International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Secure Public Auditing for Cloud Data Storage

Madhumati B. Shinde[1], Dr. Sonkamble Sulochana B[2]

[1]P.G. Student, Computer Department, JSPM, NTC, RSSOER, Narhe, Savitribai Phule Pune University, Pune, India

[2]Professor, Head of Computer Department, JSPM, NTC, RSSOER, Narhe, Savitribai Phule Pune University Pune, India

Abstract: ***Cloud computing has become a very popular buzzword. Today most of the business depends on the cloud, it realized as a pillar for IT industry.The most obvious advantage from the use of cloud computing systems and technologies is cloud computing is the most cost efficient method to use, maintain and upgrade. Due to this increasing economical condition with minimum maintenance & operational costs related to IT software and infrastructure. By using cloud storage, users store their data at remote location, data can delivered to the user on-demand for high-quality applications and services. Cloud storage is nothing but large shared configurable computing resources, that can minimize the burden of local data storage and maintenance on the user.The cloud has changed the way application software and databases are stored has. Now days they are stored in cloud data centers by using storage servers. The major concern in cloud data storage is the security of the data which is stored on cloud. The new phenomenon which is used to store and manage data without capital investment has brought many security challenges which are not thoroughly understood. This paper focuses on the security and integrity of data stored in cloud data servers. The data integrity verification is done by using a third party auditor who is authorized to check integrity of data periodically on behalf of client.***

Keywords: Data storage, privacy preserving, public auditability, cloud computing, cloud service provider (CSP),TPA(Third Party Auditor),CSS(Cloud Storage Server**).**

1. Introduction

Cloud computing is one of the most modern technical topics today, it has broad-ranging effects across IT, Information Architecture, Business, Software Engineering, and Data Storage. Cloud Computing is an pioneering technology that is revolutionizing the way we do computing. Cloud Computing is an internet based technology that uses distant servers to preserve data and applications. Cloud computing allows customers/clients and businesses to use different applications without installation and access their personal files at any location with internet access. This technology allows for much more efficient computing by centralizing data storage, processing and bandwidth. This Cloud computing technology generally has three segments: "Application", "Data Storage Space" and "Connectivity." Each segment deserves a different purpose and offers different products for businesses and individuals around the world. These segments defined as cloud computing stack which is referred as Software as a Service, Platform as a Service and Infrastructure as a Service. SaaS model is used to deliver software applications above the web through that clients can use the software application. PaaS is the model which delivers hardware and the set of tools and services designed to enable coding and deploying those applications quickly and efficiently. IaaS provides various infrastructure components like servers, storage, networks, and operating systems required for the application.

Cloud computing has become a very popular buzzword. Today most of the business depends on the cloud, it realized as a pillar for IT industry due their essential characteristics such as on-demand self-service provisioning, broad network access, rapid elasticity i.e scale in and scale out capabilities. Among the various services provided by the cloud is cloud data storage. This allows the client(Data Owner) to move the content on remote data store hence relive the burden of local data storage on the client. Although this new data storage paradigm envisioned as a promising service platform for the Internet , it brings about many challenging design issues which have intense influence on the security and performance of the overall system. Data outsourcing in cloud has two important issues security & integrity. Security is related to preserve the privacy of the data i.e. avoid unauthorized data access. While integrity maintains the data correctness. Our project has mainly focus on these two issues related to cloud data. The idea of public auditability used to check the data integrity.

Since cloud service providers (CSP) are separate administrative entities, data outsourcing is actually relinquishing user's eventual control over the destiny of their data. As a result, the accuracy of the data in the cloud is being put at hazard due to the following reasons. Firstly, although cloud infrastructures is much more powerful and consistent than personal computing devices, still they are facing the broad range of both internal and external threats for data integrity. Secondly, there is possibility that due to various motivations for CSP to behave unfaithfully toward the cloud users regarding their outsourced data status . For examples, CSP may rescue storage for economic reasons by discarding data that have not been or are rarely accessed, or even hide data loss incidents to maintain a reputation. Due to these reasons data owners would worry that the data could be lost in the cloud. Thus, enabling public auditability for cloud storage is of vital importance so that users can resort to a third-party auditor (TPA) to check the integrity of outsourced data and be worry free.

