ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Secure Cloud Auditor for Efficient Data De-duplication with maintaining shared data Integrity

Tejaswini Jaybhaye, Prof. D.H.Kulkarni

M.E Student, Department of Compute Engineering, S. K. N .College of Engineering, Pune, India

Assistant Professor, Department of Compute Engineering, S. K. N .College of Engineering, Pune, India

ABSTRACT: Cloud computing is the main attraction since last decade. It is one of the biggest innovative technologies; which basically provides facility such as large size data maintenance and data management by improving its potentiality in terms of data sharing and data storing capabilities. Data security is main hazard for this cloud data storage in terms of maintenance of data integrity and data deduplication on cloud. Handling these problems simultaneously is the tough task. There are 2 new SecCloud and SecCloud+ cloud auditing systems were introduced which help in providing integrity to cloud data with efficient data deduplication.

This paper is a detail description of secure cloud auditor which is used for the maintaining integrity of shared data with efficient data deduplication on cloud. This mechanism uses concept of SecCloud system where user is able to generate data tags before storing data on cloud which helps during performing audit to check integrity of data, it is also based on SecCloud+ system which provides encryption of data before uploading it, which provide integrity check and secure deduplication of encrypted data. It also performs batch monitoring to verify multiple tasks simultaneously.

KEYWORDS*:* Cloud computing, Data integrity, Auditing, Data deduplication.

I. INTRODUCTION

Internet based cloud computing has advanced computational power, which provides facility like data storing and stored data sharing. Cloud computing can also defined as shared pool having various configurable computing resources, on-demand network access and the service provider provisioned[1].Like coin, cloud has two sides, it is cost saver but on other side major concern is security. Attractive feature of cloud is provision for huge data storage, which provides some advantages to the customer such as mobility, scalable service and cost sever.

Data security is a main trouble related to cloud computing. Below mentioned types of data items included in the facility provided by the cloud service provider for security and privacy –

1) Sensitive data
2) Usage data
3) Unique device identities
4) Personally identifiable information

In cloud data storage service, 3 main entities involved:

1) Users access -
   Availability of the cloud data as per demand services.
2) Administrator controls -
   Control over file insertion, file access, file deletion and at the time of user presents in the network and trying to access the cloud data.
3) Third party auditor (TPA) checks -
   TPA check on the correctness of cloud data.

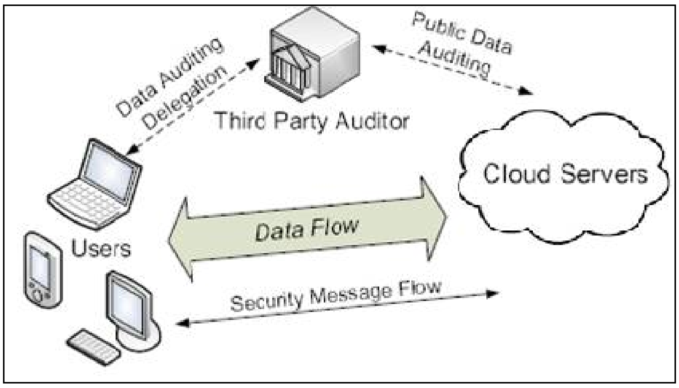


Fig.1 Architecture of Cloud Data Storage Service

Integrity is simply defined as consistency. Integrity is one of the security factors which influence the cloud performance. Data integrity defines rules for writing data in a reliable manner to keep persistent data storages. This phenomenon helps to retrieved data as it is without any changes. Preserving integrity of shared data is tough task. There are various mechanisms have been recommended [2]-[7] to preserve integrity of data. Integrity is most important security for cloud data storages because it ensure about completeness of data also provide detail information that available data is correct, easily accessible to authorized users only, data is consistent and of high quality.

There are three types of integrity constraints:

- Domain integrity
- Entity integrity
- Referential integrity

There is less control provided to the data owners on cloud, so it results into un-determination of correctness of data storage and computation related to data. Secure cloud only emphasises on the cloud data storage security and cloud computing security. On cloud, stored data safety has been compromised in several cases for monetary profit. To avoid this problem it is necessary to provide security and privacy of cloud data and it's computing by applying different techniques and mechanisms.

Data deduplication techniques used specially as a data compression technique. This technique help in eliminating duplicate copies of repetitive data, here cloud server stores single copy of data file. Benefit of this is to save network bandwidth. It also has some disadvantages such as in hybrid cloud, cloud deduplication may lead to loss of sensitive information.

Deduplication classified into two different categories, which are based data units

1) File Level Deduplication -

In this category, file is considered as a one data unit and hash value of file is used as its identifier. During deduplication it checks if two or more files have similar hash value, then it consider that files as a duplicate file having same contents hence only one copy will be stored in database.

