IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A REVIEW ON SECURELY ELIMINATE DUPLICATE DATA AT CLIENT SIDE IN PUBLIC CLOUD STORAGE SPACE

Mr. Abdul Mutkabir Shaikh
[*] Student Pursuing M.TECH. in Computer Science & Engineering,GNIET,Nagpur, India.

DOI: 10.5281/zenodo.49702

ABSTRACT

Now a day cloud computing is very popular and it is spread tremendously all over the world. Due to increasing large amount personal data in the cloud environment there are some issue for handling the bulk data in public cloud space. Data de-duplication is important technique for data compression which is used to eliminate the duplicate data in the cloud environment. In cloud environment client sent the data to public cloud at the same time client doesn't known the data, which is already repeated. Hence apply the data de-duplication technique at client side for reduce the redundancy in there data. De-duplication at Client side technique is used to identify duplicate data already at the client and save the bandwidth of data and uploading selected files to the server. Convergent encryption is another technique which is used to better protect the security of data at client side. This technique is used to encrypt the data before outsource to the data storage server and it should be authorized When a user uploads a file on the cloud, the file is split into a number of blocks. Each block of file which is encrypted using convergent key and subsequently token will be generated using token generation algorithm.After encrypting data using convergent key then cipher text is form, these cipher text send to cloud before user retain a key. The deterministic nature of encryption, when the identical data will be uploaded with same convergent key and same cipher text then de-duplication scheme prevent the duplicate data. After comparing the data base if match is found then only metadata of block store in Database profiler.

KEYWORDS: Cloud Security, De-duplication, Proof of Ownership, Cloud storage, authorized duplicate check, Convergent Encryption.

INTRODUCTION

Cloud computing is that the centralized storage for the information and it is additionally provides the web access to various computer system. Cloud computing generally focuses on increasing the efficiency of shared resources. Clouds may be classified with users into two types as private and public. Public cloud is that the cloud that is formed by Pay-as-you-use manner to overall public, and their service being sold is utility computing. Private cloud won't be available to general public, they are management over the company's data and it ensures the safety and also having the larger potential risk for data loss. Cloud user faces various security threads from inside and outside the cloud, and that they are responsible for the application level security [7]. One of the tough challenges in distributed system is to remove the bugs in a cloud storage environment. There are some same cloud services namely are Dropbox, Wuala, Memopal and Google drive which is used to stored data at remote places, apply for these services client side de-duplication scheme ([3], [6]). Data de-duplication is the one of the best data compression techniques for removing the duplicate copies of repeated data and it has been generally used to reduce amount of space for storing data and also save bandwidth in cloud environment. This concept avoids the storage of redundant data in cloud servers and reduces network bandwidth consumption associated to sending a similar contents many times. To check the level of duplicate data have two types that is file level and block level de-duplication. In file level de-duplication remove the duplicate copies of similar file and at de-duplication will be occur at block level within the occurrence of non-identical files wherever it removes duplicate blocks of data[4]. Because of this they needed minimum space and containing lot of benefit over the system. Client-side de-duplication scheme apply at Owner/User side, in that duplicate data is check first only identified before it has to be sent over the network. This will be definitely burden on the central processing unit however at similar time reduce load of the network. It is proposed for reduce bandwidth as well as minimum space required to upload the data. In this paper use the new cryptographic methodology for secure proof of ownership(POW) based on the joint use convergent encryption for improving data security in cloud storage systems, providing dynamic sharing between users and guarantee of data de-duplication([1],[2]).

MATERIALS AND METHODS

RELATED WORKS

There are some related work which related with security and privacy issues in the cloud and also we discuss here work which is similar techniques as our approach.

A. Security Analysis

Despite the significant resource saving advantages, PoW schemes bring several security challenges that may lead to sensitive data([4],[8]).

