International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.3 (2017): Page.1574-1580
www.ijcsec.com

Secure Authorized Deduplication in Hybrid Cloud

Chintu P Chacko
PG Scholar
Department of Information Technology
TOC-H Institute of Science and Technology, Arakkunnam, Kerala, India
chintupchacko@gmail.com

Abstract : *Many techniques being used for eliminating duplicate copies of repeating data, among this techniques one of the important data compression technique is data deduplication. Deduplication is widely used in cloud storage to save bandwidth and to reduce the amount of storage space. To protect the confidentiality of data along with deduplication convergent encryption technique is used before outsourcing. Problem of authorized data duplication formally addressed by the first attempt of this paper for better protection. This proposed system is different from the traditional duplication systems. For raised knowledge security this paper also present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. My goal is to implement a paradigm of the planned licensed duplicate check theme and conduct a test bed experiment using this paradigm. In this paper I have analysed the existing hybrid cloud deduplication system.*

Keywords *- Text segmentation ,Type Detection, Typed terms, Knowledgebase, Concept Labelling*

1. Introduction

Cloud computing has recently appeared as a preferred business model for utility system. The conception of cloud is to supply computing resources as a utility or a service on demand to customers over the web. As cloud computing becomes prelavent, an increasing amount of data is being stored in the cloud and shared by users with certain privileges. One of the critical challenge of the cloud storage service is the management of ever increasing volume of data. To make data management scalable in cloud computing, deduplication [1] has been a well-known technique used and has attracted more attention recently. Data deduplication is a data compression technique for eliminating duplicate copies of repeating data in storage. It is used to improve storage utilization and can 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. Although data deduplication brings a lot of benefits, security and privacy concerns arise due to attacks.

Traditional encryption, while providing data confidentiality, is conflicting with data deduplication. Traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of users will lead to different cipher texts, causing deduplication impossible. Convergent encryption [2] has been proposed to enforce data confidentiality while making deduplication feasible. Convergent encryption encrypts/decrypts a data copy with a convergent key, which is obtained by calculating the cryptographic hash value of the content of the data copy. To avoid unauthorized access, a secure proof of ownership protocol [3] is also required to provide the proof that the user owns the same file when a duplicate is found. In the model stated in this paper aims at efficiently solving the problem of deduplication with differential privileges in cloud computing. A hybrid cloud architecture consisting of a public cloud and a private cloud is considered in this model. unlike existing deduplication system private cloud is involved as a proxy to allow data owner to securely perform duplicate check with differential privileges. a new duplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the SCSP presides in the public cloud the user is only allowed to perform the duplicate check for files marked

with corresponding privileges also present an advanced scheme to support strong security by encrypting the file with differential privilege case .

2. Related Work

This section discusses the lately implemented or tested out scenarios for secure authorized deduplication. This section focuses on the implementation or research status of convergent encryption, symmetric encryption and message lock encryption. DupLess: Server aided encryption for deduplicated storage [6] for cloud storage service provider like Mozy, Dropbox, and others perform deduplication to save space by only storing one copy of each file uploaded .Message lock encryption resolves the problem of clients encrypt their file however the saving are lock. Dupless is used to provide secure deduplicated storage as well as storage resisting bruteforce attacks. Clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol in dupless server. This allow clients to store encrypted data within an existing service and achieves strong confidentiality guarantees. This shows that encryption for deduplicated storage can reach desired performance and space savings near to that of using the storage service with plaintext data [6].

Proofs of Ownership in Remote Storage Systems [3] stores only the single copy of the duplicate data. Client-side deduplication tries to identify deduplication chance so far at the client and store the bandwidth of uploading copies of current files to the server[3].To overcome the attacks the authors proposes the Proof of ownership which allows a client efficiently prove to a server that the client keep a file, rather than just some short information about it present solutions based on Merkle trees and specific encodings, and analyse their security.[5]

Twin clouds: An architecture for secure cloud computing [4] proposed an architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. The user communicates with a trusted cloud, which encrypts and verifies the data stored and operations occurred in the untrusted cloud .It separates the computations such that the trusted cloud is used for security critical operations, whereas queries to the outsourced data are processed in parallel by the fast cloud on encrypted data [4].

Most important issue in the cloud storage is utilization of the storage capacity. In Private Data Deduplication Protocols in Cloud Storage [7], there are two categories of data deduplication strategy and extend the fault-tolerant digital signature scheme on examining redundancy of blocks to achieve the data deduplication. The proposed scheme in this paper not only reduces the cloud storage capacity, but also boost the speed of data deduplication. Furthermore, the signature is computed for every uploaded file for verifying the integrity of files[7].

3. Preliminaries

In this section we go through the notations used in this paper and analyze the secure primitives used in this secure deduplication.

