INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 5, Issue 4 OCT 2016

Improving Storage Efficiency in the Hybrid Cloud Environment with Secure and Authorized Deduplication Technique

I.soundarya, M.tech,Mail Id : isoundarya57@gmail.com.

K.Ravichandra, M.tech, Associate Professor,.MailId:ravichandra.kosuri@gmail.com.

Nova college of engineering and technology, Vegavaram, Jangareddigudem, WestGodavari.

ABSTRACT – ***Cloud Computing, a platform which is providing dynamic and scalable computing environment for many kinds of applications on the web. It provides a large-scale data storage on the basis of pay-as-you-go pricing scheme. Deduplication is one of most widely used data compression and duplicate detection technique for eliminating the duplicate copies of repeated data. It has been most widely used in cloud storage management to reduce data redundancy, amount of storage space, and to save a bandwidth.To defend the confidentiality of sensitive data even as supporting deduplication, the convergent encryption technique was proposed for a data encryption earlier than outsourcing. To better protect data security, this paper makes the first tries to formally cope with the trouble of authorized data deduplication. Unique from conventional deduplication structures, the differential privileges of users are in addition taken into consideration in duplicate check besides the data itself. Security evaluation demonstrates that our scheme is comfortable in terms of the definitions special within the proposed protection model. We show that our proposed authorized duplicate check scheme incurs minimal overhead as compared to regular operations.***

Keywords - *Deduplication, authorized duplicate check, confidentiality, hybrid cloud*

I. INTRODUCTION

In cloud computing environment data de-duplication is a most essential data compression technique for eliminating and reducing identical copies of same data. This mechanism is primarily employed to increase the effective usage of storage space and additionally it is applied to enabledata transmission over network in de-duplication methodology identical data are notice and hold on throughout method of study. As this technique continues different copies of same data are matched to the hold on copy and whenever a matched data found, the identical data is then replaced with a tiny low reference that self-addressed to hold on data. A hybrid cloud may be a combination of personal cloud and public cloud which is most crucial that resides on a personal cloud and therefore the data which is well accessible can be resided on a public cloud hybrid cloud is useful for responsibility, extensibility and quick readying and price saving of public cloud with a lot of security troubles with non-public cloud. The complicated challenge of cloud storage or cloud computing is that the agreement of huge volumeof data duplication may be a method of eliminating of duplicate data in deduplication techniques redundant data removed exploit single instance of the info to be hold on. Within the previous recent system the info is encrypted back to outsourcing it on the cloud or network. These encoding needs, most time similarly as storage space demand to encrypt the info if there's great deal of information at that point encoding method becomes complicated and demanding. By mistreatment de-duplication technique in hybrid cloud the encoding technique become less complicated. As we tend to all of is aware of that the network has great deal of information that being shared by several users. Several massive networks uses data cloud to store the data and share that data on the network. As cloud computing becomes rife, an increasing quantity of information is being stored within the cloud and shared by users with specified privileges, that summarize the access rights of the hold on data. One major challenges of cloud storage services are that the management of the improving volume of information. To create data management ascendable in cloud computing, deduplication has been a famous technique and has attracted a lot of and a lot of focus recently. Data deduplication may be a specialized data compression technique for removing duplicate copies of continuance data in storage. The technique is employed to boost storage usage and might even be applied to network data transfers to reduce the amount of bytes that has to be sent. Before keeping multiple data copies with constant content, deduplication remove redundant datacopies by keeping only single physical copy and referring different redundant data thereto copy. Deduplication will be happened at either at a file level or the block level. For file level deduplication, it eliminates duplicate copies of constant file. Deduplication may happen at the block level that removes duplicate blocks of information that occur in non-identical files.

II. RELATED WORK

A. Security of Deduplication Systems

With the arrival of cloud computing, secure data deduplication has attracted a good deal interest these days from studies community. Yuan et al. was proposed a deduplication technique in the cloud storage to lessen the storage length of the tags for integrity test. To enhance the safety of deduplication and defend the information confidentiality, Bellare et al. has confirmed a way to defend the information confidentiality by reworking the predicatable message into the unpredicatable message. Of their system, every other third party called key server is delivered to generate the file tag for replica check. Stanek et al. provided a unique encryption scheme that provides differential security for popular facts and unpopular data. For popular data that aren't specifically sensitive, the conventional traditional encryption is executed. some other two-layered encryption scheme with more potent protection while assisting deduplication is proposed for unpopular statistics. In this manner, they finished higher trade-off between the efficiency and protection of the outsourced information. Li et al addressed the keymanagement issue in block-stage deduplication by means of distributing those keys across more than one server after encrypting the documents.

