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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

A Secured and Authenticated Mechanism for Data De-duplication using Hybrid Cloud

Ajinkya Borkar[1], Hemant Borude[1], Tushar Budhlani[1], Tejas Mulay[1,] Prof.Savita Lonare[2]

Students, Dept. of Information Technology, Dhole Patil College of Engineering,

Pune, Maharashtra, India[1]

Assistant Professor, Dept. of Information Technology, Dhole Patil College of Engineering, Pune, India[2]

ABSTRACT:ACurrent era is a cloud computing era. The use of cloud increasing daily. Cloud computing is nothing but the sharing of resources and pay as per we use. Now a days use of cloud computing is increasing rapidly. But the problem with cloud computing is every day data is get uploaded on the cloud. The increasing similar data. So to reduce the used size of cloud deduplication is best method in the data deduplication approach duplicate data is removed from the cloud. This will helps to save storage space and bandwidth also. To remove the duplicate data we have proposed a novel method in which user have assigned some privilege according to that duplication check is perform. We have used a hybrid cloud architecture to achieve the cloud data deduplication. Experimental result of our method shows that proposed method is more secure and consumes less resources of cloud. We have shown that proposed scheme has minimal overhead in duplicate removal as compared to the normal deduplication technique.

KEYWORDS:Authorization, data, security, privilege, deduplication, credentials, cloud.

I. INTRODUCTION

Cloud computing has wide range of scope now a days. Cloud provides large amount of virtual environment hiding the platform and operating systems of the user. User get use of resources. User have to pay as per the use of the resources of the cloud. Now cloud service providers are offering cloud services with very low cost and also with high reliability. Large amount of data is get uploaded on the cloud and shared by millions of the users. Cloud providers offer different services such as infrastructure as a service, platform as a service, etc. User not need to purchase the resources. As the data is get uploaded by the user every day it is critical task to manage this ever increasing data on the cloud. To make well data management in the cloud computing deduplication of the data [7] is best method. This method for data deduplication check is becoming more attraction now a days. Data duplication is the technique of reducing the size of data or it is the best compression method for the data deduplication. The deduplication method have application in the data management and in the networking also to send the data over the network required small amount of data. Instead of keeping redundant copies of the same data deduplication only keep original copy and provide only references of the original copy to the redundant data. There are two methods of the duplication check, one is file level duplication check and other is block or content level duplication check. In the file level duplication check the file with same name are removed from the storage and in the block level deduplication the duplicate blocks are removed. As the data deduplication is considering the user data there must be need of the some security mechanism. It arises security and privacy concern of the user's sensitive data. In the traditional method user need to encrypt his own data by himself so there are different cipher files for each new user. To avoid the unauthorized data deduplication convergent data deduplication is proposed in [8] to enforce the data confidentiality while checking the data duplication).

The cloud provide the services as shown in the above figure such as platform, services, infrastructure as a service, and database as a service. In this we are using in cloud storage as a service. To check the authorized duplicate check we are using user credentials to check the authentication of the user. In the hybrid cloud the user credentials are present at the private cloud and data of the user is at public cloud. The hybrid cloud take advantages of both public cloud and private cloud as shown in the figure 1. In the hybrid cloud architecture there are public cloud and private cloud is there.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

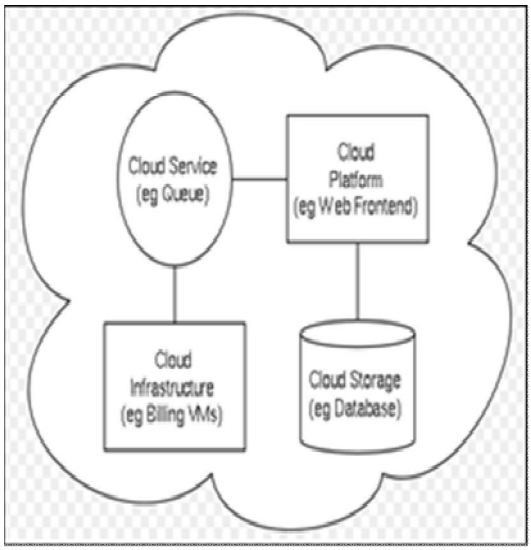

Figure 1. Cloud architecture and services

In the hybrid cloud the user credentials are present at the private cloud and data of the user is at public cloud. The hybrid cloud take advantages of both public cloud and private cloud as shown in the figure 2. In the hybrid cloud architecture there are public cloud and private cloud is there. When any user forward request to the public cloud to access the data he need to submit his information to the private cloud then private cloud will provide a file token and user can get the access to the file resides on the public cloud. In the proposed system we have used a hybrid cloud architecture. The file data duplication is check on the primary level on the file name and then deduplication is checked at the block level of the data. If user wants to retrieve his data or download the data file he need to download both of the file from the cloud server this will leads to perform the operation on the same file this violates the security of the cloud storage.

