Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

A Survey on Auditing Encrypted Data and Deduplication in Cloud Environment

Prashant Sadaphule[1], Rishabh Rapatwar[2], Uday Mahana[2], Priya Jawale[2], Chetan Magar[2]

[1]*Assistant Professor, Computer Engineering, AISSMS's IOIT, Savitribai Phule Pune University, Pune.*
[2] *Student, Computer Engineering ,AISSMS's IOIT, Savitribai Phule Pune University, Pune.*

Abstract- *Cloud systems have become extremely popular in the past few years. There has been rapid growth in the amount of data in regards to information systems in the past few years. A critical issue needs to be taken care of while considering cloud storage which is duplication. According to a recent research around 70-80% of data on cloud servers is redundant which can be a disadvantage. To solve this problem data deduplication is brought into picture. Deduplication means saving only one copy of a particular file on the cloud which proves to prevent wastage of space over cloud by deleting the duplicated data. This also helps in easy maintenance and cost effective methods of saving data over a cloud environment.*

In this paper the system works on the problem of integrity auditing and secure deduplication of data in cloud. This calls for the introduction of two systems i.e. SecCloud and SecCloud+. SecCloud is designed to help the client audit the stored data and generate data tags before client uploads the file on the server. SecCloud has reduced the computational load that the user faces during the auditing and uploading phase. SecCloud+ guarantees the confidentiality of file besides securing the deduplication and performing integrated auditing. Encryption largely helps in securing clients data which can be saved if pursued by malicious attacks this also enables integrity auditing and secure deduplication on encrypted data. The proposed system supports authorized deduplicate check in cloud architecture.

Keywords- *Cloud Computing, Cloud Storage, Data Deduplication, Proxy re-encryption, Security Proxy.*

I. INTRODUCTION

Cloud Computing is one of the widely used technologies all around the globe. Cloud computing provides Cost effective, Scalable and Hardware independent architecture for data management and storage. The increase in amount of data generated around the globe calls for popularity in multi user storage systems. To reduce the issue of data redundancy, data de-duplication technology has become a hot research topic. Data deduplication means that in a particular cloud storage environment, only one copy of same data can be stored in opposition to storing multiple copies of a single file. Data de-duplication proves to be scalable as well as more effective storage technology. Data deduplication reduces the amount of data to transmit across the network. This saves significant money in terms of storage costs and backup speed. For instance, a cloud user wants to upload a file to the server. The deduplication protocol checks for the same file and then and then only the user is permitted to upload the file to the server. If the file is already present on the server, uploading same another file makes no sense. It just increases the used data volume and raises concerns about the storage.

In this system two individual systems called as SecCloud and SecCloud+ are illustrated. SecCloud is designed to help the client audit the stored data and generate data tags before client uploads the file on the server. The computational load that the end user used to suffer has reduced to a great extent because of this design. SecCloud is also helpful in deduplication. Secure deduplication in this context means it only eliminates extra copies of data; none of the original data is lost. In this context, the term security is reffered to the prevention of leakage of side channel data. To prevent the data loss a protocol called Proof Of Ownership is introduced that allows the clients to prove to the cloud servers that they exactly own the target data. The POP protocol takes place between the Client and the cloud users. SecCloud also performs integrity auditing on the data to be stored on cloud storage servers so that it can be easily fetched whenever needed. This is done by maintaining a set of data tags for each file which is to be stored in the cloud. Apart from integrity auditing of data files SecCloud also performs Secure Deduplication on data for optimization of space over a particular cloud. SecCloud guarantees the confidentiality of file besides securing deduplication and the integration of auditing. Deterministic encryption in convergent encryption helps us achieve a method of auditing the integrity of encrypted data. The challenge to achieve integrity is to prevent the encrypted data from dictionary attacks.

