International Innovative Research Journal of Engineering and Technology
ISSN NO: 2456-1983

International Conference on Emerging Innovation in Engineering and Technology

ICEIET-2017

Secure Authorized De-duplication Data In Hybrid Cloud

S.Ramya[1], Dr.V.Govindasamy[2]

[1]PG Student/IT & Pondicherry engineering college
[2] Associate Professor/IT & Pondicherry engineering college
[1]ramiyasundaram@gmail.com

Abstract

Information de-duplication is one of the dynamic information burden procedures for the removal of replica duplicates information. It has been largely used as a part of reducing the storage space and extra data transmission. For providing the privacy of information while supporting de-duplication, the information is encrypted before outsourcing. In order to ensure information security, the paper outputs to formally address the issue of approved information de-duplication. It is not same as the predictable de-duplication frameworks, the differential benefits of clients are further considered in copy check other than the information itself. We moreover present a few new de duplication expansions supporting approved copy check in limit cloud design. Security check shows that our plan is secure concerning the explanation determined in the proposed security model. As a proof of idea, we realize a model of our proposed approved copy check plan and direct showing ground tests using our work. The proposed system approves copy check plan supports trivial overhead contrasted with ordinary processes.

Keywords—authorized duplicate check, convergent encryption, duplication privileges, de-duplication, and hybrid cloud**.**

I. INTRODUCTION

Cloud computing is a type of network based computing that provides shared computer resources data to other devices on demand. It acts as a shared pool of computing resources. Cloud computing exhibits one of key characteristics is security that can improve security focused resources and centralizations of data. Security is often good as or better than traditional system. The security is very much improved when data is spread over large area or over a greater number of devices, as well as in multi-tenant system shared by unrelated user. In addition, user access to security check logs may be complex or impossible.

Cloud computing is mainly used for storing data, retrieve data. In most of the storage system, make duplicate copies of data. The same file has been saved many times by different users in different places which can make duplication records. De-duplication is method remove replica copies in storage system. De-duplication method is used in backup and recovery application. De-duplication method is used in concern for backup and ruin recovery applications, this method is also used to make free space in storage system.

De-duplication methods have two level, file level and block level duplication [2] [3.] The file level de-duplication method also called as Single Instance Storage (SJS). Single Instance storage is used system capacity to maintain one copy of content that computer share or many user. It is a mean to eliminate replica of data and to increase efficiency. SIS is used in file system, e-mail server and backup or recovery.

Block level method large volumes of storage that can be spitted into number of block. Each block can check replica data. These blocks are avoiding a duplicate content of data. The Block-level de-duplication better than a Single Instance Storage.

J. R. Douceur et al. defines the convergent encryption algorithm where generated a convergent key is caused through hash value from the content of data. For duplicate data copies same cipher text will be created, hence convergent encryption is likely with de-duplication method. S. Halevi et al. defines the proof of ownership (POW) algorithm [11] which will provide access control.

The data in proof of ownership protocol proves that the user goes to his own file and by own file when duplicate file is create. Once the proof gets established and server will provide indicator to the same file to the user without the essentiality to upload the similar file [4].

Though, the old de-duplication system is not provisioning the difference approval duplicate check, which is important in most of the requests. In illegal de-duplication systems;

during system initialization, set of rights will be the problem for each user.

The rights are defined (or bound) to each file at the time of uploading into the cloud to specify the type of user allowed to perform duplicate check and admittance of the files. User requirement to give its own file and own rights as input before submitting the request for the file [6].

The duplicates of user own its file with right. The user have key inputs to find duplicates file for the user and access control is increase factor that giving correct rights to correct user. To know the access, the control file has to be encrypted and allows duplication check with assured rights.

If the correct user will be stimulated to public cloud on storage server. The public cloud on storage server provider (S-CSP) with exact rights and to store only one replica of the same file, the de-duplication method will be used. The privacy has to be reflected while giving access [7]. In convergent encryption method does not have difference rights [8].

