International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 10, October 2017)

Usage of a Hybrid Cloud Approach for Secure Authorized Deduplication

Vinaysagar Anchuri
Assistant Professor, Department of Computer Science Engineering,
Guru Nanak Institute of Technology, Ibrahimpatnam, Ranga Reddy, Telangana, India
anchurivinay@gmail.com

***Abstract*: - Information deduplication is one of essential information pressure procedures for disposing of copy duplicates of rehashing information, and has been generally utilized as a part of distributed storage to diminish the measure of storage room and spare data transfer capacity. To ensure the privacy been proposed to scramble the information before outsourcing. To better ensure information security, this paper makes the main endeavor to formally address the issue of approved information deduplication. Not quite the same as customary deduplication frameworks, the differential benefits of clients are additionally considered in copy check other than the information itself. We likewise exhibit a few new deduplication developments supporting approved copy check in cross breed cloud design. Security examination exhibits that our plan is secure as far as the definitions indicated in the proposed security demonstrate. As a proof of idea, we actualize a model of our proposed approved copy check plan and direct testbed tests utilizing our model. We demonstrate that our proposed approved copy check plot brings about insignificant overhead contrasted with typical operations.**

***Index Terms*: - Deduplication, authorized duplicate check, confidentiality, hybrid cloud.**

I. INTRODUCTION

To make information administration adaptable in distributed computing, deduplication has been a notable procedure and has pulled in more consideration as of late. Information deduplication is a specific information pressure system for dispensing with copy duplicates of rehashing information away. The method is utilized to enhance stockpiling use and can likewise be connected to arrange information exchanges to diminish the quantity of bytes that must be sent. Rather than keeping numerous information duplicates with a similar substance, deduplication disposes of excess information by keeping just a single physical duplicate and alluding other repetitive information to that duplicate. Deduplication can occur at either the document level or the piece level. For record level deduplication, it wipes out copy duplicates of a similar document. Deduplication can likewise happen at the piece level, which wipes out copy squares of information that happen in non-indistinguishable records. In spite of the fact that information deduplication brings a considerable measure of advantages, security and protection concerns emerge as clients' touchy information are defenseless to both insider and pariah assaults .Traditional encryption, while giving information privacy, is contrary with information deduplication. In particular, customary encryption requires distinctive clients to encode their information with their own particular keys. Accordingly, indistinguishable information duplicates of various clients will prompt distinctive figure writings, making deduplication unimaginable. Merged encryption has been proposed to implement information privacy while making deduplication possible. It scrambles/decodes an information duplicate with a united key, which is gotten by figuring the cryptographic hash estimation of the substance of the information duplicate. After key age and information encryption, clients hold the keys and send the figure content to the cloud. Since the encryption operation is deterministic and is gotten from the information content, indistinguishable ldata duplicates will produce the same joined key and henceforth a similar figure content. To avert unapproved get to, a protected evidence of possession convention is additionally expected to give the verification that the client without a doubt claims a similar document when a copy is found. After the confirmation, resulting clients with a similar record will be given a pointer from the server without expecting to transfer a similar document. A client can download the encoded document with the pointer from the server, which must be decoded by the relating information proprietors with their joined keys. Consequently, concurrent encryption enables the cloud to perform deduplication on the figure writings and the confirmation of proprietorship keeps the unapproved client to get to the record.

II. LITERATURE SURVEY

In archival storage systems, there is a huge amount of duplicate data or redundant data, which occupy significant extra equipments and power consumptions, largely lowering down resources utilization (such as the network bandwidth and storage) and imposing extra burden on management as the scale increases. So data de-duplication, the goal of which is to minimize the duplicate data in the inter level, has been receiving broad attention both in academic and industry in recent years. In this paper, semantic data de-duplication (SDD) is proposed, which makes use of the semantic information in the I/O path (such as file type, file format, application hints and system metadata) of the archival files to direct the dividing a file into semantic chunks (SC). While the main goal of SDD is to maximally reduce the inter file level duplications, directly storing variable SCes into disks will result in a lot of fragments and involve a high percentage of random disk accesses, which is very inefficient. So an efficient data storage scheme is also designed and implemented: SCes are further packaged into fixed sized Objects, which are

