International Journal of Engineering &Technology, 7 (1.7) (2018) 105-108

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Secure and duplication detection in cloud using cryptographic hashing method

J.K. Periasamy [1] *, Dr. B. Latha [1]

[1] *Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India*
**Corresponding author E-mail: jkperiasamy@gmail.com*

Abstract

De-duplication systems are adopting De-duplication strategies such as client side or server side De-duplication. Particularly, in the beginning of cloud storage, data De-duplication techniques happen to add importance to store original volumes of data in the cloud. This technique motivates the enterprise and organization to farm out data storage to cloud service providers, as proof of several case studies was done. Block level de-duplication is used to discover, removes redundancies compare with previously stored information. The file will be separated into smaller segment as given by the system size or uneven size blocks or chunks whatever we need. Using predetermined size of blocks, the system can simplify the computations of block limits, although using uneven size blocks provides improved de-duplication. Secure Cloud introduce a new concept to audit entities with continuation of a Map Reduce cloud, which is used to help the client is easy to make data tags before feeding the data and audit with the probity of data is analysed and stored in the cloud. the subject of previously finished work is fixed so that the computational load at user or auditor has huge amount of tag making. In accumulation, Secure Cloud also enables secure de-duplication. Perceive that the "validity" measured in Secure Cloud is the deterrence of leakage, the side channel information. In order to check the leakage of such side channel data, the work accept the custom and design a proof of rights protocol amid cloud servers and clients, which permit clients to confirm the cloud servers what they exactly own the object information.

Keywords: *Data Chunks; Data De-Duplication; Duplication Detection; Hashing; Secure Cloud.*

1. Introduction

Cloud computing is a technology that offers numerous services which attracts the consumers and organizations at all levels. It is deployed using three models such as: public, private and hybrid each having its own conditions and restrictions. The services offered by the cloud falls under three main categories namely: Software as a Service (Saas), Platform as a Service (Paas) and Infrastructure as a Service (Iaas). Saas is an alternative way of accessing the software which eliminates the need of purchasing the software and loading the same onto a device. It is a subscription based model where the software is hosted in the cloud and can be accessed by the users via internet. Saas examples include Sales Force, Google Apps, Office 350 etc.Paas is another category of software service where the platforms and environments are rendered as service to the developers for building applications and services over internet. Windows Azure, Force.com, Google App engine are some of the Paas examples. Iaas provides virtualized computing resources which can be accessed through WAN (Wide Area Network) such as internet. It also provides a range of services that accompany the infrastructure components. Some independent Iaas providers are Amazon Web Services (AWS) and Google Cloud Platform (GCP).In addition to these, cloud offers a massive data storage which could be accessed from anywhere in the world. Thus cloud serves to a great tool of choice for many consumers. Wherever the data is stored, a question of security arises. Hence there is a great need for establishing security in cloud. The proposed system describes an efficient way to ensure security using cryptographic techniques. Before the data is sent to the cloud, it is encrypted, dispersed into slices and finally hashing is performed. This encoded data is transmitted through cloud. In the same way, retrieval of original data from the encoded form involves decoding where the sliced data is verified for its authenticity and then encrypted file is reconstructed with the help of IDA which is then decrypted to obtain the original data.

To remove the redundant data, data duplication is used. Used in cloud storage to decrease the space of storage and to upload and download the bandwidth. There is only one copy of each file stored in cloud and that file is owned by more numbers of users. The de-duplication system progress storage is utilized with highly reliable and flexible convince. Defy of privacy for sensitive data also take place when they are outsourced by users. Thus the work constructs the first effort to take the idea of scattered De-duplication system.

Thus to increase the consistency of data, is the main objective of this, in addition to achieve the privacy of the user's data and outsourced data. Thus the security of tag consistency and integrity were attained. The implementation of this system using the Ramp secret sharing scheme, it gives the demonstration of system efficiency rather than existing one. And also, it acquires small encoding/decoding overhead compared to the network transmission overhead in regular download /upload operations.

The overview of the work is used to reduce the storage space with secure manner and never get misused. The main motive of this work is used to save memory using de-duplication. Because saving the same file such as image or a video again and again can lead to more unnecessary usage of cloud and also waste of cloud storage both client side and server side. Moreover a cloud computing can be defined as the ability to practice of using a network of remote servers hosted on the Internet to store, manage, and process the data, instead of a local server or a personal computer.

Computers, resources and other devices that are enabled to provide with shared computer processing resources on the basis of on demand by cloud computing. It is a model for enabling on-demand access to a shared pool of configurable computing resources, which can be provisioned rapidly and released with minimal management effort. A number of capabilities are provided to cloud users and enterprises to store and process their data. This can be done in either privately owned, or third-party data centres such as cloud service provider that may be located far from the user–ranging in distance from across a city to across the world by Cloud computing.

