International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 4, Special Issue 3 (March '2018)
Proceeding of - National Conference on Recent Trends in Information Technology and Communication,
Conference held at Velammal Engineering College,Chennai,India.

SECURE AND MEMORY EFFICIENT DE-DUPLICATION ON ENCRYPTED DATA IN CLOUD STORAGE

T.V.Ashwini,
Student,
Department of Information Technology,
Velammal Engineering College
Chennai,Tamilnadu,India.

M.Jelinasree,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

R.Sridharshini,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai, Tamilnadu,India.

S.Ahamed Ali,
Assistant Professor,
Department of Information Technology,
Velammal Engineering College,
Chennai.

Abstract: In cloud storage benefit, deduplication method is commonly used to decrease the space and transmission capacity prerequisite of administrations by wiping out excess information and putting away just a solitary duplicate of them. Deduplication is the best when different proprietor outsource similar information to the distributed storage, however it raises issues identifying with security and possession. Verification of possession plans permits any proprietor of similar information to demonstrate to the distributed storage server that he claims the information heartily. The various proprietors will transfer the document to the capacity under isolated classification. The verifier will check the document and afterward no one but client can get to the record . The record is produced with mystery key where cryptography system is drawn nearer is utilized. The information clients ask for a similar record which is transferred by the proprietor. To secure the classification of touchy information while supporting deduplication, the cryptography method is utilized to ensure the information privacy and it will produce a mystery key before putting away to the cloud.

Keywords: cloud computing, cryptography, deduplication

I.INTRODUCTION

Remote cloud storage have become an essential part for various applications in nowadays network, which store a large amount of data and provide the partial data needed. With the rapid growing of cloud storage services, such as cloud storage, encryption becomes an important technique for protecting the confidentiality of data. Cloud computing provides seemingly unlimited resources as services to cloud users by rearranging various resources. Cloud storage as one of the most popular cloud services enables cloud users to store tremendous amount of data in the cloud, which may exceed their own storage spaces. Data deduplication can help eliminate multiple copies of same files and improve the utilization of storage. Although data encryption provides an important guarantee for the security and privacy of clients' data, it limits the manners of the accessibility and availability of the encrypted data. Then limitation of schemes with encrypted data is that, when some special processing applications over the data are needed, such as cross-client data deduplication query, sorting over encrypted data , the schemes usually become inefficient due to the frequent date encryption and decryption operations. Thus,it is important to design efficient schemes to support secure and efficient computation outsourcing and storage outsourcing. Data deduplication enables data storage systems to find and remove duplication within data without compromising its availability. The goal of data deduplication is to store more data in less space by storing and maintaining files (blocks in fine-grained deduplication manner) into a single copy, where the redundant copies of data are replaced by a reference to this copy.

In cryptography, encryption is the way toward encoding messages or data such that lone approved gatherings can read it. Encryption does not itself prevent interception, but denies the message content to the interceptor. In an encryption scheme, the intended communication information or message, referred to as plaintext, is encrypted using an encryption algorithm, generating cipher text that can only be read if decrypted. For technical reasons, an encryption scheme usually uses a pseudo-random encryption key generated by an algorithm. It is the method possible to decrypt the message without possessing the key, but for a well designed encryption scheme, large computational resources and skill are required. An authorized recipient can easily decrypt the message with the key provided by the originator to recipients, but not to unauthorized interceptors. The purpose of encryption is to ensure that only somebody who is authorized to access data (e.g. a text message or a file), will be able to read it, using the decryption key. Somebody who is not authorized can be excluded, because he or she does not have the required key, without which it is impossible to read the encrypted information

II. RELATED WORKS

In cloud environments [1], users store their data or files in

cloud storage but it is not infinitely large. In order to reduce the requirement of storage and bandwidth, data deduplication has been applied. Users can share one copy of the duplicate files or data blocks to eliminate redundant data. Besides, considering the privacy of sensitive files, the users hope that the cloud server cannot know any information about those files. They often use certain encryption algorithms to protect the sensitive files before storing them in the cloud storage. Unfortunately, previous schemes have a security problem.