II. RELATED WORK

Ateniese et al. [1] proposed a dynamic Proof of Data Possession(PDP) schema but without insertion operation. Wang et al. [2] proposed proxy PDP in public clouds. Wang et al. proposed proxy Proof of Data Possession (PDP) in public clouds. They also improved the POR model by manipulating the classic Merkle hash tree construction for block tag authentication. Yuan et al. [5] proposed Proof of Retrieveability (POR) and Proof of Data Possession (PDP) techniques that assure data integrity for cloud storage. They also proposed Proof of Ownership (POW) that improves

storage efficiency by securely removing unnecessarily duplicated data on the cloud storage. Achieving both data integrity and storage efficiency contradicts the objectives of POW. Yuan et al. [5] proposes a novel scheme based on techniques including polynomial-based authentication tags and homomorphic linear authenticators. This scheme outperforms the POR and PDP schemes and gives an additional functionality of file deduplication. Halevi et al. [6] proposes a problem where an attacker can gain access to arbitrary-sized files of other users based on very small hash signatures of these files. Attacker can convince the cloud storage that it owns the file and the cloud storage may even give permission to the attacker to download the entire file. To overcome such attacks, Halevi et al. [6] introduced the notion of Proof of Ownership (POWs). Vanitha et al. [8] proposed a method based on Probabilistic query and periodic verification. The proposed method helps improve the performance of audit services and also audit system integrity. Hao et al. [9] presents a cryptographic solution to make the data deletion process more transparent and verifiable.

III. PROPOSED SYSTEM

A. The primary aim of the proposed system is to achieve data integrity and deduplication in cloud. A duplicate copy of a single file on the cloud server largely wastes the space. Hence there has to be some method introduced which takes care of the aforementioned problem. Two systems called SecCloud and SecCloud+ are therefore introduced.

B. SecCloud puts forward an auditing entity in cloud environment. This integrity auditing takes place in two parts which is between cloud client and the cloud and the second auditing takes place in between the auditor and cloud. The implementation of SecCloud helps the client audit the stored data and generates data tags before client uploads the file on the server. The auditing protocol helps the user and auditor to use the cloud space more ideally and also helps the user fetch the desired file in a proper manner. The integrity auditing function also helps to provide the capability of verifying the correctness of the remotely stored data. SecCloud also guides Secure Deduplication. This makes sure that during uploading or downloading the file no data is lost over the network. In an ideal SecCloud module Block level deduplication and byte level deduplication is supported as well.

C. SecCloud+ is designed as advancement in SecCloud module. Apart from securing deduplication and supporting integrity auditing in cloud environment, SecCloud+ enables the assurance of file confidentiality. File confidentiality in this context means the section of encryption and decryption of files which commute to the network towards the cloud storage and inward from the cloud storage. Encryption from the users end makes sure that the file to be uploaded is converted into cipher text which makes it difficult for the attackers to access. SecCloud+ is an advanced constructed system designed to fulfil the users demands for secure file transfer over the network, this happens since the users first prefer to encrypt their data. In this system the auditing directly takes place on the encrypted data. This also prevents the malicious attacks such as dictionary attacks and accessing user's data. Bit exchanging method and Hash algorithm are used to implement SecCloud+ module. As compared to SecCloud system, SecCloud+ is a superior advancement since it helps users prevent their data using encryption methods.

