International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Secured Deduplication using Hybrid Cloud Approach

Gaurav Kakariya[1], Sonali Rangdale[2]

[1]Siddhant College of Engineering, , Savitribai Phule Pune University, Maharashtra, India

[2]Department of Information Technology, Savitribai Phule Pune University, Maharashtra, India

Abstract: ***Today, there is wide range of scope for Cloud computing. Among those most popular is the infrastructure as a service facility of the cloud. The process of removing same files from the cloud and save the user space on the cloud is the data DE duplication in the cloud computing. In large organizations same data is stored on the different places by different users. This will increase the storage size. In the duplicate removal process one can remove the file duplicate with the original file and make space empty for the further storage. By using hybrid cloud we have proposed a novel method data DE duplication to avoid the data duplication and also to maintain the user confidentiality. The confidentiality of the user data as all credentials are stored on the private cloud, maintained the proposed system. To improve the data storage we have proposed a novel method. Proposed method insures data DE duplication securely. This method ensures the data DE duplication with securely.***

Keywords: data DE duplication, hybrid cloud, public cloud, credentials, cloud storage.

1. Introduction

Now a days the cloud computing is most famous era. There are increases rapid use of cloud computing. The amount of data over the cloud is also increasing every day. The main problem with cloud computing is the large volume of data present on it. This may leads to the space problem. In the cloud computing as per user use he have to pay. The loss of resource may happen if user uploaded the same file over the cloud. So there is need of removing duplicate files of repeating data. To utilize the storage space and also reduces the amount of bandwidth required to send the data over the internet this method useful. In the

DE duplication method duplicate chunk or pattern or file name are replace by the small data.

In the proposed system there use the hybrid cloud comprises of public and private cloud all user credentials are presents in the private cloud and data assessable publically is stored in the public cloud. Hybrid cloud provides all easy management and storage facilities an also well resources utilization. As day by day the popularity of cloud is increasing. The use therefor the data resides on the cloud is also increasing daily. In our proposed method to provide security to the user data we have add the some credentials. At the time of file upload the admin ask to user for enter the file password and at the time of file download user need to provide the password that was used by the user at the time of uploading the file on the cloud server. The figure 1. Shows the architecture of the basic cloud computing. In this architecture cloud platform cloud billing and cloud virtual machines are there. The cloud computing architecture is shown in the following figure.

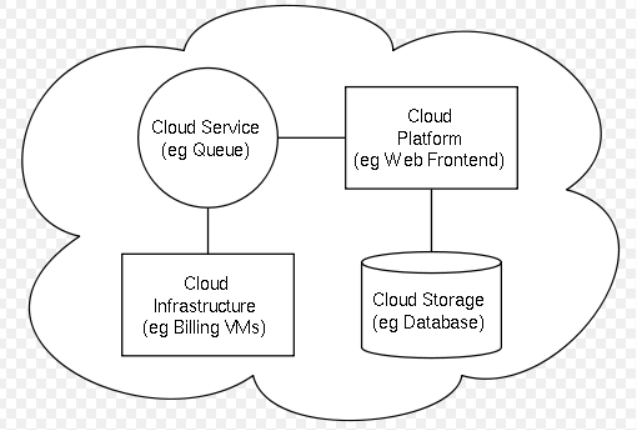

Figure 1: Cloud architecture

In the cloud computing as per user use he have to pay. There may leads to the loss of resource if user uploaded the same file over the cloud so there is need of removing duplicate files of repeating data. To utilize the storage space this method useful and also reduces the amount of bandwidth required to send the data over the internet. In the DE duplication method duplicate chunk or pattern or file name are replace by the small data.

