International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

STUDY THE PROBLEM OF SECURE DE-DUPLICATION ON CLOUD DATA, ALSO ENSURING INTEGRITY

M HARINADH[1], SYED RIYAZUL HAQ[2], S.SUDHEER REDDY[3], KANNEBOYINA MOUNIKA[4], KOMMULA PREMKUMAR[5], MOHAMMAD HASEEB UL SHOAIB[6]

[1] Assistant Professor, Department of CSE, Malla Reddy Engineering College (Autonomous), Hyderabad, India

[2] Assistant Professor, Department of CSE, Malla Reddy Engineering College (Autonomous), Hyderabad, India

[3] Assistant Professor, Department of CSE, Malla Reddy Engineering College (Autonomous), Hyderabad, India

[4,5,6] UG Student, Department of CSE, Malla Reddy Engineering College (Autonomous), Hyderabad, India

Abstract: The cloud computing technology came into existence during the21st century; outsourcing data to cloud service for storage becomes a useful yet efficient trend, which benefits in sparing efforts on data maintenance and management. Nevertheless, since the outsourced cloud storage service is not fully trustworthy, it raises security concerns on how to realize data de-duplication in cloud while achieving integrity auditing. In this work, we study the problem of secure de-duplication on cloud data, also ensuring integrity. Specifically, aiming at achieving both data integrity and de-duplication in cloud, we propose a system, namely D-cloud. D-cloud introduces an auditing entity with maintenance of the cloud, which helps generate hash value before uploading as well as audit the integrity of data having been stored in cloud. Compared with previous work, the computation by user in D-Cloud is greatly reduced during the file uploading and auditing phases. D-cloud is designed realizing the fact that users always want to encrypt their data being uploaded, and enables integrity auditing and secure de-duplication on encrypted data.

Keywords: Secure Auditing, De-Duplicating Data, D-cloud, Integrity, Encrypted Data.

I. INTRODUCTION

The use of computers has started long time back. Those days the computer was just used forperforming arithmetic and logical operations. But as the world evolved with inventions andinnovations, more and more data got generated eventually. Then there was the use of Harddriveto store useful data, which was very costly. Birth of Internet provided various technologies including Cloud Storage [1]. As we all know cloud storage is basically storing of data (Image, Videos, File, etc.) on a virtual server or we can say on a virtual database. Technically a cloudcomputing/storage is further explaining as, a system for enabling convenient on demand networkaccess to share data between computers. It is an internet based service which helps to store

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

data by managing the storage. [1] Cloud Storage provides the users ranging from cost saving from andsimplified convenience, to mobility opportunities and scalable services. According to somesurvey, the volume of data in cloud is expected to achieve many trillions of gigabytes. Even though cloud storage system has been extensively used, it ignores some important emerging needs suchas the abilities of [2] auditing integrity of cloud file and detecting duplicated files by cloud servers [9].The second problem is solved using deduplication [9]. The rapidadoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers which also causes similar (duplicate) files beingstored at multiple locations, wasting the memory resource [4]. This problem is countered by atechnology namely deduplication[6],in which the cloud servers would like to deduplicateby keeping only a single copy for each file (or block) and attach a link to this file (or block) for every client who owns it [3] . This generates the problem of transparency, which need not be compromised i.e. no two ownersof the same duplicate file should be aware of ownership by other client as well and also of the deduplication being performed on his/her data.In this paper, aiming at achieving deduplication with standard security and data integrity, we propose a secure system named Dcloud.

Dcloud generates unique hash values for each file being uploaded in the cloudstorage.This hash value is used for identifying duplicate data items. Also the hash value is usedas checksum to check the integrity of files. Dcloud also provides encrypted storing of files overthe cloud for security reasons.

II. BACKGROUND

2.1 Traditional System The traditional approach is the approach which is preferred by consumer who emphasizes more on instant acquisition of the book rather than price. The traditional approach is also seen to be the most preferred method advised by the school advisor during the enrollment into the program.

Disadvantages: The major disadvantage of existing system is the storage; it is very costly to buy hard drive nowadays (1000GB=1TB) is also not sufficient. The existing system depends upon sharing of data through physical world. The system is time consuming and expensive and have to rely on hard-drive and still not finding appropriate data. The data integrity and ownership is mainly check by HASH value, dueto this entire data to be checked which consumes a lot of time.