Convergent encryption. Convergent encryption provides secure confidentiality in data deduplication. Data owners can derive a convergent key from the original data copy and encrypts the data copy with the convergent key. And also the user can derive a tag for the data copy and this will be used to detect duplicates. We assume that the tag correctness property holds [8] here, i.e, if two data copies are same, then their tags are same. To identify duplicates the user will first send the tag to the server side to check whether identical copy has been already stored or not. Confidentiality check and tags are independently derived. Tag cannot be used to deduce the convergent key and compromise data confidentiality. The

encrypted data copy and its tag will be stored on the server side. The convergent encryption scheme can be defined with four primitive functions:

- KeyGenCE(M) → K is the key generation algorithm that maps a data copy M to a convergent key K;
- EncCE(K,M) → C is the symmetric encryption algorithm which takes both the convergent key K and the data copy M as inputs and then outputs a ciphertext C;
- DecCE(K,C) → M is the decryption algorithm that takes both ciphertext C and the convergent key K as inputs and then outputs the original data copy M; and
- TagGen(M)→ T(M) is the tag generation algorithm that maps the original data copy M and outputs a tag T(M).

Symmetric encryption. It uses a common secret key k to encrypt and decrypt information. It consist of three primitive functions:

- KeyGenSE(1^{λ}) → k is the key generation algorithm that generates k using security parameter 1^{λ} ;
- EncSE(k,M) → C is the symmetric encryption algorithm that takes the secret k and message M and then outputs the Ciphertext C; and
- DecSE(k,C) → M is the symmetric decryption algorithm that takes the secret k and ciphertext C and then outputs the original message M.

Proof of Ownership. The notion of proof of ownership (PoW) [10] enables users to prove their ownership of data copies to the storage server. PoW is implemented as an interactive algorithm run by a user and a verifier (i.e., storage server).

Identification Protocol. An identification protocol can be described in two phases: Proof and Verify. In Proof, the user can demonstrate his identity to a verifier by performing some identification proof related to his identity[9]. In Verify, the verifier performs verification with input of public information.

4. Architecture Overview

4.1 Hybrid Architecture for Secure Authorized deduplication

By using the duplication technique, to store the data, will use S-CSP are consisted as group of affiliated client at high level. The main focus is enterprise all the network. To set the data back up and disaster recovery applications for decrease the storage space, we frequently go for de-duplication. Such systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions.

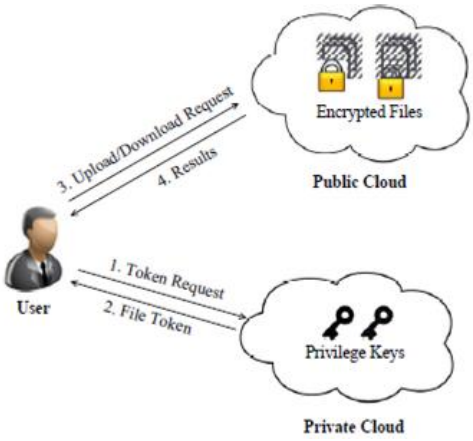

Figure.1 Architecture for Authorized Deduplication

There are three entities define in this system as shown in Fig1 [10], they are,

- Users
- Private cloud

- S-CSP in public cloud

De-duplication performed by S-CSP is by checking if the contents of two files are the same and stores only one of them. The access right of a file is defined based on the set of privileges. Each file is related with some *file tokens*, which indicate the tag with specified privileges. A user calculates and sends *duplicate-check tokens* to the public cloud for authorized duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well; else, the user further perform the block-level duplicate check and recognize the unique blocks to be uploaded. Each data copy is related with a token for the duplicate check.

S-CSP. This entity provides a data storage service in public cloud. It provides the data outsourcing service and stores data on behalf of the users. S-CSP eliminates the storage of redundant data via de-duplication and keeps only unique data to reduce the storage cost. In this paper, it is assumed that S-CSP is always online and has abundant storage capacity and computation power.

Data Users*:* This entity is used to outsource data storage to the S-CSP and access the data later. In a de-duplication system, the user only uploads unique data; it does not upload any duplicate data to store the upload bandwidth, which may be owned by the same user or different users. In the authorized de-duplication system, each user is provided with a set of privileges in the setup of the system. Each file is secured with the convergent encryption key and privilege keys to perceive the authorized de-duplication with differential privileges.

Private Cloud. Since the computing resources at data user side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user 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.

4.2 Adversary Model

Here we assume that both the public and private clouds are "honest-but-curious", but try to find out as much secret information as possible based on their possessions either within or out of the limits of privileges users would try to access data. In this paper we presume that, all the files are sensitive and needed to be fully protected against both public and private cloud. It is assume that two kinds of adversaries are considered. External adversaries who aim to extract secret information as much as possible from both public cloud and private cloud. Internal adversaries which aim to obtain more data on the file from the public cloud and duplicate-check token data from the private cloud outside of their scopes. These adversaries may include S-CSP, private cloud servers and authorized users.

