International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 3, Issue 4, July 2016 |

Hybrid Cloud Approach for Secure Authorized Deduplication

Rushikesh Naiknaware[1], Omkar Deshpande[2], Abhishek Kangle[3], Himalay Koli[4]

Dept. of IT, Rajarshi Shahu College of Engineering, Tathawade, Savitribai Phule Pune University, Pune, India[1234]

ABSTRACT: In individualized computing gadgets that depend on a distributed storage environment for information reinforcement, a fast approaching test confronting source de-duplication for cloud reinforcement administrations is the low de-duplication effectiveness because of a blend of the asset escalated nature and the constrained framework assets. Information de-duplication is one of imperative information pressure strategies for disposing of copy duplicates of rehashing information, and has been generally utilized as a part of distributed storage to lessen the measure of storage room and spare transmission capacity. To secure the secrecy of delicate information while supporting de-duplication, the merged encryption strategy has been proposed to encode the information before outsourcing. To better ensure information security, this paper makes the main endeavor to formally address the issue of approved information de-duplication. Not quite the same as conventional de-duplication frameworks, the differential benefits of clients are further considered in copy check other than the information itself. We additionally show a few new de-duplication developments supporting approved copy check in cross breed cloud design. Security examination shows that our plan is secure as far as the definitions determined in the proposed security model. As a proof of idea, we execute a model of our proposed approved copy check plan and direct tried analyses utilizing our model. We demonstrate that our proposed approved copy check plan acquires insignificant overhead contrasted with typical operations.

KEYWORDS: Deduplication authorized duplicate check, confidentiality, hybrid cloud.

I. INTRODUCTIONS

Distributed computing gives apparently boundless "virtualized" assets to clients as administrations over the entire Internet, while concealing stage and execution subtle elements. Today's cloud administration suppliers offer both profoundly accessible capacity and hugely parallel registering assets at generally low expenses. As distributed computing gets to be common, an expanding measure of information is being put away in the cloud and imparted by clients to determined benefits, which characterize the entrance privileges of the put away information. One basic test of distributed storage administrations is the administration of the perpetually expanding volume of information. To make information administration versatile in distributed computing, de-duplication has been a surely understood procedure and has pulled in more consideration as of late. Information de-duplication is a specific information pressure strategy for disposing of copy duplicates of rehashing information away. The strategy is utilized to enhance stockpiling usage and can likewise be connected to network information exchanges to decrease the quantity of bytes that must be sent. Rather than keeping numerous information duplicates with the same substance, de-duplication takes out excess information by keeping one and only physical duplicate and alluding other repetitive information to that duplicate. De-duplication can happen at either the document level or the piece level. For document level de-duplication, it takes out copy duplicates of the same record. De-duplication can likewise happen at the piece level, which takes out copy squares of information that happen in non-indistinguishable records. Distributed computing is a developing administration demonstrates that gives calculation and capacity assets on the Internet. One appealing usefulness that distributed computing can offer is distributed storage. People and ventures are regularly required to remotely file their information to stay away from any data misfortune on the off chance that there are any equipment/programming disappointments or unanticipated debacles. Rather than buying the required stockpiling media to keep information reinforcements, people and ventures can basically outsource their information reinforcement administrations to the cloud administration suppliers, which give the essential stockpiling assets to have the information reinforcements. While distributed storage is appealing, how to give security assurances to outsourced information turns into a rising concern. One noteworthy security test is to give the property of guaranteed erasure, i.e., information records are for all time blocked endless supply of cancellation. Keeping information reinforcements for all time is undesirable, as touchy data might be uncovered later on account of information rupture or mistaken administration of cloud administrators. In this way, to maintain a strategic distance from liabilities, ventures and government organizations more often than not keep their reinforcements for a limited number of years and solicitation to erase (or annihilate) the reinforcements thereafter. For instance, the US Congress is detailing the Internet Data Retention enactment in approaching ISPs to hold information

for a long time, while in United Kingdom, organizations are required to hold wages and compensation records for a long time.

II. GOALS AND OBJECTIVES

1. Unforged ability of file token/duplicate-check token. Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the S-CSP. The users are not allowed to collude with the public cloud server to break the unforged ability 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.