2) Block Level Deduplication -

In this category, file is divided into small data blocks, having fixed-size or variable size, to check deduplication hash value is computed on each data block.

One more classification criteria for dada deduplication is based on location, it further divided into 2 groups such as source-based and target-based deduplication. If data deduplicated at the client location, then it is known as source-based deduplication else target-based. In source-based deduplication, the client first provide hash value to each data segment and then uploads and sends these results to the storage provider to check similar data is already available.

II. LITERATURE REVIEW

There are different auditing mechanisms which help to check data integrity on cloud. Chandinee Saraswathy K et. all explained HLA based technique in [2] where linear combination is used for authentication and it helps in performing auditing without retrieving data present on cloud and checks its integrity. Hovav Shacham and Brent Watersy [3] introduced system with proof-of-retrievability. Here data storage centre provides proof to a verier regarding it actually storing all of user data. Two homomorphic authenticors used in this system are based on PRFs where proof-of-retrievability scheme secure in the standard model and based on BLS signatures [4], where proof-of-retrievability scheme with public variability secure in the random oracle model. Giuseppe Ateniese et all proposed provable data possession (PDP) model which used for verifying that server is processing the original data available on cloud without reading its content. [5]. Cong Wang proposed new Privacy Preserving Public Auditing technique in [7] where public auditing allows TPA and user to check the integrity of the outsourced data stored on a cloud and Privacy Preserving allows TPA to perform auditing without requesting data. SecCloud is the best protocol proposed by Jingwei Li et all in [8] which ensures security and privacy of data stored on cloud and its computing. There are different mechanisms which provide data deduplication facility for data present on cloud. M. Bellare et all designed a system [9] called as DupLESS which combines a CE-type scheme which has ability to obtain message derived keys with help of a key server (KS) which already shared with the user group. J. li et all [10] proposed Dekey mechanism which is an efficient and reliable convergent key management which helps in secure de-duplication. C. Ng et all [11] designed RevDedup system, a de-duplication system specially designed for VM disk image backup, available in virtualization environments. RevDedup system has many design goals such as - low memory usage, high storage efficiency, and high backup and restores performance for latest backups. M. W. Storer [12] proposed two models for secure de-duplicated storage - authenticated and anonymous. These two designs model demonstrate that cloud security can be combined with data de-duplication such a way that it help to provide a diverse range of security characteristics. M. Shyamala Devi et. all proposed in paper [13] how to optimize the private cloud storage backup which provides high throughput to the users of the organization by increasing the de-duplication efficiency. J. Stanek [14] had presented a new encryption scheme that guarantees semantic security for unpopular data present on cloud but provides weaker security; it has better storage and bandwidth benefits for popular data. SecCloud+ protocol proposed by Jingwei Li et all in [7] which supports in integrity auditing and also provides secure deduplication with guarantee of file confidentiality

III. SYSTEM MODEL AND ITS DESIGN GOALS

A. System Model

Fig. 2 demonstrates the system model of SecCloud system, which consists of 3 entities such as cloud clients, cloud server and auditor who helps in data auditing. The SecCloud system supporting file-level deduplication includes the following three protocols respectively highlighted by red, blue and green in Fig. 2.

1) File Uploading Protocol: This protocol allows clients to upload files via the auditor. It includes three phases:

- Phase 1 (cloud client cloud server): client performs the duplicate check with the cloud server to confirm if such a file is stored in cloud storage or not before uploading a file.
- Phase 2 (cloud client auditor): client uploads files to the auditor, and get acknowledgement from auditor.
- Phase 3 (auditor cloud server): auditor helps to generate a set of tags for file and send these tags along with file during uploading it.

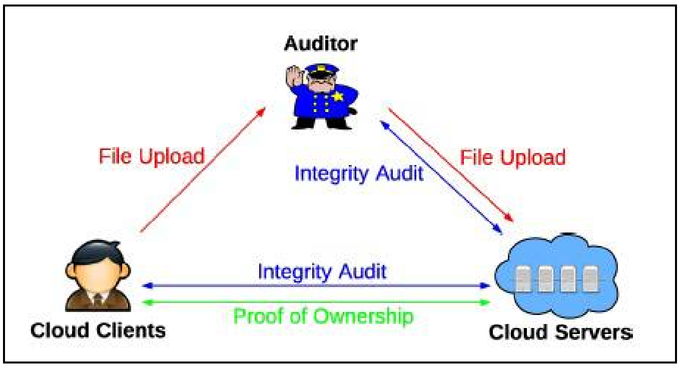


Fig. 2 the system model of SecCloud system

2) Integrity Auditing Protocol: Main aim of this protocol is integrity verification. Here, the cloud server plays the role of prover, while the auditor or client works as the verifier. This protocol includes two phases:

- Phase 1 (cloud client/auditor cloud server): verifier generates a set of challenges and sends them to the prover.
- Phase 2 (cloud server cloud client/auditor): based on the stored files and file tags, prover tries to prove that it exactly owns the target file by sending the proof back to verifier.

