ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (752-758), Month: April - June 2015, Available at: www.researchpublish.com

A Secure Backups in Cloud Servers with Authorized Deduplication by Hybrid Cloud

[1]Bharati Choudhary, [2]S.M Joshi

[1]PG Student, [2]Asst. Professor, ECE Dept., GNDEC Bidar, Karnataka, India

Abstract: **Data deduplication is a data storage saving method which has been used in cloud servers for deleting duplicate data files. It stores only one copy of the identical data file on the cloud server instead of storing number of similar data files and offers their users infinite storage space. On cloud servers, most of times users save the same data files, so same data files take huge space. Repetitious copies of data files prevented by deduplication and retain one copy of the data files. It works with encryption technique to provide security to data files before uploading to cloud server. The Cost of storing and transferring data can be greatly reduced by data deduplication. The intention of this paper is to address the problem of authorized data duplicate check and analyse security of stored sensitive data using hybrid cloud architecture which made up of public cloud and private cloud. Data confidentiality is strengthened by this hybrid cloud arrangement .Thus users protect the sensitive data and outsourced to the cloud storage securely.**

Keyword: **Cloud computing, deduplication, authorized duplicate check, Data Confidentiality, Differential Authorization, hybrid cloud.**

1. INTRODUCTION

Cloud computing is a technology in which resources of the computing assets are provided as services over the internet. In cloud all resources connected effectively to create single system image .A simple examples of cloud computing are yahoo email, Gmail email, rediff email etc. Here you don't need any additional software or hardware to use them. The idea of cloud storage is derived from cloud computing. It denotes a storage appliance retrieved over the internet via web service application program. The most widely used cloud service is depository for all common users. Simply put, cloud computing provides a storage from servers such as email, security, backup, sound, all delivered over the Internet. Total space savings up to 90-95% in backup applications. The Cloud provides a hosting media that is quick, changeable, measurable, secure, and available – while saving corporations money, time and resources. But the main point is privacy. There is no protection for data in the cloud server. One of the important problems of cloud storage service is the control of the ever-increasing volume of data. To address this problem, data deduplication technique is used. In simple words data deduplication is a process of finding and deleting repetitious data. When duplicate data is detected during backup then that data is discarded and only the pointer is created and refers to a copy of data that is already backed up. This helps to reduce the storage requirement for backup, shorten the backup window and remove network problem. Data deduplication can be used to make better the data quality and integrity which in turn reduce costs and efforts in obtaining data due to duplicate data. Storing duplicated data only once. Decreasing of duplicate data in cloud storage and save the server space. It Decreases size of occupation and decreases bandwidth.

File level deduplication and sub file deduplication are two methods of data deduplication. . File level deduplication finds and removes repetitious copies of alike files. After a file is stored, all other citation to the same file refer to the original copy. Subfile finds repetitious data within and across files. Data deduplication has turn into a key element in the backup process. It specifies that only one copy of that data is saved in the cloud. Every user, who want to use that copy linked to that single instance of copy. So it is clear that data deduplication help to decrease the size of data enter. In deduplication process, initially files are encrypted by using AES algorithm and then divided into segments. After the segment creation new and the existing data are checked for similarity by comparing fingerprints created by SHA-1Then encrypted files are uploaded. All duplicate data is deleted and data integrity check is performed.

2. PREVIOUS WORK

The differential authorization duplicate check facility is not present in old deduplication systems which are needed in many applications. In an authorized deduplication system, the user uploads their data in the cloud server. For the safeguard purpose the data owner encrypts the data file and then store in the cloud. The user can check the duplication of the file over Corresponding cloud server. The user can have capable of manipulating the encrypted data file and the data owner can check the multiple cloud data as well as the duplication of the specific file technique will be applied to store only one copy of the same file. Because of seclusion purpose, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. Existing deduplication can make it easy for outsiders to know what's already on storage servers. Earlier deduplication systems based on conventional encryption, although gives confidentiality to some extent, do not carry the duplicate check with differential privileges.

