IJDCST @June Issue- V-1, I-6, SW-103
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

# In Cloud Security Mechanism Approach for Auditing in Deduplication System

**K. Venkatesh Sharma[1], T.Sukanya[2], M. Chandra Sekhar[3]**
[1]Associate Professor, Dept. of Computer Science & Engineering, MVR College of Engineering & Technology, A.P., India.
[2]Assistant Professor, Dept. of Computer Science & Engineering, MVR College of Engineering & Technology, A.P., India.
[3]Assistant Professor, Dept. of Computer Science & Engineering, MVR College of Engineering & Technology, A.P., India.

***Abstract*** — Cloud data storage services involves four entities.(i)Administrator controls the user details, file insertion, file access, file deletion and the time of user presents in the network to access the cloud data's. (ii)Third party auditor checks the correctness of cloud data. Some techniques are used to establish the auditing concepts. (iii) Users access the cloud data as per demand services. Users retrieve more useful information from multiple repositories and no limitation to access the particular storage part in the shared pool. To create data management more scalable in cloud computing field, deduplication a well-known method of data compression to reduce duplicate copies of duplicate data in storage over a cloud. Even if data deduplication brings a lot of advantages in security and privacy concern occur as users' confidential data are liable to both attacks insider and outsider. A convergent encryption method imposes data privacy while making deduplication possible. Traditional deduplication systems based on convergent encryption even though offer confidentiality but do not maintain the duplicate check on basis of differential rights. This paper present, the plan of approved data deduplication planned to guard data security by counting discrepancy privileges of users in the duplicate check. Deduplication systems, users with differential privileges are added measured in duplicate check besides the data itself. To maintain stronger security the files are encrypted with differential privilege keys. Users are only permitted to carry out the copy check for files marked with the matching privileges to access. The user can confirm their occurrence of file after deduplication in cloud with the help of a third party auditor by auditing the data. Additional auditor audits and confirms the uploaded file on time. we present two secure systems, namely SecCloud and SecCloud+. SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients create data tags before uploading as well as audit the integrity of data having been saved in cloud. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is designed motivated by the fact that customers always want to encrypt their data before uploading, and enables integrity auditing and secure deduplication on encrypted data.

**Keywords** —***Secure auditing, SecCloud and SecCloud+, Cloud Storage, Data deduplicating.***

## 1. INTRODUCTION

cloud computing is one of the biggest innovative technology and it provides the facility of heavy data maintenance and management by improving data sharing and data storing capabilities. Those days the computer was just used for performing arithmetic and logical operations. But as the world evolved with inventions and innovations, more and more data got generated eventually. Then there was the use of Hard drive to store useful data, which was very costly. Birth of Internet provided various technologies including Cloud Storage. As we all know cloud storage is basically storing of data (Image, Videos, File, etc.) on a virtual server or we can say on a virtual database. Technically a cloudcomputing/storage is further explaining as, a system for enabling convenient on demand networkaccess to share data between computers. It is an internet based service which helps to store data by managing the storage. Cloud Storage provides the users ranging from cost saving from andsimplified convenience, to mobility opportunities and scalable services. According to somesurvey, the volume of data in cloud is expected to achieve many trillions of gigabytes. Even though cloud storage system has been extensively used, it ignores some important emerging needs suchas the abilities of auditing integrity of cloud file and detecting

duplicated files by cloud servers .The second problem is solved using deduplication. The rapidadoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers which also causes similar (duplicate) files beingstored at multiple locations, wasting the memory resource . This problem is countered by atechnology namely deduplication,in which the cloud servers would like to deduplicateby keeping only a single copy for each file (or block) and attach a link to this file (or block) for every client who owns it . This generates the problem of transparency, which need not be compromised i.e. no two ownersof the same duplicate file should be aware of ownership by other client as well and also of the deduplication being performed on his/her data.In this paper, aiming at achieving deduplication with standard security and data integrity, we propose a secure system named Dcloud. Deduplication is a method where the server saves only a single copy of each file, regardless of which clients asked to store that file, such that the disk space of cloud servers as well as network bandwidth are saved. However, trivial client side deduplication leads to the leakage of side channel information.