II. LITTERATURE SURVEY

Halevi et al. Proposed the proof of ownership the idea of "verification of proprietorship" for de-duplication framework, to such verify to validate user. The Distributed storage server owns a document itself. This POW algorithm uses Merkle Hash Tree which joins the incomplete leakage setting. Pietro and Sorniotti proposed one more plan of option same record in bit-position as the document indication.

Convergent Encryption is a combined encryption Possibilities information security in de-duplication. Bellare et al. this Convergent encryption is also known as content of hashing key, this providing identical cipher text key for identical cipher text files. Unique message-bolted encryption and surveyed its application in duplication system. Xu et al. additionally disposed to the problem and established a safe simultaneous encryption for current encryption, without seeing issues of the key-administration and piece level de duplication.

There are likewise a few traditions of focalized executions of various combined encryption differences for secure de duplication. It is realized that some commercial distributed storage suppliers.

Additionally, with the entry of distributed computing, secure information de-duplication is important factor. Yuan et al. proposed a de-duplication basis in the distributed storing to reduce the volume size of the tags for check log.

To raise the security of de-duplication and check the information secrecy, Bellare et al. established to confirm the information secrecy by altering the expected message into expected message. In their basis, another outsider called key server is familiar with create text tag for copy check. Stanek et al. presented a novel encryption combine that gives difference safety to prominent information and disliked information.

For normal information that are not particularly sensitive, the normal encryption is completed. Extra two-layered encryption idea with more stranded security while associate de duplication is proposed for disagreeable information. Beside these lines, they consummate better exchange off among the efficiency and security of the subcontracted information. Li et al. tended to the key administration problem in piece level de-duplication by transmission these keys over numerous servers in the wake of encoding the records.

ClouDedup: Secure De-duplication with Encrypted Data for Cloud Storage:

With the continuous and exponential increase of the number of users and the size of their data, data de-duplication becomes more and more a necessity for cloud storage providers. By storing a unique copy of duplicate data, cloud providers greatly reduce their storage and data transfer costs.

The advantages of de-duplication unfortunately come with a high cost in terms of new security and privacy challenges. ClouDedup, a secure and efficient storage service which assures block-level de-duplication and data confidentiality at the same time.

The security of Clouded up relies on its new architecture whereby in addition to the basic storage provider, a metadata manager and an additional server are defined: the server adds an additional encryption layer to prevent well-known attacks against convergent encryption and thus protect the confidentiality of the data.

Finally, the solution is fully compatible with standard storage APIs and transparent for the cloud storage provider, which does not have to be aware of the running de-duplication system.

Therefore, any potentially untrusted cloud storage provider such as Amazon, Dropbox and Google Drive, can play the role of storage provider.

The various surveys on the current systems are described as follows in this literature survey.

Author	Method	Feature	Result
M. Bellare, S. Keelveedhi, and T. Ristenpart	DupLESS Server-Aided Encryption for De-duplicated Storage	• Saving a Bandwidth • Determination the cross user de-duplication to reduce network Bandwidth. • Security provide for external attacker. • This provide high act	Simple storage Interface.
Ms. Samita Mokal. Prof. Nilima D. Nikam. Prof. Vaishali Londhe	Detection of File Level and Block Level De-duplication and attacks in Cloud Computing Environment	• Efficient for key management scheme for secure de-duplication • In regular upload/download operation reduce overhead using key sharing. • allocates key shares crossway numerous key server • Encrypting/decryption overhead.	Attacker module makes it more secure.
Hui Cui, Robert H. Deng, Yingjiu Li, and Guowei Wu	Attribute-Based Storage Supporting Secure De-duplication of Encrypted Data in Cloud	• User has specified access policy for sharing decrypting decryption key using attribute Based method.	It achieves stronger security compared to existing de-duplication.
S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider	Twin clouds: An architecture for secure cloud computing	• Protected computation • Store large capacity of data • Low potential • Secure execution environment	Clients provide trusted cloud to proxy. Then it manageable outsourced data and queries.
W. K. Ng, Y. Wen, and H. Zhu	Private data de-duplication Protocols in cloud storage	• Improve de-duplication speed. • Fault accepting • Reduce the cloud storage capacity.	Improve the competence of data.