A. Symmetric Encryption:

Symmetric encryption is one of the old encryption technique uses a common secret key to encrypt input plaintext data and decrypt output cipher text data. The person sending the message uses the key to encrypt it. The person receiving the message uses the same key to decrypt it, decryption is opposite of encryption. So both sender and receiver are both using the alike key, so the system is known as symmetric key encryption. Common Symmetric Key System is AES. The Advanced Encryption Standard is far more secure than DES. The key lengths can change between 128-bit and 256-bit.

B. Convergent Encryption:

The convergent encryption extracts key from the hash of plaintext and a security model for secure data deduplication. Data Encryption key extracted from data K = hash(Data) Convergent encryption lets cloud storage providers store large amounts of data at low amounts, while offering better privacy than traditional cloud storage.

The convergent encryption that consist of four functions

1. Key Generation: the users extracts Convergent key from the data files.
2. Encryption : The plain text data file encrypted with Convergent key which produces the same cipher text for the same data
3. Decryption: The cipher text is decrypted with convergent keys. To get the original plain text.
4. Tag Generation: The tag is extracted from the data file and is sent with encrypted data to server which helps to find the duplicate

Convergent encryption [1], [3] provides data secrecy in deduplication. A convergent key extracted a from input data file by user and encrypts the data file with the convergent key at the same time user also extracts a tag for the data file .the tag finds duplicates. The tag is made up of user Id, block number. If two data files are the same, then their tags are the same. To find duplicates, the user first sends the tag to the cloud server to check to same files. Both the encrypted data file and its tag will be stored on the public server.

C. Proof of Ownership:

The proof of ownership (POW) [2] protocol authorizes users to prove their possessor ship of data files to the cloud server. The POW is implemented as two sided algorithm run by user and a cloud server. PoW is in two parts and it's between two persons on common input file. In first step cloud server summaries to itself and generate sort information "v". In later step user and cloud server engage in interactive protocol where cloud server has sort information "v" and user has file "F" at the end cloud server either accepts o rejects it. At the end, the cloud server either accept or reject to denote whether the proof is passed or not.

3. PROPOSED METHODOLOGY

In the proposed system, the problem of deduplication along with differential privileges in cloud computing environment can be addressed by hybrid cloud .A hybrid cloud is made up of public cloud and private cloud

In which encrypted data and tags are stored in public cloud and convergent keys are stored in private cloud.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (752-758), Month: April - June 2015, Available at: www.researchpublish.com

