ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Secure and Authorized Data Deduplication Using Hybrid Cloud Technique

Miss.Supriya Jadhav[1], Prof. S. B. Chaudhari[2]

M.E Student, Dept. of Computer, Trinity College of Engineering, Pune, Maharastra, India [1]

Asst. Professor, Dept. of Computer, Trinity College of Engineering, Pune, Maharastra, India[2]

ABSTRACT: Today onwards cloud computing concept is the market blasting .In that uploading and downloading operation performed as well as how much data we use that much we want to pay for it . Everyday data duplication difficulty come in scenario because huge amount of data uploaded on it .To diminish the size of same data uploaded on cloud one method we used is checking data duplication which data can get uploaded check it before uploading .For data uploading MD5 algorithm is used in this technique .In this technique user has allocate a special rights according to it perform duplicate check for achieving deduplication we had used hybrid cloud i.e private as well as public cloud .Finally results proves that proposed way is consume less resources of cloud and give safer proposed way .As compared simple deduplication technique this method has very less overhead .By using token generation step in proposed method check power of user and offer they allow for accessing of data .That privilege checking done at private cloud. After that data uploading as well as downloading happen as per PoW .In that proposed system we also add one advantage that is the data partitioning means two or more public cloud we want to use for strictly more security .In that concept he know if user attack on public cloud then may he know or have privileges of single cloud at that time he/she can get information partially and this is meaningless.

KEYWORDS: Authorization; data security; privilege; duplication; credentials; public cloud; private cloud..

I. INTRODUCTION

Today, Day after day cloud computing become famous as well as popular term .It hides platform as well as OS on the basis of large amount virtual environment .As per use of resources payment want to make. Today many cloud services are available in low cost also give high reliability. Sharing of data or file forwarding by large number of user done on cloud. It has three service modules like PaaS IaaS and SaaS. Managing of data on cloud everyday becoming a critical task. Handling or managing [2] on cloud, many managing technique are use, data deduplication is one of them.

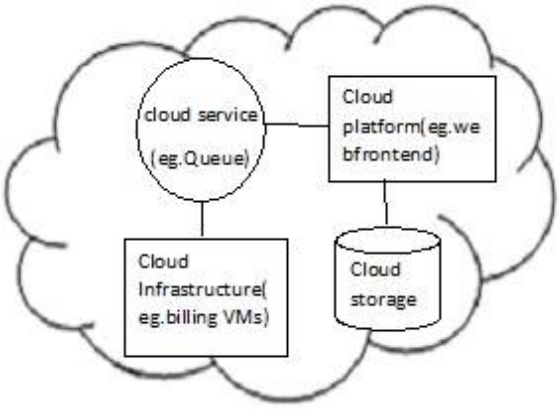

Fig 1:Cloud services & Architecture

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

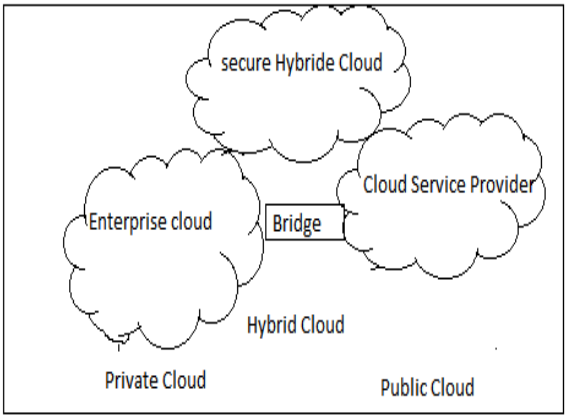

Fig 2:Hybrid cloud formation

To detect data duplication, deduplication of data is perfect or effective technique. It is simpler method, because of that it is on the way of famousing technique. Data compression is also done in that method .Data management as well as networking is done in data deduplication. It remove the copied and only keep original also references of original. First is file level and second is block level are two techniques of duplication checking. In first technique remove file of same name after checking from the storage. In second techniques removed the block which are duplicate data analysing security done in data deduplication. For users delicate data it give security as well as privacy .In traditional way many secret files created because of that need to encrypt data. Data deduplication come in scenario to avoid data duplication[5], for uploading as well as sharing of data check the user authentication in hybrid cloud .If user want to download the file he want to have PoW.