B. The Convergent Cryptography

Convergent encryption guarantees data privateness in deduplication. Bellare et al formalized this primitive as message-locked encryption, and explored its utility in area-efficient comfortable outsourced storage. Xu et al additionally addressed the trouble and confirmed a comfortable convergent encryption for efficient encryption, without thinking about issues of the important thing-control and block-stage deduplication. There also are numerous implementations of convergent implementations of various convergent encryption versions for secure deduplication it's miles regarded that some business cloud storage vendors, together with Bitcasa, additionally deploy convergent encryption.

C. Proof of Data and Document Ownership

Halevi et al proposed thenotion of "proofs of ownership" for deduplicationstructures, such that a consumer can effectively prove to the cloud storage server that he/she owns a document without uploading the document itself. Several PoW buildings based on the Merkle-Hash Tree are proposed toallow purchaser-aspect deduplication, which encompass thebounded leakage setting. Pietro and Sorniott proposed another efficient PoW scheme through selectingthe projection of a document onto a few randomly decided onbit-positions because the report proof. word that each one the above schemes do now not bear in mind records privacy. these days, Ng et al extended PoW for encrypted files, but theydo no longer deal with a way to reduce the key managementoverhead.

III. FRAMEWORK

In the projected system we have a tendency to square measure achieving the info deduplication by providing the proof of a data by the info owner. This proof is employed at the time of loading of the file. Each file uploaded to the cloud is more additionally finite by a group of privileges to specify which type of users is allowed to perform the duplicate checking and access the files. Before submitting his duplicate check request for a few file, the user has to take this file and very own privileges as inputs. The user is in a position to seek out a replica for this file if and on condition that there's a replica of this file and a matched privileges hold on in cloud.

A. Encryption Of Files

1. Here we are using common secret key k to encrypt as well as decrypt data. This will use to convert the plain text to the ciphertext and again ciphertext to plain text. Here we have used three basic functions,

2. ***KeyGenSE***:herek is the key generation algorithm that generates key κ using by using a security parameter *l*.

3. ***EncSE (k, M):*** hereC is a symmetric encryption algorithm which takes the secret κ and a message M and after that outputs the ciphertext C.

DecSE (k, C): Now*M* is our symmetric decryption algorithm that takes the secret *κ* as the input and a ciphertext *C* and then produces the original message *M*.

B. Confidential Encryption Technique

We uIt provides knowledge confidentiality in deduplication. A user derives a focused key from each original knowledge copy and encrypts the info copy with the focused key. More additionally, the user additionally derives a tag for the info copy, such the tag are going to be wont to sight duplicates.

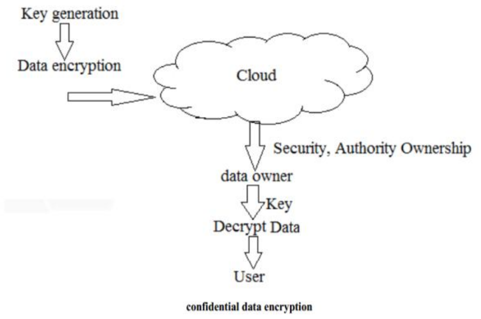


Fig.1. A System Overview

C. A Security Proof of Data

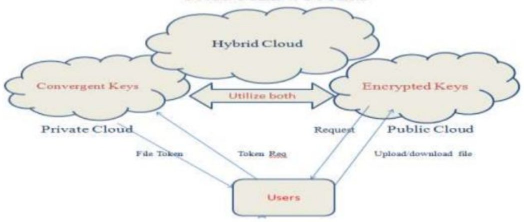


Fig.2. Architecture of Secure and Authorized Deduplication system

The user has to be compelled to prove that the info that he need to transfer or transfer is its own data. Which means he have to be compelled to give the convergent key and supportive knowledge to prove his ownership at server.

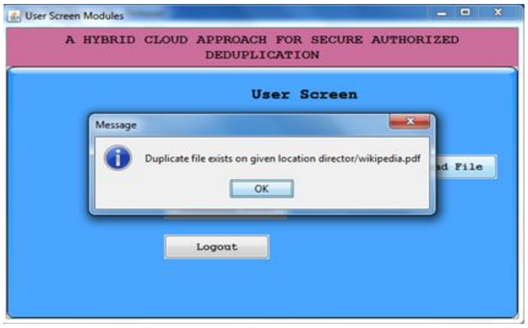
This paper we proposed the system consist of hybrid cloud. After registration the users can upload their data in to clouddata.

If you are trying to upload any similar data(In our proposed system cloud servers not allows the duplicate already uploaded data) it will not allow and shows message like as shown below.