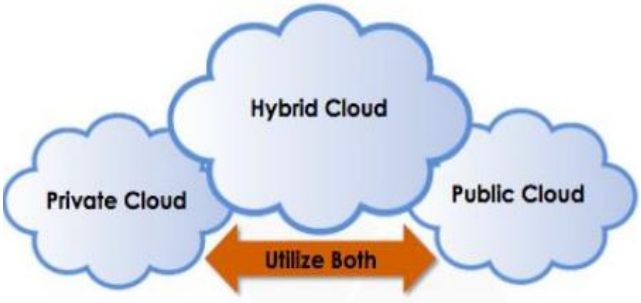

Figure 2: Architecture of hybrid cloud

The main problem in the cloud computing is continuously increasing data on the cloud server. Data DE duplication or single instancing refers to the removal of duplicate files. Duplicate files are removed keeping only one instance or

data copy on the server side but the indexing all the data is also very critical process in data DE duplication method. In simple way data DE duplication removes the duplicate files from the server side keeping only one file on the server. To provide the security to the uploaded data is encrypted before uploading over the cloud. This will take more time and to encrypt the data before store it on the cloud it is more complex. For the large size files the encryption of the file become more complex. We are using data DE duplication to remove the duplicate files and to minimize the time required to encrypt the data every time Because of this time consuming task of the data encryption. As the network consist of large amount of data, which is being shared by users and nodes in the network. The user having access to the cloud have full right of upload and download the file on the cloud server. The number of users are uploading the same data on the cloud is The main reason behind increase in the data on the cloud and due to this same files present over the cloud no of duplicate file are increasing every day. user need to download same files even if his data is present in only one file If user wants to download the data from the cloud server. The cloud will do same operation on the two copies of data files. The data confidentiality and the security of the cloud get violated Due to this. It creates the burden on the operation of cloud. The following figure shows the hybrid cloud. It comprises of private and public cloud in it.

2. Literature Survey

In this section we will see the existing methods studied about the cloud computing. The data duplication was not checked in the existing system. To access the particular file in previous system each user have assigned some privilege and only that file can user access [2], [4]. To reduce the duplicate files present on the cloud server this not useful. If device found the new file on the cloud server then it will simply refer the previous file stored on the server in the old method of the DE duplication technique. The main benefit of the inline duplication is it not required to check the duplicate data over the cloud server. On the other side this duplicate data cannot be removed by using the in line DE duplication method if data duplicate is present on the other side of the server. But some vendors have proposed a method to remove the inline DE duplication data. Post processing and inline data DE duplication are in debate. Another method of the data DE duplication is remove the duplicate data from where it is created that is the source of the duplicate file is need to find out. This is nothing but the source DE duplication [5]. This source side data DE duplication insures that the data get duplicate on the source side. This is generally happens in the file system. The file are periodically scan and hash function of the file is generated and file get removed if new created hash matches with the existing hash function. To remove the duplicate files from the secondary side with the data DE duplication technique that is from the user side file removal. The data duplication is carried out in such a way that a chunk is taken and applying the algorithm on that chunk unique id is generated that is nothing but the hash value of each file. The size of the hash function is very less as compare to the original file. But if you change the file name value of that hash is also get changed. By the software calculated hash values each chunk of data get assigned. In many case it get assume that data is identical, and chunk are also same but this is not true in al the cases. Second assumption is that the data with same hash value are same but due to pigeonhole phenomenon this is not true in all the cases the data may be identical or not. If software assume that Once the data has been DE duplicated There are two issue with this, depending upon the read back if file found to be duplicate it get removed from the database. On the basis of name of the files the existing methods not supports the secure DE duplication system. In the existing all the system duplication check is done. The filename get replace by the link to the previous file present already in the database if the file with same name is found. The check the duplication of the file on the name of that file by the convergent cipher text. By unauthorized user on the cloud Proof of ownership is avoid the file access.

3. Proposed System

We are achieving the data DE duplication by applying the proof of ownership applied by the owner of the data in the proposed DE duplication technique. The proof is applied at the time of uploading the file. The file uploaded to the cloud is bounded by some value that specifies which user can have the access to the particular file without this proof of ownership. User is not able to perform the duplication check on the file. Figure 3 shows the proposed system architecture. To access the file which can get the user from the private cloud User need to submit his proof of ownership and access level privilege. If user match the privileges of his file stored on the public cloud then User is able to find the file duplication of his own file.