These schemes did not satisfy semantic security. In this paper, a hybrid data deduplication mechanism which provides a practical solution with partial semantic security. History-Aware Rewriting algorithm (HAR) [2] improves restore performance by 2.6X –17X at an acceptable cost in deduplication ratio. HAR out performs the state-of-the-art work in terms of both deduplication ratio and restore performance. The hybrid scheme is helpful to further improve restore performance in data- sets where out-of-order containers are dominant. Deduplication is an effective technique for reducing storage costs in distinct storage environments, more specifically in backup, primary, RAM, and SSD storage systems. Although all existing deduplication systems can be classified by a common taxonomy, each storage type has different assumptions that lead to distinct design decisions. Then, as another contribution, the existing deduplication systems and classify them according to the storage type—for instance, backup, primary, RAM, and SSD was presented. Most of these systems assume immutable data and trade latency for deduplication and I/O throughput by mainly using inline deduplication approaches [3].

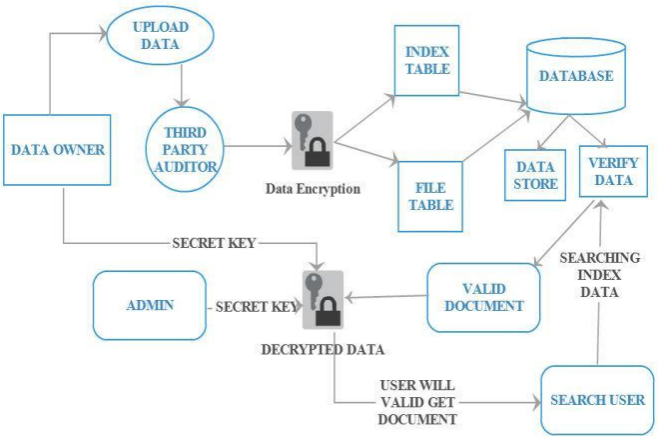
LIMITATIONS

- Duplicate data increased.
- Data leakage is more.
- Less security.

III. PROPOSED SYSTEM

We propose a scheme to de-duplicate encrypted data stored in cloud based onownership challenge and proxy encryption. It is implemented that the proposed scheme and tested its performance. It is also describing the implementation and testing environments. A MySQL database is applied to store data files and related information. It is not taken into account the time of data uploading and downloading. It integrates cloud data de-duplication with access control. We then develop and analyze a novel **"**De-duplication **"**Technique on Encrypted cloud data , For that we used trusted data checker , job of the data checker is to find given data is duplicate or not if not . Then that data forwarded to key generator section for encryption as user specified algorithm. That data is separate into multi part and generate a signature key for each part, key based on part content so we can easy to identify data duplication. Moreover we maintain data set to each data, so every time user make share only effect on data set .Data owner have rights to remove our data permanently. Focus is done on testing the performance of the deduplication procedure and algorithms designed in this scheme. The scheme provides a secure approach to protect and deduplicate the data stored in cloud by concealing plaintext from both CSP (Cloud Service Provider) and AP (Authorization Party). The security of the scheme is ensured by PRE theory, symmetric key encryption, AES and Elliptic curve Cryptography theory.

A. ARCHITECTURE



B. MODULES EXPLANATION

DATA OWNERS: Data owner can upload data's, that data are split into part data then send to trusted data checker, job of the data checker is to generate signature key from MD5 and compare with previous keys, if mismatch then that data send to Key generator Server , Job of the key generator are generate encryption key as user specified algorithm ,finally encrypt then store in Database .

OWNER DATASET: In this Module We create data owner dataset, this dataset only map owner with our upload data's , we maintain common database for effectively find duplications. The files will be uploading only once. If another data owner going to upload the same file in database means they will get the notification (the data is already uploaded in database).So data owner can save cost and time.