2.2 Proposed System All the existing applications discussed are kind of more commercial and money making, but this web application is different. Mainly this web application deals with the important factor like De-duplication, Security, Integrity and Availability. The sharing of data is easy but the one thing we should take care of is the security because we don't want anybody should see our data in the cloud without permission of the primary user. Here using Cloud storage[1], it will help the users to store their data on a network and can

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

retrieve it easily from their when it is needed and don't need to store it on hard-drive. Also the reason for making this web application is that, the data in cloud is not fully trustworthy and raise security concerns. The high cost of data storage devices and the use of data rapidly make us to use cloud storage

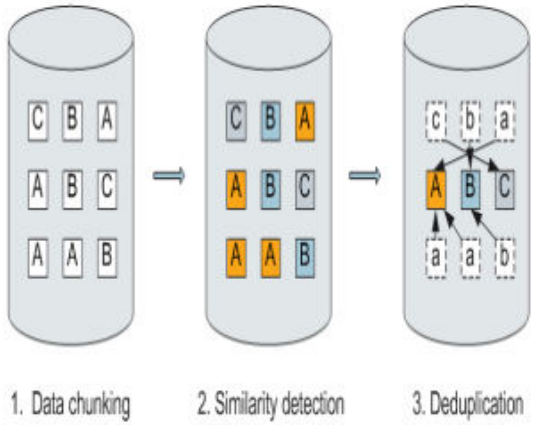
Advantages:

The main advantages of this system are as follows:

1. Security
2. Data Integrity
3. Availability
4. Accessibility.
5. Cost Saving
6. Reduction of Storage
7. File Sharing

III. RELATED WORK

Process of De-Duplication



3.1) Integrity Auditing:[1]

Cloud Storage move the user data to large data centres, which are remotely located,which causes constant data movements. During this process of uploading and downloading, the data could deliberately be tampered or getcorrupt in between. Due to this there is a lot of security issues as the user don't have control over it. Here with the help of our system, we will check the correctness of the data [10] .

3.2) De-Duplication: De-Duplication identifies and manages duplicate files in the cloud [6].As it happens with everybody; we at times upload the same file multiple times [2] . Also the same file can be uploaded by different users [4] . Due to this there is the duplication of file, resulting in wastage the scarce storage resource. [4] De-Duplication ensures duplicate data [6] is physically stored only once, and the proof of ownership of the with complete transparency provided to genuine owners.

3.3) Secured Data: The data being stored in the cloud is in an encrypted form to ensure standard security being provided [4].This helps to resist unauthorized users the access to sensitive and personal data [2] [7] . Also security is the major and important factor while storing data in cloud because the user once uploaded the data don't have any rights or authority on that file [7] .

IV. DCLOUD

System Architecture:

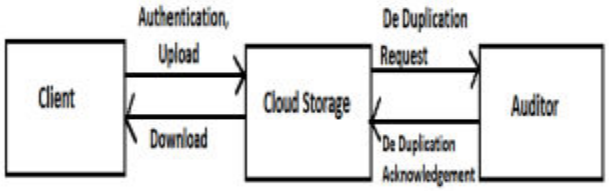


Fig: - System Structure

Fig: - System Structure

There are three main entities in the system- Client user, Cloud Server and the Auditor.

The user directly interacts with the Cloud by registration, authentication [11], upload and download. The Cloud system generates hash

value for each file being uploaded and stores the same. [9] The Cloud system checks for duplicate data items in cloudon the basis of their hash value and notifies and sends deduplication request to the Auditor [11] . The Auditor interacts with the Cloud system only [11]. Upon receiving the deduplication request, it performs suitable action over it. Through GUI the user will interact with the system. It GUI allows the use of icons or other visual indicators to interact with users.There is a database in the back end with the cloud which handle all the crucial data

Algorithms:

MD5 algorithm: Hashing is done to generate a unique hash value for each data item being uploaded [12]. This hash value is used for auditing purpose to identify duplicate files[8] . The same hash value is used as a checksum to ensure integrity of data [12] .

AES algorithm: AES algorithm is used for encrypting the user data stored in the cloud. This ensures standard security measures[9] .

V. CONCLUSION

The major goal of this web application is to help the users to store their data on the cloud with confidentiality and security. De-duplication of data is the main focus in the entire web application. Providing storage of data on a large scale with multiple file sharing. Auditing helps the user to check the integrity of the data.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp. 50–58, 2010[1] .

[2] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM, 2011, pp. 491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicate storage," in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online].

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 12:1–12:34, 2011.

[7] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of

the 4th International Conference on Security and Privacy in Communication Netowrks, ser. SecureComm '08. New York, NY, USA: ACM, 2008, pp. 9:1–9:10.

[8]https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fwww.cs.sjsu.edu%2F~stamp%2FS265%2Fprojects%2Fpapers Spr03%2FMD5.ppt

[9] AnthonyVelte& Robert C.Elsenpeter "Cloud Computing a Practical Approach", McGraw-Hill, Inc. New York, NY, USA ©2010

[10] R Sravan Kumar &A.Saxena "Data Integrity and Proofs in Cloud Storage"

[11] E. Mykletun, M. Narasimha, and G. Tsudik, "Authentication and integrity in outsourced databases," Trans. Storage, vol. 2, no. 2, pp. 107-138, 2006.

[12] A.K.Dubey, N Namdev and S.S.Shrivastava "Cloud-user security based on RSA and MD5 algorithm for resource attestation and sharing in java environment".