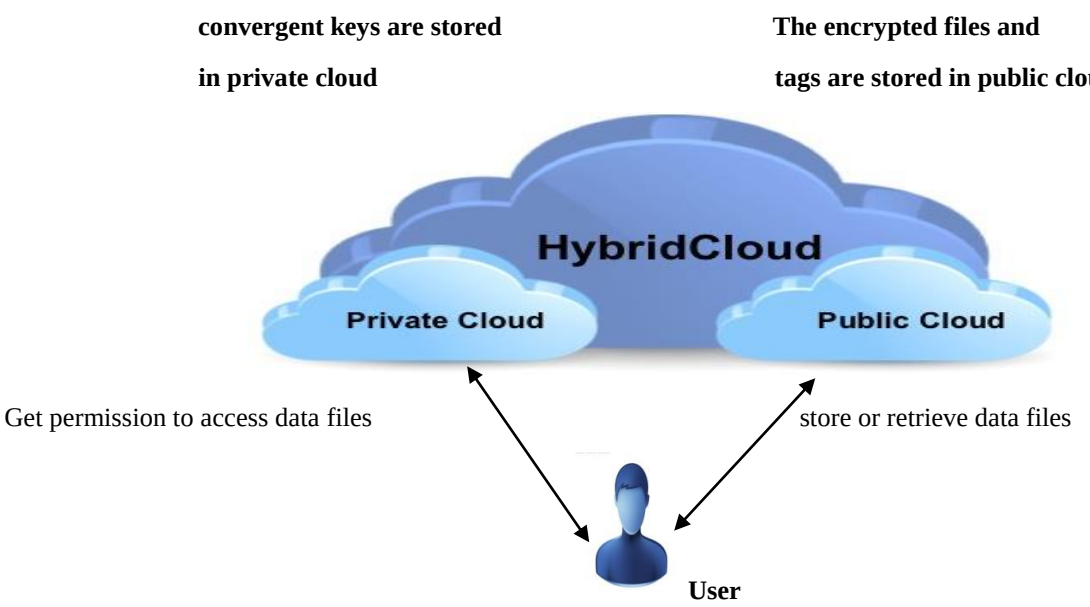

Fig 1: Architecture of Proposed System.

Using deduplication technique we can avoid or delete duplicate copies of data or files. Public cloud is used for the storage of data. The users which has authority only that user can upload and download files from public cloud. Because private cloud provide security by generating a key. When user want to download the file at that time user request to private cloud for key and then access that file.

4. SYSTEM MODEL

Let's see the architecture of our system. In our system we can use three modules.

They are:

1] User

2] Public cloud.

3] Private cloud.

1] User is an object who can upload or download the data files on public cloud when user want to upload the data file first user encrypt that file using AES encryption technique and after encryption user store or upload that file on public cloud. At the time of uploading user generate a key for that file and stored that key on private cloud server for providing the security. the user can only access the data file with the encrypted key if the user has the privilege to access the file. All the privileges are given by the particular domain control and the Data user's are controlled by the Domain control only. Users may try to access data files either within or outside the scope of their access privileges.

When user want to download file user first send the request for public cloud. Then public cloud provide the list of files that present or stored on public cloud. There will be many files stored on public cloud because public cloud do not have any security.When user select a file for downloading then private cloud ask the key which generated at the time of uploading file. When user enter the password or key if that password is matched then the user is valid user and user can access or download the file from public cloud, then decrypt the file using key which is used at the time of encryption of that file.

2] Public cloud is object used to store data files. User uploads the files in public cloud. When the user wants to download the files from public cloud, it will be ask the password or key which is generated or stored in private cloud. In public cloud all files are stored in encrypted format. If unauthorized person hack our file, without the secrete key hacker does n't access original file. Each file is protected by AES encryption key and can access by only authorized person. In our system user have to do the registration on the private cloud for storing the key with respective to file that user stored on public cloud. When user want to access that file user access respective key from private cloud and then access user files from public cloud.

3] In our system we use the private cloud with public cloud to provide a security for the data. When user upload any file he/she generates a key for that file and store that key on the private cloud. For proper management of keys we use a private cloud. Private cloud only stores the keys for different files. When user want to access the file private cloud first check for the authorized person and then and then it provide the access.

5. OPERATIONS PERFORMED ON HYBRID CLOUD

File uploading:

When user wants to upload the file on the public cloud, user first encrypt that file by using AES encryption technique. Our main purpose is to avoid the duplications of file. For that when user upload file to public cloud server, same time user generates the which is used to check duplicate.the encrypted data file and tag is stored in public cloud and secret key stored in private cloud. The user able check the duplicate data in cloud .if the tag matches with tag already stored in cloud server, it shows duplicate data present.

File downloading:

When user want to download the file from public cloud user first send request to the public cloud. Then public cloud provides the all files presented on public cloud, uploaded by the different user. User select one of the file that he want to download. Then public cloud sends a message to user that enter the key to download the file. User has to enter the key for that file. Then public cloud checks that the key is matched or not. If the key is matched then the user is valid for that file and public cloud provide the access to user for that file. After downloading of file user has to decrypt that file to get the original contents of that file.