- *Data confidentiality* **–** hash-as-a-proof schemes (e.g. Dropbox) introduce an important data confidentiality concern, mainly due to the static proof client side generation. For instance, if a malicious user has the short hash value of an outsourced data file, he could fool the storage server as an owner trying to upload the requested data file. Then, he gains access to data, by presenting the hash proof. As such, an efficient PoW scheme requires the use of unpredictable values of verifications.
- *Poison attack* **–** when a data file D is encrypted on the client side, relying on a randomly chosen encryption key, the cloud server is unable to verify consistency between the uploaded file and the proof value hash. In fact, given a pair (hashD;Enck(D)), the storage server cannot verify, if there is an original data file D, that provides a hash value hash. As such, a malicious user can replace a valid enciphered file with a poisoned file. So, a subsequent user looses his original copy of file, while retrieving the poisoned version.

This system first calculate the cryptographic hash value of the plaintext message then it will encrypts the plaintext by using its hash value that is generated as a key. Finally, the hash value (key) itself is encrypted with the key chosen by the user indeed shown below in fig 1.

Figure 1:

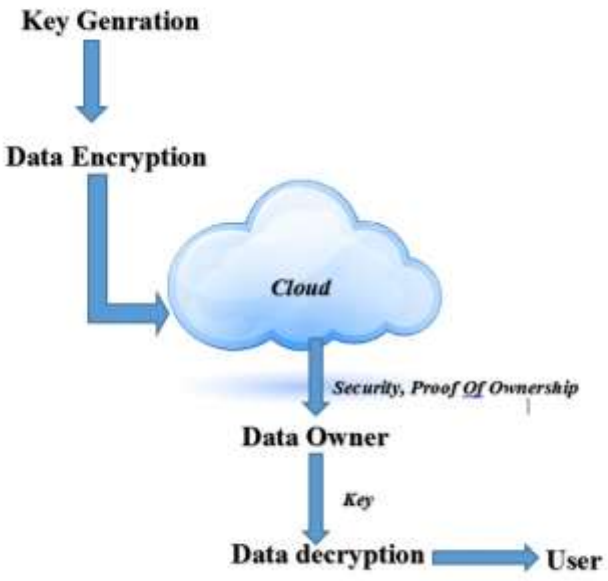

Fig 1.Encryption process

Convergent encryption scheme can be defined in with four primitive functions as shown below.
- KeyGeneration.(M) :- It is the key generation algorithm i.e. SHA-1 that generates key k using the data copy message M.

- Encryption.(k,M) :- It is the symmetric key encryption algorithm i.e AES. It takes the key k as encryption key and encrypts the plaintext message M. It produces C as the cipher text of message M.

- Decryption.(k,C) :- It is the symmetric key decryption algorithm. It takes the key k as decryption key and decrypts the cipher text C. It produces the original plaintext M from cipher text C.

- TagGen.(M) :- It is the file tag generation algorithm that maps the original data copy M and outputs a tag T(M).

B. Related Works

In order to prevent private data leakage, Halevi et al. [2] proposed the concept of Proof of Ownership (PoW), while introducing three different constructions, in terms of security and performances. These schemes involve the server challenging the client to present valid sibling paths for a subset of a Merkle tree leaves [3]. The first scheme applies erasure coding on the content of the original file. This encoded version is the input for construction of the Merkle tree. The second purpose pre-possesses the data file with a universal hash function instead of erasure coding. The third construction is the most practical approach. Halevi et al. design an efficient hash family, under several security assumptions.Recently, propose a PoW scheme over encrypted data. That is, the file is divided into fixed-size blocks, where each block has a unique commitment. The hash-tree proof is then built, using the data commitments. Hence, the owner has to prove the possession of a data chunk of a precise commitment, with no need to reveal any secret information. However, this scheme introduces a high computation cost, as requiring generation of all commitments, in every challenging proof request. In [1], the authors presented an efficient PoW scheme. They use the projection of the file into selected bit-position as a proof of ownership. The main disadvantage of this construction is the privacy violation against honest but curious storage server. In 2013, Jia et al. address the confidentiality preservation concern in cross user client side deduplication of encrypted data files. They used the convergent encryption approach, for providing deduplication under a weak leakage model[6].

