Volume 7, Issue 3, March 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Study on Auditing Approaches for Preserving Data Integrity in Cloud Storage

Anne Srijanya .K
Assistant Professor, Department of CSE, CMR Engineering College,
JNTUH, Telangana, India

Abstract*— Cloud computing provides cloud storage a service in which data is maintained, managed and backed up remotely through a centralized large data centres and made available to user over a network. Management of data in a remote data centre may not be trustworthy. For this purpose many approaches were introduced for verifying the integrity of the user`s data on the cloud storage. This paper makes comparative analysis over different approaches and choose the best approach for checking the integrity of the data.

Keywords— Cloud Computing, Privacy, Data Integrity, Auditing, TPA

I. INTRODUCTION

According to the NIST, Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics On-demand self-service , Broad network access, Resource pooling, Rapid elasticity, Measured service[1]

A. Cloud Service Models

There are actually more service models than the three (IaaS, PaaS, SaaS) widely in use today. Service models like Data Analytics as a Service and HPC/Grid as a Service are emerging as useful models. How one Selects the appropriate service model depends on factors such as availability of suitable application software Need for development and test environment, need for effective computing infrastructure control and management required distribution of data, services, and infrastructure, existence and complexity of enterprise IT infrastructure and data centre/warehouse.[2]

B. Deployment Models:

1) Private cloud: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.[1]

2) Community cloud: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.[1]

3) Public cloud: The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.[1]

4) Hybrid cloud: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds). [1]

II. CLOUD AUDIT

Cloud Audit is a specification for the presentation of information about how a cloud computing service provider addresses frameworks. The goal of Cloud Audit is to provide cloud service providers with a way to make their performance and security data readily available for potential customers. The specification provides a standard way to present and share detailed, automated statistics about performance and security.[3]

A. Need for Auditing

As **cloud computing** get increasingly complex and finds use in core enterprise applications, it is time to pay more attention to auditing. Auditing ensures that your cloud installation works per our expectations. The auditing could be done either internally by your IT or business teams, or could be done by a third party service. Regardless of who does the audit, it is important to understand the different considerations in cloud auditing.

1) Security Audit: Security should be one of the most essential aspects of any enterprise IT system. Security audit must uncover the various vulnerabilities in your cloud solution. Some of the security issues include unauthorized access, intentionally destroying data and Denial of Service (DoS). The audit should make sure the setup is sufficiently protected against the common type of attacks and has the adequate level of security that satisfies the enterprise requirements. Sufficient attention must be paid to data security issues to protect against any information leakage.

2) Performance and Reliability Audit: One of the biggest considerations to move to the cloud involves around reliability. Reliability audit must make sure that your data is available to the employees and customers 24/7. The cost of downtimes can be very high, in terms of lost employee productivity and loss of goodwill from the customers. The audit should also spell out the SLA requirements and find out if all the providers satisfy those requirements.Performance audits must identify the various metrics (time to save a document, loading time of the website landing page etc.) and verify if the cloud setup satisfies those metrics. The performance and reliability audits could also make use of stress tests to make sure the stack used is robust under severe load conditions.

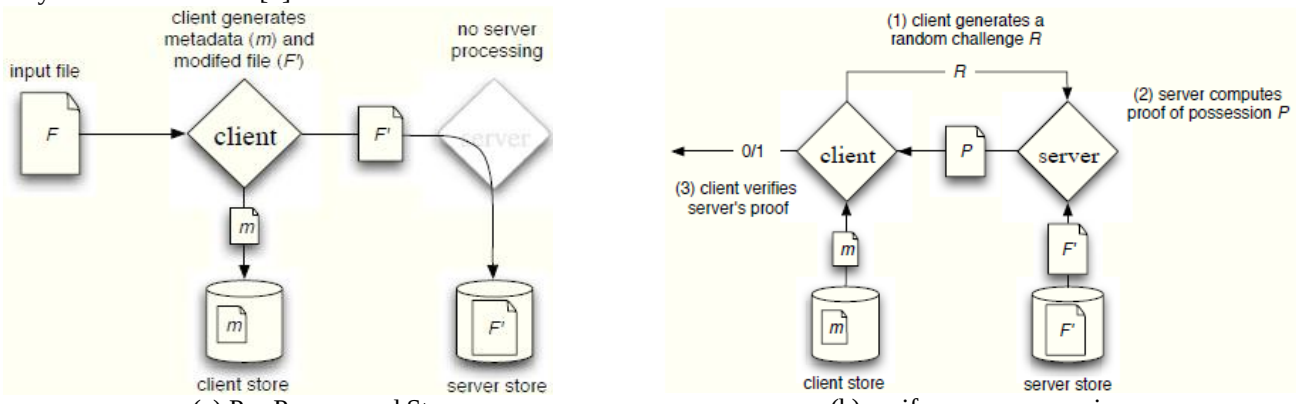
III. RELATED WORK

In the recent years , cloud storage auditing has attracted increasing attention. Following are most frequently used auditing approaches.

