of this basic construction is to issue corresponding privilege keys to each user, who will compute the file tokens and perform the duplicate check based on the privilege keys and files. In more details, suppose that there are N users in the system and the privileges in the universe is defined as P ¼ fp_1 ; . . . ; p_sg. For each privilege p in P, a private key k_p will be selected. For a user U with a set of privileges P_U, he will be assigned the set of keys fk_pi p_i 2P_U. File uploading. Suppose that a data owner U with privilege set P_U wants to upload and share a file F with users who have the privilege set. The user computes and sends S-CSP the file token and TagGen for all.

If a duplicate is found by the S-CSP, the user proceeds proof of ownership of this file with the S-CSP. If the proof is passed, the user will be assigned a pointer, which allows him to access the file. Otherwise, if no duplicate is found, the user computes the encrypted file C_F; FÞ with the convergent key kF KeyGen and uploads C_F, ff$^{0F;\ pgÞ}$ to the cloud server. The convergent key k_F is stored by the user locally. File retrieving. Suppose a user wants to download a file F. It first sends a request and the 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 F. If failed, the S-CSP sends back a signal to the user to indicate the download failure. Otherwise, the S-CSP returns the corresponding ciphertext C_F. upon receiving the encrypted data from the S-CSP; the user uses the key k_F stored locally to recover the original file F.

B. Our Proposed System Description

To solve the problems of the construction in Section 4.1, we propose another advanced deduplication system supporting authorized duplicate check. In this new deduplication system, hybrid cloud architecture is introduced to solve the problem. The private keys for privileges will not be issued to users directly, which will be kept and managed by the private cloud server instead. In this way, the users cannot share these private keys of privileges in this proposed construction, which means that it can prevent the privilege key sharing among users in the above straight forward construction. To get a file token, the user needs to send a request to the private cloud server. The intuition of this construction can be described as follows. To perform the duplicate check for some file, the user needs to get the file token from the private cloud server. The private cloud server will also check the user's identity before issuing the corresponding file token to the user. The authorized duplicate check for this file can be performed by the user with the public cloud before uploading this file. Based on the results of duplicate check, the user either uploads this file or runs POW.

Before giving our construction of the deduplication sys-tem, we define a binary relation R ¼ fp; p^0 Þg as follows. Given two privileges p and p^0, we say that p matches p^0 if and only if Rp; p^0 Þ ¼ 1. This kind of a generic binary relation definition could be instantiated based on the background of applications, such as the common hierarchical relation. More precisely, in a hierarchical relation, p matches p^0 if p is a higher-level privilege. For example, in an enterprise management system, three hierarchical privilege levels are defined as Director, Project lead, and Engineer, where Director is at the top level and Engineer is at the bottom level. Obviously, in this simple example, the privilege of Director matches the privileges of Project lead and Engineer. We provide the proposed deduplication system as follows. System setup. Furthermore, each user U is assumed to have a secret key sk$_U$ to perform the identification with servers. Assume that user U has the privilege set PU. It also initializes a POW protocol POW for the file ownership proof. The private cloud server will maintain a table which stores each user's public information pkU and its corresponding privilege set P_U. We design and implement a new system which could protect the security for predictable message. The main idea of our technique is that the novel encryption key generation algorithm. For simplicity, we will use the hash functions to define the tag generation functions and convergent keys in this section.



C. Further Enhancement

Though the above solution supports the differential privilege duplicate, it is inherently subject to brute-force attacks launched by the public cloud server, which can recover files falling into a known set. More specifically, knowing that the target file space underlying a given ciphertext C is drawn from a message space S. F$_n$g of size n, the public cloud server can recover F after at most n off-line encryptions.

III 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] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing,"in Proc. Workshop Cryptography Security Clouds, 2011, pp. 32–44.

[2] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-

[3] duplication," in Proc. 24th Int. Conf. Large Installation Syst. Admin., 2010, pp. 29–40.

[4] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. ACM Conf. Comput. Commun. Security, 2011, pp. 491–500.

[5] 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 Proc. Int. Conf. Distrib. Comput. Syst.,2002, pp. 617–624.

[6] S. Quinlan and S. Dorward, "Venti: A new approach to archival storage," in Proc. 1st USENIX Conf. File Storage Technol., Jan. 2002, p. 7.

[7] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proc. 22nd USENIX Conf.Sec.Symp., 2013, pp. 179–194.

[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int.Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.

[9] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc. 27th Annu. ACM Symp. Appl. Comput.,2012, pp. 441–446.

[10] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy-aware data intensive computing on hybrid clouds," in Proc.18th ACM Conf. Comput. Commun. Security, 2011, pp. 515–526.

[11] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," IACR Cryptology ePrint Archive,2013:149, 2013.