3) Proof of Ownership Protocol: This protocol used for verifying that the client exactly owns a claimed file. This protocol is typically triggered along with file uploading protocol to prevent the leakage of side channel information. On the contrast to integrity auditing protocol, in Proof of Ownership the cloud server works as verifier, while the client plays the role of prover. This protocol also includes two phases

- Phase 1 (cloud server client): cloud server generates a set of challenges and sends them to the client.
- Phase 2 (client cloud server): the client responds with the proof for file ownership, and cloud server finally verifies the validity of proof.

B. Design Goals

This proposed secure cloud auditor should achieve below mentioned design objectives:

- Integrity Auditing: Public verifier is able to maintain data integrity and it includes to features public verification and stateless verification.
- Secure DE duplication of Data: Public verifier should capable in identifying DE duplicate file and also notify about duplicate data to central repository. It requires that the cloud server is able to reduce the storage space by keeping only one copy of the same file.
- Cost Effective: Public verifier should easily handle a large number of auditing tasks simultaneously.
- File Confidentiality: The design goal of file confidentiality requires preventing the cloud servers from accessing the content of files.

IV. PROPOSED SYSTEM

Proposed system is combination of SecCloud system and SecCloud+ system which allows for integrity auditing and deduplication on encrypted. Below section explained detailed architecture of proposed system.

A. System Architecture

Architecture of proposed secure cloud auditor is shown in Fig.3. It consists of different modules which are responsible for different process which are required for efficient data deduplication and to check correctness of data.

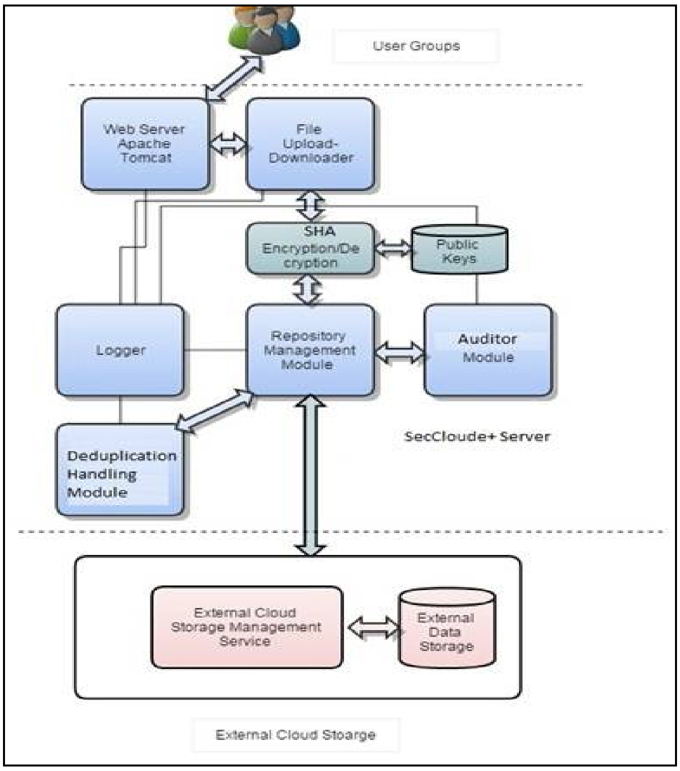


Fig.3 *System Architecture*

1) User Module:
 User module is has different small sub-modules such as Registration, File Upload, Download, Re-upload and Unblock. When new user join group he needs to register by using web server like apache. After successful registration user can upload his data on cloud storage. He can able to select data file and upload it on cloud storage. SHA algorithm converts plain text into cipher text and then stored it on cloud database. After successful uploading of data generated secret key is provided to the user.
2) Client Module / User Groups:
 In this module, a client makes use of provider's resources to store, retrieve and share data with multiple users. A client can be either an individual or an enterprise. Client can check the uploaded file or upload the new file. Client can view the deduplicate file based on this client can delete the unwanted data.
3) File Upload Download Module:
 This module is responsible for file uploading downloading activity over cloud. FTP protocol is used in this module.
4) Web Server /Apache Tomcat:
 Web server is used to host the application. Advance java is used whereas the web pages are stored on this server and it is responsible for User interactivity with web pages.
5) SHA Encryption /Decryption and Public Key Repository:
 Files uploaded by user needs to be stored in encrypted format with the help of this module to provide encryption decryption of data uploaded by user.
6) Repository Management Module:
 This module is responsible for communicating with different module present in a system. It is also responsible for giving the data to logger, Auditor module, Deduplication Handling Module.
7) Logger:
 This module generates the system logs based on user operations.
8) Deduplication Handling Module:

