The International Journal Of Engineering And Science (IJES)
|| Volume || 3 || Issue || 11|| Pages || 34-39 || 2014 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Secured Authorized Deduplication Based Hybrid Cloud

[1,] Rajashree Shivshankar Walunj, [2,] Deepali Anil Lande,
[3,] Nilam Shrikrushna Pansare
[1,2,3,]*Computer Engg. Jaihind College Of Engg,Kuran*

----------ABSTRACT----------

Data deduplication is an important technique for eliminating redundant data.Insted of taking no. of same files, it store only single copy of file.In most organizations,storage system contain many pieces of duplicate data. . For example, the same file may be saved in several different places by different users. Deduplication eliminates these extra copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. It is data compression technique for improve the bandwidth efficiency and storage utilization.Data deduplication most widely used in cloud computing.It make data management scalable and storage problem in cloud computing.Data deduplication protect the confidentiality of sensitive data.data deduplication work with convergent encryption technique to encrypt the data before uploading. . Companies frequently use deduplication in backup and disaster recovery applications.In this paper we attempt authorized deduplication check, combine with convergent encryption for providing security to sensitive data using hybrid cloud computing.

KEYWORDS : *Deduplication, authorized duplicate check, confidentiality, hybrid cloud,covergent encryption.*

Date of Submission: 30 October 2014 Date of Accepted: 20 November 2014

I. INTRODUCTION

Cloud computing technique which is most widely used todays.In that, computing is done over the large communication network like Internet.It is an important solution for business storage in low cost.Cloud computing provide vast storage in all sector like government, enterprise,also for storing our personal data on cloud. Without background implementation details,platform user can access and share different resources on cloud.The most important problem in cloud computing is that large amount of storage space and security issues. One critical challenge of cloud storage to management of ever-increasing volume of data.To improve scalability,storage problem data deduplication is most important technique and has attracted more attension recently.It is an important technique for data compression ,It simply avoid the duplicate copies of data and store single copy of data.Data deduplication take place in either block level or file level.In file level approach duplicate files are eliminate,and in block level approach duplicate blocks of data that occur in non-identical files.Deduplication reduce the storage needs by upto 90-95% for backup application,68% in standard file system.Important issues in data deduplication that security and privacy to protect the data from insider or outsider attack.For data confidentiality,encryption is used by different user for encrypt there files or data ,using a secrete key user perform encryption and decryption operation.For uploading file to cloud user first generate convergent key,encryption of file then load file to the cloud.To prevent unauthorize access proof of ownership protocol is used to provide proof that the user indeed owns the same file when deduplication found.After the proof,server provide a pointer to subsequent user for accessing same file without needing to upload same file.When user want to download file he simply download encrypted file from cloud and decrypt this file using convergent key.

PRELIMINARIES : In this section ,we first define notations used in this paper,review secure primitives .The notation used in this paper are listed in Table 1.

Convergent encryption. Convergent encryption is used to encrypt and decrypt file.User can derives the convergent key from each original data copy,then using that key encrpty data file.Also user derives tag for data copy to check duplicate data.If tag are same then both files are same. Both convergent key and tag are independently derives. Convergent encryption, also known as content hash keying, is used to produces identical ciphertext from identical plaintext files. The simplest implementation of convergent encryption can be defined

as: Alice derives the encryption key from her file F such that K = H(F), where H is a cryptographic hash function.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 ciphertext C;
[3] DecCE (K, C) ->M is the decryption algorithm that takes both the ciphertext C and the convergent key K as inputs and then outputs the original data copy M ; and
[4] TagGen(M) -> T (M) is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

Proof of ownership. proof of ownership(PoW) is a protocol enables users to prove their ownership of data copies to the storage server.PoW is implemented as an interactive algorithm run by user and storage server act as prover and verifier. The verifier derives a short value $\phi(M)$ from a data copy M . To prove the ownership of the data copy M , the prover needs to send ϕ to the verifier such that $\phi = \phi(M)$. The formal security definition for PoW roughly follows the threat model in a content distribution network, where an attacker does not know the entire file, but has accomplices who have the file. proof of-ownership is specified by a summary function S(.) (which could be randomized and takes the input file F and a security parameter),and an interactive two-party protocol ||(P V). To solve the problem of using a small hash value as a proxy for the entire file, we want to design a solution where a client proves to the server that it indeed has the file. We call a proof mechanism that prevents such leakage amplification a proof of ownership (PoW).

