ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Hybrid Cloud Approach for Secure Authorized Deduplication

Mohil .M. Shingare[1], Pankaj Birajdar[2], Hrishikesh Bibrale[3]
Student[1, 2, 3]
Department of Computer
Flora Institute of Technology, Pune, India

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 proposed system makes an 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: Data deduplication, Hybrid cloud, Proof of Ownership (POW), Confidentiality, Convergent Key

I. INTRODUCTION

Cloud computing has been deployed in a variety of data storages, data centres, network communications, data managements. Researchers introduce and define cloud computing in different aspects and terms. The effectiveness of many existing systems is strengthened by the use of proposed model.

The earlier systems were characterized by the following:-

- Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.

- Identical data copies of different users will lead to different cipher texts, making deduplication impossible.

The US National Institute of Standards and Technology defined cloud computing as a model for enabling access to a pool of resources such as servers, networks, applications, and services with low cost and minimal management. The characteristics consist of on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured pay-as-you-go services. Meanwhile the deployment models are highlighted with private cloud, public cloud, and hybrid cloud.

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. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both inside and outside attacks.

II. LITERATURE SURVEY

Security proofs for identity-based identification and signature schemes by M. Bellare, C. Namprempre, and G. Neven provides either security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly in existing literature. Underlying these is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work.

Revdedup: A reverse deduplication storage system optimized for reads to latest backups, by C. Ng and P. Lee. Deduplication is known to effectively eliminate duplicates, yet it introduces fragmentation that degrades read performance. RevDedup, a deduplication system that optimizes reads to the latest backups of virtual machine (VM) images using reverse deduplication is proposed. In contrast with conventional deduplication that removes duplicates from new data, RevDedup removes duplicates from old data, thereby shifting fragmentation to old data while keeping the layout of new data as sequential as possible.

Secure deduplication with efficient and reliable convergent key management by P. Lee, and W. Lou. Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. It makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. First it introduces a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud.

III. PROPOSED SYSTEM

The convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, we make 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 test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

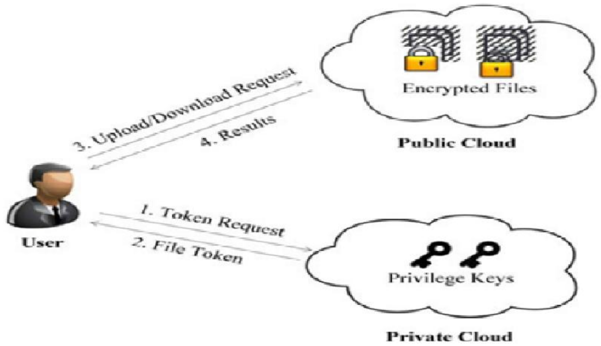

Figure.1. System Architecture

The deduplication check scheme is used to take appropriate actions over the user's queries. It is used to check the duplicate copy of the files which are been stored in the cloud by the employee. The query processing is used in following steps

- The admin gives permission to employee to create profile.
- Admin gives access specifications to employee.
- The employee can then utilize the storage space to store his file if the file is not present.
- In case a file already present then it will throw a query.
- Based on the access specification the employee can either read or write the file in cloud.