4.3 Design Description

The detailed architecture of the design is showed in figure 2 [11]. There are four different modules present in the architecture. Data Owner Module, Encryption and Decryption Module, Remote User Module, Cloud Server Module. To upload or download a file user login details are required and the details of modules are mentioned below.

Data owner module :

- Data Owner login validations.
- Upload Files.

- Manipulates Encrypted files.
- Differential Authorization.

Encryption and decryption module:

- Generate signs.
- Encrypts and uploads files.
- Decrypts and downloads files.
- Data confidentiality.

Remote user module:

- Accessing Files.
- Remote User login validations.

Cloud server module:

- Authorized Duplicate Check.
- Accessing files.

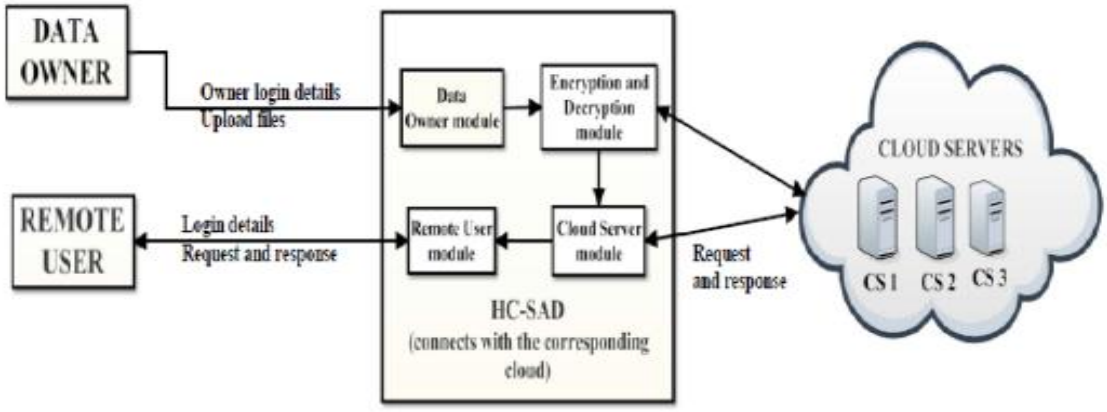

Figure 2. System Architecture Design

4.4 New Duplication System

This system address the problem of privacy preserving de-duplication in cloud computing and propose a new deduplication system supporting for, the

Differential Authorization: Authorized user is able to get individual token for his file to perform duplicate check based on privileges. Based on this assumption only the authorized user can generate token for duplicate check out of his privileges.

Authorized duplicate check: Authorized user is able to use his individual private keys to generate query for certain file and the privileges he owned with the help of private cloud, while the public cloud performs duplicate check and tells the user if there is any duplicate. The security requirements in this paper lie in two folds, including the security of file token and security of data files. For the security of file token, two aspects are defined as un-forge ability and in-distinguish ability of file token. The details are given below.

Unforgeability of file token/duplicate-check token: Unauthorized users without privileges should be prevented from generating the file tokens for duplicate check of any file stored at the S-CSP. The users are not allowed to conspire with the public cloud server to break the unforgeability of file tokens. In this system, the S-CSP is honest but curious and will perform the duplicate check upon getting the duplicate

request from users. The duplicate check token should be issued from the private cloud server in our scheme.

Indistinguishability of file token/duplicate-check token. It demands that any user without querying the private cloud server for some file token, he cannot get any useful information/data from the token, which includes the file information/data or the privilege information.

Data Confidentiality. The aim of the adversary is to retrieve and recover the files that do not belong to them. In this system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level confidentiality is defined and achieved.

5. Conclusion

The notion of authorized data deduplication was proposed in this paper to protect the data security by including differential privileges of users in the duplicate check. Also presented several new deduplication interpretations supporting authorized duplicate check in hybrid cloud architecture, in which duplicate check tokens of files are generated by the private cloud server with private keys. Analyzing the security and overhead are the future works to be done in this model.

Acknowledgements

I gratefully thank the entire faculty of the Department of Information Technology, ToC H Institute of Science & Technology for their valuable support and encouragement in working on with this project work.

REFERENCES

[1] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.

[2] 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.

[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] 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.

[5] 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 Security2012.

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

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

[8] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296– 312, 2013.

[9] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*, 22(1):1–61, 2009.

[10] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou" A Hybrid Cloud Approach for Secure Authorized De-duplication" in vol: pp no-99, IEEE, 2014

[11] Shaik Rahamathunnisa Begam, Bachina Varsha, Nittala Swapna Suhasini. "Hybrid Cloud Approach for Efficient Secure Authorized Deduplication" in vol: 04 pages 1987-1992, IJITECH, 2016

ABOUT AUTHOR

Chintu P Chacko received B.Tech degree in Information Technology from Ilahia College of Engineering & Technology,MG University,Kerala.Currently pursuing her Masters Degree in Network Computing from Toc H Institute of Science & Technology ,APJ Abdul Kalam Technological University ,Kerala, India.