ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

AVOIDING DE-DUPLICATION BY KEY GENERATION USING HYBRID CLOUD

[1]Dharambir Kumar, [2]C.Deepa
[1]Student, Faculty of Computing, Sathyabama University.
[2]Professor, Faculty of Computing, Sathyabama University, Chennai-600119, Tamil Nadu, India.
Email: dharmbir.ksingh60@gmail.com

Received on 18-02-2016 *Accepted on 25-03-2016*

Abstract

This paper is about the Avoiding De-Duplication by Key Generation Using Hybrid Cloud. The Hybrid cloud architecture is a combination of public and private clouds which bound together by standardized technology that enables data and application portability to efficiently solve the problem of de-duplication with differential privileges access permission in cloud computing.

Hybrid cloud is a Compound of two clouds First public cloud and second private cloud. Cloud services give the permission particular and businesses to use software and hardware that is handling by third parties at remote locations, social networking sites, webmail and online business applications. The cloud computing model gives permission to access the information from cloud server and computer resources from anywhere if network connection is available.

Cloud computing provides pool of resources, including data storage applications. Based on the definitions specified in the cloud computing, each file is compared with the database of cloud storage. After comparing, if a match is found in the cloud database then only avoid the de-duplicate data of the file in cloud server database. In this paper it also block unauthorized access by using a secure proof of ownership protocol, once user registered then can able to login.

The protocol uses hybrid cloud architecture to verify . Thus, prevention is achieved by de-duplication of data and security of data. Once data is secure, after login process one can login the user else not. The proposed system has been comparing with the existing system on the basis of cloud database usage, security and bandwidth.

Keywords: File level de-duplication, authorized de-duplicate verify, confidentiality, hybrid cloud, Proof of ownership.

I. Introduction:

Cloud computing that means on demand delivery of the IT resources via the internet to the user as per needed. In the centralized server room, there to be need of database server for storing data, mail server for mailing, networking for communication, firewalls for security, routers for connectivity, modem for same, switches, to setup such types of IT infrastructure; we have to need to spend lots of money on the server room. There have to need spend lots of money for server room.[9,11,21]. Management application is growing very fast for the service of providing cloud services to apply in complex group of cloud users. The data to be transmitted to the group recipients in the concept of cloud so that provide to the cost for cloud services. Means to say that the valid group subscribers are using the information frequently.[1,3,6] Several commercial applications like via net, and group user that have to allow the authorized subscribers to use the data. Now in this time cloud computing are mostly used for storage purpose, so the several users for authentication oriented applications that want to provides easily services in group like software as a services(SaaS), platform as a service(PaaS), and infrastructure as a services(IaaS). That grows rapidly.

For the file level duplication which eliminates duplicates file of data that occur in not similar files. [2][4] It means encrypts and decrypts a data copy with a secret key, the content of the data copy getting by computing the cryptographic hash value of key generation method is used, in which way cost is very high, because of all data are store on central server. So, have to need to connect to the server and then gain appropriate access, cannot able to perform any single task any user without administration permission, when having needed. Similar data copies are not allowed to upload. The secure proof of an ownership protocol is used to prevent the unauthorized users to access the data and also provide the proof to user regarding the duplicate data is found in the same file. [6][7] In cloud computing system, there is an important workload shift its cannot able to perform task for larger area. A local computer does not have capacity to do all the heavy work when it comes to run applications. But cloud computer can handle heavy load easily and automatically, it's providing large number of facility to the single user.

II. Related work:

I am presenting hybrid cloud approach for secure authorized de-duplication. The encryption for de-duplicated storage for cloud file storage, in which using two algorithms for avoiding de-duplication, first is symmetric key encryption method and second is hash based encryption for more security purpose, social networking sites, and others perform de-duplication to save space by only storing one copy of each file uploaded data in cloud server database. The first building block of file is the infrastructure where the cloud will be implemented in cloud server database. I am using

public key encryption that will provide more security, and if you have solutions to provide on related software. [16, 17, 22].

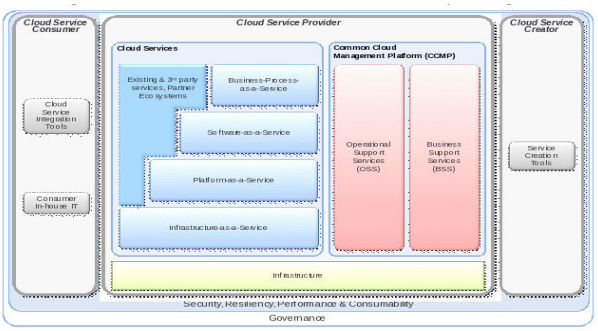

Fig: 1 cloud layer Architecture.