2. Objectives of the Propose Work

There are following objectives & motivations of the proposed work

1) **Public verification for storage correctness assurance:**
To There are following objectives & motivations of the proposed work allow anyone to perform auditing, not only for the clients that have originally stored the file on cloud servers. Public auditing capability allowed on demand for verification of the correctness of stored in cloud.

2) **Dynamic data operation support:**
Clients can perform dynamic block level operations on the data files such as insert, delete, update, while maintaining correctness of data files in the cloud .The design of system should be as effective as possible for assure the consistent integration of public verifiability and dynamic data operation support.

3) **Blockless verification:**
The challenged file blocks for data verification should not be retrieved by the verifier (e.g., TPA) during verification process for both efficiency and security concerns.

4) **Stateless verification:**
Verifier should not maintain the state information during the auditing process. Stateless verification can remove the need for state information maintenance at the verifier side between audits throughout the long term of data storage.

5) **Batch Auditing:**
Auditing can be carry out for the batch of users i.e Multi-User auditing can be supported by TPA in cloud environment.

3. Literature Review

Literature review is the most essential step in software development process. Following is the literature review of some existing technique for privacy preserving public auditing in the cloud.

In 2007 Ateniese et al. [4] firstly considered public auditability in their "provable data possession"(PDP) model.This model was designed for ensuring possession of data files on untrusted storage servers. The PDP model have used the RSA-based homomorphic linear authenticators for auditing outsourced data and advise randomly sampling a few blocks of the file. But this model has some limitations such as, the scheme that consider public auditability exposes the linear combination of sampled blocks to external auditor. The designed protocol was not provably privacy preserving ,and thus it leak user data information to the external auditor.

In 2007 Juels and Kaliski [5] defined another similar model called as Proofs of Retrievability (POR).This model used the technique of spot-checking and error-correcting codes to ensure both "possession" and "retrievability" of data files on remote archive service systems. The original file is added with a set of randomly-valued check blocks called sentinels. The verifier challenges the untrusted server by specifying the positions of a collection of sentinels and asking the untrusted server to return the associated sentinel values. But in their model , the number of audit challenges a user can perform is fixed a priori, and it is not support for public auditability is not supported in their main scheme.

In 2008 Ateniese et al [6] presented efficient PDP model to support for dynamic data that is update & delete operations on data but insert operation was not present in their mechanism. The author utilized symmetric keys, these symmetric keys to verify the integrity of data, is not public verifiable and only provides a user with a limited number of verification requests. Shah et al.[7][8] introduce the concept of TPA auditing for keeping online storage truthful .For that purpose they first encrypting the data then number of pre-computed symmetric-keyed hashes was generated over the encrypted data for storing at the the auditor. But this scheme only works for encrypted files, the auditor maintain state, which potentially brings in online burden to users.

Wang et al. [10] utilized Merkle Hash Tree and BLS signatures to support dynamic data in a public auditing mechanism. The authors proposed a dynamic auditing protocol that can support the dynamic operations of the data on the cloud servers, but this method may leak the data content to the auditor because it requires the server to send the linear combinations of data blocks to the auditor.

In 2013 C. Wang et al.[12] proposed a privacy-preserving method to carry out public auditing on the cloud information . In case of cloud computing, it is not sufficient to adopt the traditional cryptographic measures to achieve security. The reason is due to data outsourcing and the ubiquitous nature of the data. So, in this paper they opt the concept of Third Party Auditing (TPA). Homomorphic authenticator and random masking ensures that TPA could not gain any knowledge during the process of auditing. Thus, TPA is trusted and capable of accessing the cloud storage to perform auditing.

4. Existing System

In the existing system, outsourcing the data means user in fact relinquish vital control over the fortune of their data & it is in hand of CSP. The conventional cryptographic technologies used for data integrity and accessibility, cannot work properly on the outsourced data. It is not a useful solution for data justification by downloading them due to the expensive communications, especially for large size files. For securely establish an efficient third party auditor (TPA), there are following two fundamental requirements have to be met:
1) TPA should be capable to efficiently check(audit)the cloud data storage without demanding the local replica of data, and it will not put an additional on-line burden to the cloud user.
2) The third party auditing process should not bring any kind of new vulnerabilities towards user data privacy.

In the existing system, the data correctness in the cloud is being put at risk due to the subsequent reasons. Although we think that the infrastructures inside the cloud are much more dominant and trustworthy than personal computing devices, they are facing broad range of both internal (loss or destruction of data) and external (disclosure of data to unofficial users) threats for data integrity.