Public hash function. To support cross-user deduplication, all users must use the same procedure for identifying duplicate files. Hence this procedure must be public i.e. each user implement it, which means that a determined attacker can learn it.

System Model and Adversary Model : In the client deduplication scenario, the S-CSP server keeps a single copy of the original file, regardless of the number of user that request to store the file. All users that possess the original file only use link to the single copy of the original file stored on the S-CSP server. Specifically, a client user first sends the hash value of the original file to the server, and then the server checks whether the hash value already exists in its database. If the hash value is the same as an existing hash value stored on the server, the server will challenge the user to ask for the proof of possession of the original file. After a successful challenge, the client does not have to upload the file again to the server. At the same time, the server marks the client as an owner of the original file. From then on, there is no difference between the client and the users who uploaded the original file. Thus, the deduplication process saves the communication bandwidth as well as the storage space.

.

DESIGN GOALS : In this paper, we address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for:

- **Differential Authorization**. Each authorized user is able to access its individual token of his file to perform duplicate check based on authority.Under this assumption, any user cannot generate a token for duplicate check out of his access or without the aid from the private cloud server.

- **Authorized Duplicate Check.** Authorized user is able to access his/her own token from private cloud,while the public cloud performs duplicate check directly and tells the user if there is any duplicate. The security requirements considered 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 unforgeability and indistinguishability of file token. The details are given below.

- **Unforgeability of file token/duplicate-check token.** User make registration in private cloud for generating file token.Using respective file token he/she upload or download files on public cloud. The users are not allowed to collude with the public cloud server to break the unforgeability of file tokens. In our system, the S-CSP is honest but curious and will honestly perform the duplicate check upon receiving the duplicate request from users. The duplicate check token of users should be issued from the private cloud server in our scheme.

- **Indistinguishability of file token/duplicate-check token.** It requires that any user without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information and key information.

- **Data Confidentiality**. Unauthorized users without appropriate token , including the S-CSP and the private cloud server, should be prevented from access to the underlying plaintext stored at S-CSP. In another word, the goal of the adversary is to retrieve and recover the files that do not belong to them. In our system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level confidentiality is defined and achieved.

Table 1. Notations Used in This Paper

Acronym	Description
S-CSP	Storage cloud service provider
POW	Proof of Ownership
K_F	Convergent encryption key for file F
H	Hash function

II. PROPOSED SYSTEM

In our system we implement a project that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general by if we used the public cloud we can't provide the security to our private data and hence our private data will be loss.So that we have to provide the security to our data for that we make a use of private cloud also. When we use a private clouds the greater security can be provided.In this system we also provides the data deduplication . which is used to avoid the duplicate copies of data.User can upload and download the files from public cloud but private cloudprovides the security for that data.that means only the authorized person can upload and download the files from the

