Secure Auditing and Intelligent Compression in Cloud

Suzaifa, Sareen Fathima, Fathimath Kousar, Mustafa Basthikodi

Abstract— As the cloud computing technology develops throughout the decennary, externalising data to store using cloud resource becomes a trend, which is benefited in heavy data management and maintenance. Notwithstanding, since the externalized cloud depository is not fully reliable, while achieving integrity auditing it elevate security treat on how to realize single instance storage in cloud. In this work, we study the problem of integrity auditing and secure intelligent compression on cloud data. Peculiarly, directing at attaining both eliminating duplicate copies of repeating data i.e., secure deduplication and integrity of data in cloud, we propose an auditing entity with a perpetuation of a MapReduce cloud, which helps audit the integrity as well as uploading of data after clients generate data tags having been collected in cloud.

Keywords— Deduplicating Data in Cloud, prevention of duplication, duplication of data, intelligent compression, Integrity Auditing

I. INTRODUCTION

Cloud storage is a service model in which data is managed, maintained, sustained remotely and made available to users over a network. Cloud storage get more space for less money, maintains lots of hard drive storage space, mobility opportunities and to scalable service [1], [5]. These great features attract moreover users to store and deploy their confidential data to the cloud storage. Nevertheless the system of cloud storage has been wide-ranging, it fails to support some important pop up needs such as the abilities of integrity auditing of cloud files by cloud users and recognizing identical files by cloud servers.

We solve two problems. First is integrity auditing and second is secure de-duplication.

Problem of integrity auditing. Those most extraordinary assortment of cloud spare beginning with great inside limit is, larger part of the information switches through web. Also set away done a questionable space, not under control of the user toward all, which inevitably grows user's gigantic load on the balance information [1], [3]. These stresses start beginning with reality that the cloud stockpiling is feeble on security perils beginning with both outside. Also within the cloud, and the uncontrolled cloud servers may inactively cover. A rate data hardship scene from those clients to take care of their reputation.

Problem of secure deduplication. The swift advocacy of cloud services is supplemented by increasing capacities of data stored at distant cloud servers [1]. Most of them are identical, among these distant stored files.We does the convergent key of file, which is generated and controlled by using a modification on convergent encryption.

We use IaaS, providers of IaaS i.e., infrastructure as a service offer computers physical or virtual machines and other resources. IaaS clouds often offer additional resources such as a virtual-machine disk image library, block and file-based storage, firewalls, load balancers, IP addresses, VLANs, and software bundles. IaaS-cloud providers supply these resources on-demand from their large pools installed in data centers.

II. RELATED WORK

Since our slog is related to both secure auditing and secure deduplication of data, we analyse the works in both locale.

Integrity auditing. The definition of provable data possession (PDP) [2], [7], for persuading that the cloud servers owns the target files in the absence of recovering or downloading the entire data. Integrally, PDP is a probabilistic proof protocol [2][7], in which unsystematic set of chunks are sampled and asking the servers to prove that they exactly possess these chunks, and the integrity checking is retained by little number of meta data which is able to accomplish by verifier.

Secure de-duplication. Deduplication is a method where the server keeps only a one duplicate of all file, nevertheless of how numerous users request to keep that file, such that the insignificant client side deduplication leads to the discharge of side channel information [1], [6], [8]. Nevertheless, the area of cloud servers in addition web bandwidth are maintained [2], [4].

A. Existing System and its Drawbacks

Existing system has vast storage for storing information and also allows exchange of data between the cloud and user. Even though there is vast storage for data there exists some problems while doing the processes. Mainly the first problem of Integrity auditing and second is storing of multiple copies of same files i.e. Duplication of files.

B. Proposed System and its advantages

The proposed system will achieve both file de-duplication and integrity auditing. The challenge of de-duplication on encrypted is the restriction of dictionary attack. The pattern secures the problem of past slog that the computational load at client or auditor is too vast for tag creation. System also provides secure de-duplication with reliable and efficient convergent key management.

Revised Manuscript Received on April 12, 2019.

Suzaifa, PG student, BIT, Mangalore. India (E-mail: suzaifa.suzaifa@gmail.com)
Sareen Fathima, PG student, BIT, Mangalore. India (E-mail: fathimasiddiq@ymail.com)
Fathimath Kousar, PG student, BIT, Mangalore. India (E-mail: kousarfathima9@gmail.com)
Mustafa Basthikodi, Professor, BIT, Mangalore. India (E-mail: mbasthik@gmail.com)

Advantages

- While uploading and downloading the files in the cloud the files will not be modified.
- Proposed system does not allow storage for duplicate files.