Above diagram showing the services of cloud computing which is in layers .Cloud computing is a new IT delivery model accessed via the internet. It is definitely not formed in one day or day by day. Now a day this is perform big role in setup IT Company infrastructure.

There have a special case in which all users take equal possibility, that's not practical thinking. Because in real world several kinds of users in computer background; means some users are short period and some users are long period members so all need to have software. This technique is used for reduce the de-duplication storage space in cloud database and bandwidth also provide the confidentiality [4][10].Twin clouds: it means the combination of public cloud and private cloud that is called hybrid cloud Architecture for security proposed Client uses the trusted Cloud manage by the outsourced data, programs, and queries.

III. System overview:

A. Problem Statements:

The file level de-duplication method have emerge new storage overhead and less efficiency. To overcome this problem, Authorized Duplicate System is developed which avoid the duplicate copies of data and reduces space in cloud used for storage as well as data overhead in cloud storage. It also protects data and maintains confidentiality of sensitive data. [20], In order suppose a user wants to download a file called f .It first need send a request for file which is store in cloud server database and the file name found in cloud database then storage service provider (SSP) will check whether the user is eligible to download a file or not, it means user is authorized user or not, if authorized then users are able to download the files else not.[23] To avoid a wrong transaction the user will not receive the file name directly instead of the file name the key will be send to the storage service provider (SSP) for security. But there is a chance to get duplicate key same files if the key is generated two times. Because there is generated two key for one file.[14]

Existing Technique: Storage Cloud-Service provider S-CSP is used for saving the data in cloud server. [19]

Definition:

The data will be encrypted in my de-duplication system before outsourcing to the storage service provider (SSP). But it is not more secured for avoiding the duplicates, because data are storing in public cloud and key also in same cloud.

Drawback:

Each user will be issued private keys for security, because key store in public cloud, but there is a chance to get a duplicate key from the cloud database to open the same file that is store in public cloud server. [8, 13, 15, 17, 21]

- Duplicate key are easily generated for the one data which is stored in the database or cloud database server of public cloud.[16]
- There is chance of hacking the data from the cloud when the file is opened because we can't give the second proof to confirm the verification that is called private key in private cloud server. But I am using here for security purpose two algorithms that's why data is more secure.[18,21,23]
- Here there is no guarantee for the data security even though there is a one key for the single file. The key and the data will be stored in the same cloud database so inconsistent result will be occurred in cloud database server.

IV. Proposed system design:

To solve the file redundancy problem in the cloud, the de-duplication system we have introduced with hybrid cloud architecture. In this system, user can not able to share these private keys to any unauthorized user. To get the key for the file, users need to send a request to the private cloud server for private key. The final authorization for the duplicate will be checks from the public cloud before uploading the file, and generating private key for that file.

Proposed Technique: - Symmetric Encryption for secure de-duplication system in cloud server database.

Definition:

Symmetric encryption uses a common secret key *x* to encrypt and decrypt data and send it to public cloud and key will be generated in private cloud.

Advantage:

There is no chance get a duplicate key for the same file and no other authorize user will open the file because key is stored in the private cloud server using the Hybrid cloud server.

- There is no redundancy key for a single data. Only one key will be generated for the single file.

- If the file is ready to open, first it needs to run in storage service provider (SSP) to prove the file-ownership it means user is authorize or not if once user is authorize then have permission to do. If verification is done then the user will get the key token for the file to open from public cloud server database.
- The data is secured after using the hybrid cloud because it store data in public cloud server database and the key is stored in private cloud server.
- The key will be generated and stored in the private cloud database server and the data will be uploaded/downloaded from the public cloud server database. So there is no possibility of getting the inconsistent result while we are retrieving the data from cloud database server.

