iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 7 **Issue:** IV **Month of publication:** April 2019

**DOI:** https://doi.org/10.22214/ijraset.2019.4428

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com*

# Secure De-Duplication of Data on Hybrid Auditing

B. Sivasankari
*Assistant Professor, Department of Computer Science & Engineering, AEC, Chennai, Tamil Nadu*

***Abstract:*** ***The amount of data stored in storage devices are increased rapidly, as the evolution of networking techniques emerges. While storing huge amount of data, the storage servers ,may want to reduce the volume of the stored data, and the clients may want to check the integrity of their own data through low cost, since the cost of the functions related to data storage increase in proportion to the size of the data. To achieve these functionalities, privacy preserved deduplication and auditing techniques have been studied, which can reduce the volume of data stored on the server by eliminating duplicated copies and allows clients to efficiently verify the integrity of the stored files by delegating costly operations to a trusted party, respectively. We design a technique which performs both secure deduplication of encrypted data and hybrid integrity auditing of data which is stored on the storage devices. We utilize a Third Party auditor(TPA) for performing dynamic audit, in order to help the clients. The proposed scheme provides all the fundamental security requirements through cloud storage server.***
***Keywords: De-Duplication, Hybrid auditing, Third party Auditor***

## I. INTRODUCTION

The objective of the project is to remove the duplicates file while multiple users upload the same file on cloud again and again. The mechanism significantly improves cache efficiency in the inline deduplication phase and reduces the workload in the post-processing deduplication phase. The resource can be provided in-house or externally. A typical underlying requirement of private cloud deployments are security requirements and regulations that need a strict separation of an organization's data storage.The amount of data stored in storage devices are increased, along with evolution of networking techniques. While storing huge amount of data, the storage servers may want to reduce the volume of the stored data, and the clients may want to check the integrity of their data through low cost, since the cost of the functions related to data storage increase in proportion to the size of the data. To achieve these functionalities, privacy preserved deduplication and auditing techniques have been studied, which can reduce the volume of data stored on the server by eliminating duplicated copies and allows clients to efficiently verify the integrity of the stored files by delegating costly operations to a trusted party, respectively. So far experiments have been conducted on each topic, separately, whereas relatively few combined schemes, which support the two functions simultaneously, have been researched. In this paper, we design a technique which performs both secure deduplication of encrypted data and public integrity auditing of data which is stored on the storage devices. From the two functions, we perform challenge response protocols using the digital signature based homomorphic linear authenticator. We utilize a Third party auditor (TPA) for performing public audit, in order to help the clients. The proposed scheme provides all the fundamental security requirements through cloud storage server.

## II. EXISTING SYSTEM

To enable privacy-preserving public auditing for cloud data storage under the aforementioned model, our protocol design should achieve the following security and performance guarantee: Public auditability: to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional on-line burden to the cloud users; Storage correctness: to ensure that there exists no cheating cloud server that can pass the audit from TPA without indeed storing users' data intact; Privacy-preserving: to ensure that there exists no way for TPA to derive users' data content from the information collected during the auditing process; Batch auditing: to enable TPA with secure and efficient auditing capability to cope with multiple auditing delegations from possibly large number of different users simultaneously; Lightweight: to allow TPA to perform auditing with minimum communication

### *A. Advantage*

*1)* Public auditability.
*2)* Storage correctness.
*3)* Privacy-preserving.
*4)* Batch auditing.
*5)* Lightweight.

*B. Disadvantage*

*1)* Duplication of files will get increased.
*2)* File or Data colliding will occur.
*3)* Inefficient Auditing response