TABLE 1 COMPARISON OF DIFFERENT METHODS OF DATA DE-DUPLICATION IN CLOUD STORAGE.

III. OVERVIEW OF EXISTING SYSTEM

In existing system, data duplication is done by proxy. The proxy can allowed data user to access data with rights. The data owners only utilize their data and data managed by public cloud. In traditional encryption algorithm, the users use the same key for both encryptions as well as for decryption. The data is encrypted by same key for all users. So any user can upload/download the file. So the duplication is impossible and it has the low security.

Authorized duplicate check in which the duplicate-check tokens of files are generated by the private cloud server with private keys. The scheme not supports stronger security by encrypting the file with differential privilege keys. The users without corresponding privileges cannot perform the duplicate check. Unauthorized users cannot decrypt the cipher text even plan with the S-CSP.

The convergent encryption technique is used to encrypt the data with convergent key. It encrypts/decrypts the data with differential key based on the set of privileges. The private cloud is used to generate the private key. The user encrypts the data with private key and stores the data in public cloud. It will store the encrypted data as well as the tag in the storage table to identify the duplicate copies of repeating data. VAST is the storage system that is used to store the data with security.

It improves the security of the valuable and secret data. Since, it stores the data in the form of manageable pieces. In cloud computing, a virtual machine is an imitation of system. Virtual machine has key device for functionality of a physical computer. It also called vconnector.

Drawback of Existing System

- Identical data copies of different user will lead to same cipher text making de-duplication impossible.
- Low security
- Does not support the differential duplication check.
- The users encrypt/decrypt the data with same key.
- User could not identify the data easily, since it is more complex.
- Private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check.

IV. PROPOSED SYSTEM ARCHITECTURE

In this paper, we introduce document or record level de-duplication for simplicity. This documents level duplication has find duplication in different entity in cloud storage. When the copy check is done, the client performs record level duplication remove repetitive document or record in cloud storage. Hashing algorithm has been proposed impose data discretion while making de-duplication is possible. Every record has secured by encryption and decryption keys. Each record has different encryption and decryption key using hashing algorithm.

If data copy has encryption key got by computing the hash value for content of the data. After key generation, the encryption has sent to the user. In S-CSP gives information for data user and administrator. Only valid user can access its own file for downloading using private keys.

To secure by unauthorized users using proof of ownership algorithm is also need to clarify authenticated process for user really owns same file when a duplicate is found. After the authentication process is done the subsequent user with same file will be uploaded to the server which indicates the duplication file.

This method can perform multiple checking tasks concurrently and improve the effectiveness of proof for multiple checking tasks.

