IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. V. (Jul.-Aug. 2017), PP 25-29
www.iosrjournals.org

Convergent Encryption Using Deduplication on Hybrid Cloud

[*]Pusukuri Lakshmi Prasanna[1], Lella Kranthi Kumar[2]
[1]*M.Tech.IV Sem. (CSE) Lakireddy Balireddy College of Engineering, Mylavaram*
[2]*Assistant Professor, CSE Department, Lakireddy Balireddy College of Engineering, Mylavaram*
Corresponding Author: Pusukuri Lakshmi prasanna

Abstract: *This paper represents that, an important data compression technique is data deduplication. In this technique repeating data of duplicate copies are eliminated. To support deduplication we protect confidentiality of sensitive data. For encrypt the data we have been proposed convergent encryption technique. For better protection of data security, identify the problem of authorized data deduplication in first attempt. There have different traditional deduplication systems, from those systems different privileges of users are further considered in duplicate check besides the data itself. In hybrid cloud architecture, we present several new deduplication constructions for supporting authorized duplicate check. Based on the definitions we proposed security model for the purpose of security analysis. A proof of concept is to implement a prototype of authorized duplicate check and using our prototype we conduct testbed experiments.*

Indexterms*: Deduplication, confidentiality, hybrid cloud, authorized duplicate check.*

Date of Submission: 07-07-2017 Date of acceptance: 05-08-2017

I. Introduction

Cloud computing provides unlimited virtualized resources to users. Now a day's cloud service providers offers high amount of available storage and huge parallel computing resources at low costs. One of the major critical challenges of cloud storage services is the management of the ever increasing volume of data.

Dedulication is a well-known technique for make data management scalable in cloud computing. A specialized data compression technique is data deduplication. It is used in storage for eliminating duplicate copies of repeating data. For improving the storage utilization and network data transfer to reduce the number of bytes, we use data compression technique. Alternatively keeping multiple data copies with the same content, redundant data is to eliminate deduplication. It is to keep only one physical copy and referring other redundant data to that copy. Deduplication has mainly file level and block level. At this two levels dedulication can take either file level or block level. The file level dedplication is to eliminate the duplicate copies of the same file. The block level deduplication is to eliminate the duplicate blocks of the data that occur in non-identical files.

Data deduplication have a lot of benefits, security and privacy concerns arise as sensitive data of user's are susceptible to both inside and outside attacks. In traditional encryption, to provide confidentiality it is incompatible with data deduplication. Particularly traditional encryption requires different users to encrypt their data with their own keys. Making of deduplication is impossible for identical data copies of different users will lead to different cipher texts. Then we proposed convergent encryption technique, it is to enforce data confidentiality while making deduplication feasible. This technique is to encrypts or decrypts a copy of data with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After generation of key and the data encryption, users retain the key and send the cipher text to the cloud. As the operation of encryption is deterministic and is derived from the data content, identical copies of data will generate the same convergent key and hence the same cipher text. To prevent the unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user actually 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. The encrypted file can be downloaded by the user with the pointer from the server, which can only relating data owners are decrypted with their convergent keys. Thus, convergent encryption allows the cloud to perform deduplication on the cipher texts and the proof of ownership (pow) prevents the unauthorized user to access the file.

II. Problem Statement

The existing deduplication system cannot support differential authorization duplicate check, in authorized duplication system; during the system initialization each user is issued a set of privileges. Every file uploaded to the cloud is also bounded by a set of authorities to specify which kind of users is allowed to perform the duplicate check and access the files. Before submit the file of duplicate check request, the user requires taking this file and as inputs of his own authorities. If there is a copy of this file and matched authorities stored in

cloud, then only the user is able to identify a duplicate for this file. To save the cost and efficiently management, the data will be transferred to the storage server provider (s-csp) in public cloud with particular authorizes and the deduplication technique will be applied to store the same file but only one copy. Based on convergent encryption the traditional deduplication systems are providing the confidentiality to some extent, with the duplicate check. In other words, based on the convergent encryption technique there is no differential privileges have been considered in the deduplication.