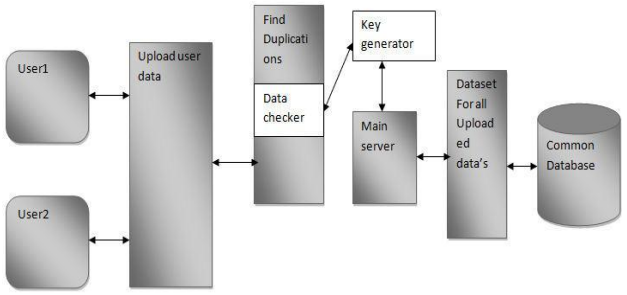
VERIFIER: In this modules, the verifier checks for the file integrity. If the file contains the same word as was in the file previously saved in the cloud then file will not store instead it shows error . The verifier will filter the file. If the file has some updation with uniqueness then verifier will accept the file and encrypt the file and stored to the cloud.

SHARED DATASET: Share Dataset is an light weight dataset that only contain mapping file metadata information, in our project we maintain one common big data database instead of unique because efficiently find duplication and memory management, if data owner share our data to client that data not replicate instead map client name. Data deduplication enables data storage systems to find and remove duplication within data without compromising its availability.

SECURITY: We are implementing **"**Dynamic Encryption key Generation**"**. It means all shared data only view with data owner permission, so we can avoid from unknown access. Social users are group members they can only view and share

the data. If want show the data mean they need to get permission to data owner then data owner will send Encryption key after they can view the data. If data owner does not provide the KEY mean user cannot view the file. data encryption provides an important guarantee for the security and privacy of clients' data, it limits the manners of the accessibility and availability of the encrypted data.

DATA METHODOLOGY:



DATA COLLECTION: In this project, the data owner will upload the file in different category like education, medical and sports. The data owner will purchase space from the different CSPs. The uploaded file should be unique then only it is accepted or else some uniqueness otherwise the verifier will not accept the file. The data user will get the file after the verification from the admin.

C. ADVANTAGES

1)Storage-based data deduplication reduces the amount of storage needed for a given set of files. It is most effective in applications where many copies of very similar or even identical data are stored on a single disk a surprisingly common scenario. In the case of data backups, which routinely are performed to protect against data loss, most data in a given backup remain unchanged from the previous backup. Common backup systems try to exploit this by omitting (or hard linking) files that haven't changed or storing differences between files. Neither approach captures all redundancies, however. Hard-linking does not help with large files that have only changed in small ways, such as an email database; differences only find redundancies in adjacent versions of a single file (consider a section that was deleted and later added in again or a logo image included in many documents).

2) Network data deduplication is used to reduce the number of bytes that must be transferred between endpoints, which can reduce the amount of bandwidth required.

3) Virtual servers benefit from deduplication because it allows nominally separate system files for each virtual server to be coalesced into a single storage space. At the same time, if a given server customizes a file, deduplication will not change the files on the other servers—something that alternatives like hard links or shared disks do not offer. Backing up or making duplicate copies of virtual environments are similarly improved.

IV. IMPLEMENTATION MODULES

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. USER MODULE

Registration:

In this module, each user registers his user details for using files. Only registered user can able to login in cloud server.

File Upload:

In this module, user upload a block of files in the cloud with encryption by using his secret key. This ensures the files to be protected from unauthorized user.

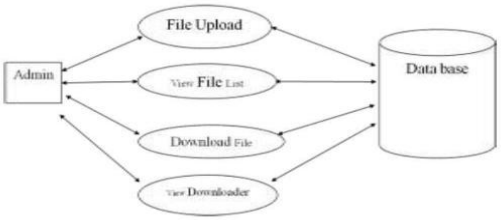
Download:

This module allows the user to download the file using his secret key to decrypt the downloaded data verify the data and re-upload the block of file into cloud server with encryption. This ensures the files to be protected from unauthorized user.