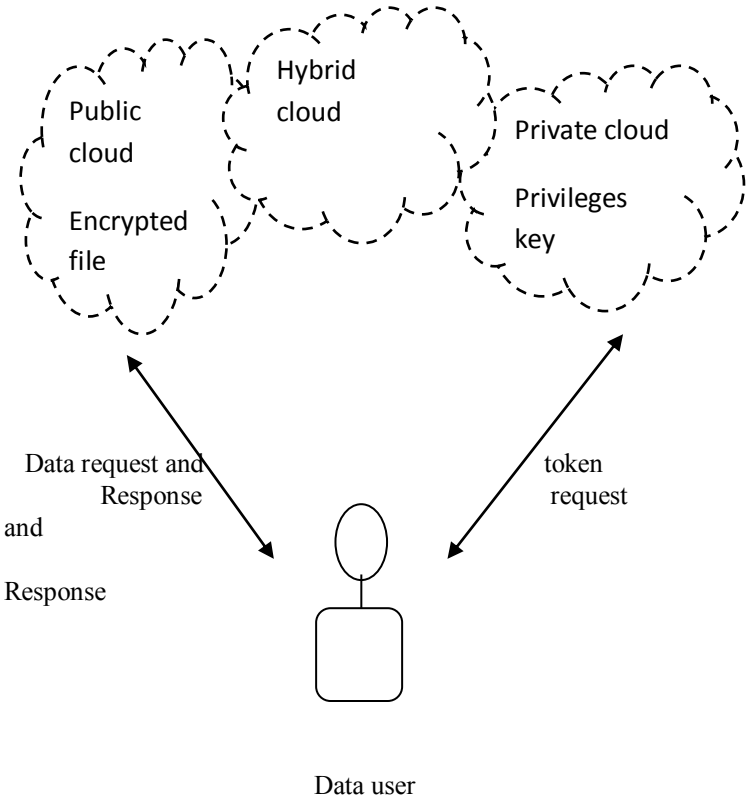

FIGURE 1 PROPOSED SYSTEM ARCHITECTURE

Hybrid Architecture for Secure De-duplication

Hybrid cloud combines private and public cloud. The S-CSP achieves both information provided by data user and de-duplications checking. Duplication checking has find duplication two records are the same in cloud. The S-CSP has permitted clients to private cloud server, an unknown client which will help encryption key for each document. In duplication method to stored data in S-CSP are consisted as collection of joined client. To set data backup and recovery application mainly used for reduced storage space.

In another word, refer a document level duplication to eliminate duplication in entire record and also to kill the volume of repetitive document. In exact, to transmission a document, client first pieces out the record level copy check. In the occurrence that the document is a copy, every one of its pieces must be the replicas, something else, the client further plays out the square level copy check and know the one of a kind places to be transferred. Every information duplicate is associated with a token for the copy check. This is an element that gives an information storing administration out in the open cloud to remove vast amount of data. There are three modules in our proposed system those are,

- User
- Private cloud
- S-CSP

A. Data Users

Data user is one of important factor in cloud storage .In this data owner uploads their document into cloud server. To solve security issue data user can encrypts files and then stored cloud server. That data user can check duplication document over cloud server.

The Data owner can have accomplished of working the cipher text data for document and the data owner can check the several cloud data find duplication in specific file. They can create remote user with admiration to registered cloud servers. And also data owner has travel to another cloud option, by this he can travel files from one cloud server to another cloud server.

B. Private Cloud

Differentiate and expected de-duplication aim in distributed computing is removal replica duplication in document or record for the use minimizes storage and bandwidth. The public cloud act as interface between data user and accessible people in general cloud.
The private clouds for the remuneration are supervised by supervised by private cloud, who answer query for submitted document and it accessible by the private cloud authorization client to submit documents and queries to be securely put away and handled separately.

C. S-CSP (storage cloud service provider)

Cloud service Provider is an entity that provides network services to the cloud. This cloud service providers are host to access by individual or group of data center using network connectivity. It also offer periodic backup and recovery is achieving. The S-CSP stores data on behalf of the user. The S-CSP used for reduce storage space by removing redundant data.

Advantages

- It makes overhead to minimal for file uploading operation.
- Strong security compared to the existing technique.
- Malicious node can be avoided.

V. RESULTS

The final results get detailed from uploading, downloading, check duplication, detecting the sign using hashing algorithm. From those results complete process of the proposed system is given. The output images as well as performance graphs are given below,

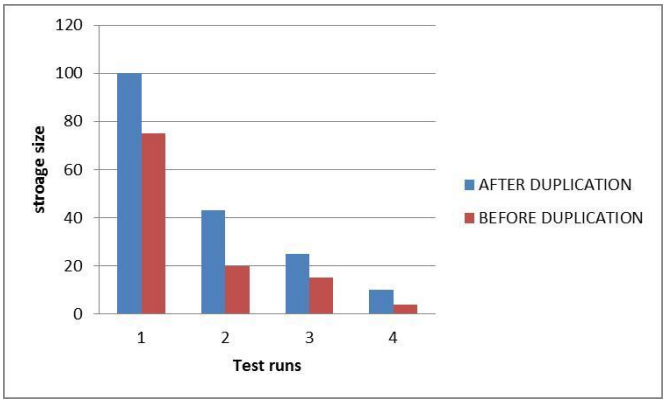

FIGURE 2 TIMES TAKEN FOR DUPLICATION

VI. CONCLUSION

The document level de-duplication was ensuring

information security for document of clients. When the copy check is done and then documents are encryption and decryption achieve by different keys. The encryption key is stored in private cloud and decryption keys are stored in public cloud. This private keys are only accessible by authenticate user. Our plans avoid attacks from insider and outsider attack and negligible overhead.

