National Conference on Computer Security, Image Processing, Graphics, Mobility and Analytics (NCCSIGMA)
International Journal of Advanced Engineering Research and Science (IJAERS) Special Issue (NCCSIGMA-16)
https://dx.doi.org/10.22161/ijaers/si.1 ISSN: 2349-6495(P) | 2456-1908(O)

Managing Deduplicate Data at Client Side in Public Cloud Storage Environment

Kranthi Kiran G[1], V Naresh Kumar[2]

[1]Department of CSE CMR Technical campus, Hyderabad, India
Email: cmr.kkg@gmail.com
[2]Department of CSE CMR Technical campus, Hyderabad, India
Email: nareshkumar99890@gmail.com

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–* *CloudSecurity, De-duplication, Proof of Ownership, Cloud storage, authorized duplicate check, Convergent Encryption.

I. 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 focus on increasing shared the cloud that is formed by Pay-as-you-use manner to overall public, and their service being sold is utility threads from inside and outside the cloud, and that they are responsible for the application level security 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 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 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 duplication ([1], [2]).

II. DETAILS OF EXPERIMENT

There is 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.

2.1. Proof of Ownership (PoW)

It is introduced by Halevi [1].It is the security protocol which allows checking server of user data using static value called as hash value. Whenever user want to upload the data file i.e. (D) to the cloud server, it have first calculate and send the hash value i.e. Hash=H (D) to the cloud storage server. In database contain list of received files compare with static hash value. If in this case match found, then data (D) is already exists remove from the cloud servers and maintain list of single copy of file. At this case cloud user as an owner of data with no need to upload file to storage server. If in this case no match found then the user has to send file data (D) to the cloud storage. This client side de-duplication, referred to as hash-as-proof [2].

2.2. Data De-duplication Technique

Data de-duplication is a specialized technique, which is used to eliminate the duplicate copies of the same data content. It is used to improve the storage space utilization and also it can be applied to network data transfer to reduce the amount of data that is to be transferred [4].

Post process de-duplication:

In this de-duplication process, new data is first stored on the storage environment and then it will analyse the data looking for duplication which is present or not. One of the potential drawbacks of post process de-duplication is that it may store the duplicate data unnecessarily for a short period of time which is an issue if the storage system is near the full capacity.

Inline de-duplication:

In this process where the de-duplication hash value calculations are created on the target device as the data enters the device in the real time. If the device spots a block of data that it already stored on the system then it does not store the new block of data.

2.3. Security and Privacy issues in the cloud

Only the authorized persons need to access the data from the cloud. In order to ensure the integrity of user authentication, the need of security mechanism usage of data in the cloud. In various cloud computing security challenges, it is the responsibility of the user to make sure that the cloud provider has taken all

Effective most widely used but when it is applied with the multiple users the cross-user de-duplication tend to have also several serious privacy problems. Cross-user de-duplication is the simple mechanisms which are used to reduce the risk of data leakage [8].

III. SYSTEM OVERVIEW

3.1. Data Encryption:

Encryption is the process of encoding the information from one format to other. Information is encrypted in such a way that only authorized users can access and read it. In an encryption process information is referred to as the plaintext. Encrypted information is the cipher text. The output of encryption is called the cipher text. Plaintext is encrypted using encryption algorithm that can be symmetric or asymmetric algorithm. Plaintext is encrypted using some keys and it is not possible to decrypt the cipher text without these keys. There are two types of Symmetric key encryption process [7].

1. Public key encryption process Symmetric key encryption: In symmetric key encryption process, the encryption and decryption key is same.
2. Public key encryption:
   In public key encryption method, the encryption key is published for anyone to use it as encryption key indeed. Only receiving parties will be having accessibility to the decryption key that enables messages to be read.