2. In distinguishability 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 or the privilege information.

3. Data Confidentiality. Unauthorized users without appropriate privileges or files, 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.

OBJECTIVE

1. To improved integrity
2. To increase the storage utilization
3. To remove the duplicate copies of data and improve the reliability.
4. To improve the security

III. LITERATURE SURVEY

Data de-duplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. De-duplication 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. Companies frequently use de-duplication in backup and disaster recovery applications, but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data de-duplication. [5]

Data de-duplication 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 de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new de-duplication 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.[3]

Data de-duplication is one of important data compression techniques which are for eliminating duplicate copies of repeating data, and has been widely used in cloud storage in order to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, papers makes the

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 3, Issue 4, July 2016 |

first attempt to formally address the problem of authorized data de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. Also present several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that the proposed scheme is secure in terms of the definitions specified in the proposed security model. [2]

This paper represents that, many techniques are using for the elimination of duplicate copies of repeating data, from those techniques, one of the important data compression technique is data duplication. Many advantages with this data duplication, mainly it will reduce the amount of storage space and save the bandwidth when using in cloud storage. To protect confidentiality of the sensitive data while supporting de-duplication data is encrypted by the proposed convergent encryption technique before out sourcing. Problems authorized data duplication formally addressed by the first attempt of this paper for better protection of data security. This is different from the traditional duplication systems. The differential privileges of users are further considered in duplicate check besides the data itself. In hybrid cloud architecture authorized duplicate check supported by several new duplication constructions. Based on the definitions specified in the proposed security model, our scheme is secure. Proof of the concept implemented in this paper by conducting test-bed experiments. [6]

The popularity and widespread use of Cloud have brought great convenience for data sharing and data storage. The data sharing with a large number of participants take into account issuers like data integrity, efficiency and privacy of the owner for data. In cloud storage services one critical challenge is to manage ever increasing volume of data storage in cloud. To make data management more scalable in cloud computing field, de-duplication a well-known technique of data compression to eliminating duplicate copies of repeating data in storage over a cloud. Even if data de-duplication brings a lot of benefits in security and privacy concerns arise as user's sensitive data are susceptible to both attacks insider and outsider. A convergent encryption method enforces data confidentiality while making de-duplication feasible. Traditional de-duplication systems based on convergent encryption even though provide confidentiality but do not support the duplicate check on basis of differential privileges. This paper presents, the idea of authorized data de-duplication proposed to protect data security by including differential privileges of users in the duplicateCheck. [7]

IV. EXISTING SYSTEM APPROACH

From the above writing review we have inferred that current information de-duplication frameworks, the private cloud are included as an intermediary to permit information proprietor/clients to safely perform copy check with differential benefits. Such design is functional and has pulled in much consideration from specialists. The information proprietors just outsource their information stockpiling by using open cloud while the information operation is overseen in private cloud. We display a propelled plan to bolster more grounded security by scrambling the document with differential benefit keys. Along these lines, the clients without comparing benefits can't perform the copy check. Moreover, such unapproved clients can't decode the figure message even connive with the S-CSP.

V. ALGORITHM USED

1) Convergent Encryption

Convergent encryption provides data confidentiality in de-duplication. A user (or data owner) derives aconvergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Note that both the convergent key and the tag are independently derived and the tag cannot be used to deduce the convergent key and compromise data confidentiality. Both the encrypted data copy and its corresponding tag will be stored on the server side.

A 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 ciphertextC;
3. **DecCE(K, C)!**M is the decryption algorithm that takes both the ciphertextC and the convergent key K as inputs and then outputs the original data copy M; and

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 3, Issue 4, July 2016 |

4. **TagGen(M)!T (M)** is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

2) Proof Of Ownership

The notion of proof of ownership(POW) enables users to prove their ownership of data copies to the storage server.Specifically, POW is implemented as an interactive algorithm (denoted by POW). The verifier derives a short value $\phi(M)$ from a data copy M. To prove the ownership of the data copy M, the properneeds to send $\phi$to the verifier such that $\phi = \phi(M)$.

PSEUDO CODE

Step1:Calculate the two convergent key values
Step2: Compare the two keys and files get accessed.
Step3: Apply de-duplication to eradicate the duplicate values.
Step4: Ifany other than the duplicates it will be checked once again and make the data unique.
Step5: That data will be unique and also more confidential the authorized can access and data is stored.

