IJMTST | Volume: 2 | Issue: 06 | June 2016 | ISSN: 2455-3778

Secure Auditing and Deduplicating Data on Cloud

H A Hingoliwala[1] | Pratik R Kamble[2] | Prafull R Pansare[3] | Pratik S Raut[4] | Dipti Salunkhe[5]

[1]Assistant Professor, Department of Computer Engineering, JSCOE, Hadapsar, Pune, India
[2,3,4,5]UG Students, Department of Computer Engineering, JSCOE, Hadapsar, Pune, India

ABSTRACT

Cloud computing is a technology that used for storing and accessing. The data on remote location. It is totally internet-based.it is self-service and on Demand technology. That's why now days it's mostly used and popular term. Enterprises and organizations used cloud storage for access data to third-party. As like, the single user also use the confidential data anywhere, anytime on earth It is now becoming business standard. Its simplify users accessibility. It is cost saving and flexible for better performance on internet. But is also occur drawbacks like security and integrity on data. Like many times the data is already available on storage but it contain slightly difference. So overcome this problems we introduce two secure system, namely seccloud and seccloud+. Seccloud is used for generating tags on data before uploading and seccloud+ is maintain the integrity auditing and secure de-duplication on data because every customer wants to encrypt their data before uploading. Data integrity and storage efficiency are tw o important aspect of cloud storage. Proof of Retrievability (POR) and Proof of Data Possession (PDP) techniques Assure data integrity for cloud storage. Proof of Ownership (POW) improves Storage efficiency by securely removing unnecessarily duplicated data on the Storage server. Cloud computing is one of the most talked about IT trends today.in cloud more application availability on the cloud. Also cloud increased growth in the market. Cloud is best technology for increased development and more innovation to make hybrid cloud adoption.

KEYWORDS: *Seccloud , seccloud+, integrity auditing ,secure de-duplication , proof of ownership convergent encryption.*

I. INTRODUCTION

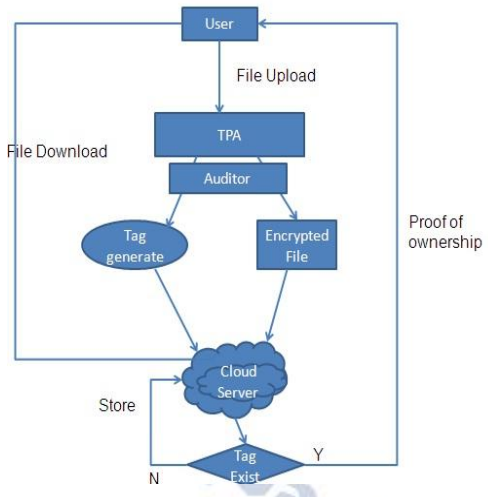
Cloud storage is a model of internet enterprise storage where data is stored in virtualized pools of storage which is hosted by third-party. Cloud storage provide offers for customer which generated more benifit for cloud companies, like popularity, more user. Even though now days cloud storage system has been smart option for work. And also it is affordable, but it has certain limitation .The main problem of client data management and maintenance which is able to Relief by cloud server storage system of cloud is different from another storage System. The first problem is integrity auditing, i.e when we uploaded data it upload various manner like packets tokens which is less secure because if any packet loss while transmitting its occur problem for client. As well as it's to easy for a professional Attacker to attack. So its most important that maintain the integrity of data on storage system. The data is transferred via internet and stored in uncertain domain not the under control of client [1]. The another uncontrolled cloud server may passively hide the any problem related data for their reputation. It is more important that cloud server might even actively and deliberately discard rearely accessed data files belonging to an ordinary file. The second problem is secure deduplication. In cloud storage among these remote stored files, most of them are already on storage. According to recent survey by EMC, 70% of files are duplicated copies .Because its helps to cloud servers paid more for space from client. That's the one of the reason why many cloud server are store duplicate copies of data. And Its more risky to available

duplictate copies of data in storage. Stored data is various manner lie confidential password, banking detail, personal information,it is open invitation for attacker .In cloud server, server store every single file link with the who ask for the file. Cloud server needs to verify whether the user actually owns the file before creating a link for user.