3.2. Convergent Encryption Technique:

Convergent encryption technique is also known as content hashing keyword method. It is a cryptosystem that produces identical cipher text from the identical plaintext data. This is mainly used in cloud computing to remove duplicate elements or files from storage environment without the provider having to access to the encryption keys. It generates the file tag, which is used to detect the duplicate files in the storage environment [8]. This system first calculates the cryptographic hash value of the plaintext message then it will encrypt 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.

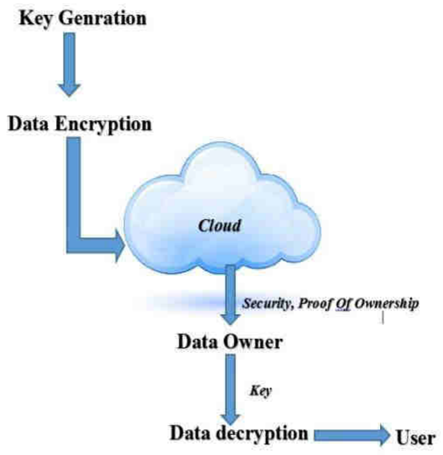


Fig 1: Confidential Encryption

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).

IV. SYSTEM DESIGN

Cloud services namely as Dropbox, and Google drive in which user upload and download file from the remote places, when the user want to upload, download, update and delete file from the cloud environment at the same time it required first authorized from the web server then user have permission to upload the file to web server for that purpose it use the proof of ownership algorithm [1]. When the user upload file it divide into number of blocks each block size is 4KB. Each block is encrypted with convergent key then it form cipher text .In database of cloud storage given block compare with all blocks if match is found it store only meta data of block[4]. Below Fig.2 System Architecture show systematic procedure of storing data without reputing one.

4.1. User:

It have first authorized from web server.

4.2. Web server:

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

4.2. Web client:

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 generate the response like cipher text, token and private key for each block.

4.3. 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.

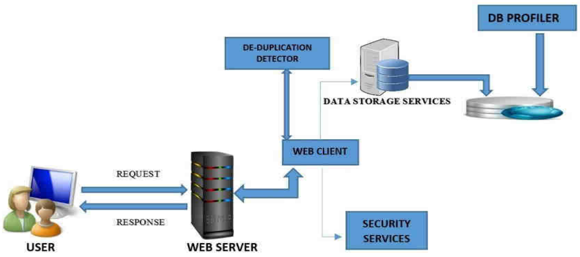


Fig. 2: Architecture of cloud data storage

4.4. Duplication detection:

Security Service generates TiBi Token on basic of 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.

4.5. 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 store in the database. The data storage server contains 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 are used in cloud for deployment purpose. 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 user at that time only metadata of file will be store in the database. So it helps 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.

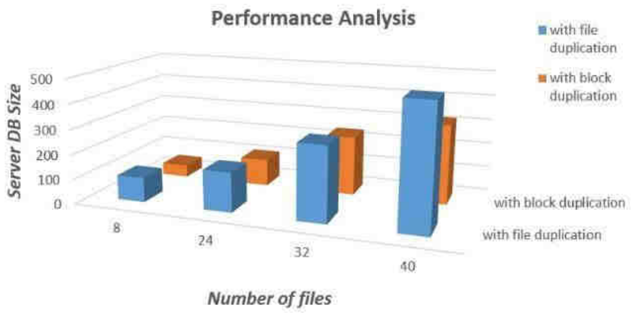
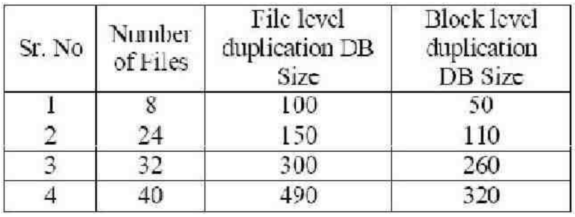


Fig.3: Comparison between file level duplication and block level duplication

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.

Table.1: 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.

V. CONCLUSIONS

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. This 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.