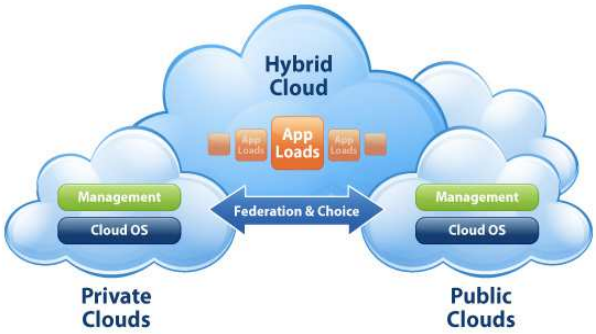
International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 10, October 2017)

actually the storage units in the storage devices, so as to speed up the I/O performance as well as ease the data management. Primary experiments have demonstrated that SDD can further reduce the storage space compared with current methods .. With the advent of cloud computing, secure data deduplication has attracted much attention recently from research cloud storage to reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality, Bellare et al. showed how to protect the data confidentiality by transforming the predicatable message into unpredicatable message. In their system, another third party called key server is introduced to generate the file tag for duplicate check. Stanek et al. presented a novel encryption scheme that provides the essential security for popular data and unpopular data. For popular data that are not particularly sensitive, the traditional conventional encryption is performed. Another two-layered encryption scheme with stronger security while supporting deduplication is proposed for unpopular data. In this way, they achieved better trade between the efficiency and security of the out-sourced data. Liet al. addressed the key management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files.

III. OVERVIEWOF THE HYBRID CLOUD CONCEPTS

3.1 HYBRID CLOUD

A hybrid cloud is a cloud computing environment in which an organization provides and manages some resources in-house and has others provided externally .For example, an organization might use a public cloud service, such as Amazon Simple Storage Service(Amazon S3) for archived data but continue to maintain in house storage for operational customer data



The concept of a hybrid cloud is meant to bridge the gap between high control, high cost "private cloud" and highly callable, flexible, low cost "public cloud". "Private Cloud" is normally used to describe a VMware deployment in which the hardware and software of the environment is used and managed by a single entity. The concept of a "Public cloud" usually involves some form of elastic/subscription based resource pools in a hosting provider datacenter that utilizes multi-tenancy. The term public cloud doesn't mean less security, but instead refers to multi-tenancy.

The concept revolves heavily around connectivity and data portability. The use cases are numerous: resource burst-ability for seasonal demand, development and testing on a uniform platform without consuming local resources, disaster recovery, and of course excess capacity to make better use of or free up local consumption.

VM ware has a key tool for "hybrid cloud" use called "vCloud connector". It is a free plugin that allows the management of public and private clouds within the vSphere client. The tool offers users the ability to manage the console view, power status, and more from a "workloads" tab, and offers the ability to copy virtual machine templates to and from a remote public cloud offering.

IV. HYBRID CLOUD FOR SECURE DEDUPLICATION

At a high level, our setting of interest is an enterprise network, 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 *rolebased* 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 accessible time period covers 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. 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 semitrusted 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

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 10, October 2017)

A. Architecture For Authorized Deduplication:

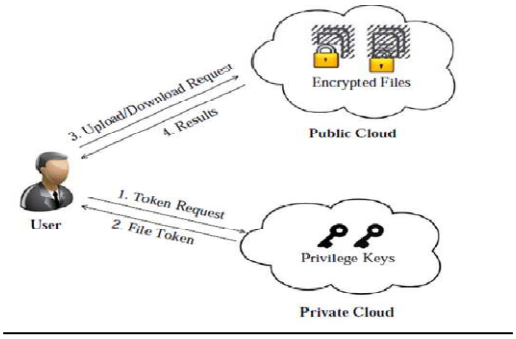


Fig.2 Architecture for Authorized deduplication

In this paper, we will 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, 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.