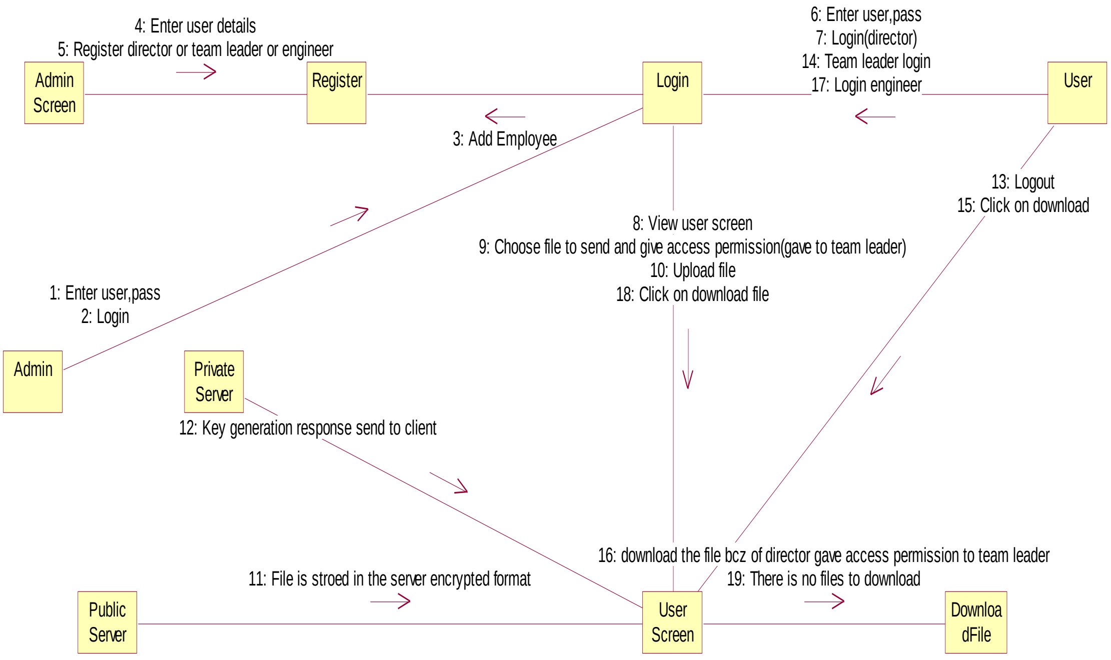
4: Enter user details
5: Register director or team leader or engineer
Admin Screen
Register
Login
6: Enter user,pass
7: Login(director)
14: Team leader login
17: Login engineer
User
3: Add Employee
13: Logout
15: Click on download
8: View user screen
9: Choose file to send and give access permission(gave to team leader)
10: Upload file
18: Click on download file
1: Enter user,pass
2: Login
Admin
Private Server
12: Key generation response send to client
16: download the file bcz of director gave access permission to team leader
19: There is no files to download
11: File is stroed in the server encrypted format
Public Server
User Screen
DownloadFile

Figure.2. Collaborative Diagram

At a high level, our setting of interest is an enterprise net-work, consisting of a group of affiliated clients (for example, 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. Such systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions. There are three entities defined in our system, that is, users, private cloud and S-CSP in public cloud. The S-CSP performs deduplication by checking if the contents of two files are the same and stores only one of them. The access right to a file is defined based on a set of privileges. The exact definition of a privilege varies across applications. For example, we may define a role-based privilege according to job positions (e.g., Director, Project Lead, and Engineer), or we may define a time-based privilege that specifies a valid time period (e.g., 2014-01-01 to 2014-01-31) within which a file can be accessed. A user, say Alice, may be assigned two privileges “Director” and “access right valid on 2014-01-01”, so that she can access any file whose access role is “Director” and accessi-ble time period starts from 2014-01-01. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denote the tag with specified privileges. A user computes and sends duplicate-check tokens to the public cloud for authorized duplicate check.

Users have access to the private cloud server, a semi-trusted third party which will aid in performing deduplicable encryption by generating file tokens for the requesting users. We will explain further the role of the private cloud server below. Users are also provisioned with per-user encryption keys and credentials (e.g., user certificates). We only consider the file-level deduplication for simplicity. In another word, we refer a data copy to be a whole file and file-level deduplication which eliminates the storage of any redundant files. Actually, block-level deduplication can be easily deduced from file-level deduplication, which is similar to. Specifically, to upload a file, a user first performs the file-level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well; otherwise, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. Each data copy (i.e., a file or a block) is associated with a token for the duplicate check.

Software resources required

Operating System	Windows
Technology	Java and J2SE
Front end	Swings & AWT
Java Version	J2SDK1.6

Hardware resources required

System	Pentium IV 2.4 GHz.
Hard Disk	40 GB.
Floppy Drive	1.44 Mb.
Ram	512 Mb.

IV. ADVANTAGES AND DISADVANTAGES

ADVANATAGES:-

- Convergent encryption ensures data privacy in deduplication.
- “proofs of ownership” (PoW) for deduplication systems, such that a client can efficiently prove to the cloud storage server that he/she owns a file without uploading the file itself.
- Data is kept highly confidential.

DISADVANATGES:-

- The proposed system checks only the file name and not the content within that file while checking for deduplication.

V. CONCLUSION

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 test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

VI. REFERENCES

[1]. OpenSSL Project, (1998). [Online]. Available: http://www.openssl.org/

[2]. P. Anderson and L. Zhang, “Fast and secure laptop backups with encrypted de-duplication,” in Proc. 24th Int. Conf. Large Installation Syst. Admin., 2010, pp. 29–40.

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

[4]. 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.

[5]. M. Bellare, C. Namprempre, and G. Neven, “Security proofs for identity-based identification and signature schemes,” J. Cryptol.,vol. 22, no. 1, pp. 1–61, 2009. [10] GNU Libmicrohttpd, (2012). [Online]. Available: http://www.gnu.org/ software/libmicrohttpd/

[6]. 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.