VI. SYSTEM ARCHITECTURE

1. Secret Sharing Scheme:

Secret sharing scheme performs two operations namely Share and Recover. The secret is divided and shared by using Share. With enough shares, the secret can be extracted and recovered with the algorithm of Recover. The input to this module is file. It performs dividing of file into fixed size blocks or shares. These blocks are then encoded and allocated on cloud server at different nodes. When user request for file these blocks are decrypted and by combining these blocks file is given to user.

2. Tag Generation:

In this tag similarity is considered a kind of semantic relationship between tags, measured by means of relative co-occurrence between tags, known as J. coefficient. The input to this block is file blocks. This module assigns tags to each block for duplication check. The output of this module is blocks with tag assigned.

3. Convergent Encryption Module

Traditional encryption, while providing data confidentiality, is incompatible with data de-duplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making de-duplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making de-duplication feasible. It encrypts/ decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text.

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 3, Issue 4, July 2016 |

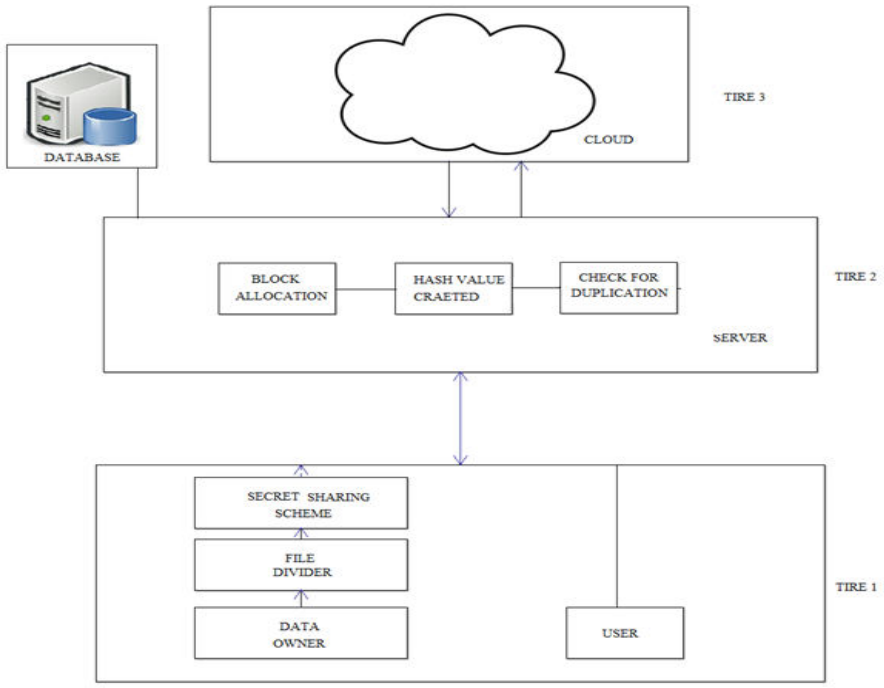

Fig 01. System Architecture.

VII. IMPLEMENTATION RESULT

We implement a prototype of the proposed authorized deduplication 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 manages the private keys and handles the file token computation. A Storage Server program is used to model the S-CSP which stores and deduplicated files. Our implementation of the Client provides the following function calls to support token generation and deduplication along the file upload process.

1. FileTag(File) - It computes SHA-1 hash of the File as File Tag;
2. TokenReq(Tag, UserID) - It requests the Private Server for File Token generation with the File Tag and User ID;
3. DupCheckReq(Token) - It requests the Storage Server for Duplicate Check of the File by sending the file token receivedfrom private server;
4. ShareTokenReq(Tag, {Priv.}) - It requests the Private Server to generate the Share File Token with the File Tag andTarget Sharing Privilege Set;
5. FileEncrypt(File) - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining
6. (CBC) mode, where the convergent key is from SHA-256 Hashing of the file;
7. FileUploadReq(FileID, File, Token) – It uploads the File Data to the Storage Server if the file is Unique and updates the
8. File Token stored. Our implementation of the Private Server includes corresponding request handlers for the tokengeneration and maintains a key storage with Hash Map.
9. TokenGen(Tag, UserID) - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1