VII. FUTURE ENHANCEMENT

Cloud De-duplication mainly focuses storing and retrieval information. We plan to define extra distinctive process such as delete and edit for upload files. And also we boot efficiency. In our existing system provided security issues in file accessing. So, they are adding anomaly algorithm. In our future work, we focus on access control for authenticate user for handle its own file and also achieve with editing and deleting operation. Then our proposed method avoids attacks from malicious user.

REFERENCES

[1] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. *In Proc. of SENIX LISA*, 2010.

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided encryption for de-duplicated storage," in Proc. 22nd *USENIX Conf. Sec. Symp.* , pp. 179–194, 2013.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message- locked encryption and secure de-duplication," *in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn,* pp. 296–312, 2013.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided encryption for de-duplicated storage," *in Proc. 22nd USENIX Conf. Sec. Symp.,* pp. 179–194, 2013.

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message- locked encryption and secure de-duplication," *in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn,* pp. 296–312, 2013.

[6] M. Bellare, C. Namprempre, and G. Neven, "Security proofs for identity-based identification and signature schemes," *J. Crypto. vol. 22,* no. 1, pp. 1– 61, 2009.

[7] Bugiel,S Nurnberger, ASadeghian, T.Schneider, "Twin clouds: An architecture for secure cloud computing,"*inProc.Workshop cryptography Security Clouds*, 2011.

[8] A D. Ferraiolo and R. Kuhn, "Role-based access controls," *in Proc. 15th NIST-NCSC Nat. Comput. Security Conf.,* pp. 554–563, 1992.

[9] Waraporn Leesakul, Paul Townend, Jie Xu" Dynamic Data De-duplication in Cloud Storage" *IEEE 8th International Symposium on Service Oriented System Engineering* 2014.

[10] C. Ng and P. Lee, "Revdedup: A reverse de-duplication storage system optimized for reads to latest backups," *in Proc. 4t Asia-Pacific Workshop Syst*., Apr. 2013.

[11] W. K. Ng, Y. Wen, and H. Zhu, "Private data de-duplication protocols in cloud storage," *in Proc. 27th Annu. ACM Symp. Appl. Comput.,* pp. 441–446, 2012.

[12] J. Xu, E.-C. Chang, and J. Zhou, "Weak leakage-resilient client side de-duplication of encrypted data in cloud storage," *in Proc. 8th ACM SIGSAC Symp. Inform, Comput. Commun. Security,* pp. 195–206,2013

[13] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with de-duplication," *IACR Cryptology ePrint Archive,* 2013.

[14] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy-aware data intensive computing on hybrid clouds," *in Proc.18th ACM Conf. Comput. Commun. Security*, pp. 515–526, 2011.

[15] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: The least-authority file system," *in Proc. ACM 4th ACM Int. Workshop Storage Security Survivability*, pp. 21–26, 2008.

[16] Quinlan, Sean, and Sean Dorward. "Venti: A New Approach to Archival Storage." *In FAST, vol. 2, pp. 89-101*. 2002.

[17] R. D. Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for de-duplication," *in Proc. ACM Symp. Inf., Comput. Commun. Security,* pp. 81–82, 2012.

[18] Li, J., Chen, X., Li, M., Li, J., Lee, P.P. and Lou, W., "Secure de-duplication with efficient and reliable convergent key management". *IEEE transactions on parallel and distributed systems*, pp.1615-1625, 2014.

[19] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data de-duplication scheme for cloud storage," *Tech. Rep. IBM Research, Zurich, ZUR 1308-022,* 2013

[20] *Proc. 3rd ACM Workshop Secutiry Cloud Comput.*, pp. 169–177, 2011.
[21] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data de-duplication," *in Proc. 4th ACM Int. Workshop Storage Security Survivability,* pp. 1–10, 2008.

[22] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: The least-authority file system," *in Proc. ACM 4th ACM Int. Workshop Storage Security Survivability,*pp.21.26,2008.