Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

Data Content Verification for Effective Cloud Storage

Praveen Kumar A M
Department of MCA
PES College of Engineering
Mandya, Karnataka, India

Dr. H. P. Mohan Kumar
Department of MCA
PES College of Engineering
Mandya, Karnataka, India

***Abstract*— Data Deduplication is a web application technique which is used in reducing storage space and increasing bandwidth. This Deduplication system is mainly important in utilizing storage space of cloud by reducing reliability. Security issues like data confidentiality and data consistency are also achieved by this application. This is achieved by sharing data by secret sharing scheme in distributed system. In earlier systems convergent encryption scheme was used. In this system security analysis demonstrates that system is secure as per definitions specified. As a proof this proposed system is demonstrated that overhead is limited.**

The main aim of "data Deduplication in cloud by secure auditing" is to provide data integrity and security in storing data, and also achieves data Deduplication methodology. All the files that have to be shared is stored in public cloud as memory can be used efficiently. In this project all users who have to upload the file should register and then by using his /her login details if it is valid then only users can upload the file. Every day batch files are run using the file comparer when the same files are uploaded many times it refers to consuming of memory of cloud hence the redundant data is not necessary then that data has to be destroyed as it consumes memory while it is not necessary. Hence to destroy that data or deduplication has to be done by secure auditing technique. This can be achieved by using file streams like size of the file contents etc if data is matched then the recently uploaded file will be deleted.

Keywords — Encryption Technique, Data content verification, Text Deduplication.

I. INTRODUCTION

Survey process is needed to require the information of the software. Survey also consists of studying the present system and also study about the tools used for developing of the software. To develop any software proper understanding of the software and also survey process is essentially required. Below information gives about the resources collected during this process.

Therefore, if a user wants to upload a data which is already stored in the cloud, the cloud provider will show deduplication not allowed. Deduplication can reduce storage needs by up to 90-95% for backup applications[11] and up to 68% in standard file systems [18].Users require the protection of their data and confidentiality along with low ownership costs and flexibility which guarantees through encryption. Unfortunately, deduplication and encryption are two conflicting technologies. While the aim of deduplication is to detect duplicate data and store them only once, the result of encryption is to make two identical at a in distinguishable after being encrypted. This means that if data are encrypted by users in a standard way, the cloud storage provider cannot apply deduplication since two identical data will be different after encryption. On the other hand, confidentiality cannot be guaranteed and data are not protected against attackers in cloud storage providers if data are not encrypted by users Convergent encryption.

Convergent encryption By linking network resources to gather, cloud computing provides a big pool of resource. It has desirable properties, such as scalability, elasticity, fault-tolerance, and pay-per-use. Thus, it has become a promising service platform. Data storage service is the most important and popular cloud. Cloud users have some personal or confidential data which will be uploaded to the data center of a Cloud Service Provider and allow it to maintain these data. Since we cannot avoid intrusions and attacks towards sensitive data at CSP, it is prudent to assume that CSP cannot be fully trusted by cloud users.

But the same or different users may upload duplicated data in encrypted form to CSP, especially for scenarios where data are shared among many users. Although cloud storage space is huge, data duplication greatly wastes network resources, consumes a lot of energy, and complicates data management. The development of numerous services will further make it urgent to deploy efficient resource management mechanisms. Here the practical issue is how to manage encrypted data storage with deduplication in an efficient way. However current industrial deduplication solutions cannot handle encrypted data Existing solutions for deduplication suffers from some security weakness i.e. brute-force attacks.

Data Deduplication is a web application technique which is used in reducing storage space and increasing bandwidth. This Deduplication system is mainly important in utilizing storage space of cloud by reducing reliability. Security issues like data confidentiality and data consistency are also achieved by this application. This is achieved by sharing data by secret sharing scheme in distributed system. In earlier systems convergent encryption scheme was used. In this system security analysis demonstrates that system is secure as per definitions specified. As a proof this proposed system is demonstrated that overhead is limited.