In de-duplicate data, when a user wants to upload a data file that already exists in the cloud storage, the cloud server executes a checking algorithm to see whether or not this user actually possesses the whole file i.e. it checks the file attribute. If the user passes the checking, he/she can directly use the file existed on the server without uploading it again.

To overcome such problems cloud server uses proofs-of-ownership protocol, which let a client efficiently prove to a server that the client holds a file, rather than short-information about it. In this a file have different ownership which introduce rigorous security defination. For working daynamic data proof-of-retrievability protocol used . Because dynamic data operation can be vital importance to storage outsourcing services.

II. System Flow



As shown in fig cloud client upload the data on server. While uploading file the Auditor check that The file is already available on server using different mechanism. If file is not available then Auditor takes the file. also other client side(who owns data) its perform some security level task for conforming the user(who wants data) actually owns the data

III. Related Work

A. *Existing System:*

Earlier client upload data file on cloud in plain text format. And wants to maintain the integrity and security on that plain data file. Customer always choose the safest and cheapest method for the data storage and transformation on cloud. but that's not possible to give all feature in such minimum amount .every system has some drawbacks and various problems.

Existing system drawback

1. It is very difficult to audit the files huge and large amount of data in cloud using integrity auditing.
2. Lots of Duplicate files in cloud

The number of security problems that are faced by cloud computing are

- Data issues
- Privacy issues
- Infected application

B. *Proposed System:*

To solve this problem on existing system we propose two secure system. Which generate better And Efficient system for accessing massive data on cloud. In this, firstly encrypted the plain data file and perform integrity auditing on that encrypted file.

- SecCloud

SecCloud system has achieved both integrity auditing and file deduplication.in this process server doesn't known the contain in the file. So here acceptance confidientially from sever is less secure .In other words, the functionalities of integrity auditing and secure deduplication are only imposed on plain files.

- SecCloud+

SecCloud+, which is used for maintaining integrity auditing and managing deduplication on encrypted files. In other Word perform the operation on the secure file. i.e encrypted files which encrypted by SecCloud over the plain text file. System Model Compared with SecCloud, our proposed SecCloud+ involves an Additional trusted entity, Namely key server, which is responsible for assigning clients with Secret key (according to the file content) for encrypting files.

IV. PROTOCOL

We used different protocol to define the working of system. In this we used 3 different protocols on data file.

 File Uploading Protocol

In this protocol client upload data files with the help of Auditor . for uploading data file it fulfill
The requirement of uploading protocol. Specifically, the file uploading protocol includes three Steps:

• Phase 1 (cloud client → cloud server): in this phase when client upload the data file, firstly checks the duplication of file. If file is already present then perform POW(Proof Of Ownership) On that file.

• Phase 2 (cloud client → auditor): After performing Client side operation client send data file to auditor, and receives a receipt from auditor.

• Phase 3 (auditor → cloud server): after recive data file from client auditor preform some task like-helps generate a set of tags for the uploading file, and send them along with this file on cloud server

 Integrity Auditing Protocol

This protocol work on the maintaining integration of data file. i.e perform the verification on data file
This protocol includes two Phase:

• Phase 1 (cloud user/auditor → cloud server): verifier (i.e., client or auditor) generates a set of challenges and sends them to the prover (i.e., cloud server).

• Phase 2 (cloud server → cloud user/auditor): on the basis of stored file, prover (i.e., cloud server) tries to prove that it exactly owns the target file by sending the proof back to verifier (i.e., cloud client or auditor). At the end of this protocol, integrity verification is done on file, if the verifier output become true.