III. SYSTEM ANALYSIS

A. Functional Requirements Analysis

1) Module1- Cloud user

- A customer will be a substance that necessities ought to outsource data stockpiling of the S-CSP what's all the more right the data sometime later.
- Done a stockpiling system supporting de-duplication, those customer best transfers fascinating data At doesn't exchange any duplicate data ought to save the exchange transmission capacity, which may make had Toward a similar customer or separate customers.
- In the approved de-duplication, every client has an arrangement of privileges key in the setup of the system. Each record is ensured with the private and open key. Utilizing this key just the approved de duplication should be possible.
- Methods used in this module doGet(), doPost()- to process request and response from server. getQueryString(), getSession(), etAttribute(), toString() are the attribute used in this module.

2) Module2- Auditor

- Audit is only Evaluator. He transfers the information and reviews their information and act like authentication expert. Inspector can have the combine of open key and furthermore the private keys.
- Open key is utilized to encode the information and in addition private key is utilized for unscramble the scrambled information. The point of this work is to give the accuracy of the remotely put away information.
- The general population check can done anybody however not only the customers initially put away of the document to perform confirmation.
- Auditor module uses function doGet(), doPost()- to process request and response from server. getParameter()- to get auditorname and password, equalsIgnoreCase()- to check the right admin or not, sendRedirect()- to link to webpage.

3) Module3- Secure De-duplication system

- We consider a couple sorts about insurance we need secure, that is, i) un-fashion limit of copy check token: there need help two sorts of enemies, that is, external enemy What's more inside adversary.
- Similarly as exhibited beneath, the external adversary camwood be viewed as an inside enemy with no whatever advantage.
- On a customer need the advantage p, it obliges that those enemy can't form and yield a significant duplicate token for whatever practical advantage p′ ahead any record Q, the place x doesn't coordinate p′. Besides, it additionally obliges that assuming that those enemy doesn't make about token with its character or advantage beginning with private cloud server, it can't mold What's more yield a generous duplicate token for p for any Q that need been questioned.
- Methods used-getConn() to connect database,getInputStream() to get input, getName() to get file name.

a) AES Analysis

For show day cryptography, AES will be comprehensively grasped furthermore supported secured nearby both hardware what's all the more programming. Till date, no valuable cryptanalytic strike against AES need been kept running over. Also, AES require inborn versatility from guaranteeing enchantment length, which allows A level from asserting 'future-sealing' against progression in the ability with perform comprehensive enchantment seeks.

b) KEY triple DES

Before using 3TDES, customer To start with creates Furthermore passes on a 3TDES key K, which includes around three various des keys K1, k2 Also K3. This suggests those genuine 3TDES keyneed period 3×56 = 168 chances. That encryption plan might be outlined Likewise takes after.

IV. SYSTEM DESIGN

A. General Architecture

- Input Design is the way toward changing over a client arranged portrayal of the contribution to a PC based framework. This plan is essential to maintain a strategic distance from blunders in the information input process and demonstrate the right course to the administration for getting right data from the electronic framework.
- It is accomplished by making easy to use screens for the information passage to deal with extensive volume of information. The objective of planning info is to make information section simpler and to be free from blunders. The information section screen is outlined such that every one of the information controls can be performed. It likewise gives record seeing offices.
- When the information is entered it will check for its legitimacy. Information can be entered with the assistance of screens. Proper messages are given as when required so that the client won't be in maize of moment. Hence the goal of info configuration is to make an information format that is anything but difficult to take after.
- System architecture is shown in Fig 1., for the system.

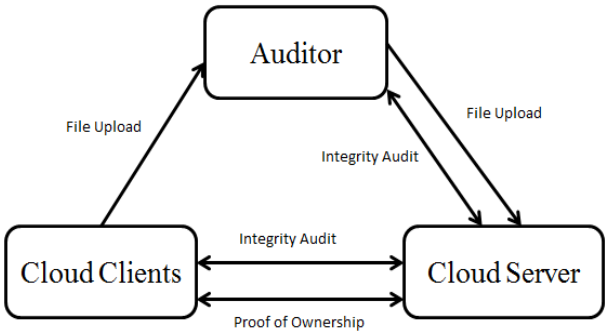

Fig 1. System Architecture

B. Sequence diagram

A succession Diagram demonstrate the diverse members and the association between them by the arrangement of messages, A grouping outline demonstrates the collaboration of the framework with the on-screen character like an utilization case yet concentrates more on the exchange of messages between the members, below Fig 2., shown for sequence diagram.

