ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

A Survey on Data Confidentiality using Hybrid Cloud for Secure Data Deduplication

Swapnali Patil[1], Prof. Vikas B. Maral[2]

Student, Department of CE, K J College of Engg. &Management Research, Savitribai Phule Pune University, India.[1]

Professor, Department of CE, K J College of Engg. &Management Research, Savitribai Phule Pune University, India.[2]

ABSTRACT: Today, cloud is gradually growing because of storage, sharing, restoration. Cloud storage allow us to store our data on remote server. As this type of storage require more resources hence costly, so it is important to limit the storage requirement. This leads to reduce the data duplication. More importantly user always prefer to store confidential data on cloud, so storage sever can not manipulate those files. Here data deduplication comes into the picture, which mainly focuses on removing the duplicate data. This will help on freeing the memory space and bandwidth. This method considers the authorisation on files given by the user and act respectively, each user has given unique token ID. As per research previous schemes did not provided full proof security mechanism in cloud storage. So taking this into the consideration we propose the hybrid data deduplication mechanism which will also provide improved security.

KEYWORDS: Data deduplication, Convergent key encryption, Data security, Proof of Ownership.

I. INTRODUCTION

Deduplication has been well-known technique to make cloud storage data management which gives more attention currently. Deduplication is used for storage management by eradicating duplicate data. Uncommon data is actually maintained on cloud.Inessential data is removed by using the pointer to unique data. Shortly data deduplication is method by which data is compressed for eliminating identical copies of the data in storage system. Example is email scheme might comprise ten copies of look a like 1Mb attachments. If data is not compressed then all ten copies are stored, demanding 10Mb memory space although using deduplication 1 copy of attachment is in reality stored; all other consequent sequence is pointed back to the one saved copy. Here 10Mb demand could be reduced to only 1Mb. On the contrary that, keeping multiple copies of reduplicating data, deduplication wipe out inessential data by putting one natural data copy. Deduplication viewed at three levels: file and block level and byte level, In file level, it eliminates duplicate copies of files. Block level will eliminates indistingushable blocks which can be present in different files[16].

Despite the fact, data compression import more advantages, data armour and isolation concerns come to light as sensitive information is impressionable to inside attack and out-side attacks. Classical data encryption, while presenting data confidentiality, is inappropriate with deduplication. Especially widely used encryption wish to distinct users to encode data with intrinsic keys. In this way, matching copies of distinctive users to encrypt, cause dissimilar cipher texts, forming deduplication unbearable. Convergent encryption designed to impose confidentiality during the time of constructing deduplication workable. It encodes data with convergent key, which can be retrieved by cryptographic hash value of data. User retains the key after key preparation and encryption and send encrypted text to cloud. As encryption action is pre-determined and extracted from the data, look a like data accomplishes the matching convergent key hence the alike cipher text. To restrict unofficial eruption, a impregnable POW protocol is required to afford the validate user easily retains the correlative file when a identical is found. Consequently, consecutive users will be provided a reference from the server without requiring to transfer the same file. From the pointer to the file, user downloads encoded file from the public cloud, which is decoded by correlative content holder with convergent keys.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Hence convergent encryptions concedes cloud to function deduplication on encoded text and the POW avoids the unofficial user to burst the file[7].

II. RELATED WORK

ChunI Fan, ShiYuan Huang, and WenChe Hsu [1], IEEE proposed the work in which they focused the on dropbox remote cloud storage. Using dropbox users upload/download the data from any device from anywhere after connecting to the cloud. In this paper author concentrated on data deduplication technique for keeping redundant copies of data on dropbox so that it will reduce consumption of storage space of dropbox. To provide solution for limited space of dropbox they proposed the system which manages storage space efficiently when there are frequently used files. They have provided security on cloud storage, data deduplication for bandwidth consumption, Also handled conflict between data deduplication and data security by providing convergent key encryption.

Jingwei-Li, Jin-Li, Dongqing-Xie and Zhang-Cai[2], In this paper author described problem arrived in the secure cloud system such as capability of analysing the cloud data by client and finding the duplicate files by cloud server. For this problem they proposed the solution and developed the system using the Sec-Cloud and Sec-Cloud+. Sec-Cloud introduces the concept of auditing body with mapReduce and generate data tags while uploading the files also audit that data. Along with this, customer encode data before uploading, SecCloud+ is used for allowing the auditing and applying deduplication on encoded data.

Pierre Meye, Philippe Raipin, Fre de ric Tronel, Emmanuelle Anceaume[3], In this paper, author described the security concerns for the systems which are performing inside user and client-side deduplication. They proposed model results in saving of bandwidth and reduced memory. This paper will describe deduplication and confidentiality along with its issues and backup time observation. The aim of the paper is to save both cloud space and bandwidth, for secure storage service. They consider file level deduplication, besides their system can be enlarged to the block level. They have proposed approach that combines intra user and inter user deduplication style by suggesting deduplication agent in between client and storage server.

