International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Data Integrity Proof in Cloud Computing using TPA with Privacy Preserving

Rushikesh P. Dhanokar

Computer Engineering, J.S.P.M, Tathawade, Pune, India, 411033

Abstract: ***Cloud Storage such as Google, Yahoo, and Amazon EC2 are cheap and sensible option for both home and professional users for storing their large amount of data remotely on Cloud Server. Customers often use these services to store valuable data such as email, family photos and videos, and disk backups. Cloud Computing is the brood architecture of computing. Cloud storage provide fringe benefit like on-demand network access to a shared pool of configurable computing resources, user do not need to maintain data as the data is maintained by the cloud service provider, user is billed based on usage, though the data is stored remotely it can be handle from anywhere by internet. These advantages make cloud storage very widespread but still it is suffering from internal and outside threats. The data loss incidents may take place because of network and software bugs, the CSP (Cloud Service Provider) may fur the data loss incidents for maintain reputation. To check the integrity of outsourced data in this paper allow a third- party auditor with privacy preserving. TPA also to perform multiple data files by batch auditing. User data privacy should maintain by the third party auditing process.***

Keywords: Data Integrity, Third Party Audit, Privacy preserving, Batch auditing, CSP.

1. Introduction

Cloud storage such as Amazon, Yahoo, Google, Microsoft, and Mozy.com allows clients to store their data on remote storage. The data is stored remotely on remote storage and it can be accessed through the internet connection between client's machine and remote database on cloud. Storing data on cloud gives clients number of advantages like client don't have to maintain the data as it is maintained by the cloud service provider, pay only that they used, client can access his data from anywhere with the help of internet and he do not need to carry the physical data storage devices, enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources. Though these advantages make cloud storage a very economical option for storing data it has some drawbacks like the data loss incidents may take placed. Lots of internal and external threats are there and the data storage of client may be kept hidden from client to maintain reputation, there may be bugs in the network path or in the software. [1]. Cloud computing concerns some security like Confidentiality: unauthorised person cannot get any stored information. Integrity: Ensuring that information held in a system is a proper representation of the information intended and that it has not been modified by an unauthorised person. Availability: Ensuring that information processing resources are not made unavailable by malicious action. Non-Repudiation: Ensuring that agreements made electronically can be proven to have been made.

As clients have limited capacity and they may do only uploading and downloading data from cloud storage. User downloads all data in order to check integrity of stored data it is very costly and tedious task. In the proposed system a Third Party Auditor (TPA) is introduced who will verify the data integrity of the client's data stored on cloud storage. TPA audit data when user needed. TPA has more potential than user and beneficial for cloud provider like audit result from TPA gives more values for Cloud base service platform and also they fulfill the cloud computing concerns. [2, 3]

Third party auditor (TPA), so met a) TPA audit cloud data storage without the local or original copy of data, and should not put any additional on-line burden to the cloud user; b) The third party auditing process should preserve user data privacy. To handle this problem in this paper we use homomorphic linear authentication (HLA) .By integrating HLA with random masking our protocol guarantees that third party auditor could not learn anything about data content stored in cloud server during auditing processes.

2. Related Work

Ateniese et al. [2] in Public audit ability was first considered their model for provable data possession [PDP] for ensuring the storage correctness of the data files on the servers. He allows a client that has stored data at an untrusted server to verify that the server possesses the original data without retrieving it. They utilize the RSA-based homomorphic linear authenticators for auditing outsourced data and suggest randomly sampling a few blocks of the file. One of the schemes which they had proposed is based on public audit ability. It generates proof for possession by randomly sampling the blocks of data files, but this way the linear combination of the blocks may reveal the data to the third party auditor. So their protocol was not fully privacy preserving. Juels et al. [3], The simplest Proof of retrivability (POR) scheme can be made using a keyed hash function $h_k(F)$. In this scheme the verifier, pre-computes the cryptographic hash of F using $h_k(F)$ before archiving the data file F in the cloud storage, and stores this hash as well as the secret key K. To check if the integrity of the file F is lost the verifier releases the secret key K to the cloud archive and asks it to compute and return the value of $h_k(F)$. By storing multiple hash values for different keys the verifier can check for the integrity of the file F for multiple times, each one being an independent proof. POR model by possession and

retrievability of remote data files on archive servers are ensured by using spot-checking and error-correcting codes. Their scheme does not support public auditability and the user can perform fixed number of audit challenges.

