ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

A Survey of Techniques and Tools Used For Cloud Data Integrity Verification

Princelly Jesu. A[1], Ramesh Kumar. S[2]

PG Scholar, Department of Information Technology, Karunya University, Coimbatore, Tamil Nadu, India[1]

PG Scholar, Department of Information Technology, Karunya University, Coimbatore, Tamil Nadu, India[2]

ABSTRACT: Data outsourcing on cloud is emerging as a research area with different tools and techniques. Data security is becoming one of the major concerns of cloud data owners. Cloud data servers use both internal and external methods to examine the outsourced data and are concerned with developing methods to explore the unique types of techniques for secure data computing. To achieve the high security on outsourced data numerous techniques were used. These tools provide various benefits for the secure data outsourcing and data verification over cloud environment. This survey focuses on the various tools and techniques used for cloud data verification and finally provide an outline to overcome the problems of those techniques.

KEYWORDS: Cloud computing, Data security, Integrity verification, Public auditing, Privacy.

I. INTRODUCTION

Cloud computing is a large group of virtual servers are networked to allow users to store, share and access resources virtually [1]. In the flexible cloud environment, data outsourcing became very popular among the cloud users. Secure data accessing is the major task of all cloud providers. Clients can join with the cloud to receive more reliable services, so that they can access data from anywhere and at any time. These cloud services are segregated into four types, such as cloud storage service, software as a service (SAAS), platform as a service (PAAS), and infrastructure as a service (IAAS) [2]. Among all the above services, the survey handles the data storage and outsourcing services from the cloud. Even though there are numerous services available in the cloud, it has several challenging security issues.

Data security on cloud includes several aspects such as secrecy, reliability and accessibility. In this survey, we will focus on various techniques involved with data integrity or reliability. In simple terms data integrity can be understood as the maintenance of intactness of any data during transactions like storage, sharing and retrieval. In general, the data integrity is referred as the process of maintaining the whole undamaged data. Due to several aspects, outsourced data may suffer integrity problems on cloud. Finding and protecting the outsourced data along with the complete integrity verification is an active research area. And there are huge research problems belong to this area have been highly concentrated in the literature.

Data integrity process:

The data integrity verification can be performed in two ways; internal and external verification. In this survey, we will focus on integrity protection and verification from external entities. The general framework of data integrity verification is shown in fig 1. Here there are seven steps involved in the data integrity process and three types of roles are there. One is Client as well as data owner, cloud server and verifier. Client stores the data on the cloud storage area via CSP (cloud service provider). TPA or verifier verifies the data integrity of outsourced data on cloud server.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