SYSTEM MODEL

In practice, the cloud service provider provides a web interface for the client to store data into a set of cloud servers, which are running in a cooperated and distributed manner. In addition, the web interface is used by the users to retrieve, modify and restore data from the cloud, depending on their access rights. Moreover, the cloud service provider relies on database servers to map client identities to their stored data identifiers and group identifiers.Figure 2 show a descriptive network architecture for cloud storage. It relies on the following entities for the good management of client data:
1.Users*:* the users are able to access the content stored in the cloud, depending on their access rights which are authorizations granted by the client, like the rights to read, write or re-store the modified data in the cloud. These access rights serve to specify several groups of users.

2. Web server: Web server is give request and response to user. It also store, process and deliver the web pages to client. When User want to upload file to cloud storage through web server. The communication between client and server takes place using the Hypertext Transfer Protocol.

3. Web client: A client makes use of provider's resources to store, retrieve and share data with multiple users. A client can be either an individual or an enterprise Web client connect to server and retrieve the web pages. Data owners want to divide file into multiple blocks. For each block it perform encryption operation and produced like cipher text, token and private key for each block.

4. Security Services: After all the response has been generated PKi is store into internal database of security service. The main Idea behind to hide PK is provide security to Cipher Text(Bi) , So no one else can used the key and try to decrypt the block.

Figure 2:

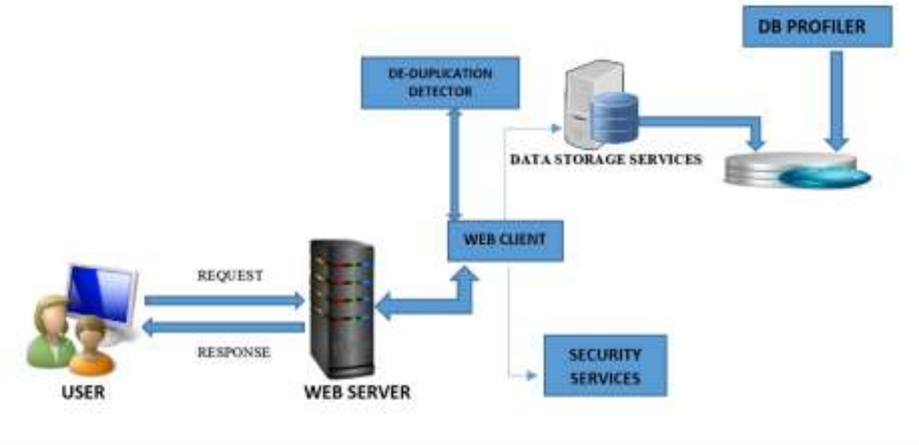

Fig. 2: Architecture of cloud data storage

5. Duplication detection: Security Service generates TiBi Token on basic on Bi , If the same Bi comes in then it will generate the same TiBi. Token generation algorithm is used. Then it will store the TiBi to the Own Security database. Now next time after generation of the code it will cross verify with the exiting token data and send back the notification accordingly.

6. DB Profiler: It stores uploaded data, shared data, all list of users, and sequence of token of blocks. It also stores all encrypted data and metadata of file also store in the database. The data storage server contain all the uploaded files and DB profiler store all the metadata of the file.

Case 1: When file F1 & file F2 are different the all the data will be store in the database in different blocks

Case 2: If the file F1 is equal to file F2 it stores only one file in the database avoid duplication of the data.

Case 3: If file F1 is belongs to file F2 then it compare the blocks with data storage and only different blocks of both file will be store in the database. For execution of Authorized duplicate system, first start different services which is used in cloud for deployment purpose.

RESULTS AND DISCUSSION

The authorized de-duplication system used to avoid duplicate copies of data within the given cloud. Proposed system implemented as file level de-duplication and block level de-duplication. In the file level de-duplication compare the file in available database and remove the duplicate one. In the block level de-duplication compare each block available in database, suppose the file is already stored in the database which same file uploaded by another user at that time only metadata of file will be store in the database. So it help to reduce the storage space of data and proper space utilization. The data will be store in encrypted format so it also maintains security because each block contains their own token, cipher text and private key. The database size will be reduced by using this technique. The proposed system has been compared with the existing system on the basis of database usage, and security using proof of ownership.