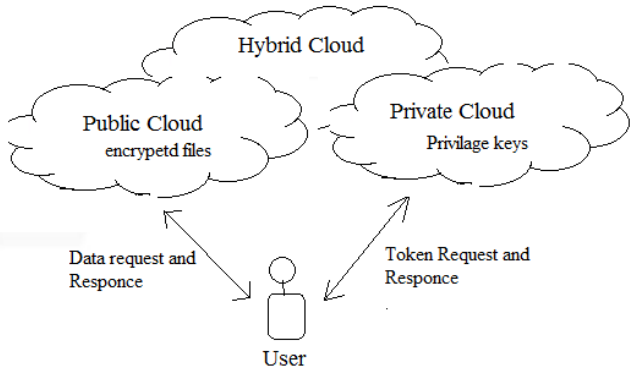

Figure 3: System Architecture

3.1 Encryption of Files and data:

In this to encrypt and decrypt the text we are using the common key. To convert the plain text to cipher text and get plain text back from the cipher text this secrete key k is used. To achieve this we have used three basic function KeyGenSE: it is the key generation algorithm which generate a key to generate the key value. Second one is EncSE (k, M): C is the second function which takes Message M as an input and apply key K on it and generate the cipher text C.

3.2 Confidential Encryption Method:

Usinf this we generate the confidentiality check. By using a convergent key each original file get encrypted. Also user can also generate the tag which are helpful to remove the duplicate files in the proposed system. To generate the key the key generation algorithm is used and this generated key is used to encrypt the data and then proof for the data ownership is need to provide. Then the data is get encrypted and stored on the cloud storage. User need to provide his proof of ownership at the time of downloading the data and also need a key to get original content back. This method ensures that the duplication is not in the file. The confidential encryption method is shown in.

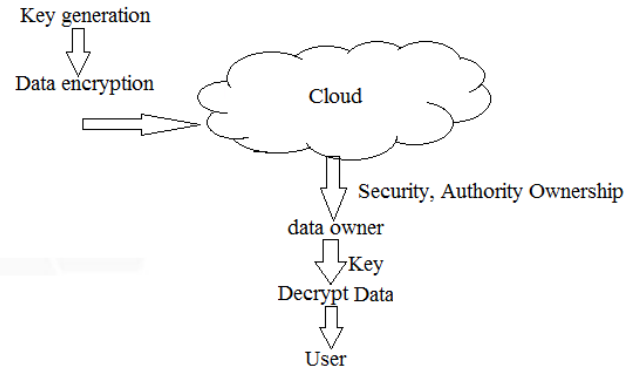

Figure 4: Confidential data encryption

3.3 Proof of Data

User need to provide his proof of ownership that means he need to submit his file details and convergent key at the time of data upload and data download. We are using data duplication check by using the file content in the proposed system. At the content level of the file the duplication is check. At the time of uploading or downloading the data from the cloud User needs to provide his proof of ownership. We have introduced novel approach to solve the problem of the existing file duplication check which can check the file duplication at the content level of the file. The user credential is managed at the private clout side. User needs to send the token request to the private cloud to get the assess to the particular file and then the according to the privilege set at the private cloud user get the token for the file. At the private cloud and at the public cloud the authentication of the user is done.

4. Experimental Analysis

We will do this experiment with the text file. File we are using MD5 algorithm to generate the hash function of the uploaded. This algorithm generate the hash value of the file get uploaded on the cloud. The file content level duplication will check by our proposed system. Our method is independent of the file size. To complete the experiment we have taken three system one of which is client Which acts as a user and performs the operation such as file upload, download etc. to manage the key and generate the token for the user file access The second system acts as Private Cloud is used. Third is Server system which contains data of the user on it and acts as C-CSP.