In this paper, efficiently solving the problem of deduplication with differential authorizes in cloud computing we use hybrid cloud architecture. The architecture consist a public cloud and a private cloud. The private cloud is involved as a alternate to allow data owner or users securely perform duplicate check with differential authorizes. The data owners only outsource their storage of data by applying 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. Security analysis demonstrates that our system, then we specifies a proposed security model. Then we implement a prototype of the proposed authorized duplicate check and testbed experiments are conducted to evaluate the overhead of the prototype.

III. Secure Deduplication On Hybrid Cloud

At this level consisting of a group of affiliated clients who will use the s-csp and store data with deduplication technique. For the data backup and accident recovery applications are used in the deduplication can be frequently used. Such those systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions. In our system there have three entities are defined. That is users, private cloud and s-csp performs deduplication by checking if the two files of content are the same and only one of them is stored.

A file of access right is defined based on a set of authorizes or privileges. Each privilege is represented in the form of a short message. This message is called “token”. Some file tokens are associated with each file, which denote the tag with specified privileges. To the public cloud a user can computes and sends duplicate check tokens for authorized duplicate check.

In private cloud server the users can access to this server, a third party which will support in performing deduplicable encryption by generating file tokens for the requesting users. In this paper, file-level deduplication only considered for simplicity. We assign a copy of data to be a whole file and file level deduplication which is the elimination of the storage of any redundant files. Compared to file-level deduplication, the block-level deduplication can be easily deduced. Particularly, to a file upload, 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 copy of data can be associated with a token for the duplicate check.

- S-csp: This is an entity in public cloud. It provides a data storage service. Data outsourcing service and stores data by providing s-csp on behalf of the users. It is to reduce the cost of storage and s-csp eliminates the storage of redundant data via deduplication and also it keeps only unique data. In this paper, we assume that the s-csp is always online and has sufficient data storage capacity and have computation power.
- Data user: A user is an entity that wants to outsource data storage to the s-csp and access the data later. In storage system support of deduplication, the user can only upload unique data but the user does not upload any duplicate data to save the band width for upload, which may be owned by the same user or different users. Convergent encryption key is to be protected each and every file. In the file authorized duplication system, each user is issued a set of authorizes in the setup of the system.
- Private cloud: In cloud computing compared with the architecture of traditional deduplication, private cloud is a new entity introduced for user’s secure usage of cloud service. Especially for the computing resources at data user or owner side are blocked and the public cloud is not fully trusted in practice, private cloud is able to provide data user or owner with an execution status and foundation working as an interface between user and public cloud.

In cloud computing this is a different architecture for data deduplication. It consists of twin clouds they are public cloud and private cloud. Recently this hybrid cloud settings has attracted more and more consideration. For example, Amazon s3 use public cloud server for data archiving but continue to maintain in-house storage for operational customer data.

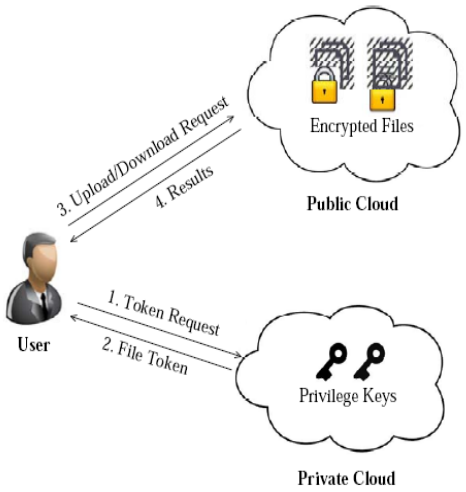

Fig: Architecture for Authorized Deduplication

IV. Proposed System

In our proposed system, another advanced deduplication system is authorized duplicate check supported. In this new deduplication system, for solving the problem we introduced hybrid cloud architecture. The private keys authorizes will not be expressed directly to the users, private cloud server is to be kept and handled. In this proposed system the users cannot share these private keys of authorizes. To get a file token, the user wants to send a request to the private cloud server. For some files we perform the duplicate check, the user wants to get the token of file from the private cloud server. This private cloud server will also be check the identity of users before submitting the relating files token to the user. The authorized duplicate check for this file can be performed by the user with the public cloud before uploading the file.