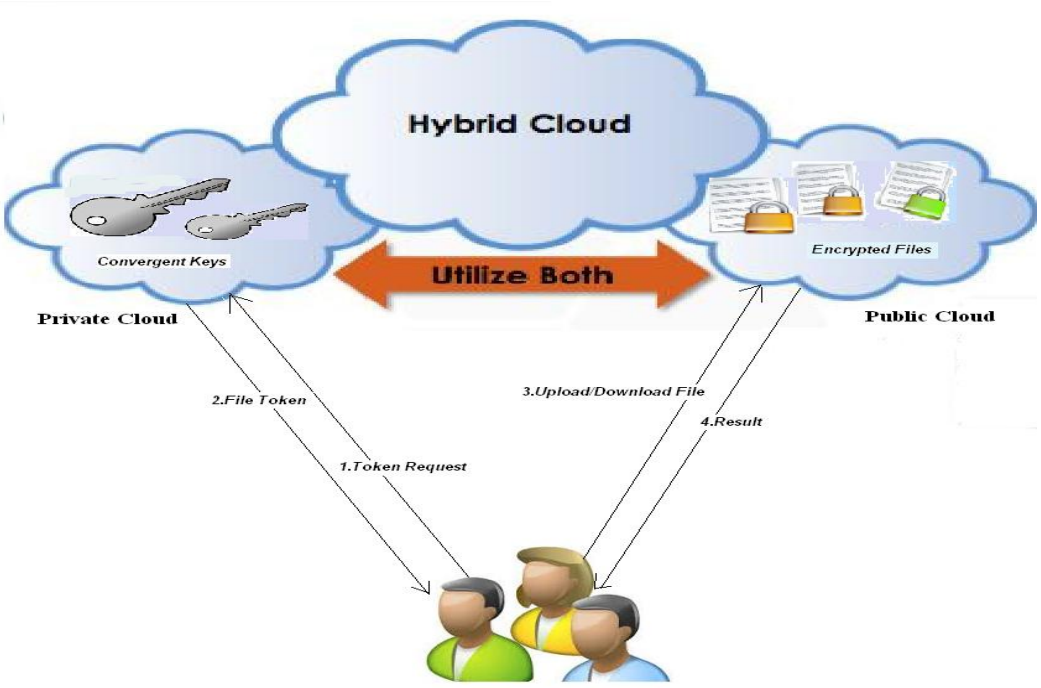

Fig 1:Architecture of Authorized Deduplication

public cloud.for that user generates the key and storeed that key onto the private cloud. at the time of downloading user request to the private cloud for key and then access that
Particular file.

System Model :Now we see the architecture of our system. in our architecture there are three modules .
[1] user
[2] public cloud
[3] private cloud.etc
First if the user want to upload the files on the public cloud then user first encrypt that file with the convergent key and then sends it to the public cloud at the same time user also generates the key for that file and sends that key to the private cloud for the purpose of security.In the public cloud we use one algorithm for deduplication. Which is used to avoid the duplicate copies of files which is entered in the public cloud. Hence it also minimizes the bandwidth. that means we requires the less storage space for storing the files on the public cloud.In the public cloud any person that means the unauthorized person can also access or store the data so we can conclude that in the public cloud the security is not provided. In general for providing more security user can use the private cloud instead of using the public cloud.User generates the key at the time of uploading file and store it to the private cloud.When user wants to downloads the file that he/she upload,.he/she sends the request to the public cloud. Public cloud provides the list of files that are uploads the many user of the public cloud because there is no security is provided in the public cloud.When user selects one of the file from the list of files then private cloud sends a message like enter the key!. User has to enter the key that he generated for that file. When user enter the key the private cloud checks the key for that file and if the key is correct that means user is valid then private cloud give access to that user to download that file successufully. then user downloads the file from the public cloud and decrypt that file by using the same convergent key which is used at the time of encrypt that file.in this way user can make a use of the architecture.

III. ROLES OF ENTITIES

S-CSP: The purpose of this entity to work as a data storage service in public cloud.On the half of the user S-CSP store the data.The S-CSP eliminate the duplicate data using deduplication and keep the unique data as it is.S-SCP entity is used to reduce the storage cost.S-CSP han abundant storage capacity and computational power.When user send respective token for accessing his file from public cloud S-CSP matches this token with internally if it matched then an then only he send the file or ciphertext C_f with token, otherwise he send abort signal to user.After receiving file user use convergent key K_F to decrypt the file.