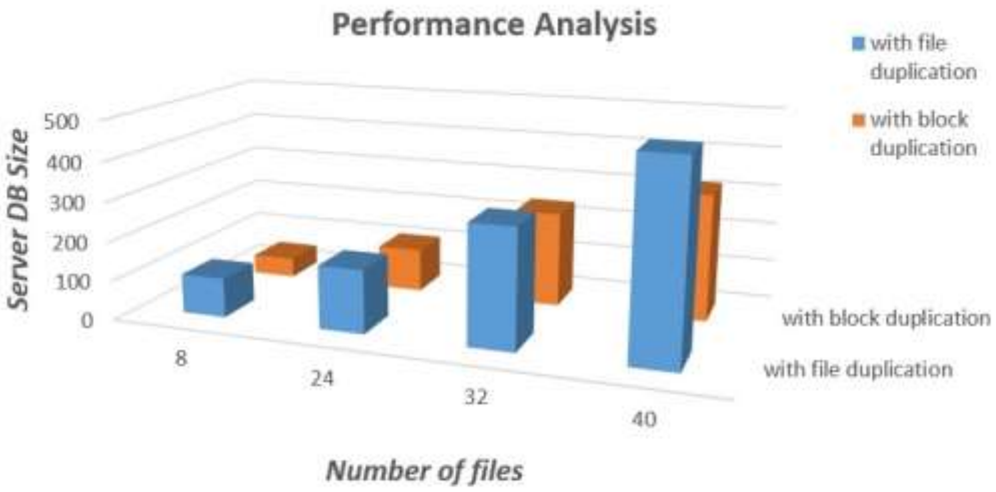

Fig.3 Comparison between file level duplication and block level duplication

The above Fig.3 show expected output from proposed system. X axis shows number of files and Y axis shows total database size It shows the storage space in the database system, comparison between file level duplication as well as block level duplication space required for storage data have different size. In case of file level duplication large space as compare block level duplication. For this purpose, use the block level duplication for reducing storage space in the database.

Table1.Actual Result Comparison

Sr. No	Number of Files	File level duplication DB Size	Block level duplication DB Size
1	8	100	50
2	24	150	110
3	32	300	260
4	40	490	320

The above Table1.Shows the database usage for file level duplication and block level duplication. The file level duplication having extra storage space as compare to block level duplication. The block level duplication having less storage space and also provide extra security using proof of ownership concept.

CONCLUSION

As the number of user's increases the amount of data that is stored in the storage environment increases and risk of the data also increases. In this paper, the main idea is reduce redundant data with the help of de-duplication scheme apply at client side with use encryption algorithm for secure upload/download file from public cloud. This helps in eliminating duplicate copies of repeating data, reduces storage space used and saves bandwidth in cloud storage. In this system have minimal overhead in entire upload and download process and is negligible for moderate file size. These system can be used by the client to manage his data stored in the cloud servers. . In proposed system proof of ownership protocol has been applied, it will help to implement better security issues in cloud computing environment.

REFERENCES

[1] 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.

[2] 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, CCS ’11, pages 491–500, New York, NY, USA, 2011. ACM.

[3] Nesrine Kaaniche and Maryline Laurent.A Secure Client Side Deduplication Scheme in Cloud Storage. IEEE Environments’6TH INTERNATIONAL CONFERENCE ON NEW TECHNOLOGIES, MOBILITY AND SECURITY, 2014.

[4] Raakesh and Varun Raj, Eliminating. Redundancy in File System Using Data Compression and Secured File Sh.aring International Journal of Computer Science and Information Technologies, Vol. 6 (3), 2015.

[5] K.Naga Maha Lakshmi and A.Shiva Kumar.Secure Data Deduplication and Data accessing among Multi-cloud Framework. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDSVOLUME 2, ISSUE 10, OCTOBER 2015, PP 687-693, 2015.

[6] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, ASIA CCS ’13, pages 195–206, New York, NY, USA, 2013. ACM.

[7] G. Kakariya and S. Rangdale. A Hybrid Cloud Approach For Secure Authorized Deduplication.International Journal of Computer Engineering and Applications, Volume VIII, Issue I, October 14

[8] K.Naga Maha Lakshmi and A.Shiva Kumar. Secure Data Deduplication and Data accessing among Multi-cloud Framework. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS,VOLUME 2, ISSUE 10, OCTOBER 2015, PP 687-693