- Provable Data Possession (PDP) Model
- Proof of Retrievability (POR) Model
- Trusted Third Party Auditing (TPA)

A. Provable Data Possession (PDP) Model

G. Ateniese, R.Burns, R.Urtmola, J.Herring, L.Kissner, Z.Peterson and D.Song introducing provable data possession (PDP) that allows client to stored data at an un trusted server to verify that server possesses the original data without retrieving it.PDP generates probabilistic proofs of possession by sampling random sets of blocks from the server. It reduces I/O costs. The client have a constant amount of metadata to verify the proof .The challenge/response protocol minimizes network communication. It transmits a small and constant amount of data.PDP supports to public databases such as digital libraries, astronomy/medical/legal repositories, archives etc. .PDP schemes have drawback is that it works only for static databases[4].



(a) Pre-Process and Store (b) verify source possession

Fig: 1 Protocol for Provable Data Possession [4]

G.Ateniese, R.D.Pietro, L.V.Mancini and G.Tsudik worked on an efficient PDP mechanism based on symmetric keys. It support update and delete operations on data but insert operations are not available in it. It exploits symmetric keys to verify the integrity of data ,it is not public verifiable. It having drawback ,it provides a user with a limited number of verification requests[5]

Dynamic provable data possession (DPDP), which extends the PDP model to support provable updates on stored data developed by C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia. Consider a file F consisting of n blocks, it define an update as inserting a new block or modifying an existing block or deleting any block. An update operation describes the most general form of modifications a client may wish to perform on file. DPDP solution is based on variant of authenticated dictionaries, where rank information used to organize dictionary entries. It supports efficient authenticated operations on files at block level such as authenticated insert and delete. Provable storage system enables efficient proofs of a whole file system, enabling verification at different users and same time not having to download the whole data[6].

B. Proof of Retrivabality (POR) Model

A.Juels and B.S.Kaliski describes POR which allows a server to convince a client that can be retrieve a file that was previously stored at the sever.POR scheme uses disguised blocks (called sentinels) hidden among regular file blocks in order to detect data modification by the server. The goal of POR is to accomplish these checks without users having to download the files themselves.POR provides quality of service guarantees means a file is retrievable within a certain time bound. POR protocol encrypts F and randomly embeds a set of randomly valued check blocks called sentinels .The use of encryption renders the sentinels indistinguishable from other file blocks. The verifier challenges the prover by

specifying the positions of a collection of sentinels and asking the prover to return the associated sentinel values. If prover has modified or deleted a substantial portion of F, then with high probability it will also have suppressed a number of sentinels[7].

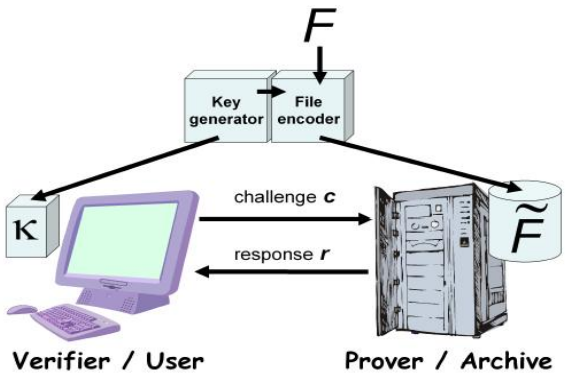


Fig.2. Schematic of a POR system[7]

HovavShacham and Brent Waters focuses on a proof of retrivability system, in that a data storage centre convinces verifier that he is actually storing all of a client's data. The central challenge is to build systems that are both efficient ad provably secure means that it should be possible to extract the client's data from any prover that passes a verification check. there are two scheme. First scheme is built from BLS signatures and secure in the random oracle model. It has the shortest query and response of any proof of retrievability with public verifiability. Second scheme ,which builds on pseudorandom functions(PRFs)and is secure in the standard model. It has shortest response of any proof of retrievability scheme with private verifiability. Both scheme depends on homomorphic properties to aggregate a proof into one small authenticator value[8].

C. Third Party Auditing (TPA)

C. Wang, Q. Wang, K. Ren, and W. Lou first presented Privacy preserving public auditing system for data storage security in cloud computing ,where TPA can perform the storage auditing without demanding the local copy of data. Homomorphic authenticator and random masking technique are used to guarantee that TPA would learn any knowledge about the data content stored on the cloud server during the efficient auditing process. It not only eliminates the burden of cloud user from auditing but also soften the user's fear of their outsourced data leakage. Consider TPA may concurrently handle multiple audit sessions from different users for their outsourced data file, it can extends privacy preserving public auditing protocol into multiuser setting ,where TPA can perform the multiple auditing tasks in a batch manner i.e. simultaneously[9].