6. SIMULATION RESULTS

The following different screen figures showing how uploading and downloading data files in cloud servers with secure authorized deduplication technique

Figure 1: Screen showing data owner registration

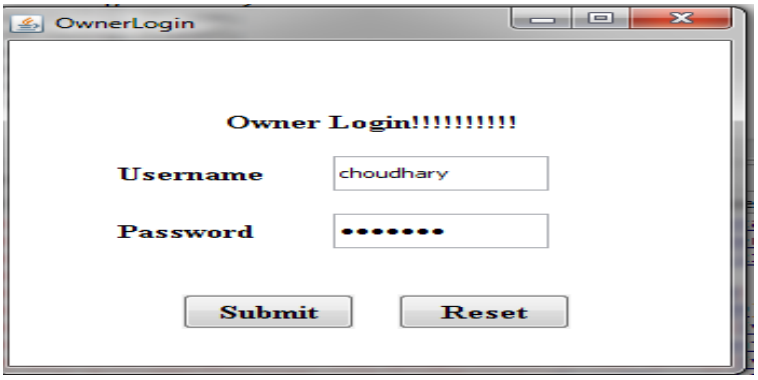

Figure 2: screen showing dataowner login

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (752-758), Month: April - June 2015, Available at: www.researchpublish.com

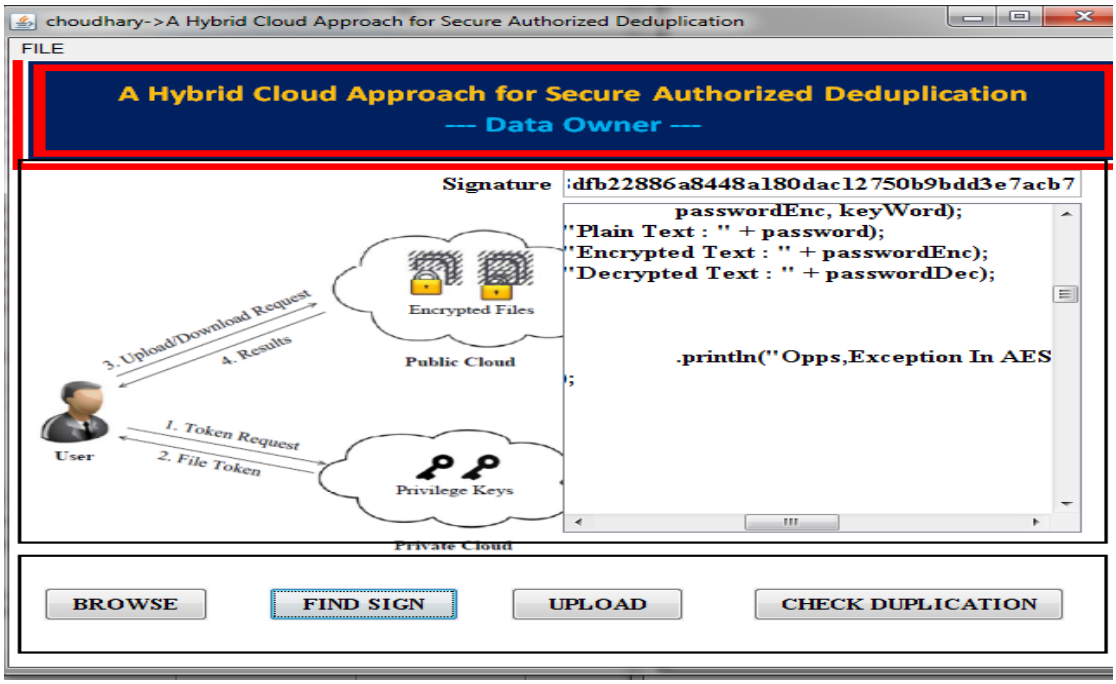

Figure3: Screen showing finding digital signature

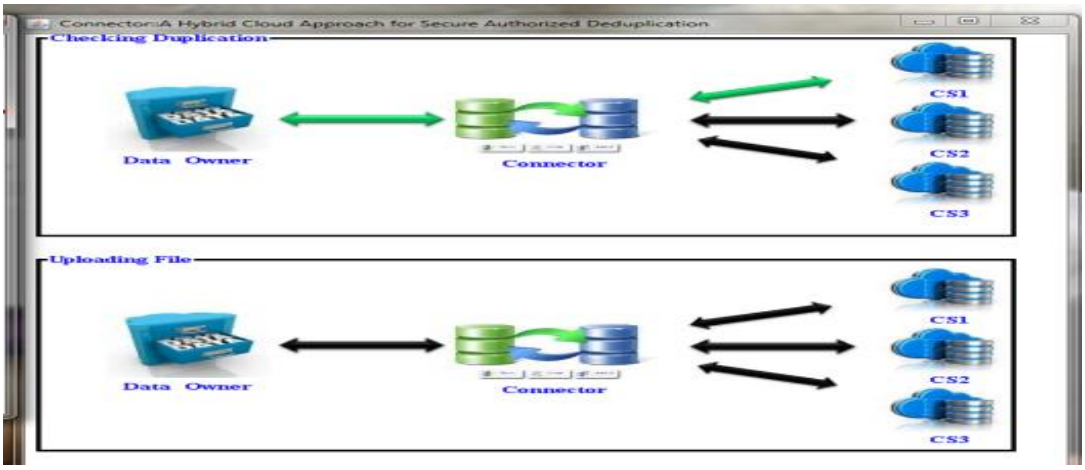

Figure 4: Screen showing checking duplicate file

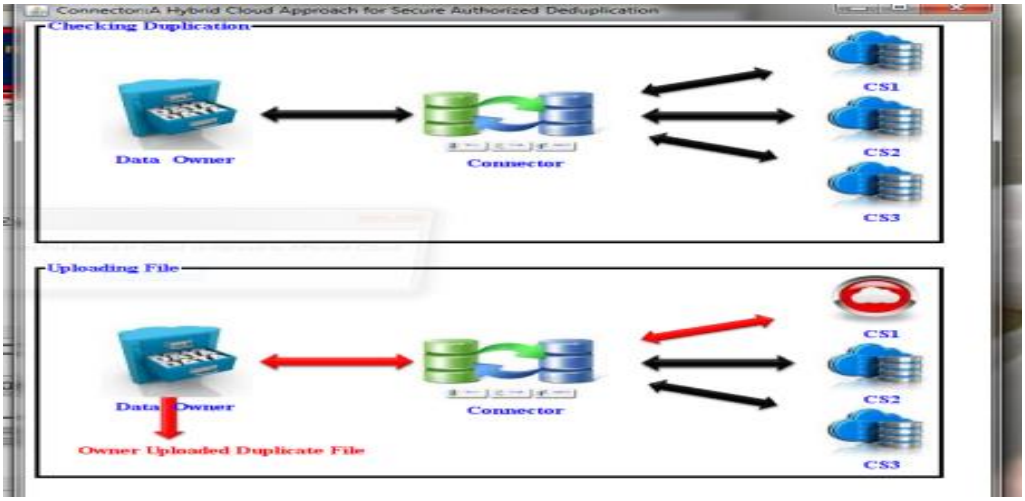

Figure 5: screen showing cs1 having duplicate file

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (752-758), Month: April - June 2015, Available at: www.researchpublish.com