Re-upload:

This module allow the user to re-upload the downloaded files of blocked user into cloud server with resign the files(i.e.) the files is uploaded with new signature like new secret with encryption to protected the data from unauthorized user.

Figure: 1 User Module



B. ADMIN MODULE

View Files:

In this module, public auditor view the all details of upload, download, blocked user, reupload.

File Upload:

In this module, admin can also upload a block of files in the cloud with encryption by using his secret key.

This ensures the files to be protected from unauthorized user.

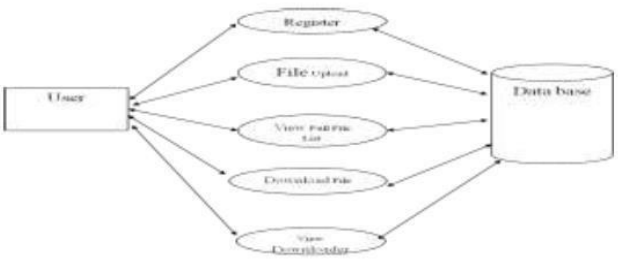


Figure 2 : Admin Module

C. TPA (Third Party Authorization) Module

In this module, third party user views the details of uploaded files in the cloud. But he could not able to read the content of the file and could not download the file.

Download:
This module allows the admin to download the file using his secret key to decrypt the downloaded data and verify. This ensures the files to be protected from unauthorized user.

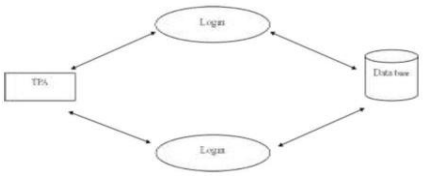


Figure 3: TPA module

V.CONCLUSION

Managing encrypted data with deduplication is important and significant in practice for achieving a successful cloud storage service, especially for big data storage. In this paper, a practical scheme to manage the encrypted big data in cloud with deduplication based on ownership challenge was proposed. This scheme can flexibly support data update and sharing with deduplication even when the data holders are offline. Encrypted data can be securely accessed because only authorized data holders can obtain the symmetric keys used for data decryption. Extensive performance analysis and test showed that this scheme is secure and efficient under the described security model and very suitable for big data deduplication. The advanced construction is motivated by the fact that customers always want to encrypt their data before updating and allows for integrity auditing and secure deduplication directly on encrypted data

VI.REFERENCES

[1]. T. Jiang, X. Chen, Q. Wu, J. Ma, W. Susilo, and W. Lou, "Towards efficient fully randomized message-locked encryption," in InformationSecurity and Privacy - 21st Australasian Conference, ACISP 2016,Melbourne, VIC, Australia, July 4-6, 2016, Proceedings, Part I, 2016.

[2]. Dropbox, "Dropbox," https://www.dropbox.com/, your stuff, anywhere.

[3]. Google, "Google drive," http://drive.google.com, all your files, readywhere you are.

[4]. NetApp,"Netapp,"http://www.netapp.com/us/products/platf orm-os/dedupe.aspx, universal Storage System.

[5]. C. Batten, K. Barr, A. Saraf, and S. Trepetin, "pstore: A secure peer-topeerbackup system," MIT Laboratory for Computer Science, progressreport, 2001.

[6]. M. Storer, K. Greenan, D. Long, and E. Miller, "Secure data deduplication,"in Proc. of the 4th ACM International
Workshop on StorageSecurity and Survivability, VA, USA, Oct. 2008, pp. 1–10.

[7]. L. Marques and C. Costa, "Secure deduplication on mobile devices," inProc. of the 2011 Workshop on Open Source and Design of Communication,Lisboa, Portugal, Jul. 2011, pp. 19– 26.

[8]. D. X. Song, D.Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy, CA, USA, May 2000, pp. 44–55.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.