ISSN(Online): 2319 - 8753
ISSN (Print) :2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

Enabling Secure and Dependable Client Side Deduplication in Large Cloud Computing

Anushree R [1], Dr Shylaja B S [2]

P.G. Student, Department of Computer Network and Engineering, Dr AIT College, Bangalore, Karnataka, India[1]

Professor & HOD, Department of Information science and Engineering, Dr AIT College, Bangalore, Karnataka, India[2]

ABSTRACT-Since from recent years, lot of corporate and private users outsource their data to cloud storage providers. Cloud storage services use deduplication which means elimination of redundant data. So it reduces the storage space and bandwidth consumption of data storage services, but brings many security and privacy implications. There are some issues in cloud computing, but security and privacy are most important issues which need to be considered for the public cloud environments. A new client-side deduplication scheme has been proposed for securely sharing and storing data via the public cloud. The proposed method has twofold. Firstly, higher confidentiality is being ensured from unauthorized users. That is, every consumer computes a per data key in order to encrypt the data which intends to store in the cloud. So the data owner manages every data access. Secondly, metadata file is being integrated with access rights, so approved user can access the file which is encrypted with his own private key only.

KEYWORDS: Cloud Storage, Confidentiality,Data Security, Deduplication, Proof of Ownership**.**

I. INTRODUCTION

These days, the explosive development of digital contents keeps on raising the demand for network capacity and additional storage, and also an expanding requirement with less cost for the use of storage and network bandwidth in order to transfer data.

Regardless of these critical preferences in saving resources, deduplication brings numerous security issues, significantly because of the multi ownership challenges. Case in point, a few attackers target either the utilization of bandwidth or the privacy. For instance, a client may check whether another client has effectively transferred a record, by attempting to outsource the same document to the cloud.

As of late, to moderate these concerns, numerous efforts have been made to propose diverse security models. These different schemes are called Proof of Ownership system (PoW). The storage server is permitted to check a client data ownership based on hash esteem. Despite that the current PoW schemes have tended to different security properties, regardless still require a cautious thought of potential attacks, for example, leakage of data and poison attacks.

So a new cryptographic strategy has been proposed for secure Proof of Ownership (PoW). In order to overcome security issues in storage, this method use convergent encryption and also the Merkle-based Tree. This method is successful in providing dynamic sharing between users. Using the Merkle-based Tree for the data which is encrypted derives an identifier which is unique. This identifier allows checking the presence of the same data in remote cloud servers. And thus efficient data deduplication is achieved.

Paper is organized as follows. Section II describes related work.Then design of the module components and system architecture is described in Section III. Section IV presents experimental results showing results for upload and download delay. Finally, Section V presents conclusion alone with future enhancement and references.

ISSN(Online): 2319 - 8753
ISSN (Print) :2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

II. RELATED WORK

In the paper [1], Pairing-based cryptography method has become a one of the highly active research area. This method define bilinear maps or pairings and also shows how new cryptosystem is being developed with new functionality. The required pairing exist in hyper elliptic curve which is only one known mathematical setting The only curve which is used in practice is elliptical curve that are the most simplest case. All existing implementations of pairing-based cryptosystems are built with elliptic curves. Likewise, a brief outline of elliptic curve and capacities known as the Tate and Weil pairings from which cryptographic pairings are determined.

In the paper [2], Deduplication which is a broadly utilized procedure as a part of capacity administrations, since it manages an extremely efficient utilization of assets being particularly successful for customer level storage administrations. Deduplication has been indicated to experience the ill effects of a few security shortcomings, the most serious ones empowering a pernicious client to get ownership of a document it is not qualified for. Standard answers for this issue oblige clients to demonstrate ownership of information before its transfer. Lamentably, the plans proposed in the writing are extremely burdened on either the server or the customer side.

In the paper [3], the farsite dispersed document framework gives accessibility by recreating every record onto numerous desktop PCs. Since this replication devours noteworthy storage room, it is critical to recover utilized space where conceivable. Estimation of more than 500 desktop document frameworks demonstrates that about 50% of all devoured space is involved by copy documents.

In the paper [4], data deduplication and different systems for decreasing storage utilization assume an imperative part in moderately dealing with today's touchy development of information. Upgrading the utilization of capacity is a piece of a more extensive technique to give a proficient data base that is receptive to element business necessities.