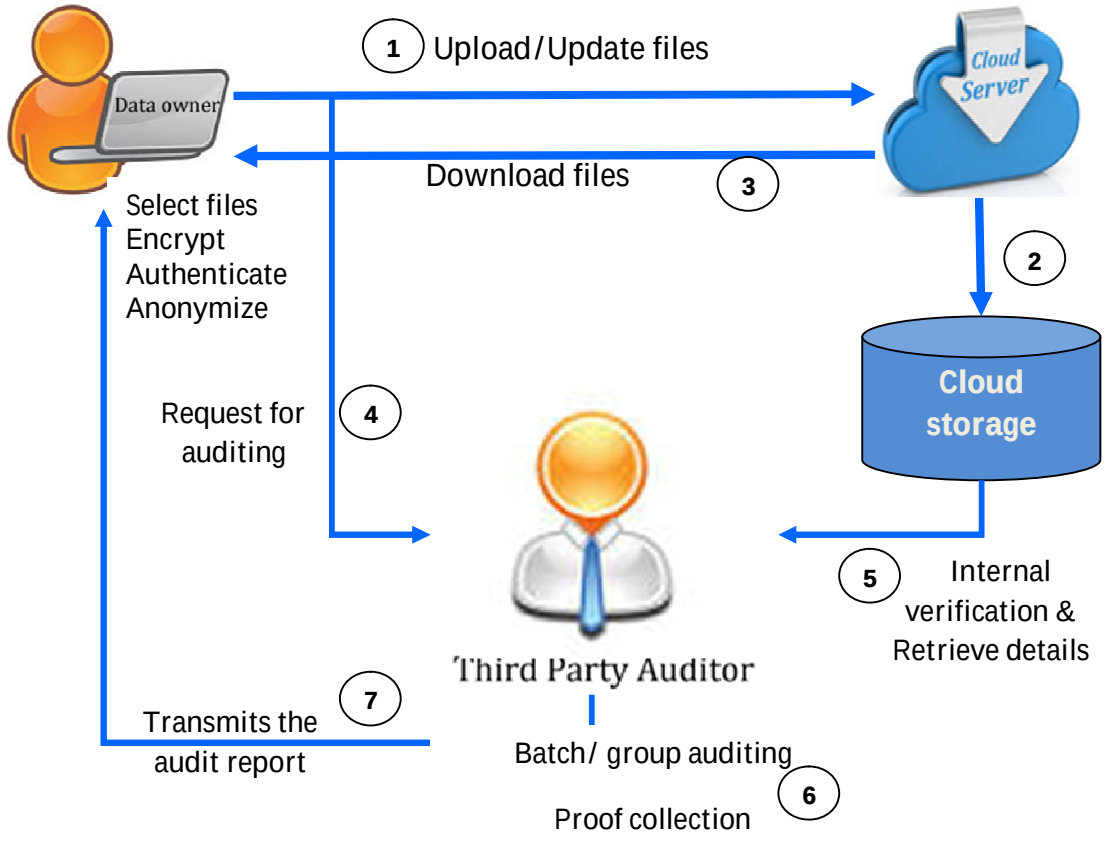

Fig 1: cloud data outsourcing and integrity verification process

The main process of a data integrity verification scheme can be analyzed in the following steps:
Step 1: the client initially performs encryption before uploading in to the cloud storage. The client can specify several security options such as authentication and anonymization.
Step 2: the uploaded data will be stored with metadata, which is prepared by the client at the time of verification in the cloud storage. The cloud service provider provides the storage services to the clients.
Step 3: the data owner can access their files from the cloud at any time with their authentication. At the time of download, the client can perform the internal verification for their content. This verification needs the metadata named as homomorphic linear authenticator (HLA) which is based on the homomorphic signatures.
Step 4: the client can perform external audit from the TPA (third Party Auditor). In such cases, the TPA is considered as semi trusted. In this time, the client may perform privacy and anonymization techniques to protect the data and report from the TPA.
Step 5: with the use of internal audit report and metadata from the cloud server, the TPA performs batch or group auditing. The TPA also prepares and transmits the verification proof to the client at the time of audit.
Step 6: The TPA performs the verification process without knowing the whole content. So the TPA should not demand the local copy of data. And this should not create any new vulnerability to the clients.
Step 7: finally the TPA transmits the audit report to the client with attestation. This attestation scheme is helpful to track the verifier in future.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

II. VARIOUS APPROACHES USED FOR INTEGRITY VERIFICATION

In cloud environment, data storage and outsourcing increases tremendously. To deal with the integrity problems associated with the outsourced data, there are several traditional systems deployed numerous techniques. This chapter shows the various approached used for integrity verification.

Reed–Solomon code: Reed–Solomon coding is very widely used in large cloud storage systems to correct the burst errors associated with media defects. These codes are an important group of error-correcting codes.

Checksums: A checksum is a count of the number of bits in a cloud data outsourcing process unit that is included with the unit so that the clients can check whether the same number of bits arrived. If the total bit count matches, then this is assumed as the process is not affected by any integrity problem.

Message authentication code (MAC): In cloud security message authentication code (MAC) is a tiny section of information used to authenticate outsourced data. This also helps to verify sender authenticity, which achieves data integrity.

Digital signatures: Digital signatures are the public key primitives of message authentication over the cloud. A digital signature is a popular approach that authenticate and verifies the client for outsourced digital data.

Even though there are several approaches are implemented, the data owner in the cloud still needs a method to verify their data stored remotely on a semi-trusted cloud server.

III. LITERATURE REVIEW

There are huge researchers were in the development of data verification methods and proposed Remote Data checking protocols to overcome the integrity problems in outsourced cloud data. Those protocols frequently verify the outsourced data in the un trusted storage server in the cloud without retrieving the original data. The remote data checking is divided into two types; Probabilistic and deterministic protocols.

1) **Deterministic Verification Protocols:**
   In In Deterministic verification protocols, the examiner checks the all data blocks of the outsourced file in single at each server. If the file size is small, this kind of protocol gives good result. This type of protocol is not suitable for big data environment.
2) **Probabilistic Verification Protocols:**
   In Probabilistic verification Protocols, the verifier checks the Integrity of random subset of the data chunks. This is widely used by the clients, who have huge data size and dynamic datasets.

The following chapter describes the various approaches and techniques are developed from the above methods.

Wang, Qian, et al [3] developed a new scheme, which supports both public auditability and dynamic data operations, but prior works only relies on any one operation. In specific, in order to achieve efficient data dynamics, the authors improved the existing proof of storage models by manipulating the classic Merkle Hash Tree construction for block tag authentication. This also developed with the intension to support efficient handling of multiple auditing tasks when the client size is huge; this paper also developed and utilizes the technique of bilinear aggregate signature to optimize the multi client settings. The main advantages of this approach are increases the performance of TPA by providing multiple auditing tasks simultaneously.

Wang, Cong, et al [4] had proposed a privacy preserving public auditing system for data storage security in un-trusted cloud server. They used the **homomorphic linear authenticator** and random masking to protect data from semi-trusted TPA. These types of mechanisms empower both security and privacy in the cloud environment. They further extended the privacy preserving public auditing protocol with batch audit, where numerous auditing tasks handled together.

Hongwei, Peng Zhang, and Jun Liu [5] proposed a new public auditing protocol. Using this protocol, the client can check the data loss and security threats of the outsourced data without retrieving the data. This protocol reduces the computation cost for clients. This protocol handles numerous attacks and protects data from that. The attacks are tamper attack, loss attack and curiosity attack. They have used BLS short signature scheme and the homomorphic hash function to build the secure public auditing protocol.

Zhu, Yan, et al [6] proposed a construction of dynamic audit services for outsourced data on un-trusted storage servers. They also presented a method for periodic sampling audit to improve the performance of TPA's. This proposal is developed from a simple schedule, which helps to periodically manage all audit tasks. They presented the audit

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

service based on the following approaches such as random sampling, fragment structure and index technique with hash table. This paper has a small and constant amount of overhead, which minimizes computation and communication costs for the clients.

Yang, Kan, and Xiaohua Jia [7] had designed an auditing framework for cloud storage systems and proposed an efficient and privacy preserving auditing protocol [PPAP]. This protocol supports the data dynamic operations, which is efficient and provably secure in the random model. This auditing protocol also supports batch auditing, which suitable for both multiple clouds and multiple owners. This is not relayed with any trusted controller. This protocol protects the data privacy against the auditor by combining the cryptography method with the bilinearity property of bilinear paring, rather than using the mask technique.

Wang, Boyang, Baochun Li, and Hui Li [8] had proposed a novel privacy-preserving mechanism that supports public auditing on shared data stored in the cloud. They utilized ring signatures to handle the integrity verification. This mechanism needs metadata to audit the precision of shared data over the cloud. This mechanism maintains the signing procedure on each block to keep the data private from public verifiers. This verifies the integrity without retrieving the whole file from the cloud server. This paper also supports batch audit process, which works simultaneously. This paper overcomes the problem of traditional ring signature schemes, which cannot be directly used in the public auditing systems. Because those signature schemes are not supported blockless verifiability.

Liu, Chang, et al. [9] presented a formal analysis and development fine grained data updates and its requests. They developed an enhanced version, which reduces the communication cost for small update verification requests. They proposed a public auditing scheme, which is based on both BLS signature and Merkle hash tree (MHT) that can support fine-grained update requests. This scheme supports updates with a size that is not restricted by the size of file blocks. This provides extra flexibility and scalability. It provides 'authorized auditing" where the unauthorized audit requests are eliminated.

Liu, Qin, Guojun Wang, and Jie Wu[10] had proposed a novel consistency as a service (CaaS) model, which consists of a large data cloud and multiple small audit clouds. In the CaaS model, a data cloud is maintained by a (CSP) cloud service provider, and a group of users that constitute an audit cloud can verify whether the data cloud provides the promised level of consistency or not. They proposed a two-level auditing architecture, which only requires a loosely synchronized clock in the audit cloud. They developed a heuristic auditing strategy and reveal the violations. They uses Local and global Consistency Auditing algorithms. Finally this helps to find the malicious CSP in the cloud environment.

Wang, Huaqun [11] had proposed a new remote data integrity checking model named as ID-DPDP on distributed multicloud storage. This eliminates the need of certificate management. This ID-DPDP is based on the bilinear pairings in random oracle model. The protocol is more flexible besides the high efficiency. Based on the client's authorization, the proposed ID-DPDP protocol can realize private verification, delegated verification and public verification.

Liu, Chang, et al [12] had developed a novel public auditing scheme named MuR-DPA(multi-replica dynamic public auditing. They incorporated the new scheme named as authenticated data structure (ADS) based on the Merkle hash tree (MHT), which is also known a MR-MHT. This configuration allows efficient verification of updates for multiple replicas. With the scheme of MR-MHT, they also designed a novel public auditing protocol for verification of all replicas at once.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Table1. COMPARISON OF VARIOUS INTEGRITY VERIFICAITON TECHNIQUES

Author	Descriptions	Techniques	Merits	Demerits
Wang, Qian, et al [3]	supports both public auditability and dynamic data operations	Merkle Hash Tree (MHT) construction for block tag authentication	Handles multiple audit tasks.	Increases server overhead.
Wang, Cong, et al [4]	Privacy preserving public auditing protocol	homomorphic linear authenticator and random masking	Handles multiple audit tasks. Increases the performance.	Not suitable for large scale systems.
Hongwei, Peng Zhang, and Jun Liu [5]	Proposed a public auditing protocol	BLS short signature scheme and the homomorphic hash function	supports public verification, dynamic update and privacy preserving	They didn't Considered of pollution attack
Zhu, Yan, et al [6]	proposed a dynamic audit services	fragment structure, random sampling, and index-hash table (IHT)	Constant amount of overhead, which minimizes computation and communication costs.	Not suitable for high dimensional data environment.
Yang, Kan, and Xiaohua Jia [7]	Designed an auditing framework for cloud storage systems and propose a privacy-preserving and efficient storage auditing protocol.	Cryptography method with the bilinearity property of bilinear paring	Supports Dynamic operations. Multiple clouds but also multiple owners. Improves the performance. Suitable for large scale servers.	Data integrity not fully verified.
Wang, Boyang, Baochun Li, and Hui Li [8]	a novel privacy-preserving mechanism supports public auditing on shared data	New Ring signature algorithm with blockless verifiability. New homomorphic authenticable ring signature (HARS)	No need to retrieve full file.	Cannot trace the verifier. Possible for data forgery at the time of auditing. No guarantee for data freshness.
Liu, Chang, et al. [9]	proposed a public auditing scheme	BLS signature and Merkle hash tree	Lower overheads for big data applications.	Server-side protection not fully handled.
Liu, Qin, Guojun Wang, and Jie Wu[10]	Proposed a novel consistency as a service (CaaS) model, which finds malicious CSP.	heuristic auditing strategy (HAS) Directed acyclic graph (DAG)	Finds the malicious CSP Finds the violations effectively	Expensive and increases the overhead.
Wang, Huaqun [11]	proposed a new remote data integrity checking model named as ID-DPDP	bilinear pairings in random oracle model	Flexible and improves the efficiency	Verification delay occurs
Liu, Chang, et al [12]	Presented a multi-replica dynamic public auditing (MuR-DPA)	MuR-DPA(multi-replica dynamic public auditing).	Verification of cloud data storage with multiple replicas.	Works only with constant-sized integrity proofs

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

	scheme, which is a newly designed authenticated data structure to handle replica data in the cloud.	authenticated data structure MHT	less communication overhead. security against dishonest CSP	

V. CONCLUSION

This paper provides a survey of the various techniques involved with cloud data verification. In this paper various methods for integrity verification for outsourced data on cloud are discussed. It is observed that various methods and techniques for integrity checking are presented. The selection of the methods may depends on the type of data and its size. Finally the survey summarizes the overall drawbacks of all the methods with various considerations.

REFERENCES

[1] Conway, Gerry. "Introduction to Cloud Computing." (2011)..
[2] Luo, Jun-Zhou, et al. "Cloud computing: architecture and key technologies."*Journal of China Institute of Communications* 32.7 (2011): 3-21.
[3] Wang, Qian, et al. "Enabling public auditability and data dynamics for storage security in cloud computing." *Parallel and Distributed Systems, IEEE Transactions on* 22.5 (2011): 847-859.
[4] Wang, Cong, et al. "Privacy-preserving public auditing for secure cloud storage." *Computers, IEEE Transactions on* 62.2 (2013): 362-375.
[5] Hongwei, Peng Zhang, and Jun Liu. "Public data integrity verification for secure cloud storage." *Journal of networks* 8.2 (2013): 373-380.
[6] Zhu, Yan, et al. "Dynamic audit services for outsourced storages in clouds."*Services Computing, IEEE Transactions on* 6.2 (2013): 227-238.
[7] Yang, Kan, and Xiaohua Jia. "An efficient and secure dynamic auditing protocol for data storage in cloud computing." *Parallel and Distributed Systems, IEEE Transactions on* 24.9 (2013): 1717-1726.
[8] Wang, Boyang, Baochun Li, and Hui Li. "Oruta: Privacy-preserving public auditing for shared data in the cloud." *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*. IEEE, 2012.
[9] Liu, Chang, et al. "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates." *Parallel and Distributed Systems, IEEE Transactions on* 25.9 (2014): 2234-2244.
[10] Liu, Qin, Guojun Wang, and Jie Wu. "Consistency as a Service: Auditing Cloud Consistency." *Network and Service Management, IEEE Transactions on* 11.1 (2014): 25-35.
[11] Wang, Huaqun. "Identity-Based Distributed Provable Data Possession in Multicloud Storage." *Services Computing, IEEE Transactions on* 8.2 (2015): 328-340.
[12] Liu, Chang, et al. "MUR-DPA: top-down levelled multi-replica merkle hash tree based secure public auditing for dynamic big data storage on cloud."*IEEE Transactions on Computers* 1 (2014): 1-1.