## 2. RELATED WORK

**Cloud Clients:** Cloud Clients have large data files to be stored and rely on the cloud for data maintenance and computation. They can be either individual consumers or commercial organizations. **Public Storage:** Public Storage is an storage disk which permit to store the users data which contains authorization and not permit to upload the duplicate data. Thus save storage space and bandwidth of transmission. This uploaded data is in encrypted form, only a user with individual key can decrypt it. **Private Cloud**: A private cloud acts as a proxy to allow both data owner and user to strongly perform duplicate check along with disparity permissions. **Auditor:** Auditor is a TPA work as proficiency and capabilities where cloud users do not have to faith to assess the cloud storage service reliability on behalf of the user upon request. The set of permissions and the symmetric key for each privilege is allocates and stored in private cloud. The user registers into the system, permissions are
assigned to user according to identity given by the user at registration time; means on basis of situation which access by the user. The data owner with permission can upload and share a file to users, further the data owner performs identification and sends the file tag to the private server. Private cloud server checks the data owner and computes the file token and will send back the token to the data owner. The data owner throws this file token and a request to upload a file to the storage provider. If duplicate file is found then user needs to run the PoW protocol with the storage provider to prove that user has an ownership of respective file. In the PoW result; if proof of ownership of file is approved then user will be provided a pointer for that file. And on the next case; for no duplicate is found for the file, the storage provider will be come again a signature for the result of that proof for the particular file. To upload file user sends the privilege set as well as the proof to the private cloud server in the form of a request. The private cloud server verifies the signature first on receiving the request for the user to upload file.

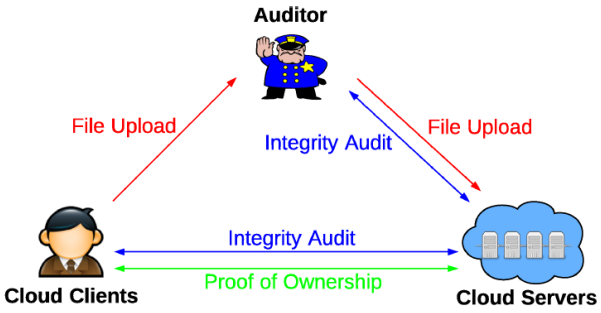
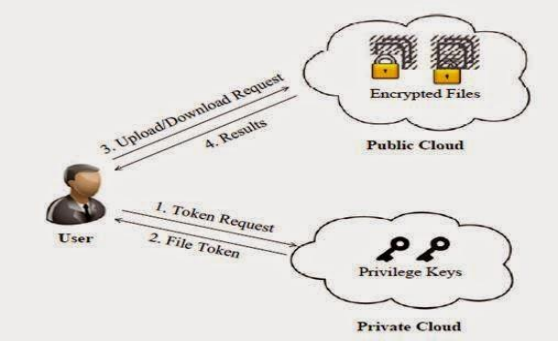



Fig. 1 Overview of System Architecture

## 3. PROPOSED METHODOLOGY

We proposed authorized deduplication system with auditing, in which we used three model entities i.e cloud user, private cloud and public storage. Cloud user is a data user which performs both upload/download file on public storage. A Private Cloud is used as a private one which controls the private keys and handles the file token calculations. A Public storage wills stores and checks duplicate files present in it.We implement cryptographic process of hashing and encryption/decryption methods for storage purpose to provide cloud computing environment. To support deduplication with authorization, the tag of a file will be determined by its privilege. For sustaining authorized access for user, a secret key will be bounded with a privilege to make a file token for it. Consider, the token is only allowed to access by user with privilege defined by the users itself. In another words, the token could only computes by the users with privilege. The token generation function could be easily denotes as a

cryptographic hash function. The user with a set of privileges will assign the set of keys as Binary relations defined. If the value matches along with given two privileges, such represented as based on the background of function which include a common concept in relation of hierarchical system. More exactly, hierarchical relation is that when matches only when a higher – level privilege occurs. The target file space underlying given ciphertext is drawn from a message space of size, the public cloud server can get well after almost off-line encryptions.




All the existing applications discussed are kind of more commercial and money making, but this web application is different. Mainly this web application deals with the important factor like De-duplication, Security, Integrity and Availability. The sharing of data is easy but the one thing we should take care of is the security because we don't want anybody should see our data in the cloud without permission of the primary user.