- *S-CSP.* This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.
- *Data Users.* A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized deduplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.
- *Private Cloud.* Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner 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.

Notice that this is a novel architecture for data deduplication in cloud computing, which consists of a twin clouds (i.e., the public cloud and the private cloud). Actually, this hybrid cloud setting has attracted more and more attention recently. For example, an enterprise might use a public cloud service, such as Amazon S3, for archived data, but continue to maintain in-house storage for operational customer data. Alternatively, the trusted private cloud could be a cluster of virtualized cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implement a remote execution environment trusted by the users.

V. ALGORITHMS USED

In this section, we use two types of algorithms,
1). For file uploading.
2). For file downloading.

FOR UPLOADING A FILE

BEGIN
Step –1 Read file
Step –2 Cloud server checks for duplication
Step –3 Sends duplication response whether the file already exists or not
Step – 4 If the file does not exist
4.1 Display "file does not exist"
Step – 5 Then it uploads the file
Step – 6 If the file already exist
6.1 Display "file already exist"
END

FOR DOWNLOADING A FILE

BEGIN
Step –1 Read file
Step –2 Cloud server checks for duplication
Step –3 Sends duplication response whether the file already exists or not
Step –4 If the file exist –
4.1 Display "file exist"
Step –5 then it downloads the file
Step –6 If the file does not exist –
6.1 Display "file does not exist"
END

VI. IMPLEMENTATION

We implement a prototype of the proposed authorized De-duplication system, in which we model three entities as separate C++ programs. A *Client* program is used to model the data users to carry out the file upload process. A *Private Server* program is used to model the private cloud which

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 10, October 2017)

manages the private key and handles the file token computation. A *Storage Server* program is used to model the S-CSP which stores and de-duplicates files. We implement cryptographic operations of hashing and encryption with the OpenSSL library [1].We also implement the communication between the entities based on HTTP, using GNU Libmicrohttpd [10] and libcurl [13]. Thus, users can issue HTTP Post requests to the servers. Our implementation of the **Client** provides the following function calls to support token generation and de-duplication along the file upload process.

- *FileTag(File)* – It computes SHA-1 hash of the File as File Tag;
- *TokenReq(Tag, UserID)* – It requests the Private Server for File Token generation with the File Tag and User ID;
- *DupCheckReq(Token)* – It requests the Storage Server for Duplicate Check of the File by sending the file token received from private server;
- *ShareTokenReq(Tag, {Priv.})* – It requests the Private Server to generate the Share File Token with the File Tag and Target Sharing Privilege Set;
- *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; and
- *FileUploadReq(FileID, File, Token)* – It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.

Our implementation of the **Private Server** includes corresponding request handlers for the token generation and maintains a key storage with Hash Map.

- *TokenGen(Tag, UserID)* - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1 algorithm; and
- *ShareTokenGen(Tag, {Priv.})* - It generates the share token with the corresponding privilege keys of the sharing privilege set with HMAC-SHA-1 algorithm.

Our implementation of the **Storage Server** provides de-duplication and data storage with following handlers and maintains a map between existing files and associated token with Hash Map.

- *DupCheck(Token)* - It searches the File to Token Map for Duplicate; and
- *FileStore(FileID, File, Token)* - It stores the File on Disk and updates the Mapping.