II. RELATED WORK

There are some available techniques of deduplication and cloud storage of data .Post processing is the first deduplication technique [3]. In that first storing of data on cloud after that processed by deduplication. This method doesn't require waiting to calculating hash function and speed doesn't get downgrade of storage. Disadvantage of this method is without storing of whole data also device get full means its storage capacity is low.This is useless method it store data and then check for duplicate. Second is inline duplication check [5].In this deduplication done on file level while new entries .In that block level or file level done but not before adding new entries. But post processing and inline methods have similar output. The other method is source data deduplication check. This techniques are used to check file context before storing them on cloud server .Deduplication check of source is done on the resource side .Periodic scan is performed on file to generate hash file whose hash value is used to check hash value of new file before uploading it on the cloud server .If match found with hash value then it indicates that file contents are same. Then instead of uploading new file on cloud server is provided. Chunk level check of duplication is another technique tagging is done on every file is assign with on unique identification which is generated by software. If it founds same blocks then it keeps only single and removes other blocks. MD5 algorithm we used for file duplication checking and AES algorithm to maintain security until now both of this algorithms were not used in single system.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

A. **Different methods comparison**

Sr no.	Title	Technique used
1	"Hybrid cloud Approach For secure Authorized Deduplication"	Deduplication is only based on file name .For storing data used single cloud.
2	"Secure deduplication with Efficient and Reliable convergent key Management"	File uploaded by single user hold independent master key for encrypting files.
3	"A survey of indexing Techniques for scalable record linkage and deduplication."	In that used indexing techniques for removing duplicate data entries.
4	"DupLESS:server Aided encryption for Deduplicated storage"	In DupLESS system use brute force attack for resisting duplicate storage

For these checking many authors' works some of them are mention in the table.

III. PROPOSED SYSTEM

In proposed system we used authenticated way for checking of duplication. In file checking of duplication. In file level duplication checking at the time of uploading of data is done and also decide file level privilege access .Firstly user want to submit PoW for sending duplicate check request. Only all conditions of PoW get satisfied then request accepted for uploading file.

3.1 System Architecture

Fig. 3 shows user along with private as well as public cloud .For user credentials private cloud is used and the public cloud to managed data storage .For every transaction user has to send request to generate token. If user credentials on public cloud matches with credentials on the private cloud then user only get access to check duplicate data.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

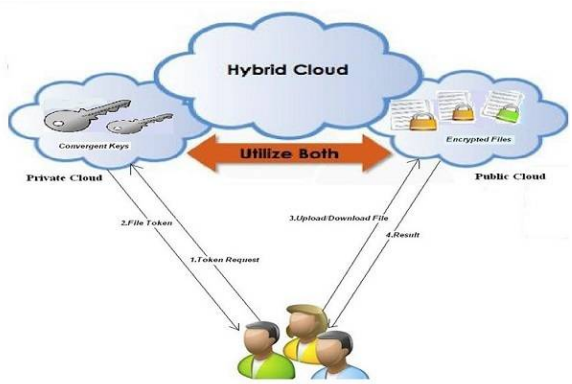

Fig 3. Whole system overview

3.1.1 File Encryption:

The key received from the private cloud That secrete key e are using for encryption , the same key is useful for encrypting plain text in cipher text also for decryption.
Following are the 3 core functions to encrypt and decrypt data.
KeyGenSE:K is algorithm for key generation. Secrete file is generated by these algorithm using parameter of security.
EncSE(K,M):M is message text and K-Secreat key cipher text is generated by using K and M.
DECSE(K,C): C-cipher text
K-key of encryption there by using K and C plain text is generated.

3.1.2:Credential data encryption:

CDE ensures confidentiality of data duplication. Convergent key generation is done by user based on original data and also the encryption of plain text. User adds unique identification tag to identify and eliminate duplicate data. CK generation algorithm is useful for key generation algorithm is useful for key generation and same key is useful in data encryption .User can upload encrypted data on cloud server which make sure both data security and authority.

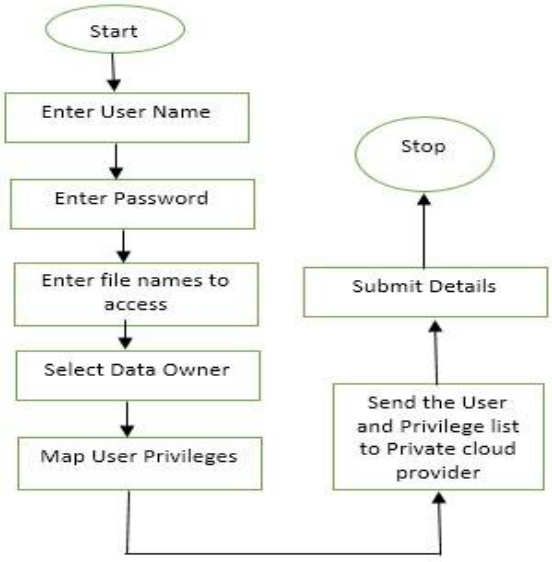

Fig 4.Token generation process

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

3.1.3 Proof of data:

To upload and download data it is mandatory to have proof of it convergent key generation is needful for uploading the file .We are using MD5 algorithm to generate hash value of data .Hash value varies with respect to variation of data

IV. MATHEMATICAL MODEL

S={R,T,P,H,O}
R=Registration for specifying size of user on cloud.
T=Generation and forwarding token through mail for user activation.
P=User privileges
H=Calculation of hash function.
D=Content matched which are uploading data of existing database.
R={r0,r1} where
r0=For registration authority information provider.
r1=Information validity for registration authority.
r2=user id as well as cloud id
r0->t1
T={t1,t2}
t1=Token send through mail to user.
t2=Get credentials to user
D={d0,d1,d2,d3} where
d0=key and name of file get
d1=encrypt file and hash function generation
d2=matching content checking by using upload button
T2->d3
d3=download/upload/update file by giving key/token.