Shacham and Waters [5] system to improve a proof-of-retrievability in compact proof of retrievability should be possible to mine the client's data from any cloud storage that passes a verification check. Give the first proof -of-retrievability schemes with full proofs of security against random in the strongest model, that of Juels and Kaliski [3]. The client's query and server's response are both extremely short by using BLS signatures. This scheme allows public verifiability not only just the file owner but anyone can act as a verifier. PRFs is secure in the standard model, allows only private verification. Schemes rely on homomorphic properties to aggregate a proof into one small authenticator value. Homomorphic linear authenticators are used built from secure BLS signatures which are publicly verifiable. Their approach is not privacy preserving due to the linear combination of the blocks and this equation can be solved by third party auditor.

Online storage integrity Shah et al. [6] [7] introduces a third party auditor. They are encrypting the whole data and taking keys to hide it from auditor moreover both this symmetric-keyed hashes and encrypted data are store on cloud and auditor. Auditor checked the integrity of data file. This scheme requires the auditor to maintain state of every key, works on encrypted files and when all the keyed hashes are used is affected by online burden on users.

Further Wang et al. [8] additional feature partial dynamic data storage and combine BLS-based HLA with MHT for supporting full data dynamics. In this scheme symmetric key cryptography is used but with limitation on number of audits. The protocols discussed above are not privacy preserving as both require the linear combination of sampled blocks as input

3. Proposed System

There are mainly three different models involved that are cloud user, cloud service provider or cloud server and third party auditor. Cloud users contain data that has to be stored in the cloud by registering to particular cloud storage. Cloud server and cloud service provider are same they to take data from users and it is hub of storage space and computational resources. Third party auditor is auditor on behalf of user will do the verification of the data integrity of the data stored on cloud storage. User relay on cloud service provider to store data and to ensure that data are being correctly stored and maintained. User's data may be damage by both internal and external attack at cloud server [2]. So to proof of integrity of storage data on cloud and minimize the overload of user we use a third party auditor. Public audit ability, storage correctness, privacy preserving, batch auditing are the design goals to be achieved.

Client, TPA and cloud server task:

Client:
- The clients generate signature and metadata by using AES and HMAC algorithm.
- The Client will login to system through the client login id and password.
- The client then will be allowed to upload the file. The file is split into four parts.
- HMAC option is used when client wants to upload the file by generating metadata. The metadata would hash code and encrypted lengths of the parts of the file.

TPA:
- The TPA will login to the system through the TPA login.
- The TPA will then select the Client for whom the verification needs to be carried out.
- Then the files will be selected for the verification purpose.
- The TPA verify using AES or verify using HMAC.TPA verify data using Verification of proof.

Cloud Server:
- Get and stored data from registered user
- Give the verification of proof to TPA.
- Update and maintain the user's data.

The paid cloud like Amazon or Google cloud is designed to communicate with the cloud storage server by web. The Client will login to the web if he is register to use the cloud storage .If not then he can register to use the storage from the web. After successful login the client will be able to see all the file details and will be able to download the file if required.
Fig 1 shows the basic architecture of proposed system.

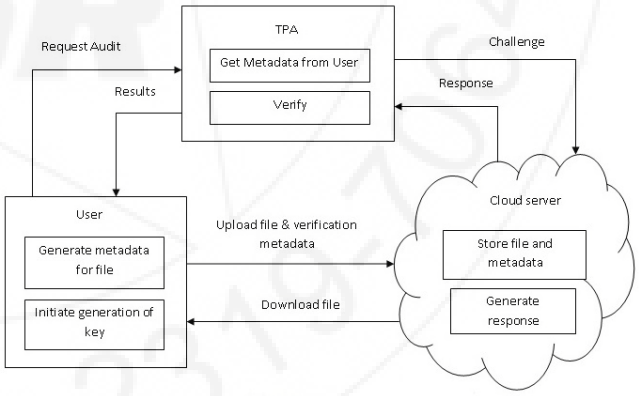

Figure 1: System Architecture

The system consists of four algorithms.
1) Key Generation(k_gen): is a key generation algorithm it initiated by the user.
2) Metadata Generation(m_gen): The user will generate verification metadata.
3) Generation of Proof(p_gen): is used to generate proof of data storage correctness by server when it challenge.
4) Verification of Proof(v_prf): is run by third party auditor to audit the proof.