In our proposed deduplication system we provide system setup, upload file, retrieve file.

System setup: In the system setup we are using a symmetric key for each privilege user and this will be sent to the private cloud. The identification protocol is defined, where proof and verifies are the proof and verification algorithm respectively. Further imagine every user have a secrete key to perform the identification with servers. Privilege set is pow protocol for the file ownership proof. A table was maintained by the private cloud server, it will store every user's public information and related privilege set. Upload file: Suppose a data owner wants to upload a file, then the data owner needs to interact with the private cloud before performing duplicate check with the s-csp. Mainly the data owner performs an identification to prove its identity with private cloud server will find the corresponding privileges of the user from its stored table list.

If duplicate file is found, the user needs to run pow protocol pow with the s-csp to prove the file ownership. The user sends the privilege set for the file as well as the proof to the private cloud server. Then request was received, the private cloud server first verifies the proof from the s-csp. If duplicate file is not found, a proof from the s-csp will be returned. The user sends the proof to the private cloud server, then request was received, the private cloud server first verifies the proof from the s-csp. Then the private cloud server computes and sends with the signature to the S-csp.

Retrieve file: For suppose a user wants to download a file, first it will sends a request and the file name to the s-csp. Then the file request and name was received the s-csp will check whether the user is suitable to download the file or not. If it fails, the s-csp sends back and signals to the user to indicate the download failure. Otherwise the s-csp returns the related cipher text.

Flow chart:

The private cloud was managed to the private keys for the privileges, who answer the file token requests from the users and the interface offered by private cloud. And it allows the user to submit files and queries to be securely stored and computed. To solve the problems of unauthorized deduplication of file we introduced hybrid cloud architecture. The users will not permits directly to the private keys for privileges; private cloud will be kept and managed the private keys. The user needs to get the file token from the private cloud server will find the related authorizes of the user from its stored table list and send to the user can upload the files. In the same way user can download the file from cloud storage.

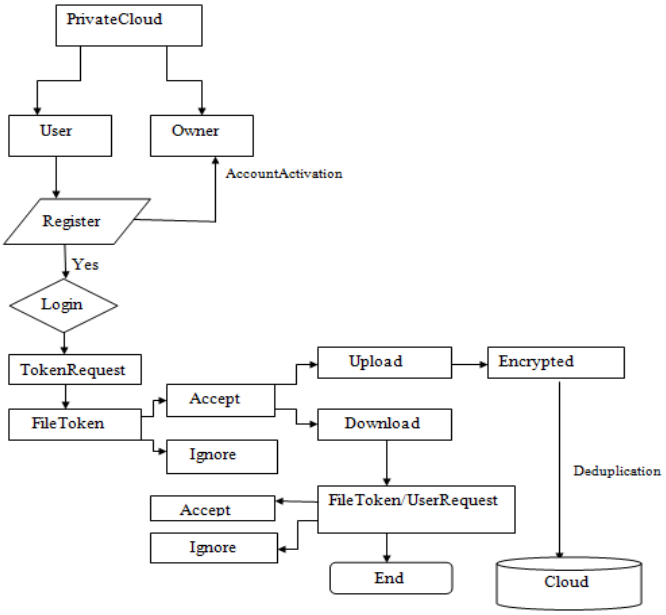

Fig: Flow chart of proposed deuplication system

V. Related Work

Secure deduplication: Now a day's secure data deduplication has more attracted and much attention from research side. Yuan and Yu were proposed a deduplication system in the cloud storage to reduce size of the storage for integrity check. For improve the security of deduplication and protect the confidentiality of data by transforming the predictable message into unpredictable message.