V. COMPARISON OF PROPOSED AND EXISTING SYSTEM

Sr no.	Existing system	Proposed system
1	It cann't provide support for differential authentication to duplicate check.	It provide strong support for differential check.
2	When system get initializes issue starting set of credentials.	In that system firstly outsourcing of data has been done.
3	In that only one copy of user data.	The different credentials of user consider check of duplication.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

VI. EXPERIMENTAL RESULTS

Proposed system should avoid user to upload duplicate data .Encryption method used to upload duplicate data will stored on cloud downloading or uploading data operations unable to do by malicious user which user have PoW only that can modify data. The main goal of system is checking file duplication at the time of upload also avoid wastage of space of cloud

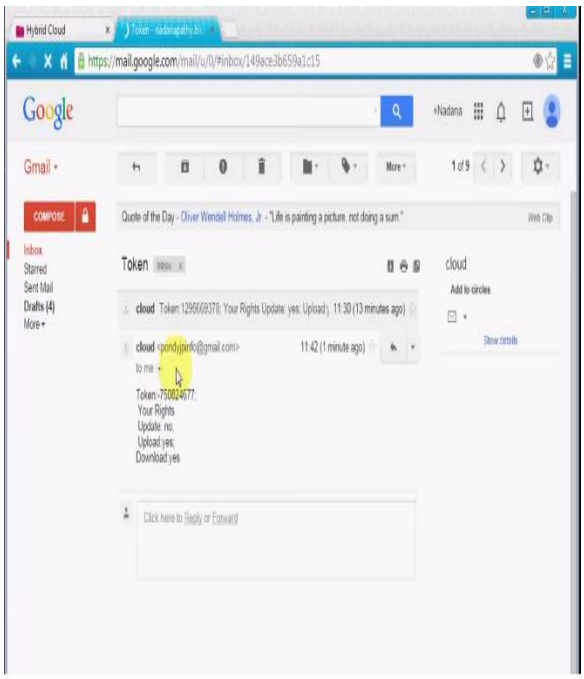

Fig 5.Cloud user activation mail

Here main aim is to check duplicate file at the time of uploading file on cloud storage and also remove wastage space of cloud. This will strictly avoid unauthorized data uses .At the time of registration token of activation mail is sent to mail id .In mail include rights which are provided by private cloud and token.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

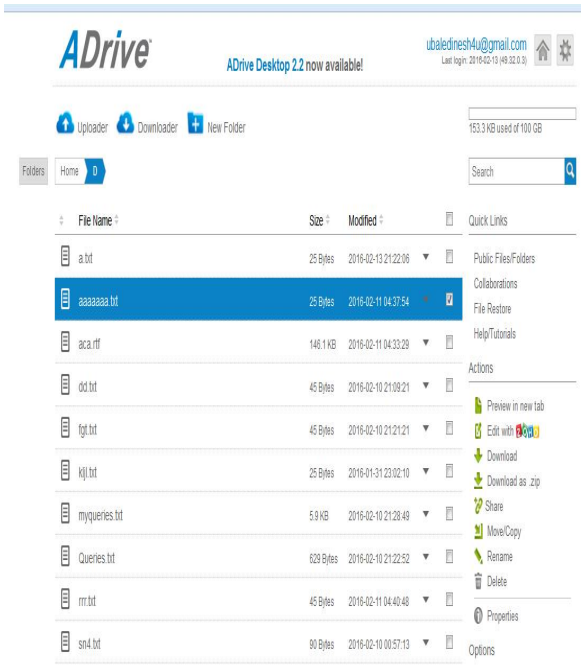

Fig 6.Files on public cloud

In this project we are using ADrive as a public cloud .This is freely available uploaded files after checking content level duplication get stored on public cloud . Which user has PoW can download data or file through public cloud.

Fig 7.Downloaded of files

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

VII. CONCLUSION

In that proposed system shown that completion of secure deduplication and proposed system's data deduplication method is authorized .Also for private files we have proposed token generation technique .As a PoW with the convergent key user submit privilege .We try to solve critical part of storing of data on cloud server it can tolerate by many methods .Proposed method give surety of secure deduplication data.

REFERENCES

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

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

[3] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEE Transactions on Parallel and Distributed Systems*, 2013.

[4] 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 and Communications Security*, pages 491–500. ACM, 2011.

[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEE Transactions on Parallel and Distributed Systems*, 2013.

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

[7] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant TermSuggestion in Interactive Web Search Based on ContextualInformation in Query Session Logs," J. Am. Soc. for Information science and Technology, vol. 54, no. 7, pp. 638-649, 2003.

[8] 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.

[9] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012.

[10] 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 Security*, pages 81–82. ACM.

[11] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.

[12] 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.

[13] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29:38–47, Feb 1996.

[14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In *Technical Report*, 2013.

.