5. Results

Figure 5: Home Page

Figure 6: User Registration

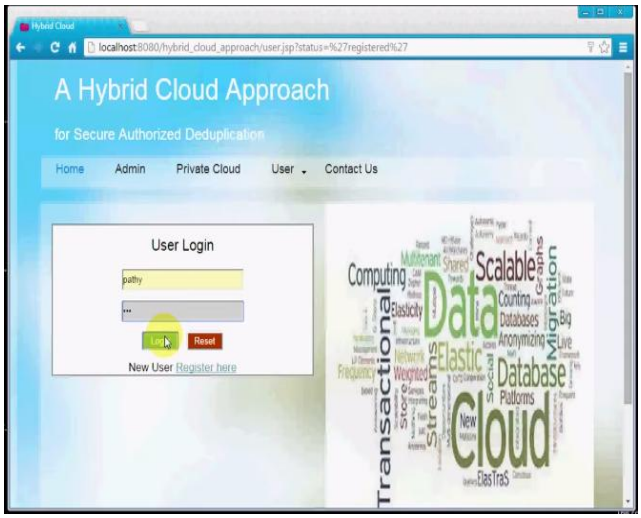

Figure 7: User Login

Figure 8: Private Cloud Login

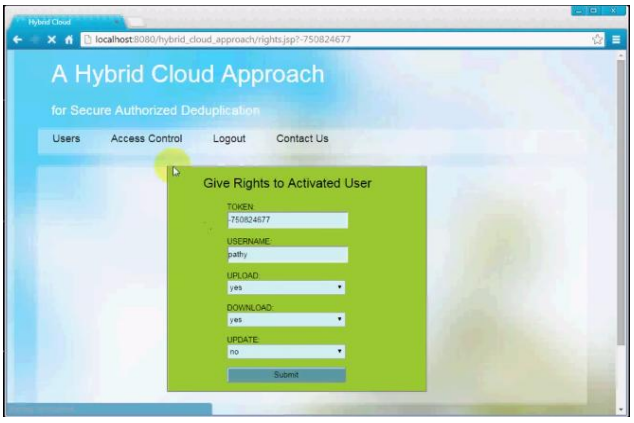

Figure 9: User activation by Admin

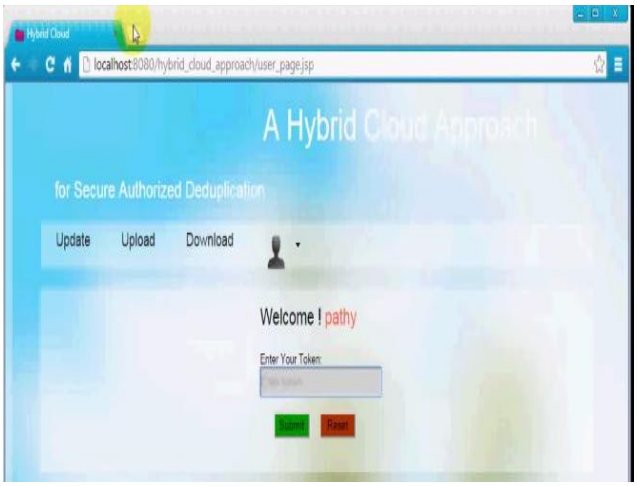

Figure10: User Window

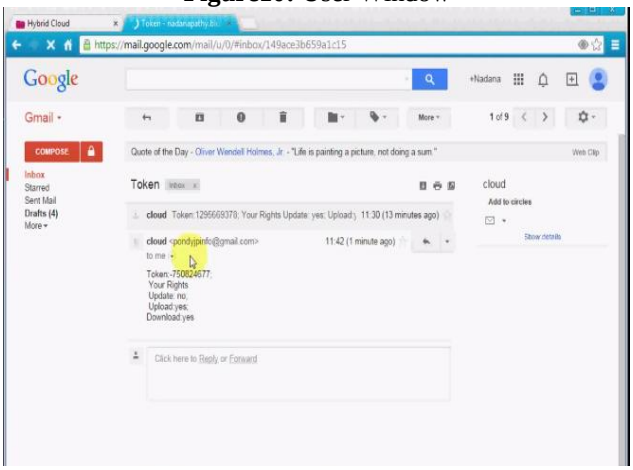

Figure 11: Activation token mail

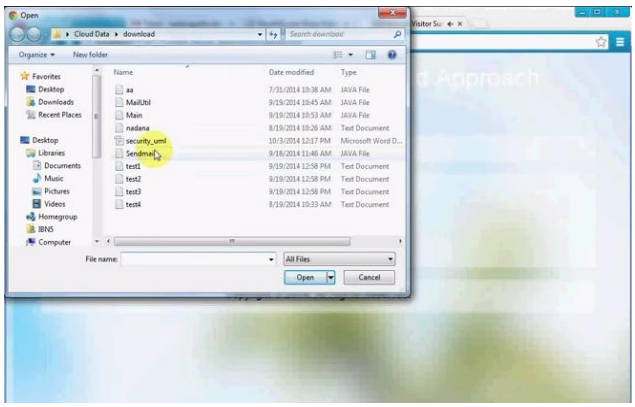

Figure 12: File Upload

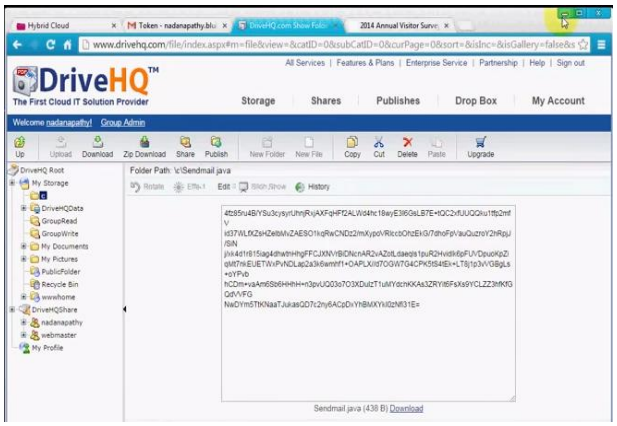

Figure 13: File on drive HQ

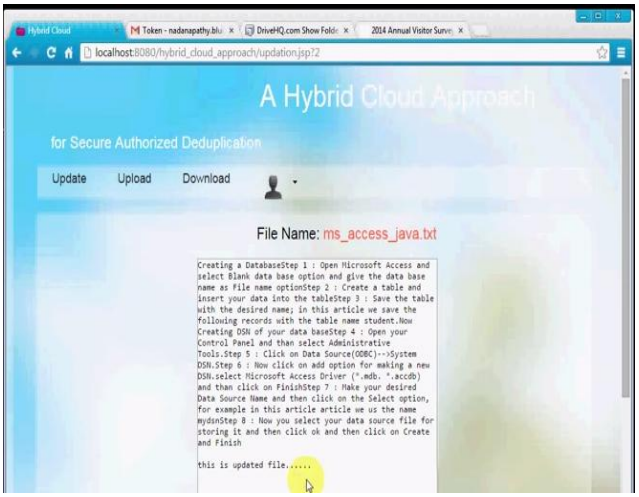

Figure 14: File Update

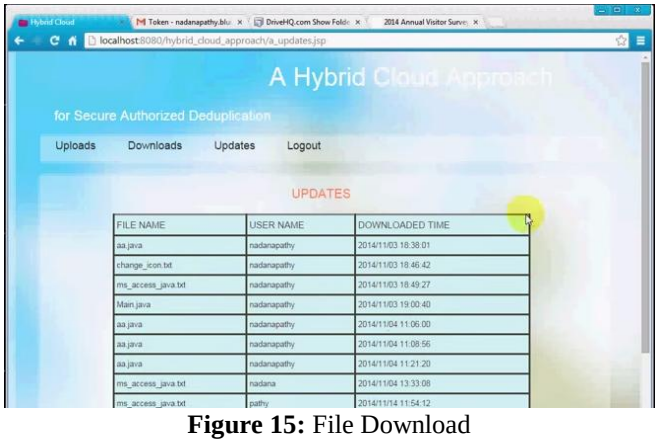

Figure 15: File Download

6. Conclusion

The file DE duplication is addressed, in this paper. To securely check the duplication we have proposed a novel method. Cloud computing has reach to his maturity level. That means the cloud is used by commercial user in very large amount. It does not mean that all the problems of cloud computing are solved totally. According to the situation this problems are only tolerated. Due to this cloud computing is more research part. To remove the duplicate files from the cloud storage we have proposed novel scheme. This method insures that the proposed method successfully remove the duplicate files from the cloud storage.

7. Acknowledgment

Author would like to take this opportunity to express our profound gratitude and deep regard to my guide prof. Sonali Rangdale for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout my project work. His perceptive criticism kept me working to make this project in a much better way. Working under him was an extremely knowledgeable experience for me.

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.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," J. Am. Soc. for Information science and Technology, vol. 54, no. 7, pp. 638-649, 2003.
[7] 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.
[8] 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.
[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 Security, pages 81–82. ACM.
[10] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002.
[11] 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.
[12] 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.
[13] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.
[14] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.