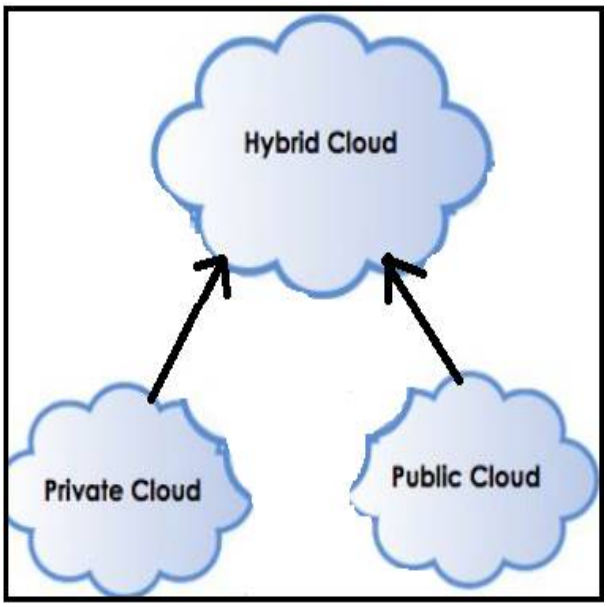

Figure 2. Hybrid Cloud Architecture.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

II. LITERATURE SURVEY

In the existing methods of cloud storage and data deduplication. First method of the data deduplication is post processing method [3] in which data is first store on the storage device and then duplication check is applied on the data. The use of this method is there is no need to wait for calculating the hash function and the speed of storage not get downgrade. The main drawback with this system is that if storage capacity of the device is low then the file storage may get full. The post processing method is not useful at all because it checks the file after storing it on the cloud server. The another method of the duplication check is the inline duplication check [5] as it check the duplication of the file when new entries are to be added to the database. Before adding the new entry or new data to the database it will checks for the block level duplication of the file. Till this method have drawback such as each time need to calculate the hash function which may lead to slower throughput of the storage device. But the some of the vendors have proof that the inline and post processing data duplication check have same output. Another method of duplication check is source duplication check in which the file duplicate contents are checks for duplication before storing it on the cloud server. Third method of deduplication is source data deduplication in which data duplication is done at the side of the source. The file duplication is check before it get uploaded on the cloud server. If new file is to be added to the cloud server and it get match the hash function of the old file then it only remove the new file and just provide hard link to the old file resides on the cloud server.

Another method of the duplication calculation is chunk level duplication checker. In this for each chunk identification is get assigned generated by the software. For the preprocessing file checking we have to make some assumption that identification is same then data is also same but this is not true in all the cases due to the pigeonhole principal. It will produce wrong result that if for two blocks of the data same identification number is get generated it simply remove the one block of the data.

III. PROPOSED SYSTEM

For the data duplication check in the proposed system we are doing duplication check in authenticated way. For the file duplication check proof of ownership is also set at the time of file upload the proof is added with the file this proof will decide the access privilege to the file. It will define who can perform duplication check of the file. Before sending the request to for the duplicate check Request to the cloud user need to submit his file and proof of ownership of the file. The duplicate check request get only approved when there is file on the cloud and also privileges of the user are there.

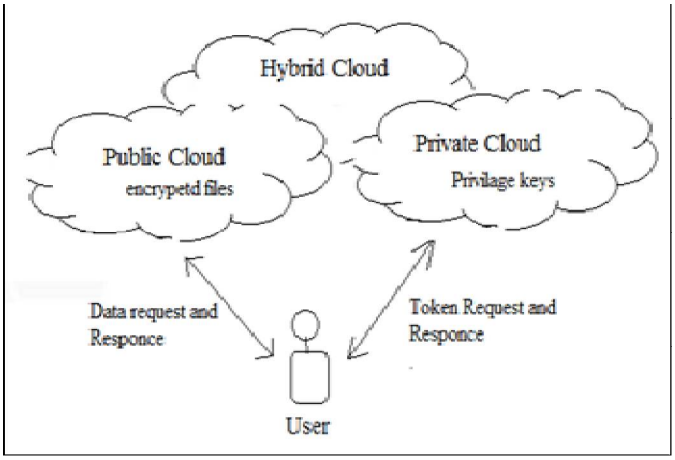

Figure 3. Overview of the System

A. *System Architecture*

It Figure 3 shows the proposed system architecture which comprises of public cloud, private cloud and user. In the proposed system architecture shown in Figure 3. There are one public cloud and one is private cloud. Public cloud contains all data of the user such as files and private cloud consist of user credentials. For each transaction with the

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

public cloud user need to take token for the private cloud. If the user credentials stored at the public cloud and private cloud are get matched then user can have assess for the duplicate check. Following operations are need to be done in the authenticate duplicate check.