IV. SYSTEM ARCHITECTURE

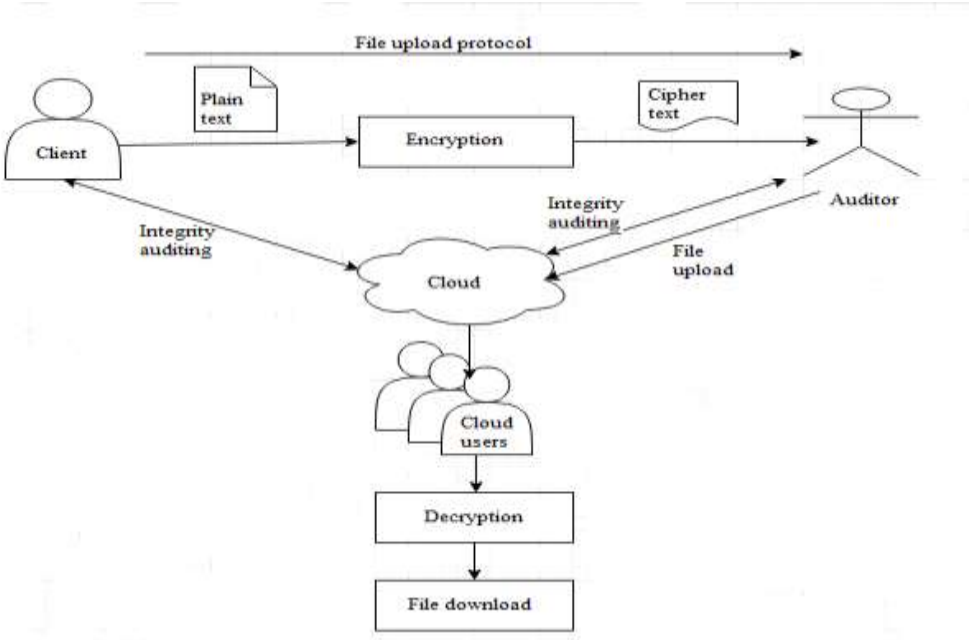

Figure 1.Architecture diagram

The System consist of 3 modules in total as follows

a) Cloud Client (Users)
b) Auditor
c) Cloud

A client has to register and login to upload its files on the cloud server. The registration details are maintained in a database on the cloud. The file uploaded by the client is in encrypted format. An encryption process called BEM (Bit Exchanging Method) is implemented to encrypt the file. The uploaded file is first audited by the Auditor, then and then only it is uploaded on the cloud. User downloads the file in encrypted format. User must enter correct key to decrypt the file. Auditor uses Secure Duplication Protocol that sends the status of a file being duplicated or not for all the files being uploaded by the client. If the file is duplicated then the auditor doesn't give permission for the file to be uploaded on the cloud. The auditor gives activation for non duplicated files only. The file is then uploaded on the cloud.

V. ALGORITHMS

Bit exchanging method:

Bit exchanging method is an encryption or decryption method which uses bit shift and XOR operation on the secret file.

The bit exchanging method used to encrypt the data is as follows:

Steps:

1. Read bytes from the secret data. Convert each given byte to 8 bits of data.
2. Use one bit right shift method to converted bits.
3. Divide the 8 bits into two blocks of 4 bits each, one on the left and other on the right and
4. Perform XOR operation on both the blocks
5. Repeat the same thing for all the bytes in the given file.

Hash algorithm for file level deduplication:

1. Start
2. Declare variable
3. Initialize variable
4. Read the file name
5. Read the file name till the end of file title
   Generate hash value from strBUFF[FILENAMESIZE]
   if (FirstFile)
   Consider node as root element
   Inc FileCtr
   Else
   Search the generated hash in BST
   If (Find Hash == True)
   Compute the node
   Add the node to a linked list
   Change the Endlink of SLL
   Else
   Add the node in BST
   Inc The FileCounter
6. Calculate Deduplication Ratio
7. Display the Result for each file iteration
8. END

VI. IMPLEMENTATION MODULE

A. User Module: In this module a user registers with the cloud to upload its files to the server. The registration details are stored and maintained in a database. Then using the login credentials, user login to the cloud. Auditor also uses his login credentials to login to the cloud database.

B. File uploading/downloading protocol- The uploading/downloading protocol allows the client to upload or download a file. The uploaded and downloaded files are in encrypted format. A method called BEM (Bit Exchanging Method) is used for the encryption and decryption process. Decrypting a downloaded file requires a key to be entered, and then only the file will be decrypted. To be specific, the protocol is takes place in following three steps.

Phase 1-Consider a Client wants to upload a file named abc.txt to the server. In the first phase a hash number is given to the file. The hash number of the file is checked against the hash index in the database of the server. Hash algorithm performs this operation. If the file turns out to have a hash number already present in the index then another protocol called Proof of Ownership (POP) is ran between the client and the cloud storage server.

Phase 2- Client 1 uploads file abc.txt to the auditor, and receives a receipt from auditor for the same file.

Phase 3- A set of data tags are generated by the auditor and are sent along with the file to the cloud server.