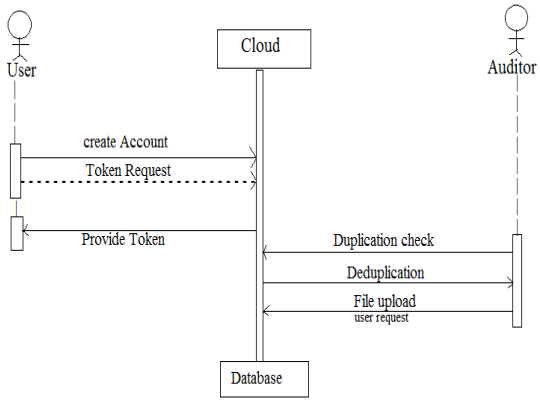

Fig 2. Sequence Diagram

C. Algorithms

1) AES(advanced encryption standard)

Encryption process

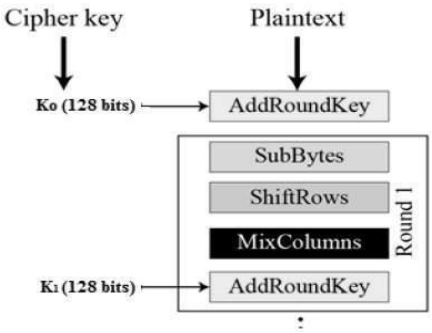

Fig 3. Encryption process

Fig 3. shows encryption process. Here, we consider on delineation of a normal round of AES encryption. Each round contains the four sub-approaches. The essential round change resembles the accompanying.

Decryption Process

The interpreting procedure is the reverse of the encoding procedure. An each round has the includes the four strategies:

- Byte substitution
- Mix rows
- Mixing of columns
- Add round key

Since sub-approaches in each round would done inverse way, Dissimilar to for A Feistel Cipher, the encryption and unscrambling counts require to be freely actualized, despite they require help out and out about related.

2) 3-KEY triple DES

Before using 3TDES, customer To start with creates Furthermore passes on a 3TDES key K, which includes around three various des keys K1, k2 Also K3. Fig b.shows DES. This suggests those genuine 3TDES key need period 3×56 = 168 chances. That encryption plan might be outlined Likewise takes after −.

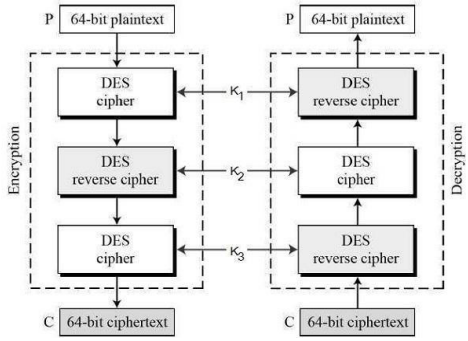

Fig 4. DES encryption and unscrambling

D. Data flow diagram

- The DFD is called as air pocket outline. It is a straightforward graphical method that can be utilized to speak to a framework, input information to the framework, different handling completed on this information, and the yield information is created by this framework.
- The information stream outline (DFD) is a standout amongst the most essential displaying apparatuses. It is utilized to demonstrate the framework segments. These segments incorporate the framework procedure, the information utilized, an element that interfaces with the framework and the data streams in the framework.
- DFD indicates how the data travels through the framework and how it is adjusted by a progression of changes. It is a graphical procedure that demonstrates the data stream and the changes that are connected as information moves from contribution to yield.
- DFD is otherwise called bubble diagram. A DFD might be utilized to speak to a framework at any level. DFD is spoken to as levels that speak to expanding data stream and utilitarian detail. Below Fig 5. shows data flow diagram.

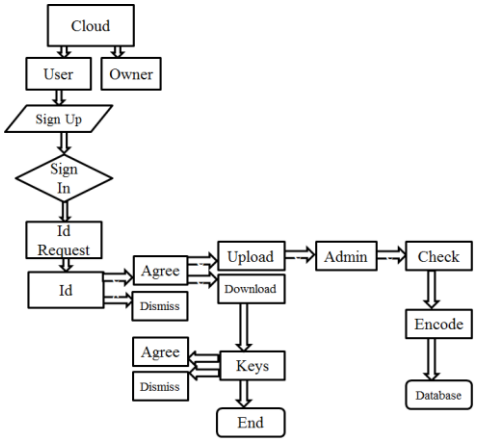

Fig 5. Data flow diagram

V. EXPERIMENTATION AND RESULT ANALYSIS

A. Unit Testing

These aides in checking the yield of inward capacity appropriately and gives the processing plant yield .It is utilized for just single –single yield after fulfillment of each capacity. Know decision appendages and internal code stream should make affirmed. It is utilized for result of single output. It is done then subsequently the realization around an one of a kind unit When blend. Unit tests perform basic tests at part level Also test a specific advantages of the business procedure , application, or structure setup. Unit tests ensure that each intriguing method for advantages of the business philosophy performs perfectly of the recorded judgments and holds clearly described data sources also required impacts