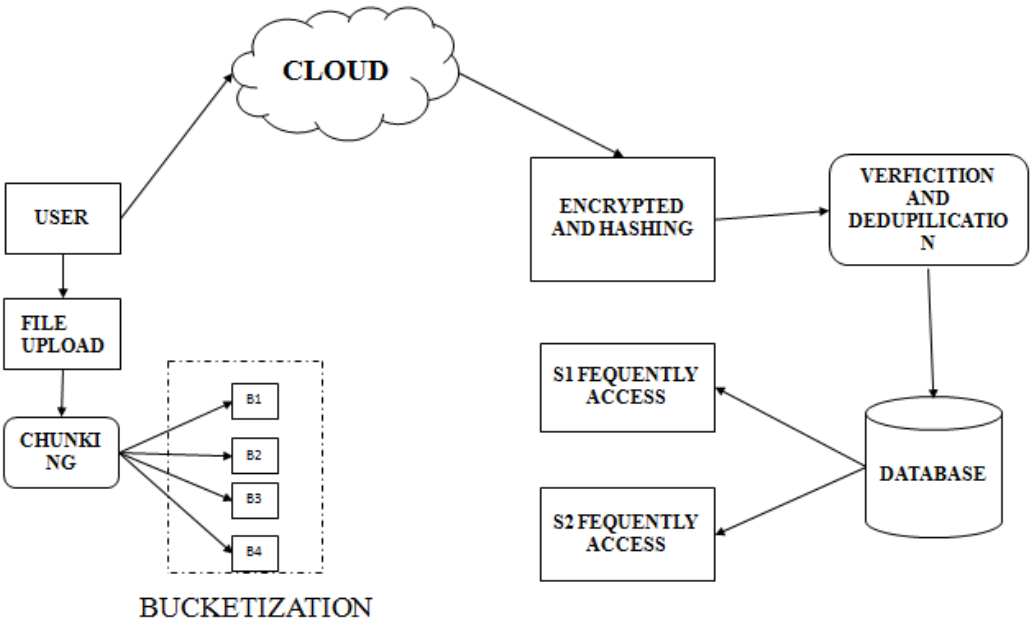
## III.PROPOSED SYSTEM

The proposed system consists of the entities: Client (or user) Outsources data to a cloud storage. Encrypted data is first generated, and then uploaded to the cloud storage to protect confidentiality. The client also needs to verify the integrity of the outsourced data. To do this, the client delegates integrity auditing to the TPA. Cloud Storage Server (CSS) will provide data storage services to users. Deduplication technology is applied to save storage space and cost. We consider that the CSS may act maliciously due to insider/outsider attacks, software/hardware malfunctions, intentional saving of computational resources, etc. During the deduplication process, the CSS carries out the protocol to verify that the client owns the file. Moreover, in the integrity audit process, it is necessary to generate and respond to a proof corresponding to the request of the TPA. TPA (Third Party Auditor) can performs integrity auditing on behalf of the client to reduce the client's processing cost. Instead of the client, the auditor sends a challenge to the storage server to periodically perform an integrity audit protocol. TPA is assumed to be a semi-trust model, that is, an honest but curious model. Under the assumption, it is assumed that the TPA does not collude with other entities. A client and a CSS perform secure deduplication, and a TPA is placed between the client and the CSS to execute integrity auditing instead of the client.

*A. Advantage*

*1)* TPA periodically checks the integrity of the data stored in the CSS.
*2)* To reduce the user overhead, the TPA performs periodic integrity audits.
*3)* Deduplication of the data stored in the CSS.
*4)* Increased in performance.

## IV.IMPLEMENTATION




*A. User Interface Deisgn*

In this Module User interface is created so that the data owner will upload the data to the server. The main objective of this module is to store and share the data by uploading the file to the remote machine. Data is Hashed and applied XOR functionality and then finally stored in the main server.

*B. Cloud Server*

Data owner will upload their data to the cloud server and request for a particular file is send to cloud server. Both the upload and the file request are handled by the main Cloud Server. During the file request is processed main server will communicate with the data owner and the files are retrieved only after the approval given the data owner.

*C. MHT(Merkle Hash Tree)*

In this module, data is encrypted, Hashed and XOR is applied sonly then the data is uploaded to the server. Once the data is uploaded, the entire data is chunked into multiple parts using Merkle Hash Tree algorithm and stored in separate data servers. In cryptography and computer science, 2a hash tree or Merkle tree is a tree in which every leaf node is labeled with the hash.

*D. Deduplication*

In this module user will upload more number of file on cloud on the same time many people will upload same file in different file name. So that more number of space will occupied in cloud. For that we implement duplicate detection by reading the content of the text file. And also we separate the file by frequency accessing and infrequent file. We will maintain the keyword in index in encrypted form.