C. Integrity Auditing Protocol-This protocol helps the user and the auditor verify the integrity of the system. Any entity other than the cloud server can initialize the verification. In this protocol auditor logins to the cloud server using the login credentials. A secure duplication protocol that sends the status of all the files being uploaded on the cloud is used by the auditor to audit and approve the non duplicated files to store on the cloud. Auditor doesn't give uploading permission to the duplicated files. Activation is given to non duplicated files only. Auditor along with giving activation to the file being sent by the client, also audits the file storage. This protocol has two key players, the prover and the verifier. Cloud server is the prover and the auditor or client is the verifier.
Phase 1- Verifier or auditor produces a series of challenges for the file abc.txt and sends them to the prover.
Phase 2- This phase consists of prover trying to prove the verifier that the target file is owned by them. This is done by sending a proof of target file. The proof of generated tags for the file abc.txt is sent to the verifier.

D. File level Deduplication: Deduplication is the ability to reduce the required storage capacity by only storing the unique data. There are three levels of deduplication to which the data can be optimized. They are File level, block level and byte level. It defines the minimal data fragment that is checked by the system for redundancy. File level deduplication is being used using hash algorithm. Hash algorithm analyzes the file title for every file being uploaded on the cloud. It generates a hash number for every analyzed title. This hash number is then stored in an index. Every new file uploaded to the auditor gets a hash number and is checked against this index for duplication. Two duplicated files have same hash numbers.

E. Proof of ownership protocol- This protocol is used to prove that the client owns the targeted file. This protocol guarantees the ownership of the file. The protocol can be illustrated in two phases.
Phase1-A series of challenges are produces by the cloud server for the file abc.txt and sent to the client.
Phase2-The client responds with the proof of ownership. The cloud server verifies the validity of proof for the file which user is trying to upload over the cloud.

VII.CONCLUSION

In this system two modules have been proposed called as SecCloud and SecCloud+. their primary aim is to achieve data integrity and deduplication of data on cloud servers. SecCloud takes cares of integrity auditing in cloud environment, this also takes care of secure deduplication in cloud storage. As compared to the previous works the implementation module is more secure since before uploading any data in SecCloud+ the data is encrypted before uploading it on the cloud server, this also allows for integrity auditing and secure deduplication before hand on encrypted data. Encryption in SecCloud+ helps save important data which is passed over the network. These two systems largely optimize cloud storage and helps maintain a set of tags for already stored file. Proof of ownership protocol helps the user to get the acknowledgment regarding a particular file he is trying to upload. If the user is authorized the user can download the file and access the data.

VIII.REFERENCES

[1] Ateniese.G, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z.Peterson, and D. Song,"Provable Data Possession At Untrusted Stores",*In Proc. 14th ACM Conf. Computer and Comm. Security* (CCS"07), pp. 598-609, 2007

[2] Wang.Q, C. Wang, K. Ren, W. Lou, and J. Li,"Enabling Public Audit Ability And Data Dynamics For Storage Security In Cloud Computing", *In IEEE Trans. Parallel Distributed Systems*, vol. 22, no. 5, pp. 847-859, May 2011.

[3] Yan Cong Wang, Student Member, IEEE, Sherman S.M. Chow, Qian Wang, Student Member, IEEE, Kui Ren, Member, IEEE, and Wenjing Lou, Member, IEEE,"Privacy-Preserving Public Auditing For Secure Cloud Storage"

[4] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, "A view of cloud computing", In *Communications of the ACM (CACM)* Vol. 53 No. 4 April 2010.

[5] Jiawei Yuan, Shucheng Yu,"Secure and Constant cost public cloud storage auditing with Deduplication", In *Communications and Network Security (CNS), 2013 IEEE Conference* Oct 2013

[6] Shai Halevi , Danny Harnik , Benny Pinkas , and Alexandra Shulman-Peleg, "Proofs of Ownership in Remote Storage Systems", *18th ACM conference on Computer and communications security* oct 2011

[7] N.Vidhya, P.Jegathesh ,"Secure file sharing of dynamic audit services in cloud storage", *International Journal of Research in Engineering and Technology* Volume: 03 Issue: 05 May 2014

[8] M.Vanitha, Ar.Sivakumaran, L.Priyadharshini ,"A Study on Secure Storage of Dynamic Audit Services in Cloud", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* Vol. 1, Issue 1 July 2012

[9] Feng Hao, Member, IEEE, Dylan Clarke, Avelino Francisco Zorzo ,"Deleting Secret Data with Public Verifiability", *IEEE Transactions on Dependable and Secure Computing* (Volume: 13, Issue: 6, Nov.-Dec. 1 2016).