After duplicates are recognized, then the duplicates will be truncated and remaining chunks of data sent to the storage server. There are different types of overheads will be occur on the storage period. To avoid these having been abandoned, a technique is used called as in memory filter and inherit the data locality to reduce the frequency of disk lookups, is proposed. The scattered chunks located in different storage place are bundled together into a single TCP connection and this method is named as bundling. The duplicate chunks are usually clustered and data locality preserved for already detected chunk is utilized to reduce the number of duplication questions raised in storage period. Moreover, the messaging back scheme helps the sender to recognize the duplicate hash value that has been sent from the receiver. The redundant chunk transfer will be reduced using Messaging Back Scheme.

2. Related works

Jiawei Yuan and Shucheng Yu [1] deal with two important requirements for cloud storage such as Data integrity and storage efficiency. Storage efficiency is improved by Proof of Ownership (POW). It securely removes unnecessarily duplicated data on the storage server. Non-trivial duplication of metadata is due to the trivial combinations of the two techniques which in turn contradicts the objectives of POW. Tremendous computational and communication costs have been introduced recently which also prove to be insecure. To secure data integrity a new technique by auditing with storage de-duplication for cloud storage was introduced. Here we solve this open problem with a magnificent scheme based on techniques which include polynomial-based authentication tags and homomorphic linear authenticators. De-duplication of both files and their corresponding authentication tags is possible in this design.

Shai Halevi, D. Harnik, B. Pinkas and A. Shulman-peleg [2], they explained about Cloud storage. All customers are storing their data and personal information in cloud. Client-side de-duplication try to identify de-duplication opportunities in the client side itself to save the bandwidth of uploading copies of existing files to the server and also identify attacks. In coordinate to our work the division of these problems was introduced in the wild with respect to drop box file synchronization service. To get rid of the problems the work has introduced the notion of proofs of ownership which helps the client to prove the server that the client owns the file rather than the minimal information.

The work introduces a model for provable data possession (PDP) that supports to check large amount data in remotely. Performance is increased compare the previous results [3-5].

The work introduce Dekey [5], User no need to maintain any keys in their own but instead securely distribute the convergent key shares across multiple servers. As a proof of concept, the work implements Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments.

The objective of this scheme lie in constructing a hybrid cloud framework, using convergent encryption algorithm to encrypt original files, and introducing differential privacy mechanism to resist against the side channel attack. Performance evaluation shows that our scheme is able to effectively save network bandwidth and disk storage space during the processes of data de-duplication. Meanwhile, security analysis indicates that our scheme can resist against the side channel attack and related files attack, and prevent the disclosure of privacy information. It provides differential privacy mechanism to resist against the side channel attack. It prevents the disclosure of privacy information. [8].

This work based on attribute-based encryption (ABE) to encrypt duplicated data and stored in the cloud while it is also supporting secure data access control. To preserve cloud data confidentiality and user privacy, cloud data are often stored in an encrypted form. The authors evaluate the scheme's performance based on analysis and implementation. Results show the efficiency, effectiveness, and scalability of the scheme for potential practical deployment. It provides flexibility to support secure data access control. The performance is based on analysis and implementation of several data. [9].

In the big data era, all smart devices may be generated duplicated copies of data. Although convergent key encryption has been explicitly adaptable for secure De-duplication, it is one of the challenging to efficiently manage data as well as keys to search encrypted data. This work explain the above problem by presenting an efficient scheme for big data outsourcing with secure De-duplication, named BDO-SD. Specifically, with the convergent encryption technique, data owners can outsource their data to the cloud server with data De-duplication. A user can query encrypted data while preserving its privacy. It provides data confidentiality, query privacy and key security. The security provided by is scheme is not sufficient to protect the data from the eavesdropping attacks. The maintenance cost to save the data in cloud is high. [11].

3. Proposed system

To provide better fault tolerance, this work introduces the distributed cloud storage servers into De-duplication systems. To further protect the data, the compatible secret sharing technique is used association with distributed system. The uploaded file is divided into small amount of chunks of data and also applied security to the chunks of data.

4. Methodology

The system design describes the overall structure of the system and the nodes that are participating in the system.

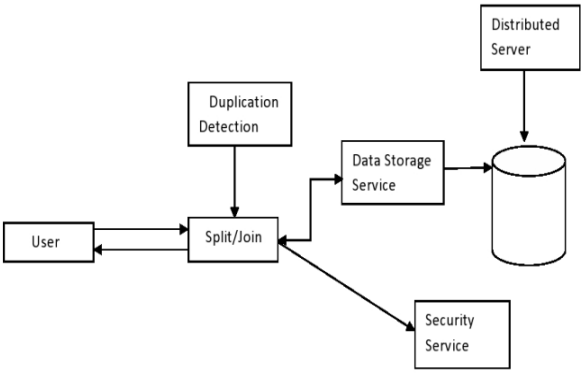

Fig. 1: De-Duplication and Security.