B. Encryption of File

ThereTo encrypt the user data we are using secrete key resides at the private cloud. This key is used to convert plain text to cipher text and again for the decryption of the user data. To encrypt and decrypt we have used three basic functions as follow:

KeyGenSE: In this k is the key generation algorithm which can generate the secrete file by using security parameter.

EncSE (k, M): in this formulae M is the text message and key is the secrete key by using this both we have generated a cipher text C.

DecSE (k, C): Here C is the cipher text and k is the encryption key by using cipher text and secrete key we have to generate plain text.

C. Confidential Encryption of data

This ensures a data confidentiality in the duplication. User derives a convergent key from each original data and encrypt the data copy with the generated convergent key. User also add the tag for the data so that the tag will helps to detect the duplicate data.

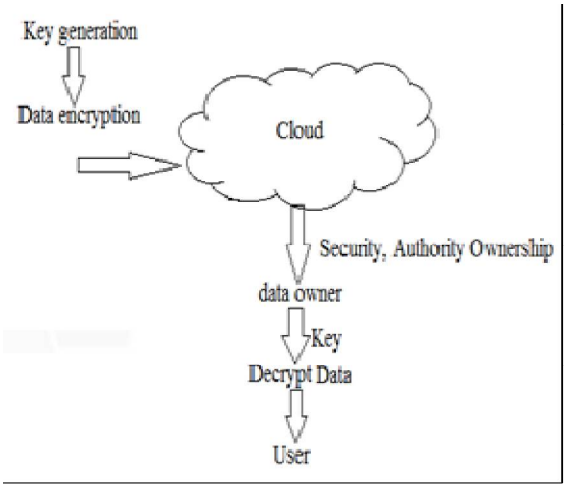

Figure 4. Confidential data encryption

By using convergent key generation algorithm key is get generated this key is used to encrypt the user data. This will ensures the security, ownership and authority of the data.

D. Proof of Data

At the time of file upload and download user need to provide proof of the data. User need to submit his convergent key which was generated at the time of file upload. To generate the hash value of the data we have used MD5 massage digest version 5 algorithm to generate the hash value of the user data. If there is any change in data occur the hash value of that data get changed.

IV. GRAPHICAL ANALYSIS

A. File Size:

To evaluate the effect of file size to the time spent on different steps, I upload 100 unique files of particular file size and record the time break down. Using the unique files enables us to evaluate the worst-case scenario where I have to upload all file data. The average time of the steps from test sets of different file size are plotted in Figure 10.12. The time spent on downloading, encryption, upload increases linearly with the file size, since these operations involve the actual file data and incur file I/O with the whole file. In contrast, other steps such as token generation and duplicate check only use the file metadata for computation and therefore the time spent remains constant. With the file size increasing from 10MB to 120MB.

.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

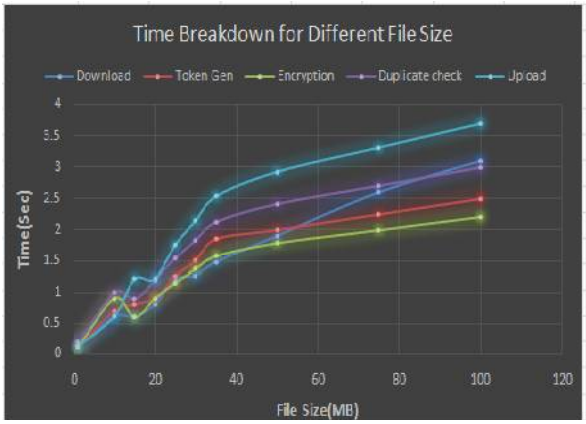

Figure 5 : Time Breakdown for different Fie size

B. Number of Stored Files:

To evaluate the effect of number of stored files in the system, I upload different number of unique size files and record the breakdown for every file upload. From Figure shown above, every step remains constant along the time token checking is done with a hash table and a linear earch would be carried out in case of collision.