Authorized Duplication System:

This system uses private cloud and storage cloud service provider(S-CSP) in public cloud for storing data. The storage cloud service provider(S-CSP) accomplishes de-duplication by checking the private key in private cloud, if the contents of two files are identical then only one of them is stores in cloud database server.[15][13] The access right file is describing based on a set of privileges Access permission. The accurate definition of a privilege means Authorized users. File Token means each privilege is represented in the form of a short message, like when I upload/download file from cloud, each file is related with some file tokens, which denotes the File tag with specified privileges. A user computes as well as sends duplicate check tokens for file to the public cloud for authorized duplicate check token[13][7].

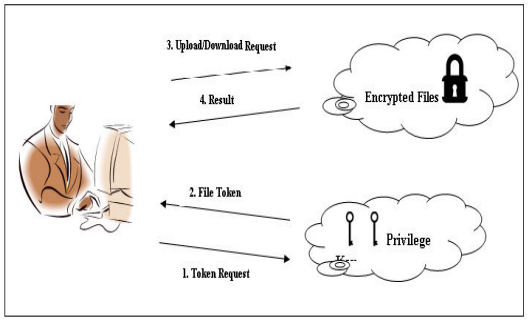

Fig.2 Authorized Duplicate system for Cloud.

Encryption of Files:

Here we are using the common secret key x for encrypt as well as decrypt data. It will use to convert the plain text to cipher text and again convert in to the cipher text to plain text. Here we have used three basic functions,

Key Gen SE: the key generation algorithm is x, that generates *x* using security parameter and storing key in private cloud.

EncSE (x, *M*): the symmetric encryption algorithm is d, that takes the secret *x* and message *M* and then outputs the cipher text *d.*

DecSE (x, *d*): the symmetric decryption algorithm is M, that takes the secret *x* and cipher text *d* and then outputs the original message *M*. [27]

Hash Based Encryption :(md5)

In this authenticate users and MD5 hashing for hiding information from user and generating key. This model is used security for whole cloud computing structure. It provides data confidentiality in de-duplication. The user derives a private key from each original data copy and encrypts the data copy with the private key it means secure key. [18, 23]

Symmetric Key Encryption: Symmetric encryption uses a common secret key x to encrypt and decrypt information from cloud in both ways. A symmetric encryption scheme consists of three basic primitive functions: KeyGenSE (1λ) = x it is the key generation algorithm that generates x using security parameter 1λ.

ncSE(x,M)=the symmetric encryption algorithm is d, that takes the secret x and message M and then out- puts the cipher text d.DecSE(x, C) = the symmetric decryption algorithm is M, that takes the secret x and cipher text d and then out- puts the original message M. [24, 25, 26]

Security Model for Cloud Storage:

When user uploaded file to cloud storage database, data store in public cloud and key will generated in private cloud server. it divides into multiple blocks .Security service received data and start encryption using advanced encryption standards. After encryption, generate token for unique identification of block in public cloud. [22, 24]

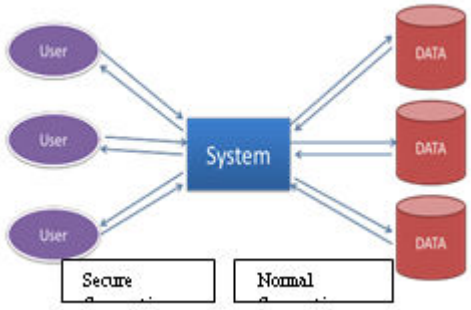

Fig.3 Proposed Security Model.

Tester for validation of authorized user.

Security service executer for checking duplicate copies over cloud server.

Block level encryption which provides the registration and further details to fig.4. user regarding duplication system

in cloud.

1. Read the data form uploaded request that file is already in uploaded in the cloud.
2. The generated key is store in private cloud that's why data is more secure in cloud server.
3. Security Service receives the file and performs encryption using algorithms.
4. Each file is generating key, cipher text and private key.
5. In security services, the hash table used to maintain the sequence of Files and gives the original file because file store with private key in cloud server.

Token Generation Algorithm for secure storage:

The key generation algorithm used for generating the key for uniquely identification of file and maintain the proper sequence of the storage file block at the time of downloading the given file. The following Fig.4 show the steps of generating key for file.