Figure 6: screen showing user login for downloading file

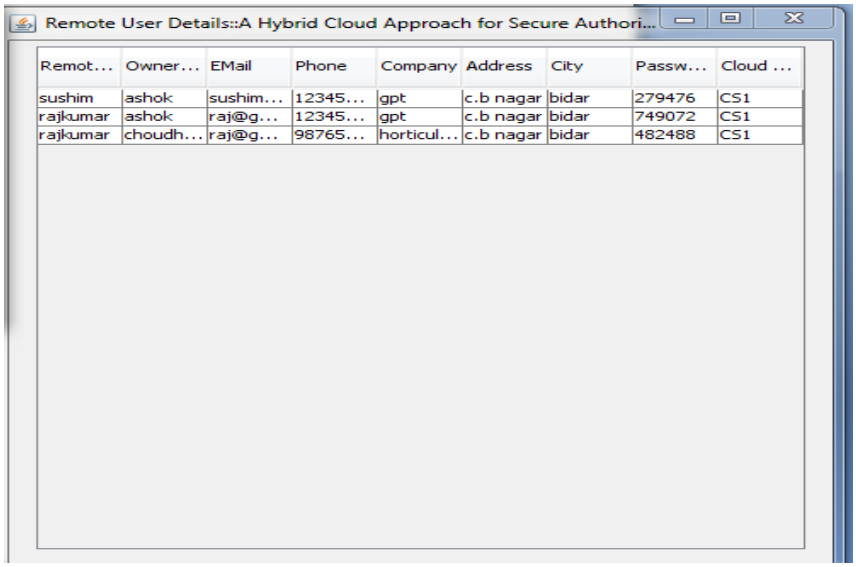

Figure 7: screen showing secret key for downloading file

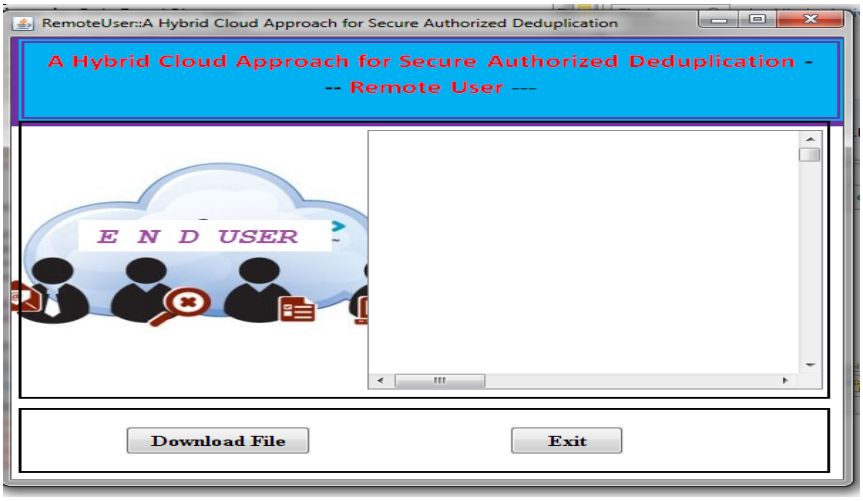

Figure 8: shows Downloading screen

The cloud server owner can add the different users, After adding user should register himself and login .Every user can upload the files onto the cloud with access permissions: Before uploading the file on to cloud, user encrypts the file and generate tag for it .there is facility to perform duplicate check to see duplicate files in cloud then click on upload file to upload the selected file on the cloud. Before uploading the file on to cloud the server will checks whether the uploading files available at server or not, if it is not available then it uploads on the cloud. If it is available then it tags to the existed file but not uploads it again.User can download the files from the cloud: To download the file,user has to get secret key from private cloud.

7. CONCLUSION

In this paper, we provided with the fundamental concepts about deduplication, I presented new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture . Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept,. We showed that our authorized duplicate check scheme incurs minimal overhead compared to other schemes. In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. The encrypted data file and tag are stored in public cloud. The secret key is stored in private cloud.without key anyone cannot access our file or data from public cloud.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. "A Hybrid Cloud Approach For Secure Authorized Deduplication," In IEEE Transactions on Parallel and Distributed Systems, 2014.

[2] 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.

[3] Open SSL Project. http://www.openssl.org/.

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

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart."Server aided encryption for deduplicated storage," In USENIX Security Symposium, 2013.

[6] M. Bellare, S. Keelveedhi, and T. Ristenpart."Message-locked encryption and secure deduplication," In EUROCRYPT, pages 296–312, 2013.

[7] M. Bellare, C. Namprempre, and G. Neven "Security proofs for identity-based identification and signature schemes," J. Cryptology,22(1):1–61, 2009.