VII. EXPERIMENTAL RESULTS

The final results of the designed system are given below. From those results we get the detailed information to Check de-duplication and upload the files, Fetching the Signs using Hashing Algorithm, Checking for Duplication, file uploading, file downloading and attacker trying to attack(block) the cloud. Detailed procedure of the proposed system is given. Based on this we confirm that securely authorized de-duplication is successfully achieved with hybrid cloud approach. The output images given as below,

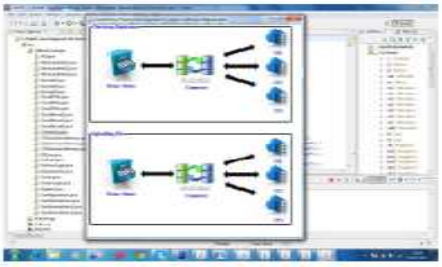


Figure 3. Checking duplication and uploading the files

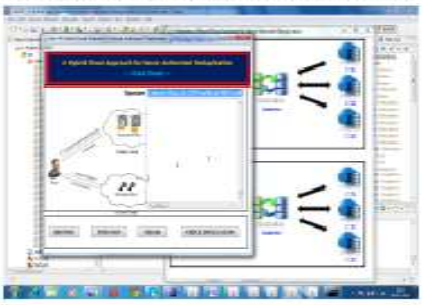


Figure 4. Fetching the Signs using Hashing Algorithm

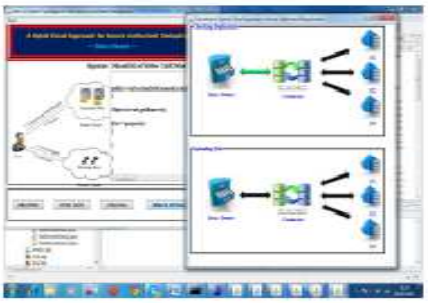


Figure5(a). Checking for Duplication.

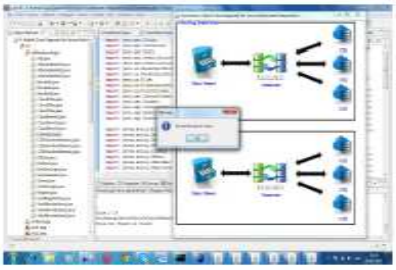


Figure 5(b). File not found in the cloud after checking for duplication

Figure6. File uploading

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 4, Issue 10, October 2017)

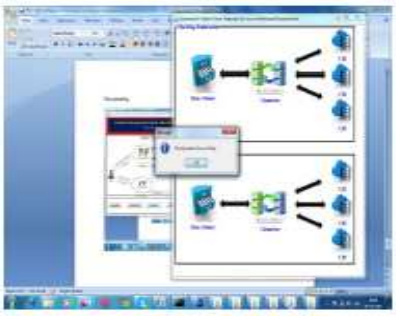


Figure7. Successfully uploaded the file

Figure8. Downloading files from the server

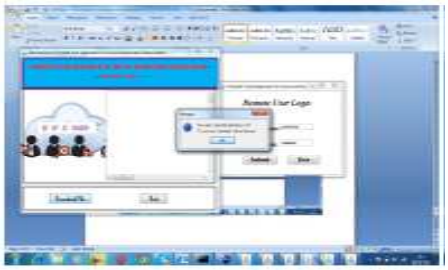


Figure9. Attacker trying to attack the cloud (Blocked)

Figure10. Successfully downloading the file if a valid user is logging in

VIII. FUTURE SCOPE

It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provide us with sufficient memory. It provides authorization to the private firms and protect the confidentiality of the important data.

REFERENCES

[1] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
[2] M. Bellare, S. Keelveedhi, and T. Ristenpart.Message- locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013..
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart.Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
[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] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
[6] Bugiel, S., N¨urnberger, S., Sadeghi, A.-R., Schneider, T.: Twin Clouds: An architecture for secure cloud computing (Extended Abstract). In: Workshop on Cryptography and Security in Clouds (WCSC 2011), March 15-16 (2011)
[7] Chung, K.-M., Kalai, Y., Vadhan, S.: Improved delegation of computation using fully homomorphic encryption. In: Rabin, T. (ed.) CRYPTO 2010. LNCS, vol. 6223, pp. 483–501. Springer, Heidelberg (2010)
[8] Cloud Security Alliance. Top threats to cloud computing, v. 1.0 (2010)