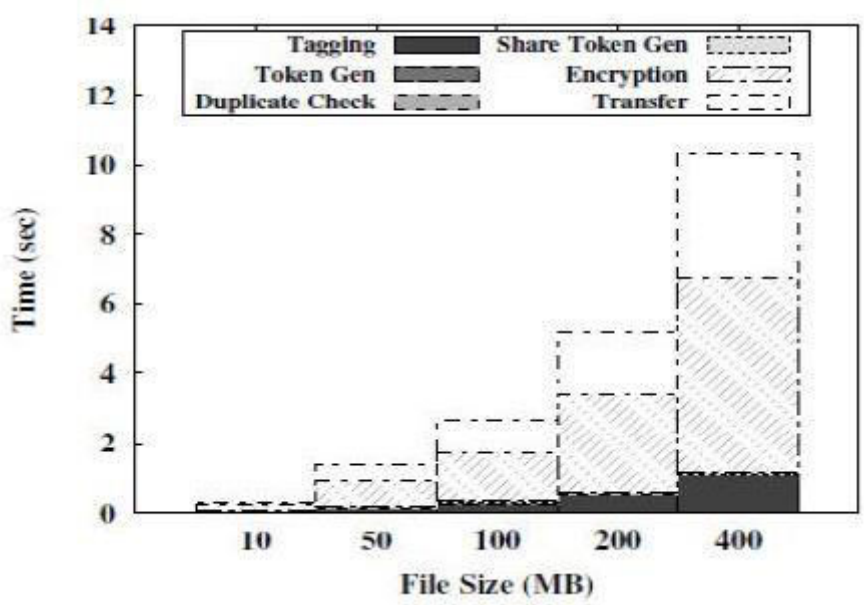

Fig 02 Time Breakdown for Different File Size

VIII. CONCLUSION AND FUTURE WORK

A few new de-duplication developments supporting approved copy check in half and half cloud engineering, in which the copy check tokens of documents are produced by the private cloud server with private keys. Security examination exhibits that our plans are secure as far as insider and pariah assaults indicated in the proposed security model. As a proof of idea, we executed a model of our proposed approved copy check plan and direct proving ground investigates our model. We demonstrated that our approved copy check plan acquires insignificant overhead contrasted with united encryption and system exchange.

REFERENCES

1. S. Plank LihaoXu Optimizing Cauchy Reed-Solomon Codes for Fault-Tolerant Network Storage Applications The 5th IEEE International Symposium on Network Computing and Applications (IEEE NCA06), Cambridge, MA, July, 2006
2. Mr Vinod B Jadhav Prof Vinod S Wadne Secured Authorized De-duplication Based Hybrid Cloud Approach International Journal of Advanced Research in Computer Science and Software Engineering
3. Abdul Samadhu, J. Rambabu, R. Pradeep Kumar, R. Santhya Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized De-duplication International Journal for Research in Applied Science and Engineering Technology (IJRASET)
4. JadapalliNandini, Rami reddyNavateja Reddy Implementation De-duplication System with Authorized Users International Research Journal of Engineering and Technology (IRJET)
5. Sharma Bharat, Mandre B.R. A Secured and Authorized Data De-duplication with Public Auditing International Journal of Computer Applications (09758887)
6. Wee Keong Ng SCE, NTU Yonggang Wen SCE, NTU Huafei Zhu Private Data De-duplication Protocols in Cloud Storage SAC12 March 2529, 2012, Riva del Garda, Italy. Copyright 2011 ACM 9781450308571/12/03
7. Shweta D. Pochhi, Prof. Pradnya V. Kasture Encrypted Data Storage with De-duplication Approach on Twin Cloud International Journal of Innovative Research in Computer and Communication Engineering
8. Backialakshmi. N Manikandan. M SECURED AUTHORIZED DE-DUPLICATION IN DISTRIBUTED SYSTEM IJIRST International Journal for Innovative Research in Science and Technology— Volume 1 — Issue 9 — February 2015
9. BhushanChoudhary, AmitDravid A Study On Secure Deduplication Techniques In Cloud Computing International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) Volume 3, Issue 12, April 2014