N. Jayapandian, Dr. Zubair Rahman, I. Nandhini [4], IEEE proposed work to protect the data confidentiality the authorized deduplication was proposed by divergent privileges of users. Duplicate check tokens are created with private keys in private cloud. They provide deduplication constructions bearing validate duplicate check in hybrid cloud construction. In proposed model, security of insider and outsider attacks are secured by security analysis which provides the protection to their schemes. As a proof, they performed a prototype of their suggested authorised duplicate check plan and run tested demonstrations on their prototype. They manifest that their system incurs small overhead estimated to other encryption and network transfer.

N. Lakshmi Pritha, N.Velmurugan, Dr. S.Godfrey Winster, A.Vijayaraj[5], In this paper author proposed the approach for development of secure cloud access system for data storing and confidential data sharing. For that author proposed the use of hybrid cloud architecture. They have provided the solution to secure system that use of secrete keys for the files, But as number of files increases it will increase the number of secrete keys. They have proposed the key-aggregate crypto system, in that many secrete keys can be combined to get a single secrete key. For sharing the file user can create the secrete key so that while accessing that file it need this secrete key. They proposed work by considering, user encode the data before uploading and validating for analyse and data deduplication on encrypted data.

Ramanjaneya Reddy D , Assoc. Prof Swathi Y[6], proposed the work for data integrity and data deduplication for data on cloud. They have proposed two main modules called SecCloud and SecCloud+. In that SecCloud is used for determining the entity by considering the MapReduce cloud. It will create information tags and check purity of information and save both the data on cloud. Along with this SecCloud will provide Proof Of Ownership for secure deduplication.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

III. PROPOSED SYSTEM

Our focus of the system is on enterprise network, which includes category of related users that uses he Secure CSP with Data Deduplication technique. Along with reduction of storage space, backup data and application of disaster recovery are also achievable. The three main modules of this system are private cloud, S-CSP in public cloud and end users as shown in fig 1. This deduplication can be performed by Secure CSP by comparing data contents of two files and if system founds same data content it will keep only one of them.

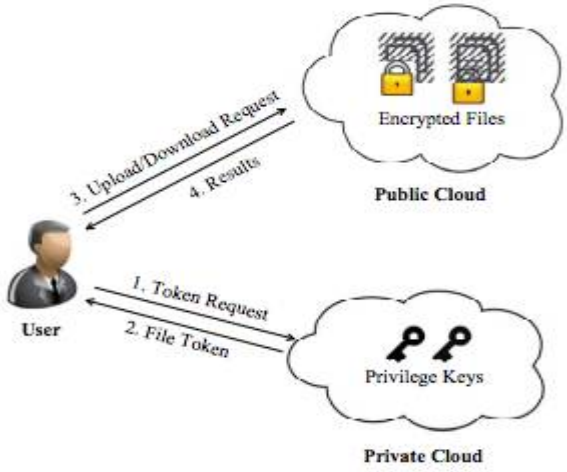

Fig 1. System Architecture for Authorized Data Deduplication

In this system, when the user wants to transfer the file on the shared cloud at that time, user have to secure the file with the one time key sent by private cloud and after that user will transfer data to cloud shared among multiple users. For data security purpose actual data and secret key(HashTags of file) will be on different clouds, secret keys will be stored on private cloud only. We are using an algorithm for deduplication on public or shared cloud to minimize the consumption of bandwidth, it will result in avoiding duplicate copies of data. Whenever user wish to download the uploaded file he needs to send request to public cloud. User will select the file to download from list of files provided by public cloud. As soon as user selects the file to download private cloud sends message to for entering private key. At the same and private cloud will send message one time key to the authorised user through message. User will enter that one time key. Private cloud will check whether the user is valid or not. If the user is authorised person then, private cloud gives entry to download the file. After downloading the file from public cloud user will decode the file with the help of same convergent key used for encrypting file. This is use of data deduplication architecture.

IV. PSEUDO ALGORITHM

Here, we use two types of algorithms, Uploading File and Downloading File from server.

Algorithm for File Uploading file

- Select the file to upload
- Encrypt the file using one time key
- Generate the Hash Tag for the file
- Send generated HashTag to private server
- Check if file exists on public server if exits, display message that file exits on server
- Otherwise encrypt the file using hashTag and upload encrypted file to public server,
- Save the HashValue of uploaded file on private server

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

• Display success message

Algorithm for downloading the file
- Get list of files from public server
- Select the file to download
- Download the file from server after verifying identity by entering one time key
- If user is not valid is not allowed to download the file
- Decrypt the file file using HashTag

V. PROPOSED RESULT

Data security is very important thing in now-a-days. Data Deduplication in hybrid cloud will provide data security along with storage management. In the proposed result, we will improve the data redundancy and data security over existing approaches. And it will provide great extent of data confidentiality and efficient use of storage space.