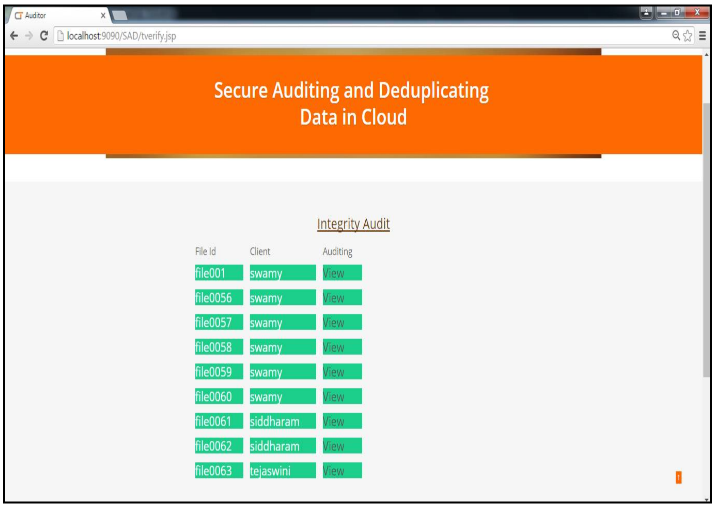
This module traces the various files uploaded by the users and also having capability to identify DE duplicate file. It also notifies the duplicate files to central repository in this module, clients uploaded files can be stored in cloud database. Clients can view the file from the database based on the DE duplicate factor it can be very secure

9) Auditor module / SecCloud + Module:

It gives the audit report and checks overall integrity of the system. It gives pass or fails results by auditing data. The strength of this system is it will not decrypt any data for auditing instead it audits on encrypted data.

V. RESULTS

This section demonstrate the implementation results and performance analysis is carried out on the basis of time required for file uploading with effective data deduplication and time required for overall auditing time to check data integrity by implementing proposed mechanism. Graphically presented analysis is most feasible analysis of mechanism with the existing public auditor.

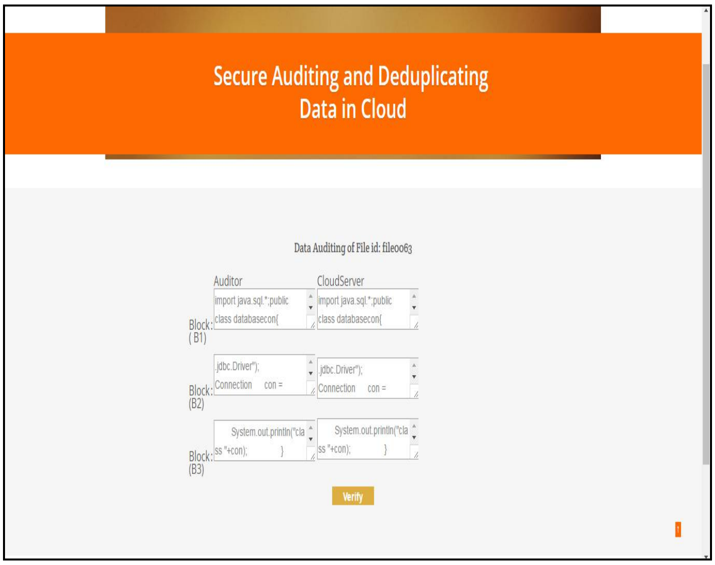


Secure Auditing and Deduplicating Data in Cloud

Integrity Audit

File Id	Client	Auditing
file001	swamy	View
file0056	swamy	View
file0057	swamy	View
file0058	swamy	View
file0059	swamy	View
file0060	swamy	View
file0061	siddharam	View
file0062	siddharam	View
file0063	tejaswini	View