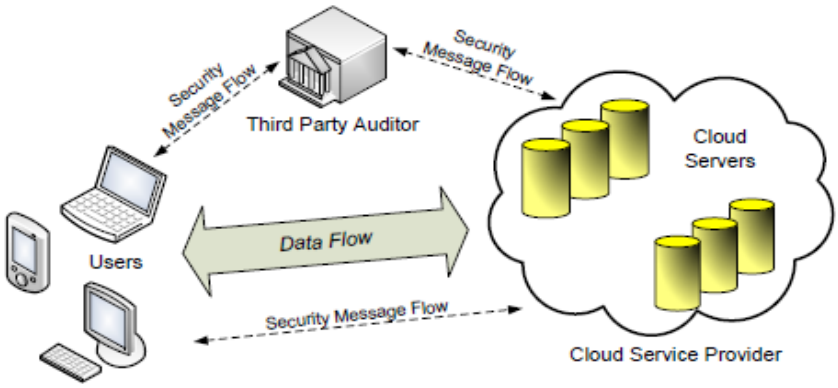


Fig. 3 Cloud data storage architecture [6]

Later on many TPA mechanisms were presented, those mechanisms will be studied and explored in the future work.

IV. COMPARISON OF AUDITING SCHEMES

The following table shows the comparison between PDP, POR and TPA mechanisms in different perspectives.

Table I Comparison of Auditing Schemes for Cloud Storage

	PDP Model	POR Model	TPA Mechanism
Goal	Allow a verifier to efficiently, periodically and securely validate that a remote server is not cheating the verifier	To Remotely audit the integrity of data stores in cloud, without keeping local copy of original files	To verify the correctness of cloud data on demand without retrieving the copy of whole data or introducing additional online burden to the cloud users

<table>
<tr><td>Approach or Idea</td><td>Is a technique for ensuring possession of files on trusted storage</td><td>Is a technique for ensuring both possession and retrievability</td><td>Is a privacy preserving mechanism</td></tr>
<tr><td>Uses or Utilizes</td><td>Uses RSA based Homomorphic tags for auditing outsourced data</td><td>Uses spot checking and error correcting</td><td>Uses primary key or public key to authenticate user</td></tr>
<tr><td rowspan="5">Limitations</td><td>1) Do not consider the case of dynamic data storage
2) direct extension of this scheme from static storage to dynamic case may suffer design and security problems
3) Does not support fully dynamic operations</td><td>1) Introduction of pre computed sentinels prevent the development of realizing dynamic data updates</td><td rowspan="5">1) To overcome these limitations in PDP and POR models Trusted TPA mechanism can be used.
2) In TPA Adversary can obtain zero knowledge information from auditing scheme.
3) TPA will keep our data safe and provide integrity.
4) Supports public auditing, privacy protection,Dynamic auditing and batch auditing</td></tr>
<tr><td colspan="2">Does not support public auditing, privacy protection, Dynamic auditing and batch auditing</td></tr>
<tr><td colspan="2">These model are not provably privacy preserving and this may leak user data information to the external auditor</td></tr>
<tr><td colspan="2">Cloud service provider may potentially reveal user data to auditors or adversaries during auditing</td></tr>
<tr><td colspan="2">System imposes a priori bound on the number of queries a client can perform</td></tr>
</table>

V. CONCLUSIONS

Although PDP and POR schemes aim at providing integrity verification for different data storage systems, the problem of supporting both public auditablity and data dynamics has not been fully addressed and the System imposes a priori bound on the number of queries. in Third Party auditing scheme , the TPA could not learn any knowledge any knowledge about the data content and can handle multiple audit sessions. After wang et al.[12] many TPA mechanisms were presented, those mechanisms will be studied and explored in the future work.

REFERENCES

[1] The NIST Definition of Cloud Computing Special Publication 800-145, Peter Mell Timothy Grance, September 2011
[2] *IEEE eLearning Library* Cloud Service and Deployment Models
[3] http://searchcloudsecurity.techtarget.com/definition/CloudAudit
[4] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at Untrusted Stores," Proc.
[5] 14th ACM Conf. Computer and Comm. Security (CCS '07), pp. 598-610, 2007.
[6] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," Proc. Fourth Int'l Conf. Security and
[7] Privacy in Comm. Networks (SecureComm'08), 2008.
[8] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession," Proc. 16th ACM Conf. Computer and Comm.
[9] Security (CCS'09), pp. 213-222, 2009.Juels and B.S. Kaliski, "PORs: Proofs of Retrievability for Large Files," Proc. 14th ACM Conf. Computer and Comm. Securit y (CCS'07),pp. 584-597, 2007.
[10] H. Shacham and B. Waters, "Compact Proofs of Retrievability," Proc. 14th Int'l Conf. Theory and Application of Cryptology and Information
[11] Security: Advances in Cryptology (ASIACRYPT '08), pp. 90107, 2008.
[12] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEEINFOCOM, pp. 525-533, 2010.
[13] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou and Jin Li, "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing"