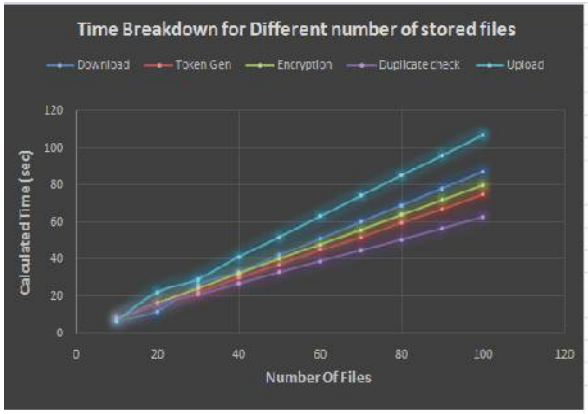

Figure 6. Time Breakdown for different number of stored file

V. RESULTS

This system should prevent user from uploading duplicate data on cloud. Data stored on cloud must be in secure encrypted format. Malicious user not able to upload or download data on cloud. The user who has proof of ownership only that user can modify data.

Fig 6. Home

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

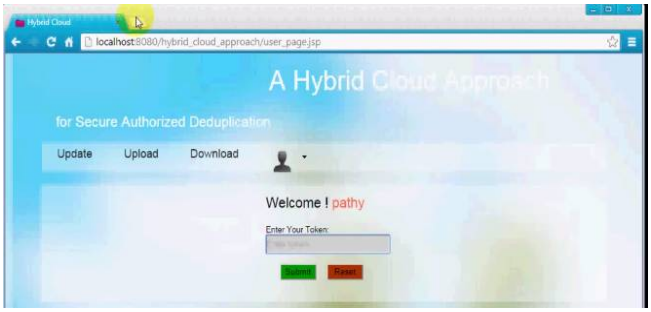

Fig 7. User Window

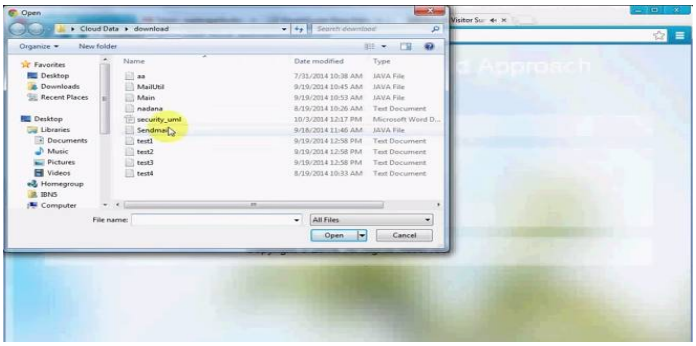

Fig 8. File Upload

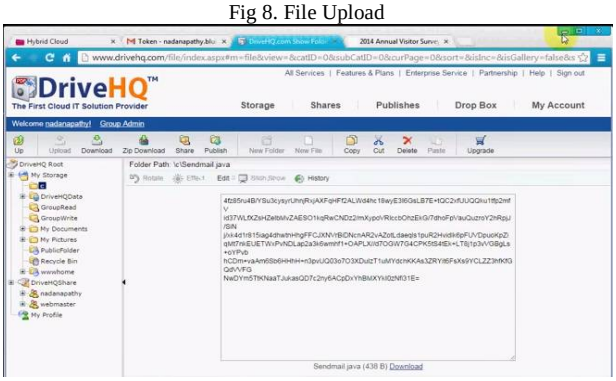

Fig 9. File on drive HQ

VI. CONCLUSION

This paper shows that the proposed method for data deduplication is authorized and securely duplication of the file is done. In this we have also proposed new duplication check method which generate the token for the private file. As a proof of ownership of the data user need to submit the privilege along with the convergent key. We have solved more critical part of the cloud data storage which is only tolerated by different methods. Proposed methods ensures the data duplication securely.

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.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

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.

BIOGRAPHY

Mr.AjinkyaBorkarpursuing his Degree Course in Bachelor of Engineering from Dhole Patil College of Engineering, Pune, Maharashtra, India

Mr.HemantBorudepursuing his Degree Course in Bachelor of Engineering from Dhole Patil College of Engineering, Pune, Maharashtra, India

Mr.TusharBudhlanipursuing his Degree Course in Bachelor of Engineering from Dhole Patil College of Engineering, Pune, Maharashtra, India

Mr.TejasMulaypursuing his Degree Course in Bachelor of Engineering from Dhole Patil College of Engineering, Pune, Maharashtra, India

Prof. SavitaLonareis anAssistantProfessor in Department of Information Technologyin Dhole Patil College of Engineering, Pune, Maharashtra, India.