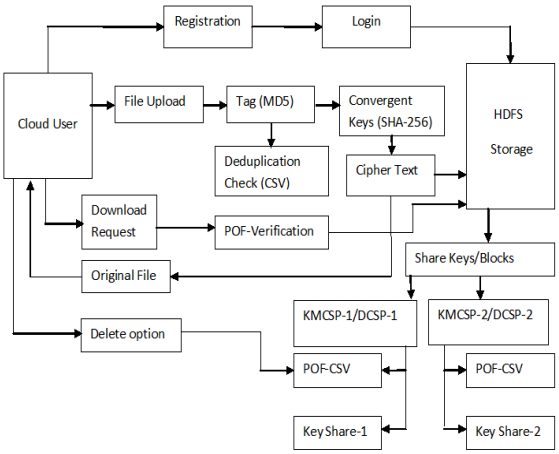

Fig. 2: Architecture Diagram.

4.1. De duplication

To eradicate the duplicate copies of repeating data a specialized data compression technique called data de-duplication is introduced. In the Block level De-duplication, the size of block will be defined as parameter.

4.2. Data sharing

In this module the algorithm is specified in the figure 2.

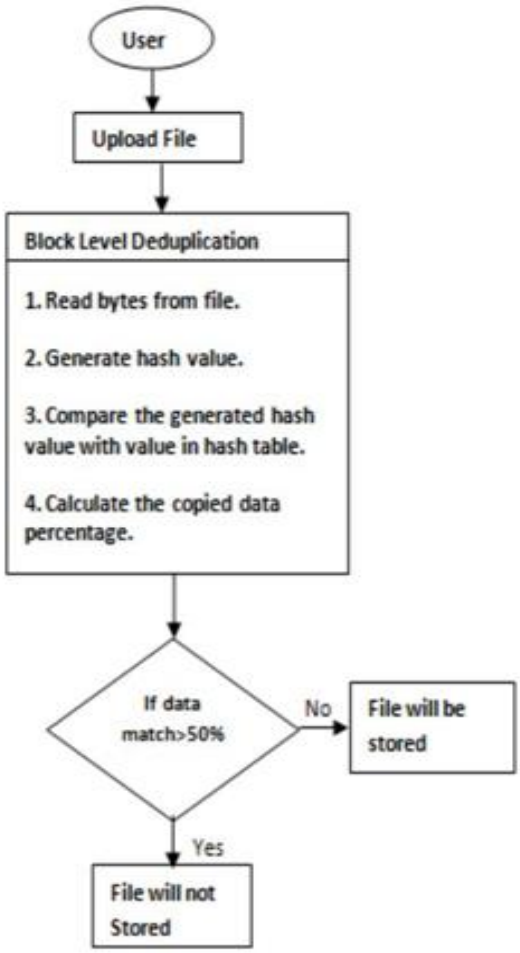

Fig. 3: Steps of De-Duplication.

Step 1: Read the bytes from the given file.

Step 2: Generate the hash value

Step 3: Compare the generated hash value with already stored value in hash value table.

Step 4: Calculate the copied data percentage.

Step 5: if the percentage is greater than 50

Step 6: File will be stored

Step 7: Otherwise, File will not be stored.

4.3. Distributed storage

In the method of De-duplication, in the operation of assay different blocks of data, or byte instructions, are detected and saved. After that those separate blocks of input are checked in which duplicate occurs, the repeated block is replaced with a reference that makes way to the saved block of input .The match frequency is dependent on the block size, and therefore the amount of data that must be stored or transferred is significantly reduced.

4.4. User revocation

User revocation is performed by the group manager via a public available revocation list, based on which group members can encrypt their data files and ensure the confidentiality against the revoked users. Let the group manger update the revocation list each day even no user has being revoked in the day. In other words, the others can verify the freshness of the revocation list from the contained current date URL. In addition, the revocation list is bounded by a signature sig declare its validity. The signature is generated by the group manager with the BLS signature algorithm, i.e., finally, the group manager mi-grates the revocation list into the cloud for public usage.

4.5. File restored

The data owner (i.e., the member who uploaded the file into the server) can have rights to delete any of his own file that is stored in the cloud. The group manager sends his deleted file ID with signature ID data to the cloud. Then the cloud will delete the file automatically.