B. Test Cases

TABLE I. LOGIN

Test case	1
Test name	Loading login details
Test Item	Data from user
Input	User details
Excepted output	Login successful
Output	Excepted output
Results	Positive

TABLE II. UPLOAD

Test case	2
Test name	Upload data
Test Item	Upload text file
Input	Data in table format
Excepted output	Upload data in cloud
Output	Excepted output
Results	Positive

TABLE III. DOWNLOAD

Test case	3
Test name	Download file
Tested Item	Number of text file download
Input	Downloaded file
Excepted output	Should display the downloaded file
Results	Positive

TABLE IV. UPDATE

Test case	4
Test name	Update
Tested Item	Update the text file
Input	Should provide public and private key
Excepted output	Updated
Results	Positive

TABLE V. REGISTRATION

Test case	5
Test name	Registration
Test Item	Insert details
Input	Text format special characters
Excepted output	Save
Output	Excepted output
Results	Positive

TABLE VI. DELETE

Test case	6
Test name	Delete
Test Item	Delete file
Input	Delete
Excepted output	Alert sent to owner
Output	Excepted output
Results	Positive

TABLE VII. DE-DUPLICATION

Test case	7
Test name	De-duplication
Test Item	File upload
Input	Upload
Excepted output	Avoids duplication
Output	Excepted output
Results	Positive

The experimental analysis is done using NetBeans IDE and cloudHQ, snapshot of work is given from Fig 6 to Fig 19