In the paper [5], the quick selection of cloud services has moved system data sharing and progressing working costs for force, cooling, and work can likewise be lessened in light of the fact that there is less gear to work and oversee. Expanding the proficiency and adequacy of their storage surroundings helps organizations uproot imperatives on information development, enhance their administration levels, and better influence the expanding amount and mixed bag of information to enhance their aggressiveness storage. This paper method helps to minimize data transmission and space expected to transfer and store copied information. But the current arrangements, then again, were as of late discovered to be helpless against assaults that empower the assailants to get full access to the whole record put away on the server. And hence secure and dependable Client side deduplication is proposed to address all this issues.

III. SYSTEM ARCHITECTURE

Cloud Computing aim to drive the design of the next generation data centers by architecting them as a network of virtual services, so that users can access and deploy applications from anywhere in the world on demand at competitive costs. The figure 1 shows the high level architecture of the proposed methodology. There are basically four main entities:

Data provider

In this module, the data provider is responsible for creating Remote user by specifying user name. Data provider will automatically generate the password. The Data provider uploads their data to the cloud server. For the security purpose the data providers encrypts the data file, then divides the file, generate meta data(hmac) based on content of file and then finally stores in the cloud in parts (splited form). The provider keeps a copy of Meta data for checking dedupliction.

Cloud Server

The cloud server is responsible for data storage and file authorization for an end user. The data file will be stored in user data base and Backup DB in blocks with their tags such as file name, secret key, hmac1, hmac2, hmac3, hmac4, hmac5 and owner name. The data file will be sending based on the privileges. If the privilege is correct then the data

ISSN(Online): 2319 - 8753
ISSN (Print) :2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

will be sent to the corresponding user and also will check the file name, end user name and secret key. If all are true then it will send to the corresponding user or he will be captured as attacker.

Cloud Data Backup

Cloud Data backup is nothing but the Backup Database, The data backup start processing only when client requests for fetching the data which is stored previously in the cloud storage. The data backup has the following messages during its processing:

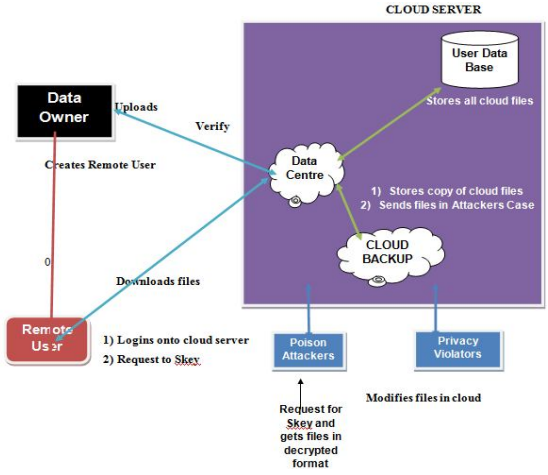

Figure 1: High level Architecture diagram

ClientRequestBackup: This message will contain requested data URL that the client wants to fetch. After receiving the client request to fetch the data CSP checks for the ownership of requested file and in response produce ResponseBackup message.

ResponseBackup*:* This response message of CSP contains the encrypted file in splited Meta data form. Once after receiving the ResponseBackup message, the client first retrieves the metadata file in splitted form and deciphers the data using secret key.

Data Consumer (End User)

The data consumer is nothing but the end user who will request and fetch file contents from the corresponding cloud servers. If the file name and secret key, access permission is correct then the end is getting the file response from the cloud or else he will be considered as an attacker and also he will be blocked in corresponding cloud. If he wants to access the file after blocking he wants to unblock from the cloud.

ISSN(Online): 2319 - 8753
ISSN (Print) :2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

IV. EXPERIMENTAL RESULTS

Figures In this section the results from extensive experimental evaluation is presented.

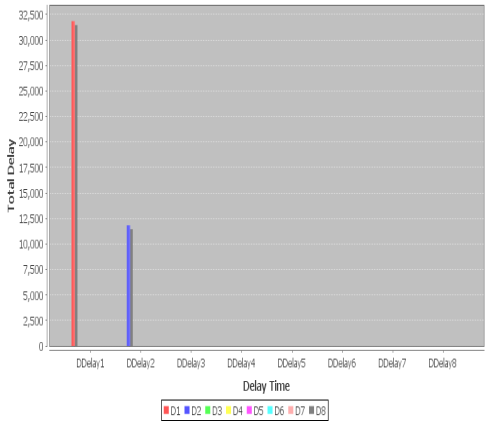

Figure 2 : Download delay time

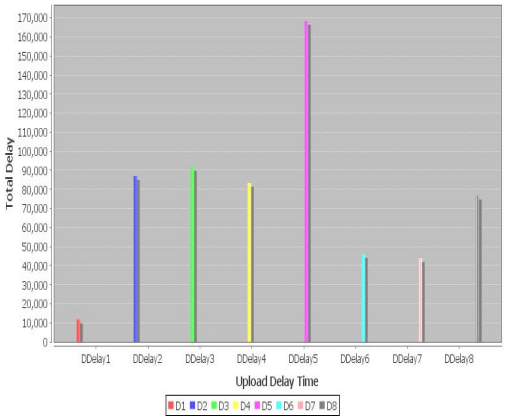

Figure 3 : Upload delay time

By analyzing Figure 2 and 3,the results can be concluded that:

- The time required to upload a given file in the cloud is always greater than the time required to download it from remote servers.
- For a given data size greater than $5*10^4$ bits, the time required in order to upload the file in the cloud increases by increasing the size of the file.
- For a given data size greater than $5*10^4$ bits, the time required in order to downloadload the file in the cloud increases by increasing the size of the file.
- Since mata data is generated for each file it saves respective delay for uploading the same file which checks for deduplication.

Our idea helps users to deploy private cloud on their datacenters. It is evaluated using Java programming language and private cloud severs that is Rackspace. This method is implemented and shown that minimizing utilization of resources that is bandwidth consumption and also storage. This also helps to achieve securely transmittion of file between client and server where deduplication along with security is achieved.

ISSN(Online): 2319 - 8753
ISSN (Print) :2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2015

V. CONCLUSION

In order to achieve security and efficiency in cloud storage services, convergent cryptography is used. Encryption alone will not overcome security issues. And hence this proposed procedure will encrypt the file to be uploaded, split the encrypted file and then generate the meta data based on content. Meta data is different for different contents and is not readable. Single copy of mac file is kept in client side so that file deduplication can be checked which helps for preexistence of data in server. So that it saves bandwidth utilization and storage in the server. For efficiency purpose file is also uploaded to the backup server. Remote user normally download file from user database after entering decrypting key. While file is attacked the user downloads file from backup data base. The attacker's details can also be retrieved from server. This proposed solution is very much resistance towards unauthorized user and data disclosure while transferring a file to cloud. So the proposed method is successful in achieving both deduplication check and secured sharing. Since SHA 512 algorithm is used in generating meta data of 512 bytes which is highly impossible for brute force technique.

At long last, cloud information storage security is still brimming with difficulties and of foremost significance, and numerous examination issues stay to be recognized.

VI.FUTURE ENHANCEMENT

Here the data owner can upload the file securely whatever the size but the only the text format. In future work it can be implemented for the audio, video, and other type of file which can become great work.

REFERENCES

[1] L. Ben. On the implementation of pairing-based cryptosystems, 2007.
[2] R. Di Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security,ASIACCS '12, pages 81–82, New York, NY, USA, 2012. ACM.
[3] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system.In Proceedings of 22nd International Conference on Distributed Computing Systems (ICDCS , 2002.
[4] M. Dutch. Understanding data deduplication ratios. SNIA White Paper, June 2008.
[5]. 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 , CCS '11, pages 491–500, New York, NY, USA, 2011. ACM.
[6] D. Hankerson, A. J. Menezes, and S. Vanstone. Guide to Elliptic Curve Cryptography. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.
[7] D. Harnik, B. Pinkas, and A. Shulman-Peleg. Side channels in cloud services: Deduplication in cloud storage. IEEE Security And Privacy,8(6):40–47, 2010.
[8] R. C. Merkle. A digital signature based on a conventional encryption function. In A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology, CRYPTO '87, pages 369–378, London, UK, UK, 1988. Springer-Verlag.
[9] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In Proceedings of the 27th Annual ACM Symposium on Applied Computing , SAC '12, pages 441–446, New York, NY, USA, 2012. ACM.
[10] M. W. Storer, K. Greenan, D. D. Long, and E. L. Miller. Secure data deduplication. In Proceedings of the 4th ACM International Workshop on Storage Security and Survivability, StorageSS '08, pages 1–10, New York, NY, USA, 2008. ACM.
[11] C. Wang, Z. guang Qin, J. Peng, and J. Wang. A novel encryption scheme for data deduplication system. In Communications, Circuits and Systems (ICCCAS), 2010 International Conference on, pages 265–269, 2010.
[12] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. InProceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ASIA CCS '13, pages 195–206, New York, NY, USA, 2013. ACM. IEEE Transactions on Mobility and Security (NTMS) in Cloud Computing,Issue Date:April.2.2014