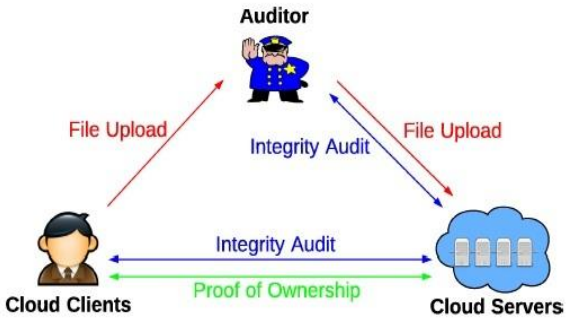
 Proof of Ownership Protocol

It is an interactive protocol which run on cloud server to verify the client, in this client play the role of prover to cloud server for its own claimed file. This protocol also includes two phase.

• phase 1 (cloud server → user): cloud server generates a set of challenges and sends them to the client. Server paly important role here, because the basic principle is file can access only authorized user .if any unauthorized get the file access it become dangerous for client as well as cloud server also.

• phase 2 (user → cloud server): ones the client responds with the proof for file ownership, and cloud server finally verifies the validity of proof. The access granted for data file without any challenges.

V. WORKING OF SYSTEM MODEL



MODULES DESCRIPTON:-

Cloud Clients

- ✓ Cloud Client is fixed hardware or software which perform accessing and storing data on virtual server. It is important because cloud services is useless if client not used. It may be any device like computer, mobile, browser etc.

Cloud Servers

- ✓ Cloud Server is nothing but virtual pool server which provide different services for different client with the help of internet. It is service – oriented architecture which provide high-capacity network, low-cost,hardware-virtualization. It is work platform independent. Its support vaious devices with their compatibility.

Auditor

- ✓ Auditor is the system or manual software which helps clients upload and audit their out- sourced data maintains a MapReduce cloud and acts like a certificate authority. This assumption presumes that the auditor is associated with a pair of public and private keys. Its public key is made available to the other entities in the system.

ADVANTAGE OF SYSTEM:

1. It provides the Integrity auditing by clustering the files with removing the duplicate files.

2. The duplicate files are mapped with a single copy of the file by mapping with the existing file in the cloud

VI. Algorithm

A. Bilinear Map and Computational Assumption

Definition 1 (Bilinear Map): Let G and GT be two cyclic multiplicative groups of large prime order p. A bilinear pairing is a map e : G × G → GT with the following properties:

• Bilinear: e(ga 1 ; gb2) = e(g1; g2)ab for all g1; g2 ∈R G and a; b ∈R Zp;

• Non-degenerate: There exists g1; g2 ∈ G such that e(g1; g2) ≠ 1;

• Computable: There exists efficient algorithm to compute

e(g1; g2) for all g1; g2 ∈R G.

The examples of such groups can be found in supersingular elliptic curves or hyperelliptic curves over finite fields, and the bilinear pairings can be derived from the Weil or Tate pairings.

For more details, We then describe the Computational Diffie-Hellman problem, the hardness of which will be the basis of the security of our proposed schemes.

Definition 2 (CDH Problem): The Computational Diffie-Hellman problem is that, given g; gx; gy ∈ G1 for unknown x; y ∈ Z∗ p, to compute gxy.

B. Convergent Encryption

Convergent encryption provides data confidentiality in deduplication. A user (or data owner) derives a convergent key from the data content and encrypts the data copy with the convergent key. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. Formally, a convergent encryption scheme can be defined with four primitive functions:

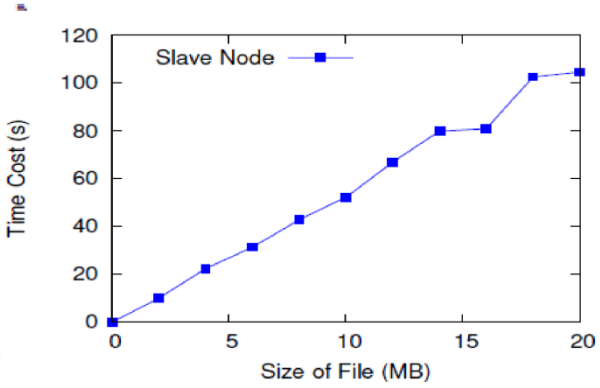
• KeyGen(F) : The key generation algorithm takes a file content F as input and outputs the convergent key ckF of F;