Fig 6. Registration Page

Fig 7. Server login

Fig 8. User login

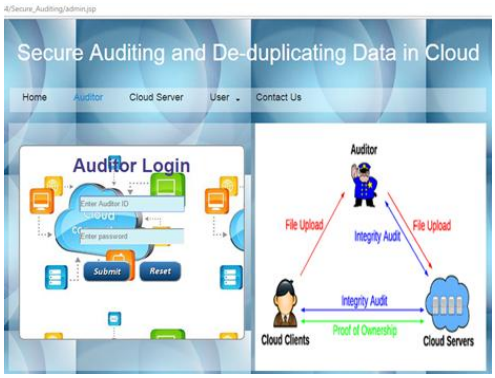

Fig 9. Auditor login

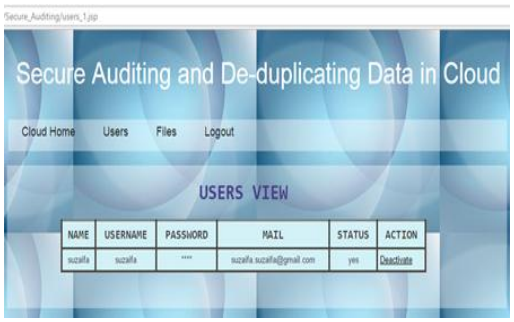

Fig 10. Activating account in cloud

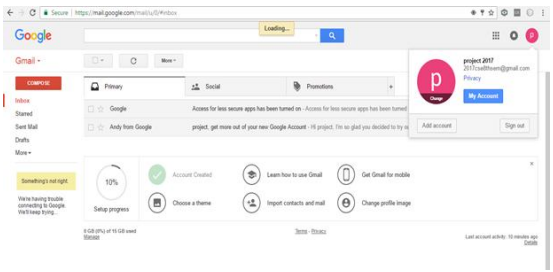

Fig 11. Email which sent token id

Fig 12. Entering token id

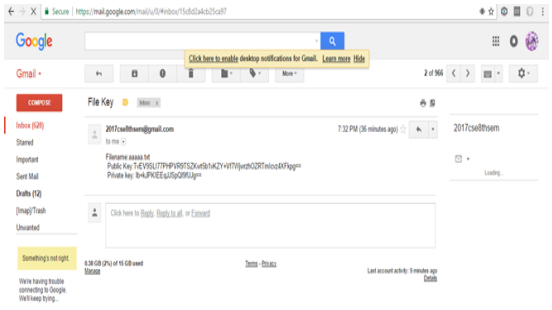

Fig 13. File key

Fig 14. To Upload choose file

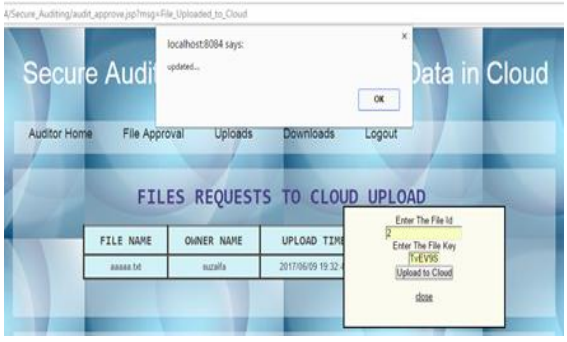

Fig 15. Auditor approval to upload file

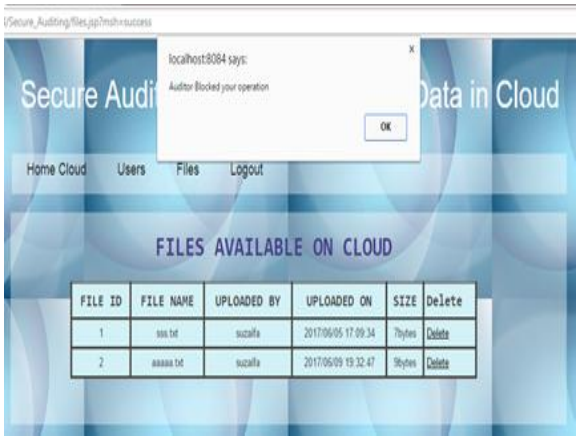

Fig 16. To show other user delete the file in cloud

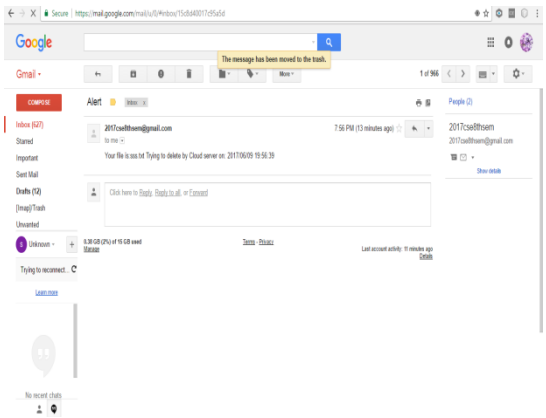

Fig 17. Alert message sent to user

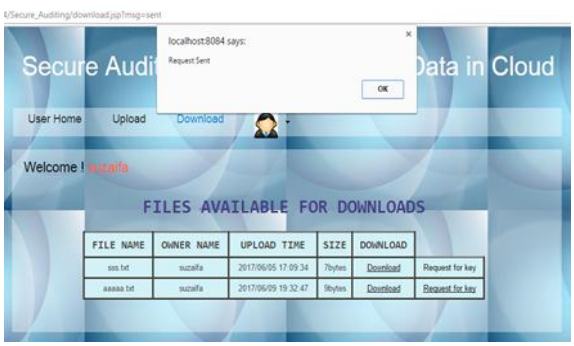

Fig 18. To download files

Fig 19. Enter public and private key to download

VI. CONCLUSION

This structure settles those issue in that those computational load throughput, so we are utilizing AES for the encryption and 3-KEY triple DES to protect against attacks , we can transfer content record, not other files. Our proposed system successfully achieved Integrity auditing and secure de-duplication in cloud, which are main factors in achieving security in cloud. We propose a system for plainly examining integument looking under encoded share of the information.

REFERENCES

1. Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai," Secure Auditing and Deduplicating Data in Cloud." IEEE Transactions on Computers, 2016, 65.
2. J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.
3. H. Shacham and B. Waters, "Compact proofs of retrievability," in Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryp- tology, ser. ASIACRYPT '08. Springer Berlin Heidelberg, 2008, pp. 90–107.
4. S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Server- aided 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]. Avail- able: https://www.usenix.org/conference/usenixsecurity13/technical - sessions/presentation/bellare
5. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp. 50–58, 2010.
6. S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Server- aided 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]. Avail- able: https://www.usenix.org/conference/usenixsecurity13/technical - sessions/presentation/bellare
7. G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, ser. SecureComm '08. New York, NY, USA: ACM, 2008, pp. 9:1–9:10
8. J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure dedupli- cation with efficient and reliable convergent key management," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 6, pp. 1615–1625, June 2014.