Input: File as a input

a. Web browser client request for key to private cloud server.
b. Web services validate key for private cloud server.
c. Return key to web server.
d. Web client got key as a for Output **downloading file.**

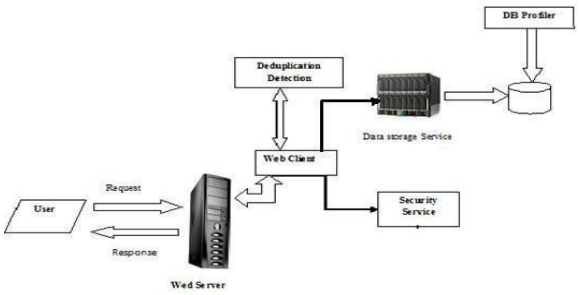

Fig – 4

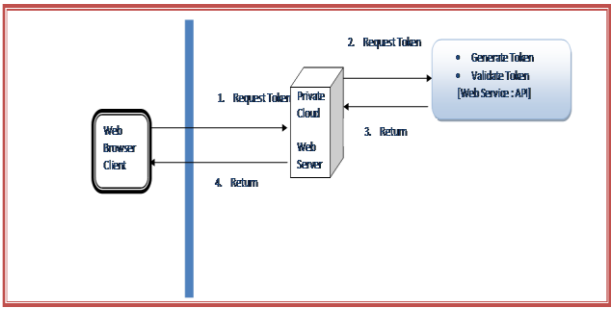

Fig. 5 key Generation for Security.

User Authentication :(proof of ownership)

To support authorized de-duplication, first need proof for the file owner it means authorized or not. The key of a file *x* will be determined by the file *x* and the privilege. To show the difference with traditional notation of key generation

technique, I call it file by the key instead. To support authorized access, a secret key will be bounded with a privilege *p* to generate a key for the file. [22, 23, 24]

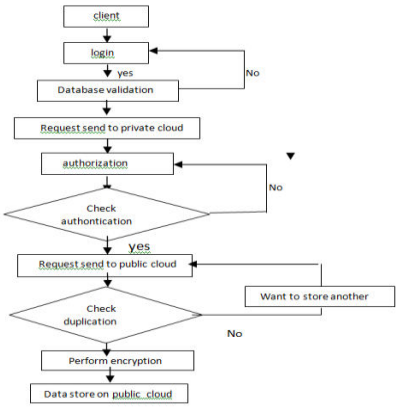

Fig: 6: User Authincation.

Mathematical Model Statement

Let x be a key system that find out duplicate copies of the file using Authorized de-duplication system in hybrid cloud.

S=[C, B, T, P, O].

CBi=Set of cipher text block for file.

T = Token [16-Bit unique key for file].

P=Private Key used for encryption & description both purpose.

O=Output consist reduce cloud server database size.

ALGORITHMS USED FOR UPLOADING/ DOWNLOADING:

In this section, I am use two

Types of algorithms,

a). For file uploading.

b). for file downloading.

A). FOR UPLOADING A

FILE:

BEGIN

Step –1 Read the file Uploading.

Step –2 Cloud server checks File for

Duplication in cloud server database.

Step –3 Sends de-duplication response whether the file already found or not on the cloud server database.

Step – 4 if the file is does not found

On the cloud server database Display “(file does not exist)”.

Step – 5 Then the uploads file on cloud

server database.

Step – 6 if file is already found on cloud server then Display “(file already exist)”.

End

B). For Downloading A File:

BEGIN

Step –1 Read file Downloading file.

Step –2 Cloud server checks for de-
Duplication.

Step –3 Sends de-duplication response
Whether the file already found or not on cloud server database.

Step –4 if file is found on cloud server
Database Display "(file
Exist)".

Step –5 then the downloads file from
Cloud server database.

Step –6 if file is does not found on cloud server then Display "(file does not exist)".

End

A. REMOTE USER MODULE :

a. Remote User login validations.

b. Accessing Files for Remote User.

B. CLOUD SERVER MODULE:

a. Authorized Duplicate Check.

b. Accessing Files for Cloud.

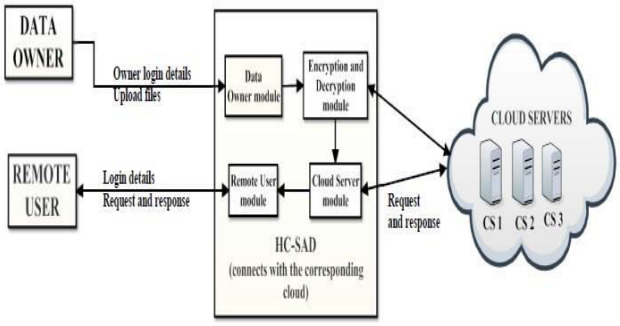

Fig-7: System Architecture design for cloud.