IV. EXPERIMENTAL RESULTS

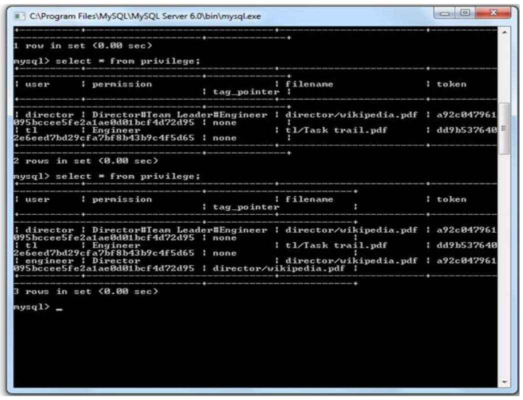
In our experiments, users registered as Director, Engineer and Team Leader. When director upload one file he must give privilege to another user ether engineer or Team Leader. If he gives the permission only for Engineer then engineer only can access that file. And incase Director Upload same file for second time then our system detects the file already existed.

The given screen describes that the if any file existed in the system, it display the message i.e.,



Duplicate file exists on given location director/wikipedia.pdf

When you are uploaded any data the data will be stores in the encrypted format and same tag_pointer generated for duplicate file.



```
+----------+-------------------------------+----------------------+------------
-----------------------------+----------------+
1 row in set (0.00 sec)

mysql> select * from privilege;
+----------+-------------------------------+----------------------+------------
-----------------------------+----------------+
| user     | permission                    | filename             | token
                              | tag_pointer |
+----------+-------------------------------+----------------------+------------
-----------------------------+----------------+
| director | Director#Team Leader#Engineer | director/wikipedia.pdf | a92c047961
095bccee5fe2a1ae0d01bcf4d72d95 | none        |
| tl       | Engineer                      | tl/Task trail.pdf    | dd9b537640
2e6eed7bd29cfa7bf8b43b9c4f5d65 | none        |
+----------+-------------------------------+----------------------+------------
-----------------------------+----------------+
2 rows in set (0.00 sec)

mysql> select * from privilege;
+----------+-------------------------------+----------------------+------------
-----------------------------+-------------------------------+
| user     | permission                    | filename             | token
                              | tag_pointer                   |
+----------+-------------------------------+----------------------+------------
-----------------------------+-------------------------------+
| director | Director#Team Leader#Engineer | director/wikipedia.pdf | a92c047961
095bccee5fe2a1ae0d01bcf4d72d95 | none                          |
| tl       | Engineer                      | tl/Task trail.pdf    | dd9b537640
2e6eed7bd29cfa7bf8b43b9c4f5d65 | none                          |
| engineer | Director                      | director/wikipedia.pdf | a92c047961
095bccee5fe2a1ae0d01bcf4d72d95 | director/wikipedia.pdf |
+----------+-------------------------------+----------------------+------------
-----------------------------+-------------------------------+
3 rows in set (0.00 sec)

mysql> _
```

V. CONCLUSION

In context our paper, the main aim of authorised deduplication of a data is used to ensure the data security by including several differential privileges of several users in the process of duplicate checking. The hybrid cloud architecture is proposed to support the operation of authorised duplicate checking. In thisa private cloud produces the duplicate check tokens of various files, so that we can say, our system can be protected from both of several insider attacks and outsider attacks. And our authorised duplicate check model can eliminate the overhead than the convergent encryption and secure network transfer.

References

[1] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In *Workshopon Cryptography and Security in Clouds (WCSC 2011)*, 2011.

[2] 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 andCommunications Security*, pages 81–82. ACM, 2012.

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

[4] OpenSSL Project. http://www.openssl.org/.

[5] GNU Libmicrohttpd . http://www.gnu.org/software/libmicrohttpd/.

[6] 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.

[7] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15th NIST-NCSC National Computer Security Conf.*, 1992.

[8] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs ofownership in remote storage systems. In Y. Chen, G. Danezis,and V. Shmatikov, editors, *ACM Conference on ComputerandCommunications Security*, pages 491–500. ACM, 2011.

[9] libcurl. http://curl.haxx.se/libcurl/.

[10] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In *Proc. of APSYS*,Apr 2013.

[11] P. Anderson and L. Zhang. Fast and secure laptop backups withencrypted de-duplication. In *Proc. of USENIX LISA*, 2010.K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan. Sedic: privacyaware data intensive computing on hybrid clouds. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

KOSURIRAVICHANDRA : He is an Associate.Professor in Department of Computer Science and Engineering at Nova College Of Engineering &Technology.

Indla.Soundarya:She was born in Tadepalli in Guntur District, AP on July 05, 1990. She graduated from the Jawaharlal Nehru Technological University, Kakinada. Her special fields of interest included cloud computing and software engineering. Presently she is studying M.Tech (Software Engineering)in NovaCollegeOf Engineering&Technology, Vegavaram.