Proof of integrity setup into two steps:

a) SETUP: -In set up phase user initializes public and secret parameters of the system by executing k_gen algorithm and preprocess the data file F by using m_gen algorithm to generate the verification metadata. By deleting it's local copy user will upload the data file on cloud server.
b) AUDIT- In audit phase TPA send audit message or challenge to the cloud server to checking the stored data integrity. Random masking Homomorphic linear authenticator technique is used. Cloud server give the proof by recalled the data file F as it is and TPA will then verify the proof. TPA cannot originate the user's data content due to lack of all the necessary information to build up a correct group of linear equations so we preserve privacy.

Properties of proposed system:

- Data storage validation: TPA verified storage correctness without the access to original copy of data.
- Auditing Support: The cloud server pass the verification not keeping the data storage intact.
- Efficient: The communication and computation costs are not very high for auditing task.
- Preserve Privacy property: Original data is not shared with the TPA, metadata with encrypted form is shared. So it difficult for TPA to gain any knowledge about data.
- Batch audit: TPA has capacity to verify multiple files at a time.

4. Conclusion

The paper takes place privacy preserving with public auditing data integrity in the cloud storage system. Secure cloud storage is proposed and implemented using two algorithms AES and HMAC which verifies the data integrity. Our system not only reduces the load on client but also reduce fear of their outsourced data leakage. In this system TPA cannot audit only one client at a time but also many clients simultaneously by using batch auditing scenario. The system is totally secure and highly efficient. The algorithm is partially homomorphic encryption so using fully homomorphic encryption can be a future enhancement. Also a full fledge deployment of the application on public like handle large amount of data cloud can be an important future enhancement.

5. Acknowledgment

I would like to thank my guide Prof. G. S. Mate for her help and guidance throughout this project and the semester, without them this would not have been possible. I would also like to give special thanks to Dr. P. K. Deshmukh for his constant input to my project.

References

[1] CloudSecurityAlliance,"TopThreatstoCloudComputing,"http://www.cloudsecurityalliance.org, 2010
[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z.Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc.14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 598-609, 2007.
[3] A. Juels and J. Burton, S. Kaliski, "PORs: Proofs of Retrievability for Large Files," Proc. ACM Conf. Computer and Comm. Security (CCS'07), pp. 584-597, Oct. 2007.008.
[4] M.A. Shah, R. Swaminathan, and M. Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," Cryptology ePrint Archive, Report 2008/186, 2
[5] H. Shacham and B. Waters, "Compact Proofs of Retrievability," Proc. Int'l Conf. Theory and Application of Cryptology and Information Security: Advances), in Cryptology (Asiacrypt vol. 5350, pp. 90-107, Dec. 2008.
[6] M.A. Shah, M. Baker, J.C. Mogul, and R. Swaminathan, "Auditing to Keep Online Storage Services Honest," Proc. 11th USENIXWorkshop Hot Topics in Operating Systems (HotOS '07), pp. 1-6, 2007.
[7] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing," IEEE Trans. Parallel and Distributed Systems, vol. 22,no. 5, pp. 847-859, May 2011.
[8] C. Wang, Q. Wang, K. Ren, and W. Lou, "Towards Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Service Computing, vol. 5, no. 2, 220-232, Apr.-June 2012.
[9] F. Sebe, J. Domingo-Ferrer, A. Martı´nez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient Remote Data Possession Checking in Critical Information Infrastructures," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 8, pp. 1034-1038, Aug. 2008.

Author Profile

Rushikesh P. Dhanokar, have completed Bachelor of Engineering in Computer Science and Engineering from SCOE, Sudumbre, Pune in 2012. Currently, He is pursuing Masters of Engineering in Computer Science and Engineering from J.S.P.M`s Rajarshi Shahu College Of Engineering, Tathawade, Pune. His research interests take in data security in cloud computing.