Drawback of the existing system

1. Cloud Storage system provides the user for safe and consistent place to save important data and documents. However, in some cases user's files are not encrypted before store on some open source cloud storage systems. i.e. TPA

demands retrieval of user data, here actual privacy is not preserved.

2. The storage service provider i.e storage server can effortlessly access the user's files. This brings a big anxiety about user's privacy. The user has no ultimate control over the software applications including secret data. User has to totally depend on the providers for maintenance and administration.

5. Proposed System

This section presents the architecture of proposed system that overcomes the drawbacks in the existing system. Proposed system can achieve the public auditing that allow the TPA to verify accuracy of data without retrieving local replica of the data. It also conserve the privacy by applying encryption scheme. The auditing scheme also support batch auditing that allow the TPA to use the audit for multiple clients. To support proficient handling of several auditing tasks, this work further explore the technique of bilinear aggregate signature to extend the main result into a multiuser setting, where TPA can perform multiple auditing tasks simultaneously. Wide security and performance examination show that the projected scheme is highly efficient and provably secure.

The following figure shows the system architecture of proposed system. The cloud storage consist of three different entities Cloud Client, Cloud Storage Server & TPA(Third Party Auditor).

- Cloud Client: This entity has huge amount of data, that will store on cloud storage server.
- TPA (Third Party Auditor):-This entity can work on the behalf of the client, which is responsible for checking integrity of the outsourced data.
- Cloud Storage Server:-Cloud Storage Server is the location where, user store their data.

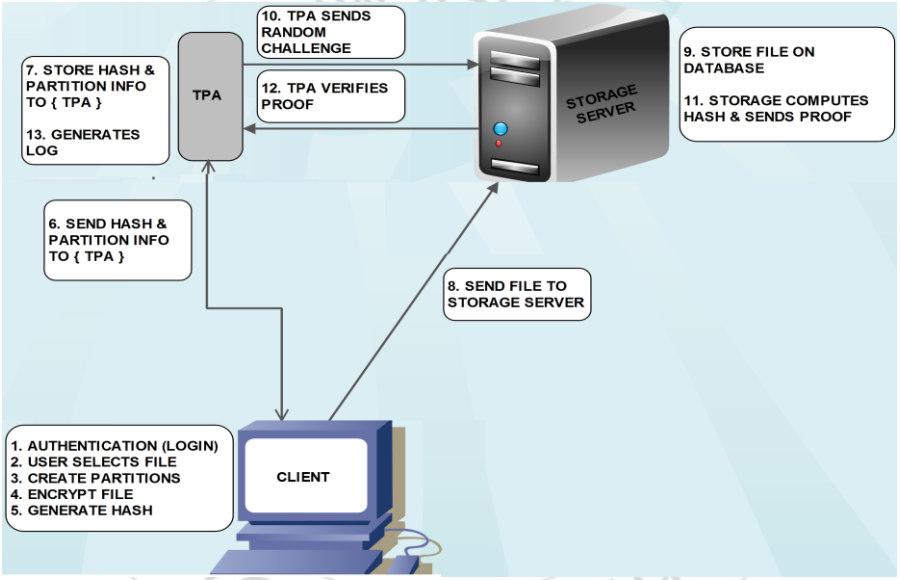

Figure 1: System architecture of cloud storage system

5.1 Methodology of Proposed System

The public auditing scheme consists of four algorithms (KeyGen, SigGen, GenProof, and VerifyProof). KeyGen is a key generation algorithm that is run by the user to setup the scheme. SigGen is used by the user to generate verification metadata, which may consist of digital signatures. GenProof is run by the cloud server to generate a proof of data storage correctness, while VerifyProof is run by the TPA to audit the proof. Running a public auditing system consists of two phases, Setup and Audit:

1. Setup Phase

The user initializes the public and secret parameters of the system by executing KeyGen, and preprocesses the data file F by using SigGen to generate the verification metadata. The user then stores the data file F and the verification metadata at the cloud server. The user may alter the data file F by performing updates on the stored data in cloud.

2. Audit Phase

The TPA issues an audit message or challenge to the cloud server to make sure that the cloud server has retained the data file F properly at the time of the audit. The cloud server will create a response message by executing GenProof using F and its verification metadata as inputs. The TPA then verifies the response by cloud server via VerifyProof.

5.2 Mathematical Model of Proposed work

Let S be the main set shown below.