Data User: A user is an entity that want to access the data or files from S-SCP.User generate the key and store that key in private cloud.In storage system supporting deduplication,The user only upload unique data but do not upload any duplicate data to save the upload bandwidth,which may be owned by the same user or different users.Each file is protected by convergent encryption key and can access by only athorized person.In our system user must need to register in private cloud for storing token with respective file which are store on public cloud.When he want to access that file he access respective token from private cloud and then access his files from public cloud.token consist of file content F and convergent key $K_{F.}$

Private Cloud: In general for providing more security user can use the private cloud instead of public cloud.User store the generated key in private cloud.At the time of downloading system ask the key to download the file.User can not store the secrete key internally.for providing proper protection to key we use private cloud.Private cloud only store the convergent key with respective file.When user want to access the key he first check authority of user then an then provide key.

Public Cloud: Public cloud entity is used for the storage purpose.User upload the files in public cloud.Public cloud is similar as S-CSP.When the user want to download the files from public cloud,it will be ask the key which is generated or stored in private cloud.When the users key is match with files key at that time user can download the file,without key user can not access the file.Only authorized user can access the file.In public cloud all files are stored in encrypted format.If any chance unauthorized person hack our file,but without the secrete or convergent key he doesn't access original file.On public cloud there are lots of files are store each user access its respective file if its token matches with S-CSP server token.

Operations performed on Hybrid Cloud

- **File Uploading :** When user want to upload the file to the public cloud then user first encrypt the file which is to be upload by make a use of the symmetric key,and send it to the Public cloud. At the same time user generates the key for that file and sends it to the private cloud. in this way user can upload the file in to the public cloud.

- **File Downloading:** When user wants to download the file that he/she has upload on the public cloud.he/she make a request to the public cloud. then public cloud provide a list of files that many users are upload on it.Among that user select one of the file form the list of files and enter the download option.at that time private cloud sends a message that enter the key for the file generasted by the user.then user enters the key for the file that he/she is generated.then private cloud checks the key for that

file and if the key is correct that means the user is valid.only then and then the user can download the file from the public cloud otherwise user can't download the file. When user download the file from the public cloud it is in the encrypted format then user decrypt that file by using the same symmetric key.

IV. IMPLEMENTATION

We implement system with data deduplication,in which we model three entities as separate programs. A Client program is used to model the data users to carry out the file upload/download process.
A Private Server program is used to model the private cloud which manages the private keys and handles the file token computation. A Storage Server program is used to model the S-CSP which stores and deduplicates files. Followings are **function calls** used in system:

- FileTag(File) - It computes SHA-1 hash of the File as File Tag;
- DupCheckReq(Token) - It requests the Storage Server for Duplicate Check of the file.
- FileEncrypt(File) - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode, where the convergent key is from SHA-256 Hashing of the file;
- FileUploadReq(FileID, File, Token) – It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.
- FileStore(FileID, File, Token) - It stores the File on Disk and updates the Mapping.

V. CONCLUSION

In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format,and also in private cloud our key is store with respective file.There is no need to user remember the key.So without key anyone can not access our file or data from public cloud.

REFERENCES

[1] Open SSL Project. http://www.openssl.org/.
[2] P. Anderson and L.Zhang. Fast and secure laptop backups with encrypted de-duplication.In Proc.of USENIX LISA, 2010.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided 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. 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 concurrentattacks. In CRYPTO, pages 162–177, 2002.
[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twinclouds: An architecture for secure cloud computing. In *Workshopon 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 distributedfile system. In *ICDCS*, pages 617–624, 2002.
[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15thNIST-NCSC National Computer Security Conf.*, 1992.
[10] GNULibmicrohttpd.http://www.gnu.org/software/libmicrohttpd/.

R.S.Walunj completd diploma in computer engg(2012). Currently she is student of Department of Computer Engineering at Jaihind college of engg,kuran in last year.She is interested to search latest information related to security purpose.

N.S.Pansare completd diploma in computer engg(2012). Currently she is student of Department of Computer Engineering at Jaihind college of engg,kuran in last year.She is interested to search latest information related
to cloud computing.

D.A.Lande completed diploma in computer engg(2012). Currently she is student of Department of Computer Engineering at Jaihind college of engg,kuran in last year.She is interested to search latest information related
to hybrid cloud computing.