In the secure deduplication system , we have a third party called key server, it was introduced to generate the file tag for duplicate check. Stanek was introduced a different encryption scheme that provides differential security for popular data and unpopular data. Traditional conventional encryption technique was performed for popular data but that are not sensitive particularly. For the unpopular data another two layered encryption scheme was proposed. This scheme was stronger security while supporting deduplication. Like that way, they accomplished the better tradeoff

between the security and efficiency of the outsourced data. Key-management issue was introduced by Li et al, his issue was in block-level deduplication by assigning these keys across multiple servers after the files encrypted.

Convergent encryption: Convergent encryption ensures data privacy in deduplication. Message-locking encryption was introduced by Bellare, and explored its application in space-efficient secure outsourced storage. The problem of the secure convergent encryption for efficient encryption was identified by Xu et al. This problem is resolved without considering issues of key-management and block-level deduplication.

Proof of ownership: Proof of ownership for deduplication was proposed by Halevi, in that system a client can effectively prove to the cloud storage server that he owns a file without uploading the file itself. Based on the Merkle-Hash Tree there have several pow constructions. It was proposed to allow the duplication of client-side, which include the bounded leakage setting, another effective pow scheme was proposed by Pietro and Sorniotti. This scheme was select the projection of a file onto some randomly selected bit-positions as the file proof.

Architecture of twin clouds: The architecture of twin cloud is public cloud and private cloud. Recently architecture was proposed by Bugiel, it consists of twin clouds for security of outsourced data to an untrusted produced cloud. Hybrid cloud technique was proposed by Zhang, it was to support privacy-aware data intensive computing. In public cloud also we identify the authorized deduplication problem.

VI. Implementation

For the proposed authorized deduplication system we implement a prototype. In this implementation model there have mainly three entities are present. They are client, private server and storage server. A client program is used to model the data users to carry out the process of uploading the files. A private server is used to model the private cloud. The server was manages the private keys and file token computation also handles. A storage server program is used to model the s-csp which stores and deduplicates the files.

By using the openSSL library we implement cryptographic operations of hashing and encryption. And also implements the communication between the entities based on hyper text transfer protocol (HTTP).

Client: In the client program implementation, it will provide some function calls to support generation of token and deduplication along with the process of file upload. In the client program SHA-1 hash function was implemented.

FileTag- For File tag SHA-1 hash was computes in the file.

TokenRequest- For the file token generation TokenRequest was requests the private server with the file tag and userID.

DuplicateCheckRequest- For the duplicate check of the file, it was request to the storage server and then it is send the file token received from private server.

ShareTokenRequest- To generate the share file token, it was requested to the private server.

FileEncryption- With the convergent Encryption by using 256-bit AES algorithm the file will be encrypts.

FileUploadRequest- It uploads the file data to be stored server if the is unique.

Private server: The private server implementation is to be includes related request handlers for the token generation and also maintains a key storage with hash map.

TokenGeneration- By using HMAC-SHA-1 algorithm associated to user's privilege keys and generates the token.

ShareTokenGeneration- By using HMAC-SHA-1 algorithm, it will generates the share token with the related privilege keys.

Storage server: The storage server implementation is to provide the deduplication and data storage.

DuplicateCheck- It searches the file to token map for duplication.

FileStorage- The file can be stored on disk and updates the mapping.

VII. Conclusion And Future Scope

In this paper, for protection of data security authorized data deduplication was proposed. The data security is includes differential authorizes of users in the duplicate check. There are several new deduplication constructions also presenting for support the authorized duplicate check in hybrid cloud architecture. Security analysis determines that our schemes are secure. The security analysis is to be secures the inside attacks and outside attacks. By using the proof of concept, we implement a prototype of our proposed authorized deduplication system and also conduct the testbed experiments. In the future the security problems will be prevent and may arise in the practical deployment of the present model. By the deduplication of data, it saves the memory and provides the sufficient memory to us. And also it protects the confidentiality of the important data.

References

[1] OpenSSL project, http://www.openssl.org/
[2] P. Anderson and L. Zhang. Fast and secure laptop backup with encrypted deduplication. In proc. Of USENIX LISA, 2010.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server-aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296– 312, 2013.
[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*, 22(1):1–61, 2009.
[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concur- rent attacks. In *CRYPTO*, pages 162–177, 2002.
[7] 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.
[8] 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.