*E. Bucketization*

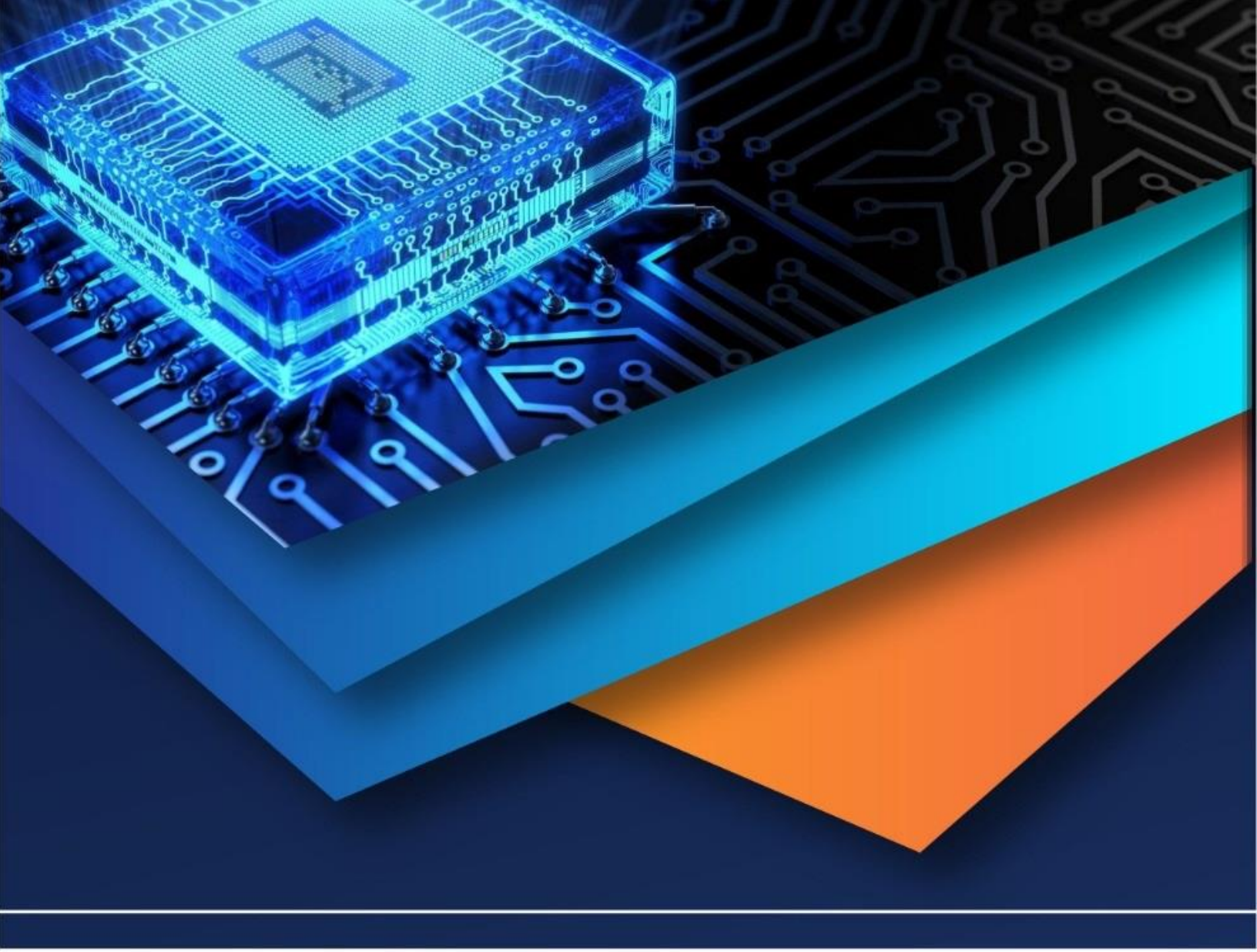
In this module we implement the chunking of text file in different part using MHT. In this we will chunk two different file one is text file another one is video. Once user upload the file user can download the file from the cloud. When user gives request to download the file system will reconstruct the chunking file and send it to the user. Same like the video will download based on time frame.

## V. CONCLUSION

We proposed a scheme to achieve both secure deduplication and integrity auditing in a cloud environment. To prevent leakage of important information about user data, the proposed scheme supports a client side deduplication of encrypted data, while simultaneously supporting public auditing of encrypted data. We used digital signature based homomorphic linear authenticator to compute authentication tags for the integrity auditing. The proposed scheme satisfied the security objectives, and improved the problems of the existing schemes.

## REFERENCES

[1] B. Zhu, K. Li, and R. H. Patterson, “Avoiding the disk bottleneck in the data domain deduplication file system,” in Proc. 6th USENIXConf. File Storage Technol., 2008, pp. 1–14.
[2] A. El-Shimi, R. Kalach, A. Kumar, A. Ottean, J. Li, and S. Sengupta, “Primary data deduplicationlarge scale study and system design,” in Proc. USENIX Annu. Tech. Conf., 2012, pp. 285– 296.
[3] D. T. Meyer and W. J. Bolosky, “A study of practical deduplication,” ACM Trans. Storage, vol. 7, no. 14, pp. 1–20, 2012.
[4] K. Srinivasan, T. Bisson, G. R. Goodson, and K. Voruganti, “iDedup: Latency-aware, inline data deduplication for primary storage,” in Proc. 11th USENIX Conf. File Storage Technol., 2012, pp. 1–14.
[5] B. Mao, H. Jiang, S. Wu, and L. Tian, “POD: Performance oriented I/O deduplication for primary storage systems in the cloud,” in Proc. IEEE 28th Int. Parallel Distrib. Process. Symp., 2014, pp. 767–776.
[6] A. Wildani, E. L. Miller, and O. Rodeh, “Hands: A heuristically arranged non-backup in-line deduplication system,” in Proc. IEEE 29th Int. Conf. Data Eng., 2013, pp. 446–457.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY