ISSN:0975-9646

Third Party Auditor: An Integrity Checking Technique for Client Data Security in Cloud Computing

Renuka Goyal, Navjot Sidhu

Centre for Computer Science and Technology,
Central University of Punjab Bathinda, India

***Abstract*— Now-a-days the concept of Cloud Computing is one of the major theories in the world of IT. Its services are now being applied to several IT scenarios. Cloud Computing is the internet based computing which provides users with a number of services. Users store their data in the cloud without the burden of local data storage. As the user no longer have physical possession of data so the integrity and security of data become the major concern in the cloud computing. Data stored on the cloud server may be get corrupted and sometimes even the cloud service provider for his own benefit like for more space on data centre can discard the user data which is not used for a longer time. In order to maintain the integrity of data, the user takes the assistance of a Third Party Auditor (TPA). The TPA checks the integrity of data on user demand and the released audit reports help the user to evaluate the risk of their services. TPA have an experience that user does not have and have capability to check integrity of data which is not easy for the user to check. This paper highlighted the basics of cloud computing, general model and different approaches used for TPA.**

***Index Terms*—Cloud computing, Data Integrity, Third Party Auditor, MAC, Digital signatures, Public- auditing.**

I. INTRODUCTION

Cloud Computing gained intention since 2007.It is the general term for anything that involves providing services on internet. It moves the data and computing from desktop to large datacenters. It is combination of parallel, grid and distributed computing.

Many big companies such as IBM, Google, Amazon, Microsoft, Yahoo and other move themselves to develop Cloud Computing. These companies have launched their own Cloud Computing infrastructures and services and achieved good application results and social impact, such as Amazon’s EC2 and S3, Google’ Google Apps, Microsoft’ Azure and so on.

According to National Institute of Standards and Technology (NIST): "Cloud Computing is a model for enabling 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”. It also defines Cloud Computing by:

- 5 essential characteristics
- 3 cloud service models
- 4 cloud deployment models

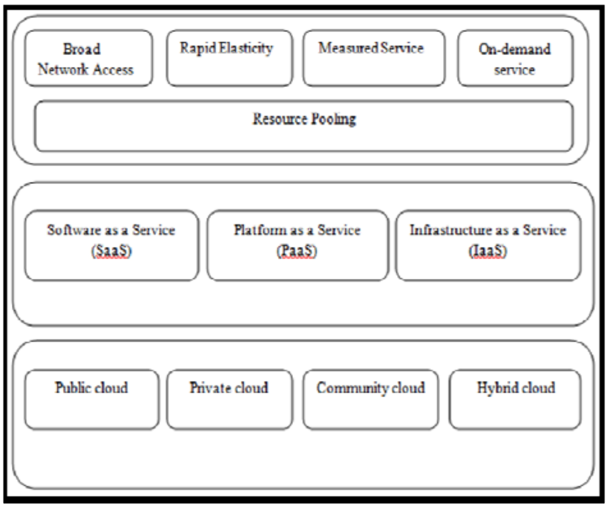

Fig. 1: Visual Model of NIST Definition of Cloud Computing

A. Cloud Computing Service Models

According to [1], [3], [6], [7] a cloud user can subscribe the following types of cloud providers: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These three types differ in the amount of control that user have over his information and how much user can expect his provider to do for him. In this IaaS is most basic and each higher model abstracts from the details of lower models.

1) Software as a Service (SaaS): In this model, users use the applications of service provider that run on cloud infrastructure. The users need not to install and run the applications on his system. The user can use these applications via any thin and thick client devices. This eliminates the user need to upgrade their applications. The user is billed according to his usage. User does not maintain the underlying the cloud infrastructure including the network, server, operating system, storage or applications. For example: Google Docs, SalesForce , SAP Business by Design etc.

2) Platform as a Service (Paas): In this model user can deploy their applications on cloud infrastructure created using some programming language, libraries and tools provided by cloud service provider. This eliminates the user need to install software and hardware required for it. User does not maintain the underlying the cloud infrastructure including the network, server, operating system, storage but has controlled over the deployed applications. For example: Force.com, Google App Engine, Window Azure etc.

3) Infrastructure as a Service (Iaas): In this model user have capability to provision processing, storage, networks and their fundamental computing resources so that user can deploy and run arbitrary software, which include operating system and applications. The user doesn't manage or control the infrastructure. He does manage or control the operating system, storage, applications, selected network components. For example: Amazon's EC2, Amazon S3, etc.

B. Deployment models in cloud computing

Enterprises can choose to deploy applications on four types of cloud that are Public, Private, hybrid or community cloud. These deployment models describe who owns, manages and is responsible for the services provided.

1) Public Cloud: The cloud infrastructure is open to use by the general public. It can be accessed by any user with an internet connection and access to cloud space.They do not know about the other users who are using the same server or network. However public clouds are less secure compared to other cloud models because public cloud is more prone to attacks. For example: Amazon, Google Apps, Window azure[5][4].

2) Private Cloud: The cloud infrastructure is used by a single organization. It is created for a specific group or organization and having access to just that group or organization.This is more secure as compared to public as only the users of organization have access. For example: eBay[1].

3) Community Cloud: The cloud infrastructure is used by a specific community of users. The community is made of two or more groups or organizations that have similar cloud requirements. For example : zimory and RightScale [8], [1].

4) Hybrid Cloud: A hybrid cloud is a combination of public or private cloud. The cloud infrastructures will be unique entities, but bound together by technology that enables data and application portability. It is created to fulfill the demand of the organization. There are not many hybrid clouds but some companies like IBM and Jupiter have introduced their base technologies for hybrid cloud [6], [7].

C. Cloud Characteristics

The following are the five key characteristics of the cloud computing that illustrate the relation to and difference from traditional computing:

1) On-demand service: The users can get computing capabilities as needed automatically. There is no need for user to directly interaction with the cloud server provider. The computing capabilities can be server time, software use, network storage etc[1],[3].

2) Broad Network Access: The Services are available over the internet via a standard mechanism that allows the users to access these services through any thin or thick client tools like desktop, laptop, PDA, mobile phone[4],[7].

3) Resource pooling: The cloud service provider employs a multitenant model to serve multiple customers by pooling computing resources like virtual machines, storage, memory ,network bandwidth, processing. The different physical and virtual resources can be dynamically assigned and reassigned according to users demand. The users have no knowledge of the exact location of the provided resources[2].

4) Rapid Elasticity: The users can rapidly scale up to use whenever needed or scale down to release whenever finished. From the user's points of view the available services are unlimited [9], [8].

5) Measured service: The cloud server can use mechanism to measure the usage of resources and services for each individual user. For both the provider and the user resources usage will be monitored, controlled, metered and reported[4].

II. Cloud Data Integrity

Data integrity means data should be correctly stored on the cloud server without any modification and if any violations i.e. if the data is get lost, altered or compromised can be detected. It must remain in the same state. But the integrity of data is at risk in cloud server [10], [11].

As the user does not have physical possession of data so the integrity and security of data become the major concern in the cloud computing. Data can get modified by other users or even sometimes cloud service provider for his own benefit can behave unfaithfully towards the users regarding outsourced data. For example cloud service providers for more space on data centre can discard the user data which has not been or rarely accessed by the user for a longer time or even can hide the data loss incidents to maintain his reputation [13].

We need to ensure the integrity by making the user capable to check over the cloud data from any unauthorized modification. One solution is to first download the files whose integrity have to check but downloading the files requires high transmission cost. So to maintain the data integrity and to minimize the storage risk it is important to take assistance of a Third party auditor (TPA) who checks the data integrity for the cloud user and helps the user in minimizing his risk.

III. Third Party Auditor

The Third party auditor is a kind of inspector. The Third Party Auditor who has resources and experience that a user does not have and check the integrity that is difficult for users to check. The auditors can understand the threats and they know best practices. The released audit report helps the user to evaluate the risk of their services. It also helps the cloud service provider in improving their cloud platform [14], [15].

A. Cloud Data Storage Model

There are three different network entities in cloud system which is users, cloud service provider and third party auditor.

1) *Users:* These are active participants. They have data to be stored in the cloud and rely on the cloud for data maintenance and computation. Both individual consumers and organizations can be the users[13], [14].

2) *Cloud service provider (CSP):* It is the most important part of cloud architecture. It has significant storage space and computation resource to store and maintain the user data.it provides all his services in pay per use manner[13],[14].

3) *Third party auditor (TPA):* It has more capabilities than the user and checks the integrity of data for the user and his audit reports helps the users in evaluating the risk. [14], [16].

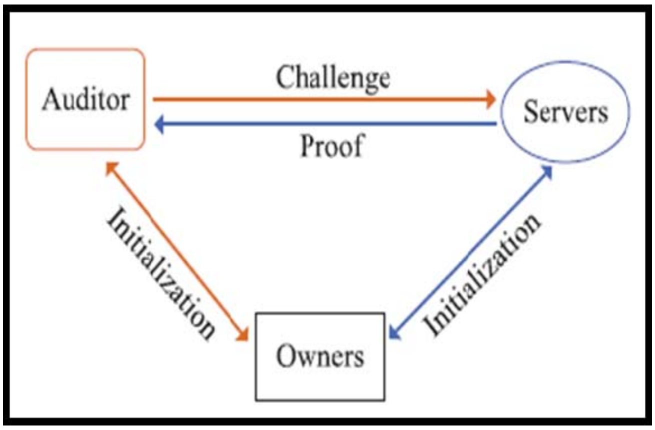

Fig. 2: Cloud Data Storage Model

B. Functions of Third Party Auditor:

According to (Patel & Patel, 2012) (Gowrigolla, Sivaji & Masilliamani, 2010) (Balakrishnan et al, 2011) standard TPA in cloud environment should take following functionalities into consideration:

1) *No data leakage or data learning:* TPA should neither learn any information about the data file from the message it receives from client/server nor leak the same to any unauthorized entity.

2) *Audit without downloading:* The TPA should audit without asking for entire file from server, not even in encrypted form. TPA should audit the user data without asking for the local copy of the data or even learning the data contents

3) *Integrity Verification:* One of the important security concerns is to verify integrity of data stored on cloud. TPA should verify the integrity of client's data stored on cloud with low communication overhead.

4) *High Performance:* Performance of TPA is also an important issue as it is a central component of the cloud system, where there are thousands of client and multiple servers. TPA should not become bottleneck of entire system and performance of overall system should not be compromised due to heavy load on TPA.

5) *Scalability:* As cloud is a completely dynamic environment, any number of users can come in or go out. Also it is expected to have huge data storage on cloud server. Functionalities of TPA should not be affected by number of cloud clients, servers, number of data files stored on the cloud or the overall size of the entire storage. TPA should offer scalable architecture which is independent on all the factors mentioned.

6) *Dynamic data operation support:* One of the main differences between the cloud computing and other online storage system is its dynamic data support & sharing. TPA should take the fact into consideration that the data stored on cloud may be used & edited by multiple users simultaneously. It must support dynamic operations on data blocks i.e. data update, append and delete.

7) *Batch Auditing:* Third Party Auditor also supports batch auditing to improve efficiency. TPA performs multiple auditing tasks simultaneously and it also reduces communication and computation cost.

C. Different Schemes used for TPA

1) *MAC Based solutions:* There are two types of MAC based solution the first solution does not ensure privacy preserving the second one suffers from auditor statefulness and other demerits.

In first solution user first divides the files into blocks and calculate the MAC for each block. Users transfer the file blocks and the MACs to the cloud service provider and share the secret key to Third party auditor. TPA demands for a random no. of blocks and theirs MACs from cloud service provider. Then TPA uses the secret key to verify the correctness of stored data on the cloud server.

Drawbacks of this system are:

- TPA requires retrieval of data blocks for verifying the correctness of data blocks which is not privacy preserving.
- It supports only for static data not for dynamic data.
- Communication and computation complexity are linear with the sample size.

To avoid the requirement of data retrieval in TPA verification, one can improve the solution as: Before outsourcing the cloud user chooses S random message authentication code keys, pre-computes S MACs for the whole data file f and publish these verification metadata (the keys and the MACs) to TPA. The TPA can reveal a secret key to the cloud server and ask for a fresh keyed MAC for comparison in each audit. This is privacy preserving as long as it is impossible to recover F However, it suffers from the following drawbacks:

- The number of times a particular data file can be audited is limited by the number of secret keys that must be fixed a priori. Once all possible secret keys are exhausted, the user then has to retrieve data in full to recompute and republish new MACs to TPA.
- The TPA also has to maintain and update state between audits, i.e., keep track on the revealed MAC

keys. Considering the potentially large number of audit delegations from multiple users, maintaining such states for TPA can be difficult and error prone.

- It can only support static data, and cannot efficiently deal with dynamic data at all.

2) Public Auditing Scheme for Third Party Auditor:

According to (Paigude & Chavan, 2013) (Yang & Jia, 2013) (Gowrigolla, Sivaji & Masilliamani, 2010) the working of TPA Consists of four algorithms: KeyGen, SigGen, GenProof, VerifyProof.

a) KeyGen: Key generation algorithm is run by the user to setup the scheme. In this the user generates his own public/private key pairs.

b) SigGen: This algorithm is run by the cloud user for the generation of verification metadata, which can be signatures or other information used for auditing

c) GenProof: This algorithm is for the generation of proof of correctness of data storage and is run by the cloud server.

d) VerifyProof: This algorithm is run by the TPA for the auditing of the proof generated by the cloud server

It is divided into two phases: Setup and Audit phase.

1. **Setup Phase:** The user initializes the public and private parameters of the system by executing KeyGen and preprocesses the data file by using SigGen to generate the verification metadata. The user then stores the data file at the cloud server, delete its local copy, and publish the verification metadata to TPA for later audit.
2. **Audit:** The TPA issues an audit message or challenge to the cloud server to make sure that the cloud server has retained the data file properly at the time of the audit. The cloud server will derive a response message from a function of the stored data file by executing GenProof. Using the verification metadata, the TPA verifies the response via VerifyProof.

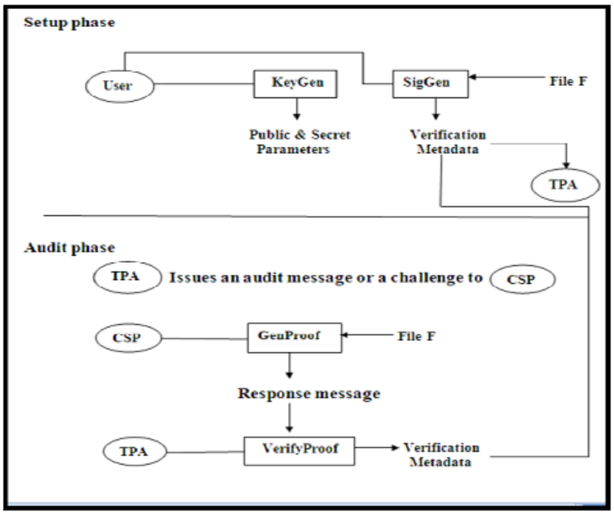

Fig. 3: Public Auditing Scheme for Third Party Auditor

3) HLA based solution: To support public auditability without retrieving the data blocks the HLA based solution is used. Like the MACs, HLAs are also some unforgeable verification metadata that is used to check the integrity of data. The only difference is that HLAs can be aggregated. It is possible to compute an aggregated HLA which authenticates a linear combination of the individual data blocks. It allows efficient auditing and consumes constant bandwidth. But this solution may reveal user data information to TPA and violates privacy preserving.

4) Privacy- Preserving Public Auditing Scheme: To achieve privacy preserving public auditing, public key based homomorphic linear authentication with random masking is used. TPA checks the integrity without demanding the actual copy of data. The linear combination of sampled blocks in the server's response is masked with randomness generated by the server. With random masking, the TPA no longer has all the necessary information to build up a correct group of linear equations and therefore cannot derive the user's data content, no matter how many linear combinations of the same set of file blocks can be collected.

5) Digital signatures based solution: K. Gonvinda proposed digital signature method to protect the privacy and integrity of data. It uses the RSA algorithm for encryption and decryption which follows the process of digital signatures for message authentication.

IV. CONCLUSION

Cloud computing provides many benefits to their user but security is major issues in cloud computing. As user store their data to cloud data centers but as user does not know the exact location of their data so integrity of data is very important. To check the integrity of data there are many solutions available. One of solution is to take the assistance of a third party auditor. Different authors provide different solutions for implementing third party auditor. Each scheme has its own merits and demerits.

ACKNOWLEDGMENT

I would like to thank my supervisor for his most support and encouragement. I sincerely thank to my parents and friends, who provide the advice and financial support.

REFERENCES

[1] H. K. Idrissi, A. Kartit, and M. El Marram, "A taxonomy and survey of Cloud computing", *IEEE Transactions*, pp.1-5, April 2013.

[2]Y. Jadeja, and K. Modi, "Cloud computing-concepts, architecture and challenges" in *IEEE International Conference on Computing, Electronics and Electrical Technologies*, March 2012, pp. 877-880.

[3] M. B. Mollah, K. R. Islam, and S. S. Islam, "Next generation of computing through cloud computing technology", in *25th IEEE Canadian Conference on Electrical & Computer Engineering*, April 2012, pp.1-6.

[4] J. J. Wang, and S. Mu, "Security issues and countermeasures in cloud computing", in *IEEE International Conference on Grey Systems and Intelligent Services*, September 2011, pp.843-846.

[5] H. Tianfield, "Security issues in cloud computing", *IEEE International Conference on Systems, Man, and Cybernetics*, October 2012, pp- 1082-1089.

[6] S. Zhang, S. Zhang, X. Chen, and X. Huo, “Cloud computing research and development trend”, in *Second International Conference on Future Networks*, January 2010, pp. 93-97.

[7] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A break in the clouds: towards a cloud definition” *ACM SIGCOMM Computer Communication Review*. vol. 39, no. 1, pp. 50-55, January 2009.

[8] T. Dillon, C. Wu, and E. Chang, “Cloud computing: Issues and challenges”, in *24th IEEE International Conference on Advanced Information Networking and Applications*, April 2010, pp.27-33.

[9] A. Behl, “Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation”, *IEEE World Congress on Information and Communication Technologies*, pp. 217-222, December 2011.

[10] D. Attas, and O.Batrafi, “Efficient integrity checking technique for securing client data in cloud computing”, *International journal of electrical & computer science*, pp. 43-48, 2011

[11] S. Balakrishnan, G. Saranya, S. Shobana, and S. Karthikeyan, “Introducing Effective Third Party Auditing (TPA) for Data Storage Security in Cloud”, *International Journal of Computer Science and Technology*, pp. 397-400, June 2011.

[12] A. Bhagat, and R.K. Sahu, “Using Third Party Auditor for Cloud Data Security: A Review”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, pp. 34-39, March 2013.

[13] T. K. Chakraborty, A. Dhami, P. Bansal, and T. Singh, (2013, February), “Enhanced public auditability & secure data storage in cloud computing”, in *IEEE 3rd International Advance Computing Conference,* February 2013, pp.101-105.

[14]B. Gowrigolla, S. Sivaji, and M. R. Masillamani, “Design and auditing of cloud computing security”, in *5th International Conference on Information and Automation for Sustainability*, December 2010, pp.292-297.

[15]S. Han, and J. Xing, “Ensuring data storage security through a novel third party auditor scheme in cloud computing”, in *IEEE International Conference on Cloud Computing and Intelligence Systems*, September 2011, pp.264-268.

[16] W. Itani, A. Kayssi, and A. Chehab, “Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures”, in *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, December 2009, pp.711-716.

[17] W. Luo, and G. Bai, “Ensuring the data integrity in cloud data storage”, in *IEEE International Conference on Cloud Computing and Intelligence Systems*, pp.240-243, September 2011.

[18]B. Makhija, V.K. Gupta, and I. Rajput, “Enhanced Data Security in Cloud Computing with Third Party Auditor”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 2, pp. 341-345, February 2013.

[19] A. Mohta, R.K. Sahu, and L. K. Awasthi, “Robust Data Security for Cloud while using Third Party Auditor”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 2, pp.1-5, Feburary 2012..

[20] M. S. Muneshwara, and A.T. Chandarl, “Monitoring the integrity of Dynamic Data Stored in Cloud Computing”, *International Journal of Engineering Research & Technology*, vol. 1 no.4, pp. 1-8, June 2012.

[21] T. Paigude, and T. A. Chavan, “A survey on Privacy Preserving Public Auditing for Data Storage Security”, *International Journal of Computer Trends and Technology*, vol. 4, no. 3, pp. 412-417, 2011.

[22] P. Syam Kumar, R. Subramanian, and D. Thamizh Selvam, “Ensuring data storage security in cloud computing using Sobol sequence”, in *IEEE International Conference on Parallel Distributed and Grid Computing, October 2010,* pp. 217-222.

[23] V. Vinaya, and P. Sumathi, “Implementation of Effective Third Party Auditing for Data Security in Cloud”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 5, pp. 382-387, May 2013.s

[24] C. Wang, K. Ren, W. Lou, and J. Li, “Toward publicly auditable secure cloud data storage services” *IEEE Network*, vol. 24, no.4, pp. 19-24, 2010.

[25]C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-preserving public auditing for data storage security in cloud computing”, *IEEE INFOCOM*, pp. 1-9, March 2010.

[26] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, “Toward secure and dependable storage services in cloud computing”. *IEEE Transactions on Services Computing*, vol.5, no.2, pp.220-232, 2012.

[27] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling public auditability and data dynamics for storage security in cloud computing” *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 847-859, 2011.

[28] K. Yang, and X. Jia, “An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing”, *IEEE Transactions on Parallel & Distributed Systems*, vol. 24, no. 9, pp. 1717-1726, 2012.

[29]Y. Zhu, H. Hu, G.-J. Ahn, and S. S. Yau, “Efficient audit service outsourcing for data integrity in clouds,” In *Journal of Systems and Software*, vol. 85, no. 5, pp. 1083-1095, May 2012.