The main aim of "data Deduplication in cloud by secure auditing" is to provide data integrity and security in storing data, and also achieves data Deduplication methodology. All the files that have to be shared is stored in public cloud as memory can be used efficiently. In this project all users who

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

have to upload the file should register and then by using his /her login details if it is valid then only users can upload the file. Every day batch files are run using the file comparer when the same files are uploaded many times it refers to consuming of memory of cloud hence the redundant data is not necessary then that data has to be destroyed as it consumes memory while it is not necessary. Hence to destroy that data or deduplication has to be done by secure auditing technique. This can be achieved by using file streams like size of the file, contents etc if data is matched then the recently uploaded file will be deleted

They cannot flexibly support data access control and revocation at the same time Most existing solutions cannot ensure reliability, security, and privacy. In practice it is hard to allow data holders to manage deduplication due to a number of reasons. First they cloud cause storage delay because data holders may not be always online or available for such a management. Second deduplication could become too complicated in terms of communications and computations to involve data holders into deduplication process. Third in process of discovering duplicated data it may in trude the privacy. Forth a data holder may have no idea about how to issue data access rights or deduplication keys to a user in some situations when it does not know other data holders due to data super distribution. Therefore, CSP cannot cooperate with data holder data storage deduplication in many situations [5]. The results show the superior efficiency and effectiveness of the scheme for potential practical deployment, especially for data deduplication in cloud storage.

II. LITERATURE SURVEY

Survey process is needed to require the information of the software. Survey also consists of studying the present system and also study about the tools used for developing of the software. To develop any software proper understanding of the software and also survey process is essentially required. Below information gives about the resources collected during this process.

Cloud storage service providers such as Drop box [2], Google Drive[3], Mozy[4], and others perform deduplication to save space By only storing one copy of each file uploaded. However if clients conventionally encrypt their data, storage savings by deduplication are totally lost. This is because the encrypted data are saved as different contents by applying different encryption keys. Existing industrial solutions fail in encrypted data deduplication. For example, DeDu [17] isan efficient deduplication system, but it cannot handle encrypted data. Reconciling deduplication and client -side encryption is an active research topic [8]. Message-Locked Encryption (MLE) intends to solve this problem [5]. The most prominent manifestation of MLE is Convergent Encryption (CE), introduced by Douceuret al.[6] and others [7], [10], [12]. CE was used within a wide variety of commercial and research storage service systems. Letting be a file's data, a client first computes a key K HðMÞ by applying a cryptographic has function H to M, and then compute cipher text C EðK; M Þ via a deterministic symmetric encryption scheme. A second client B encrypting the same file M will produce the same C, enabling deduplication. However, CE is subject to an inherent security limitation, namely, susceptibility to offline brute force dictionary attacks [13], [14]. Knowing that the target data M under lying the target cipher text C is drawn from a dictionary S ¼fM1; . . .;Mng of size n, an attacker can recover M in the time for n¼jSj off-line encryptions: for each i ¼ 1;. . . ;n, it simply CE encrypts Mi to get a cipher text denoted as Ciand returns Misu chthat C¼Ci[9].

This works because CE is deterministic can d keyless. These curity of CE is only possible when the target data is drawn from a space too large to exhaust. Another problem of CE is that it is not flexible to support data access control by data holders , especially for data revocation process, since it is impossible for data holders to generate the same new key for data re-encryption. An image deduplication scheme adopts two servers to achieve verifiability of deduplication. The CE-based scheme described in combines file content and user privilege to obtain file token with token unforgeability. However, both schemes directly encrypt data with a CE key, thus suffer from the problem as described above. To resist the attack of manipulation of a data identifier, Mayetta. Proposed to adopt two servers for intra-user deduplication and inter deduplication. The cipher text C of CE is further encrypted with a user key and transferred to the servers. However, it does not deal with data sharing after deduplication among different users[16]. Cloud Deduplication aims to cope with the inherent security exposures of CE, but it cannot solve the issue caused by data deletion. A data holder that removes the data from the cloud can still access the same data since it still knows the data encryption key if the data is not completely removed from the cloud. Bell are et al. [1] proposed DupLESS that provides secure deduplicated storage to resistbrute force attacks.