S={U,F ,P, Ep, H, SDB ,TPADB , C, R,L}

Where,

U: is cloud users they want store there data on cloud.

F : is the file that user want upload on the cloud server.

P : is the number of partitions created of the file .

Ep : is the encryption algorithm.

H : is the hash values of the encrypted partitions of the file.

SDB : is the copy of the server database.

TPADB: is the TPA database that consist of hash information as well as user who has generated the file partitions.
C : is the set of challenges send by the TPA to the server.
R : is the result generated by the cloud storage server to generate the proof of the integrity.
L : is the log record maintained at the TPA.

Functionalities:
The funcionalities of the existing system working as follows.
1.U=Register(uid,Password,Uname,Address, ContactNo,EmailID)
U=UserLogin(UID,Passward)
2. F= Selectfile (FileName)
3. P= Partitionfile (F, Psize)
4. Ep= EncryptPartitions (P, Key)
5. H = HashCalculation(Ep)
6. TPADB= StoreHashInfo (H)
7. SDB =StorePartitions (Ep ,uid)
8. C=RandomChallenges(TPADB)
9. R=ComputeResult(SDB, Ep)
10. L=AddLogEntry(Pname, PDate, State)

5.3Algorithms of proposed system

Algorithm 1: **Key Generation(KeyGen)**
Key Generation is a key generation algorithm running at the data owner side, which generates the secret key PKi upon getting input of some security parameter using AES algorithm.
Input: File (F).
Output: Private Key (PKi) [key is used for encryption and decryption mechanism..
1. Result = ValidateFormat(F);
validateFormat = { png ,.pdf .jpg , ...}

2. If the format is valid then client want to secure the data using AES Algorithm.
CipherData(Fi) = AES(Fi);

Algorithm 2: Sign Generation(SigGen)
Sign Generation algorithm used to generate digital signature of the encrypted file partitions.
Input: File partitions(P1,P2,P3...Pn)
Output:Hash values of encrypted partitions(H1,H2...Hn)

1. MessageDigest md = MessageDigest.getInstance("SHA");
MessageDigest is generic class used to provide the default functionality of the SHA-1 algorithm for the application.The getInstance() method generates the MessageDigest Object.

2. md.update(datArr, 0, datArr.length);
Updates method takes the input _le by appending a byte array at the end.

3. byte[] sha1hash = md.digest();
This method applied SHA-1 algorithm on current input message and returns the array of bytes.

4.Separate the first 4 bits of the particular byte i.e MSB(Most Significant Bits)and convert then to character.

5. Separate the last 4 bits of the particular byte i.e LSB(Least Significant Bits)and convert then to character.

Algorithm 3: VerifyProof
Verify proof algorithm run by the TPA for performing the auditing task.
Input: Random File blocks
Output:Verification Status
1. File Indices = getRandom();
This method returns the random file indices and store in the vector array.

2. Result=updateLog()
The updateLog method update the result of the verified partitions maintain the file status like file is safe or file is unsafe.

Algorithm 4:Generation proof(GenProof)
This algorithm run by the cloud storage server for the possession of the data files.
Input:File
Output:Hash of the files

1. Result=GenProof(File, Chal)

This method accept the challenge from TPA and send the proof of possession of the data files to the TPA.

6. Result

Verified Blocks (bytes)	TPA Individual Audit time(ms)	TPA Batch Audit Time(ms)
3072	1014	921
3891	1154	1014
4096	1155	999
4505	1532	1045
5734	1632	1357

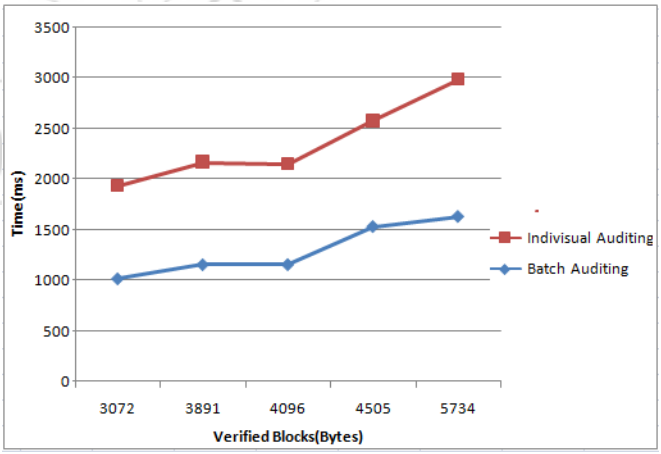

Figure 3: Comparison on auditing time between batch and individual auditing.