## 4. LITERATURE REVIEW

**Author: Qian Wang** Cloud Computing system has been predicted as the next-generation architecture of IT Enterprise. It moves the application software and databases to the centralized with large data centers, where the management of the data and services may not be fully trustworthy. This unique ensample brings about many new security challenges, which have not been well understood. Our research work examine the problem of assuring the integrity of data storage in Cloud Computing. In particular, we consider the task of allowing a third party auditor (TPA), on concern of the cloud client, to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA dismiss the involvement of client through the auditing of whether user's data stored in the cloud is truly intact, which can be important in achieving economies of scale for Cloud Computing. The support for data dynamics through the most general forms of data operation, such as block modification, insertion and deletion, is also more powerful step to - ward practicality, since services in Cloud Computing are not limited to archive or backup data only. While presiding work on ensure remote data integrity often lack the supports of either public verifiability or dynamic data operation. Proofs of Ownership in Remote Storage Systems.

**Authors: Giuseppe Ateniese** We suggest a model for provable data possession (PDP) that can be used for remote data checking: A client that has stored data at an untrusted server can verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking is lightweight and supports large data sets in distributed storage systems. The model is also robust in that it incorporates mechanisms for mitigating arbitrary amounts of data corruption.

**Author: Hovav Shacham and Brent Watersy** introduced proof-of-retrievability system. In this integrity check system, data storage centre provide proof to a verier that it is actually storing all of a client's data. Here they have explained two homomorphic authenticators the first authenticator is based on PRFs, gives a proof-of-retrievability scheme secure in the standard model. The second, based on BLS signatures [4], which give a proof-of-retrievability scheme with public variability secure in the random oracle model. Frameworks explained can allow arguing about the systems unforgeability, extractability, and retrievability with these parts based on cryptographic, combinatorial, and codingtheoretical techniques respectively.

**Author: Shai Halevi** Cloud storage systems are becoming more and more popular. A promising technology that keeps their cost down is deduplication, which stores only a single copy of duplicatin g data. Client-side deduplication attempts to identify deduplication opportunities already at the client side and save the bandwidth of uploading copies of existing files to the server. In this work we identify attacks that exploit client-side deduplication, granting an attacker to gain access to arbitrary-size files of other users based on a very small hash signature of these files. More specifically, an attacker who knows

the hash signature of a file can assure the storage service that it owns that file, hence the server lets the attacker download the entire file.

## 5. CONCLUSION

Data privacy and data security are the main issues for cloud storage. To ensure that the risks of privacy have been mitigated a variety of techniques that may be used in order to achieve privacy. This paper showcase some privacy techniques which introduced to maintain integrity of data and different methods for overcoming the issues data deduplication on untrusted data stores in cloud computing.There are still some approaches which are not covered in this paper. This paper categories the different methodologies in the literature as encryption based methods, access control based techniques, query integrity, keyword search schemes, and auditability schemes. Even though there are many techniques in the literature for considering the concerns in data integrity and data deduplication. De-duplication of data is the main focus in the entire web application. Providing storage of data on a large scale with multiple file sharing. Auditing helps the user to check the integrity of the data.

## REFERENCES

[1] 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 . ACM, 2011, pp. 491– 500.

[2] S. Keelveedhi, M. Bellare, and T. Ristenpart, ―Dupless: Serveraided encryption for deduplicated storage,‖ in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online]. Available:https://www.usenix.org/conference/usenixsecurity13/technicalsessions/presentation/bellare [3] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, ―Provable data possession at untrusted stores,‖ in Proceedings of the 14th ACM Conference on Computer and Communi cations Security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.

[4] H. Wang, ―Proxy provable data possession in public clouds,‖ IEEE Transactions on services Computing, vol. 6, no. 4, pp. 551–559, 2013.

[5] E. H. Miller, ―A note on reflector arrays (Periodical style—Accepted for publication),‖ IEEE Trans. Antennas Propagat., to be published.

[6] J. Wang, ―Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),‖ IEEE J. Quantum Electron., submitted for publication.

[7]. Govinda.K, V.Gurunathaprasad, H.Sathishkumar,"Third Party Auditing For Secure Data Storage In Cloud Through Digital Signature Using RSA", In International Journal Of Advanced Scientific And Technical Research(Issue 2, Volume 4-August 2012) Issn 2249-9954.

[8]. Ezhil Arasu.S, B.Gowri, S.Ananthi ,"Privacy-Preserving Public Auditing In Cloud Using HMAC Algorithm ",In International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, March 2013 .

[9]. Shingare Vidya Marshal ,"Secure Audit Service by Using TPA for Data Integrity in Cloud System",In International Journal of Innovative Technology and Exploring Engineering (IJITEE)ISSN: 2278-3075, Volume-3, Issue-4, September 2013.