III. PROPOSED METHODOLOGY

1. In proposed system we overcome the disadvantages of existing system.
2. Main aim in proposed system is to achieve data integrity and deduplication in cloud.
3. To achieve deduplication secure system is introduced called secCloud+.
4. secCloud+ guarantee about integrity auditing and deduplication of data.
5. It also provides security for the confidential data.

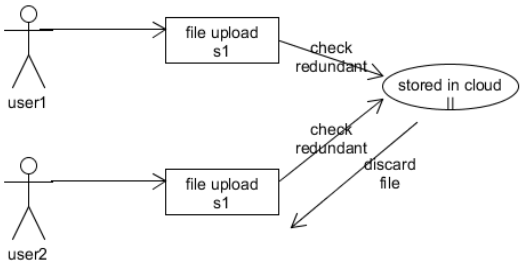


Fig 1:Proposed system

A. System Perspective

A Dekey is constructed in which users do not need to manage any keys on their own but instead securely distribute the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

convergent key shares across multiple servers. It incurs limited overhead in realistic environments we propose this new construction, which provides efficiency and reliability guarantees for convergent key management on both user and cloud storage sides. Dekey is proposed to provide efficient and reliable convergent key management through convergent key Deduplication and secret sharing.

Security analysis demonstrates that Dekey is secure in terms of the definitions specified in the proposed security model. Dekey remains secure even the adversary controls a limited number of key servers. We implement Dekey using the secret sharing scheme that enables the key management to adapt to different reliability and confidentiality levels. Our evaluation demonstrates that Dekey incurs limited overhead in normal upload/download operations in realistic cloud environments.

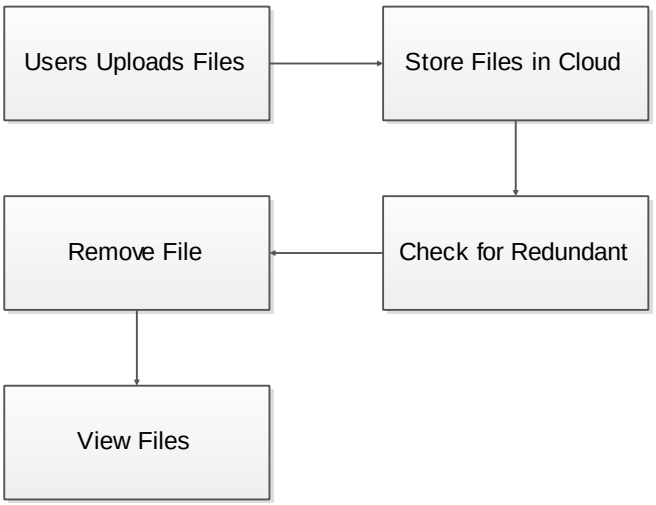


Fig 2: Block Diagram of proposed system

B. Advantages of proposed system

- Proposed system fixes issue of load on cloud that destroys the duplication files.
- To provide efficient memory storage and efficient access of data secure system secCloud+ generated.
- This secure system provides deduplication and secure auditing.
- The key feature and main advantage in proposed system is that it removes the duplicate files in cloud.

C. Architecture Diagram

Auditor: This refers to admin who have all privileges in this system. Managing users, managing files. Auditor is the one who checks whether data that user is uploading to cloud is already in the cloud and it is a redundant data. Auditor is the one who manages files and users and he needs to approve the user request for file if auditor approves the request then user can proceed otherwise not.

Cloud clients: clients refers to users who want the service of cloud. All the users before using the cloud service he has to register with his details and check for ownership to use the cloud service. If client want to upload any file to cloud client has to get permission from auditor. Auditor checks for redundant data then he approves for file upload.

Cloud Servers: Clouds servers are the storage space where all type of files will be stored that are required by the clients. Clients can download the files, upload the files to the cloud. It is a sharing space for clients.

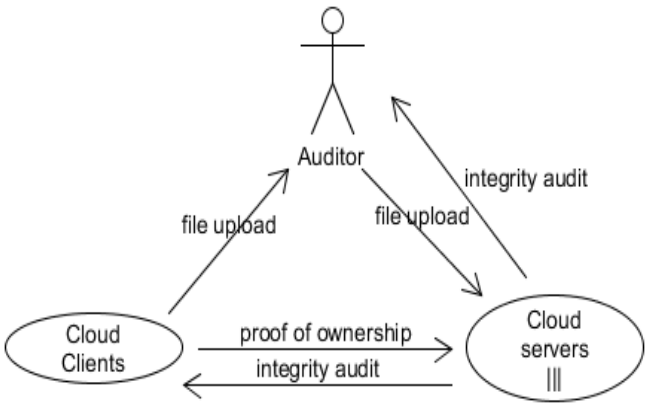
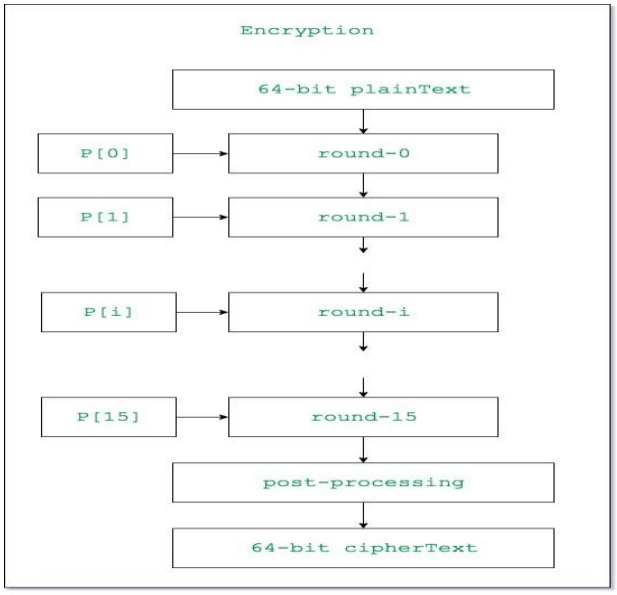


Fig 3: Architecture Diagram

D. Blowfish Algorithm

- Blowfish is a symmetric encryption algorithm, meaning that it uses the same secret key to both
- Encrypt and decrypt messages. Blowfish is also a block cipher, meaning that it divides a message up into fixed length blocks during encryption and decryption. The block length for Blowfish is 64 bits; messages that aren't a multiple of eight bytes in size must be padded.
- Blowfish consists of two parts: key-expansion and data encryption. During the key expansion stage, the in-putted key is converted into several sub key arrays total 4168 bytes.

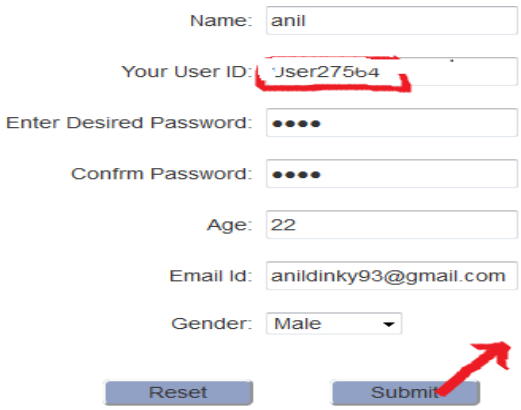


- They uploaded file will be extracted and contented will be past it is will be store in string builder

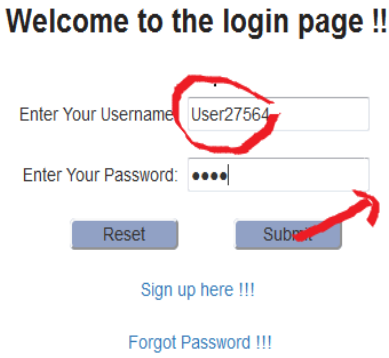
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021

- From the string builder data will be extracted it will be converted into bits
- The bits will be provided has input for the blowfish algorithm
- The algorithm obtained the bits has input and it will have the key will be generate .It is first round of blowfish,
- Once after the first iteration it will generate the passed key.
- This iteration will be continuous tell 16 rounds. After 16 rounds new key will be generated and it will be mapping to the document and it will be stored in database.

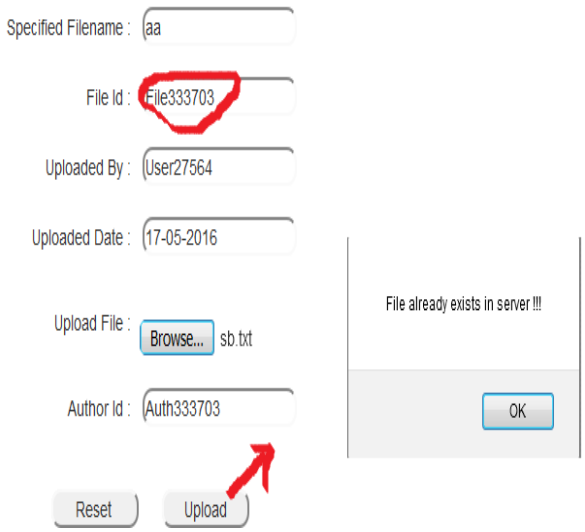
IV. EXPERIMENTAL RESULTS



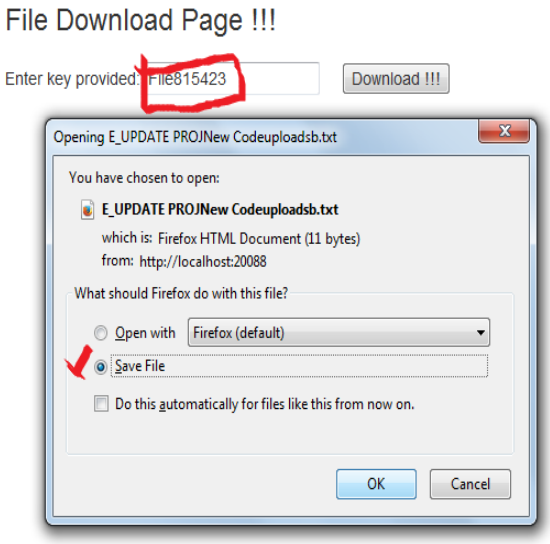
- User has to get registered by clicking on submit button after filling details. User Id is unique will be automatically generated and it is used as user name for login



- User will login by giving his unique id and password which was registered. After login home page will be presented. There user gets options like file upload, download, request file etc.

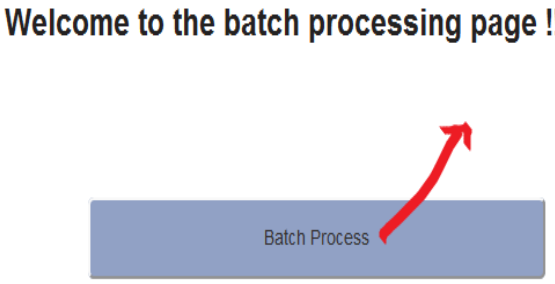


- To upload file details should be given. File Id is unique. Upload button should be pressed to upload files after details filled. No same files can be uploaded it gives error message



- This is download page where user needs any file to be downloaded he send request to particular user who uploaded file, when that user approves request current user gets key through email. That key is given to download file and select save

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 07, July-2021



- This is the work done by admin. Batch process is run to check for duplication of files. When this button is pressed if there is any same file it compares and recently uploaded file will be deleted. Duplicate files

V. CONCLUSION

In this project we have successfully implemented the software as it has to work for. Main purpose of this project is to delete duplicate files from cloud. This is achieved by comparing file streams of batch files and checking for duplicate files and deleting it. And many other process like sharing of files to cloud that is called as uploading files to cloud has been implemented. Downloading files is also implemented. This implementation requires permission from the owner who uploaded that file one has to request for file to download and if that user approves request then only others can download.

Security issues are solved by providing encryption techniques. The file will be encrypted using unique key. While sharing of file to the cloud the data will be encrypted and passed in channel. Data will be in the form of binary bits and together all will be joined and decrypted at the receiving end.

REFERENCES

[1] M.Bellare,S.Keelveedhi,andT.Ristenpart,"DupLESS:Serveraidedencryptionfordeduplicated Storage,"inProc.22USENIX Conf.Secur.,2013,pp.179–194.

[2] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[3] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] GoogleDrive[Online]Available:http://drive.google.com

[5] T.Y.Wu,J.S.Pan,andC.F.Lin,"Improving Accessing Efficiency of cloud storage using deduplication and Feedback schemes," IEEE Syst.J.,vol.8,no.1,pp.208–218,Mar.2014.

[6] G.Wallace,et al. ,"Characteristicsof backup workloadsIn production systems,"inProc.USENIXConf. File Storage Technol.,2012,pp.1–16.

[7] Y.-K.Li,M.Xu,C.-H.Ng,and P.P.C.Lee,"Efficient Hybrid in line and out–of-line Deduplication for backup storage," ACM Trans. Storage, vol.11, no.1, pp.2:1-21,2014.

[8] Ankush.R.Deshmukh,prof. R.V.Mante, Dr.P N Chatur," cloud based deduplication on encrypted data",I JIRSET, Vol. 6(1), jan2017.

[9] Open dedup. [Online]. Available: http://opendedup.org/,2016

[10] D.T.MeyerandW.J.Bolosky,"Astudyofpractical deduplication,"ACMTrans.Storage,vol.7,no.4,pp.1-20,2012,

[11] Opendedup. http://opendedup.org/

[12] TheFreenetProject, Freenet, (2016). [Online] Available :https: //freenet project.org/

[13] J.R.Douceur,A.Adya,W.J.Bolosky,D.Simon and M.Theimer,"Re-claiming space from duplicate files in a server less distributed file system",inProc.IEEEInt.Conf.Distrib Comput.Syst.,2002,pp.617–624,

[14] M.Bellare,S.Keelveedhi and T.Ristenpart,"Message-Locked encryption and secured duplication,"inProc.Cryptolo EUROCRYPT2013,pp.296–312.

[15] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart. Message–locked encryption and Secured duplication.In *Advancesin CryptologyEUROCRYPT2013*,pages296–312.Springer,2013.

[16] P.PuzioR.Molva,MOnen,andS.Loureiro"ClouDedup:Securededuplication with encrypt data for cloudstorage," in Proc.IEEE Int.Cof.CloudComput.Technol.Sci.,2013,pp.363–370,doi:10.1109/CloudCom.2013.54.

[17] Z.Sun, J.Shen, and J. M. Yong,"DeDu: BuildingA deduplication storage system over cloud computing," InProc.IEEEInt. Conf.Comput.Supported Cooperative WorkDes.2011,pp.348–355.

[18] John R Douceur, AtulAdya, William J Bolosky, P Simon,and Marvin Theimer. Reclaiming Space from duplicate Files in a Server less Distributed filesystem.In Distributed Computing Systems, 2002. Proceedings. 22nd International Conferenceon, pages 617–624.2002.

.