The Fig.3 shows the results for the auditing task. The number of auditing task & time required for auditing shown in the graph. The performance of the corresponding individual auditing is provided as a baseline for the measurement. Following the same settings for the data blocks of the file. The average per task auditing time, which is computed by dividing total auditing time by the number of tasks, is given

for both batch and individual auditing. It can be shown that compared to individual auditing, batch auditing indeed helps reducing the TPA's computation cost, as more than 15 percent of per task auditing time is saved.

7. Conclusion

Use of third-party auditing service provides a cost-effective method for users to gain trust in cloud. We assume the TPA, who is in the business of auditing, is reliable and independent entity. In this paper, public auditing system is proposed for data storage security in cloud computing along with preserving privacy mechanism. It can utilizes the homomorphism linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task, but also alleviates the users fear of their outsourced data leakage. Public auditing is able to allow clients for delegating the tasks of integrity verification to TPA while they are independently not reliable or cannot commit required resources of computation performing verifications in a continuous manner. One more important concern is the procedure for construction of verification protocols which can be able to accommodate data files that are dynamic.

For achieving data dynamics that are effective, the existing proof of storage models is enhanced through manipulation of the construction of classic SHA-I for authentication of block tag. For supporting good handling of multiple numbers of auditing tasks, the method of bilinear aggregate signature is further explored for extending the main result into a multiuser setting, where TPA is able to perform multiple auditing tasks in a simultaneous manner. Huge security as well as performance analysis proves that the proposed scheme is efficient and secure to a greater extent. As the future work, efficiently audit the integrity of shared data with dynamic groups while still preserving the identity of the signer on each block from the third party auditor and also include the features to enable dynamic operations (e.g. inserting/deleting data block) in this system.

8. Acknowledgement

I sincerely thanks to prof. S. B. Sonkamble for her continuous and constructive support for the work. I would also like to thanks to my friends for their valuable comments.

References

[1] P. Mell and T. Grance, "The NIST Definition of Cloud Comput- ing," technical report, Nat'l Inst. of Standards and Technology, 2009. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010K.
[3] Cloud Security Alliance, "Top Threats to Cloud Computing,"http://www.cloudsecurityalliance.org, 2010.
[4] 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.
[5] 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.
[6] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," Proc. Int'l Conf. Security and Privacy in Comm. Networks (SecureComm '08), pp. 1-10, 2008.
[7] M.A. Shah, R. Swaminathan, and M. Baker, Privacy-Preserving Audit and Extraction of Digital Contents, Cryptology ePrint Archive, Report 2008/186, 2008.
[8] M.A. Shah, M. Baker, J.C. Mogul, and R. Swaminathan, Auditing to Keep Online Storage Services Honest, Proc. 11th USENIX Workshop Hot Topics in Operating Systems (HotOS 07), pp. 1-6, 2007
[9] 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.
[10] 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. 847859, May 2011.
[11] D. Boneh, B. Lynn, and H. Shacham, "Short Signatures from the Weil Pairing," Proc. Seventh Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT'01), pp. 514-532, 2001
[12] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Storage Security in Cloud Computing," IEEE Transactions On Computers, vol. 62, no. 2, FEB 2013.
[13] Wang C, Wang Q et al. Privacy-Preserving Public Auditing for Storage Security in Cloud Computing, Proceedings IEEE INFOCOM'10.2010
[14] Wang B, Li B et al. Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud, IEEE Fifth International Conference on Cloud Computing, 295–302.2012
[15] 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.
[16] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession," Proc. ACM Conf. Computer and Comm. Security (CCS '09), pp. 213-222, 2009.

Author Profile

Madhumati Shinde received the B.E in computer Engineering from ADCET,Ashta, Maharashtra . She is now persuing M.E.in computer Engineering at RSSOER, Pune, Maharashtra. Her area of interest is Cloud Computing.

Mrs. Sulochana Sonkamble is HOD of Computer Science Department at Rajarshi Shahu College of Engineering and Research, JSPM NTC Pune-India. She obtained Bachelor of Engineering, Masters of Engg. And PhD in Computer Science Shri Guru Gobind Singhji Institute of Engineering and Technology, Swami Ramand Tirth Marathwada University, Vishnupuri, Nanded-India.