Fig.4 Integrity Check

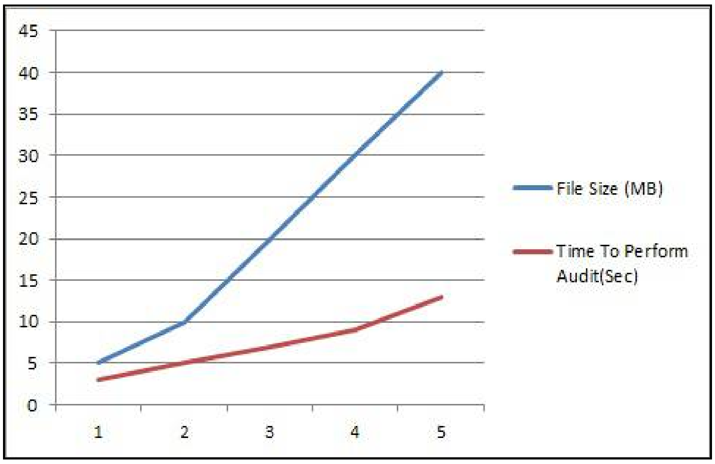

Secure Auditing and Deduplicating Data in Cloud

Data Auditing of File id: file0063

	Auditor	CloudServer
Block: (B1)	import java.sql.*;public class databasecon{	import java.sql.*;public class databasecon{
Block: (B2)	.jdbc.Driver"); Connection con =	.jdbc.Driver"); Connection con =
Block: (B3)	System.out.println("class "+con); }	System.out.println("class "+con); }

Fig.5 Duplicate Data Check

A. Impact on File Uploading time:



B. Impact on Auditing Time: In Fig.5 graphical representation of auditing time required to check integrity of files having different file format.

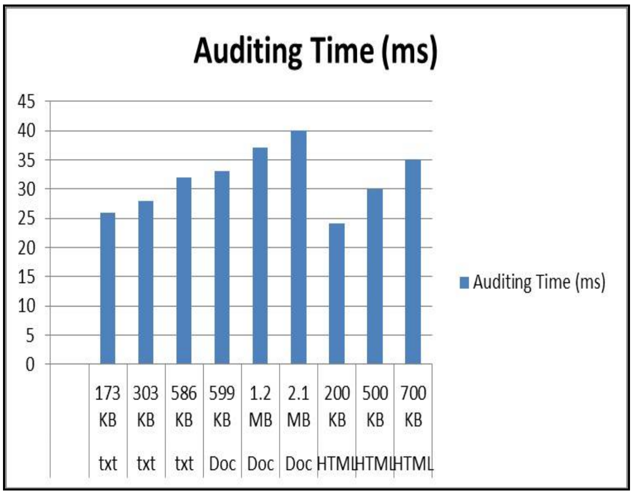


***Fig. 7**. Graphical representation of Impact of auditing on files having different file format*

VI. CONCLUSION AND FUTURE SCOPE

Secure Cloud auditor plays an important role when dealing with security aspects of cloud. In this paper, we have proposed a new auditing mechanism for data stored on cloud which provide facilities like efficient data de-duplication while maintaining shared data integrity. In this mechanism we can able to achieve properties like correctness and scalability while improving the data de-duplication. Experimental results demonstrate that it can helps in saving significant amount of computation and communication resources during data deduplication.

Future work includes introduction of concept parallel computing, to support it multi-threading environment can also useful, it helps in improving the overall auditing performance

REFERENCES

[1] P. Mell and T. Grance, “Draft NIST working definition of cloud computing”.

[2] Chandinee Saraswathy K. , Keerthi D. , Padma G. "HLA Based Third Party Auditing For Secure Cloud Storage" International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1526-1532

[3] H. Shacham and B. Waters, “Compact Proofs of Retrievability,”in the Proceedings of ASIACRYPT 2008. Springer-Verlag, 2008, pp.90–107.

[4] K. Kiran Kumar, K. Padmaja, P. Radha Krishna, “Automatic Protocol Blocker for Privacy-Preserving Public Auditing in Cloud Computing”, International Journal of Computer science and Technology, vol. 3 pp, ISSN. 0976-8491(Online), pp. 936-940, ISSN: 2229-4333 (Print), March 2012

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, “Provable Data Possession at Untrusted Stores,”in the Proceedings of ACM CCS 2007, pp. 598–610.

[6] Reza Curtmol, Osama Khan, Randal Burns " Robust Remote data Checking"

[7] C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,”in the Proceedings of IEEE INFOCOM 2010, 2010, pp. 525–533.

[8] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai "Secure Auditing and Deduplicating Data in Cloud" IEEE TRANSACTIONS ON COMPUTERS VOL: PP NO: 99 YEAR 2015

[9] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013

[10] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[11] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013

[12] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In Proc. of StorageSS, 2008

[13] M. Shyamala Devi, V.Vimal Khanna, Naveen Balaji”Enhanced Dynamic Whole File De-Duplication (DWFD) for Space Optimization in Private Cloud Storage Backup”, IACSIT, August, 2014.

[14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.