5. Performance evaluation

To learn the content of a shared file, a member does the following actions:

1) Getting the data file and the revocation list from the cloud server. After a successful verification, the cloud server responds the corresponding data file and the revocation list to the user.
2) Checking the validity of the revocation list. This operation is similar to the step 2 of file generation phase.
3) Verifying the validity of the file and decrypting it. The format of the downloaded file coincides with that given in Table.

Fig. 4: Entry Details.

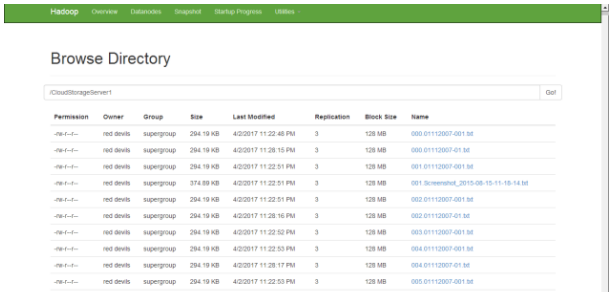

Fig. 5: The Directory of Files.

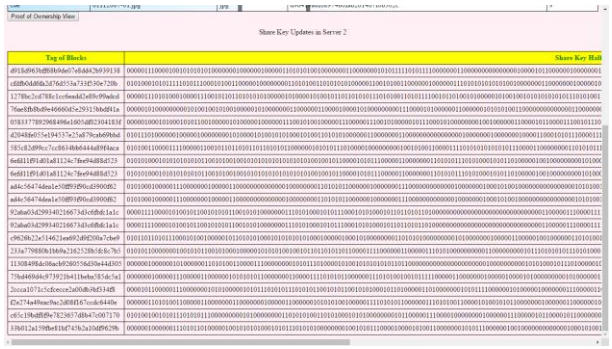

Fig. 6: Share Key Updating.

6. Conclusion

To achieve the confidentiality of the users' outsourced data which lack encryption mechanism the work have proposed a secure de-duplication system to develop the reliability of data. For constructions the proposed work to hold file level and fine grained block level data de-duplication. The security of tag consistency and integrity is achieved. Clients are helped by an auditing entity with preservation of a Map Reduce cloud that make data tags ahead of uploading and also audit the integrity of data having been stored in cloud. The work implemented in our De-duplication systems with secret key sharing scheme. The computation by user in Secure Cloud is very much reduced through the file uploading and auditing phases compared with the prior work. Before uploading secure cloud is higher construction aggravated by the detail that customers always desire to encrypt their data and allows integrity auditing and secure De-duplication straight on encrypted data.

References

[1] Jiawei Yuan, Shucheng Yu, “Secure and Constant Cost Public Cloud Storage Auditing with De duplication”, in IEEE Conference, Volume 25, Issue 1, Pp: 123-132, October 2013.

[2] Shai Halevi , Danny Harnik , Benny Pinkas and Alexandra Shulman-Peleg, “Proofs of Ownership in Remote Storage Systems”, in IEEE conference, August 2011.

[3] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson and Dawn Song, “Provable Data Possession at Untrusted Stores”, in IEEE Transactions,Pp.598-604,November 2007.

[4] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson and Dawn Song, “Provable Data Checking Using Provable Data Possession”, in ACM Transaction on Information and System Security, Section 14, Article 12, May 2011.

[5] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, "Secure De duplication with Efficient and Reliable Convergent Key Management", in IEEE transactions,Pp.370-375, June 2014.

[6] Giuseppe Ateniese, Osama Khan, Reza Curtmola Zachary Peterson, Joseph Herring, Randal Burns "Remote Data Checking Using Provable Data Possession", in IEEE Transactions, Volume 14, Pp.1-34, May 2011.

[7] Jiawei Yuan and Shucheng Yu, "Secure and Constant Cost Public Cloud Storage Auditing with De duplication", in IEEE conference, October 2013.

[8] Jun Ren, Zhigiang Yao, Jinbo Xiong, Yuanyuan Zhang, Ayong Ye, “A Secure Data Deduplication Scheme Based on Differential Privacy”, in IEEE international conference on Parallel and Distributed Systems, Dec 2016.

[9] Zheng Yan, Mingiun Wang, Yuxiang Li, Athanasios V. Vasilakos, “Encrypted Dataanagement with Deduplication in Cloud Computing”, in IEEE Cloud computing, Volume 3, Issue 2,Pp.35-41, May 2016.

[10] J. Gantz and D. Reinsel, “The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East”, in emc.com, Dec.2012.

[11] K. Periasamy and B. Latha, “The Enhancement of Storage and Bandwidth Optimization Using DataDe-Duplication, in International Journal of Applied Engineering Research Volume 9, Number 20, 2014.