V. Experimental Results: The authorized de-duplication system used to avoid duplicate copies of data in the Hybrid cloud. Proposed system implemented by using key duplication which compare the uploaded data in cloud database server, suppose the file is stored in the cloud database if the same file uploaded by another user at that time only data about file will store in cloud server database actually, so it reduce the storage space of data in cloud database and proper utilization of space. The data will be store in encrypted format in cloud database so it also maintains security because each files having secret key, their own cipher text and private key. The cloud server database size will be

reduced by using this technique. This proposed system has been compared with the existing system on the basis of database usage and overhead, and security using proof of ownership.

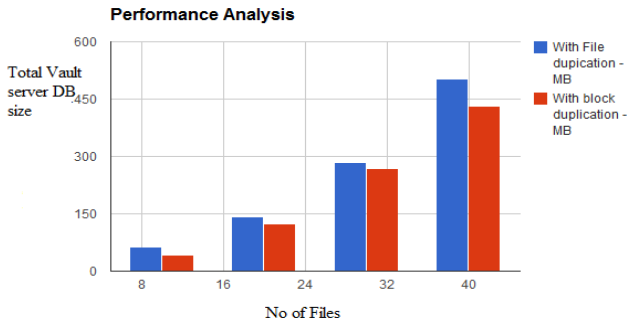

Fig.8: Comparison between Existing system de- duplication and purpose system de-duplication.

	Numb	File	Block level
1	8	80	50
2	16	12	70
3	24	14	110
4	32	28	210
5	40	49	308
6	48	510	410

Table. A: Actual Result Comparison

The above Table A. Shows the database usage for file level de-duplication using hybrid cloud. The file level de-duplication having extra storage space as compare to de-duplication. The file level de-duplication having less storage space and it is also provide extra security with the help of proof of ownership concept that is mainly used for authentication purpose.

VI. Conclusions:

In this paper, the investigation is based on official data de-duplication concepts have been proposed to protect the security of data with de-duplicate control of users with different privileges using key. Secure de-duplication occurs with the help of key generation and secure upload/download of file. It conform the user about high data security and also avoids data de-duplication in hybrid cloud storage. This helps in eliminating duplicate copies of repeating data, reduces storage space in cloud server database used and saves bandwidth in cloud storage. Convergent Encryption protects the confidentiality of the user data and the dynamic cloud background users, to give an enhanced optimal cloud model whereas old one is canalize server concepts. In this paper there is no any concept of assumptions so it does not follow the concept where the users does not have the same exit possibilities and also the concept of the cloud.

VII. References:

1. Jin Li and Yan Kit Li ˙A Hybrid cloud approach for secure authorized deduplication, IEEE Transaction on parallel and distributed system, 2014.
2. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
3. J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication .IACR Cryptology Print Archive, 2013.
4. 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.
5. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013
6. J. Xu, E.-C. Chang and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In ASIACCS,pages 195–206, 2013.
7. C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS,Apr 2013.
8. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.
9. 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 and Communications Security2012.
10. 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.
11. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer andCommunications Security, pages 491–500. ACM, 2011
12. K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan. Sedic: privacy aware data intensive computing on hybrid clouds. In Proceedingsof the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM

13. A. Rahumed , H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011

14. M. Bellare, C. Namprempre , and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 2009.

15. M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177,2002

16. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou" A Hybrid Cloud Approach for Secure Authorized De-duplication" in vol: pp no-99, IEEE, 2015

17. OpenSSL Project. http://www.openssl.org/.

18. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.

19. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.

20. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message- locked encryption and secure eduplication. In *EUROCRYPT*, pages 296– 312, 2013.

21. Bellare, M., Keelveedhi, S., Ristenpart, T.(*2013)* Message- locked encryption and secure deduplication, *In: Advances in Cryptology.*

22. Xu, J., Chang, E.C., Zhou, J., Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. *In: 8th*

23. Ronald Rivest, "MD5 Message-Digest Algorithm", rfc 1321, April 1992.

24. http://www.c-sharpcorner.com/UploadFile/shashijeevan/PublicKeyTokenGenerato08302005015104AM/ PublicKeyTokenGenerato.aspx.

Corresponding Author:

Dharambir Kumar*,

Email: *dharmbir.ksingh60@gmail.com*