Table I describes the time required in ms for uploading the file. Proposed system will be more efficient and time required to upload or download the file is depends on network speed.

TABLE I
TIME EFFICIENCY COMPARISON

	Existing System	Proposed System
Time in msec	2000	1000

Figure 2 represent the graphical comparison of time efficiency in existing and proposed system respectively. X axis represent the systems and y axis represent the memory required in KB. As the system detects the duplicate copies of data it will save the cloud storage space.

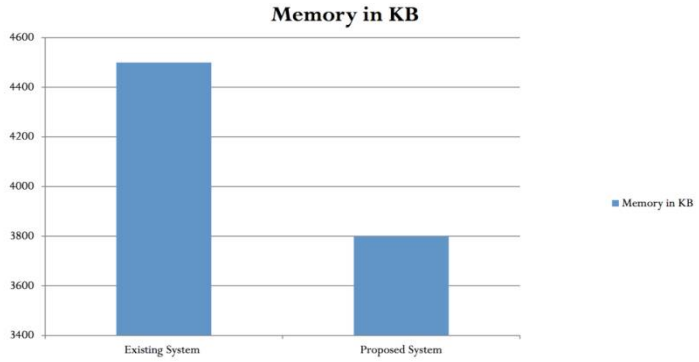

Figure 2 Expected Result of Proposed System

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

VI. CONCLUSION AND FUTURE WORK

Here, the use of hybrid cloud utilization was suggested for deduplication and to protect the data from public cloud. Shortly the aim is to achieve data protection security by one time key to the user. On public cloud information is safely stored in encoded format, as well as private cloud stores corresponding files key and. No need to recall the key. Any authorized user can send request for accessing the file. Using authorized deduplication and ordered access control method, More efficiency is provided in the cloud computing. This paper explains how storage efficiency and performance of cloud is improved.

REFERENCES

[1] Chun-I Fan, Shi-Yuan Huang, and Wen-Che Hsu , IEEE proposed the work on "Hybrid Data Deduplication in Cloud Environment", 2012
[2] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, "Secure Auditing and Deduplicating Data in Cloud ", 2012
[3] Pierre Meye, Philippe Raipin, Fre de ric Tronel, Emmanuelle Anceaume, "A secure two-phase data deduplication scheme", 2014
[4] N.Jayapandian, Dr.A.M.J.Md.Zubair Rahman, I.Nandhini [2015], IEEE "A Novel Approach for Handling Sensitive Data with Deduplication Method in Hybrid Cloud", 2015
[5] N. Lakshmi Pritha, N.Velmurugan, Dr. S.Godfrey Winster, A.Vijayaraj [2015], "Deduplication Based Storage and Retrieval of Data from Cloud Environment", 2015
[6] Ramanjaneya Reddy D , Assoc. Prof Swathi Y [1], proposed the work on "SecCloud and SecCloud+ in Secure Auditing and Deduplicating data in Cloud", 2016
[7] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, "Hybrid Cloud Approach for Secure Authorized Deduplication, Parallel and Distributed Systems", IEEE Transactions on, 2014
[8] Jin Li, Xiaofeng Chen, M Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management", IEEE transactions on parallel and distributed systems, 2014
[9] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. "Proofs of ownership in remote storage systems". ACM Conference on Computer and Communications Security,2011
[10] JinLi,YanKitLi,XiaofengChen,PatrickP.C.Lee,WenjingLou, "A hybrid cloud Approach for Secure Authorized Dedplication". 2014, IEEE
[11] https://www.starwindsoftware.com/file-level-vs-block-level-vsbyte-level- deduplication.
[12] G.Prashanthi, Z.Shobarani, "A hybrid cloud Approach for Secure Au-thorized Dedplication", 2015
[13] P.AndersonandL.Zhang. "Fast and secure laptop backups with encrypted de-duplication". In Proc. of USENIX LISA.
[14] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium", 2013.
[15] OpenSSL Project. http://www.openssl.org/.
[16] http://searchstorage.techtarget.com/definition/data-deduplication
[17] Hema Sable, Savita R.Bhosale, "Secure cloud data deduplication for remote storage", 2015

BIOGRAPHY

Swapnali Sanjay Patil is a student of Computer Engineering Department, K J College of Engg. &Management Research , Savitribai Phule Pune University. She received Bachelor of Engg. (BE) degree in 2014 from Pune University. Now she is pursuing Master's Degree (ME) from KJCOEMR.

Prof. Vikas B Maral is a professor in Computer Engineering Department, K J College of Engg. &Management Research , Savitribai Phule Pune University.