• Encrypt(ckF;F) : The encryption algorithm takes the convergent key ckF and file content F as input and outputs the ciphertext ctF;

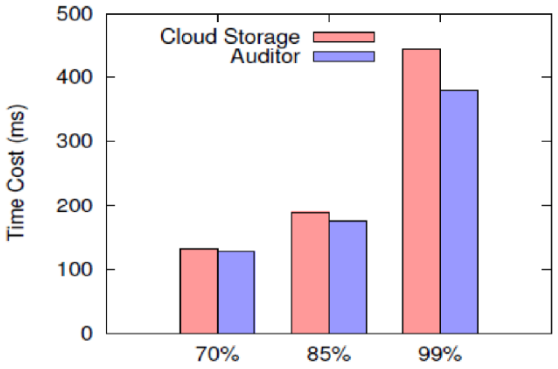
• Decrypt(ckF; ctF) : The decryption algorithm takes the convergent key ckF and ciphertext ctF as input and outputs the plain file F;

• TagGen(F) : The tag generation algorithm takes a file content F as input and outputs the tag tagF of F. Notice that in this paper, we also allow TagGen(·)

c. Performance analysis



. Tag Generation (1)



7. File Auditing

VII. Literature Survey

The author Qian wang researched the new paradigm for security challenges in cloud where the management of data is not trustworthy. For overcome problem of integrity author introduces TPA (third party auditor)model, on behalf of the cloud client ,to verify the integrity of the dynamic data stored in the cloud. TPA Eliminates the involvement of client through the auditing of whether his data is stored in secure manner[1]. The popularity of cloud increasing now days, but problem is deduplication of data .Which is helpful for attacker that exploit client-side deduplication. Allowing an attacker to gain access to arbitrary-size files of other user based on very small hash signature of a file can convince the storage of these files.[2]

The author Mihir Bellare researched on secure DupLESS system. Which provide secure duplicated storage resisting brute-force attacks. It allow clients to store their data in encrypted format with an existing service,and their services perform

de-duplication on their behalf, and yet achieves storng confidentiality guarantees[3].

The author jiawei yuan and shucheng yu improves storage security using proof-of-retrievability(POR) and proof-of-data possession(PDP) by removing unnecessarily duplicated data on storage server[4].

The author Karyn Benson, Hovav Shacham,BrentWaters propose Bilinear Map cryptographic algoritham. Which is Identity-Based Encryption. System based on. The author build bilinear map syatem that depend on weaker assumptions than the decisional-BHD assumption[5].

The author Karyn Benson, Hovav Shacham,San Diego,Brent Waters proposed convergent key management for data deduplication. Using public key and private key decide the access for file and the secure transmission using keys on multiple server[6].

VIII. Conclusion

We examine that different algorithm that helps to secure transmitting data on cloud server. Like uploading, downloading data using key ,word search algorithm. Which ensure the cloud Storage security. And to overcome all existing system proposed SecCloud and SecCloud+. SecCloud introduces an auditing entity with maintenance which helps client to tag their file/data before uploading on server as well as maintain the integrity of that data. SecCloud uses proof-of-ownership protocol for secure data de-duplication also prevent from data leakage on internet. SecCloud+ is an advanced method for SecCloud that encrypt clients data before uploading , and Allow secure integrity auditing and data de-duplication on that encrypted data.

References

[1] 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.

[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] 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.

[4] 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

[5] 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 Communications Security*, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 12:1–12:34, 2011.

[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] C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 213–222.

[9] F. Seb´e, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," *IEEE Trans. on Knowl. and Data Eng.*,vol. 20, no. 8, pp. 1034–1038, 2008.

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