ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 9, September 2016

# Survey on Cloud Data Integrity: Issues and Current Solutions

[1] Shivaraj Hiremath, [2]Sanjeev R Kunte
[1] SKSVM Agadi College of Engineering and Technology Laxmeshwar
[2] JNN College of Engineering, Shivamogga

***Abstract:* — Cloud computing is a general term for the delivery of hosted services over the internet. The cloud is a platform where data owner remotely store their data in the cloud to enjoy the high quality applications and services. Inspite of many advantages associated with cloud computing, accessing and storing data on cloud storage has some security risks. For sensitive and confidential data there should be some security mechanisms. In this paper, we present cloud data integrity issues that are preventing people from adopting the cloud and we give a survey on the solutions that have been done to minimize risks of these issues.**

***Keywords:*-- Cloud Computing, Cloud Data,; Data Integrity, Third party Auditor.**

## I. INTRODUCTION

Cloud computing now is everywhere. It is a new computing paradigm in which resources are shared as a service over the internet. According to [1], small and medium organization will move to cloud computing because it will support fast access to their applications and reduce the cost of infrastructure.

Organizations now are trying to avoid focusing on IT infrastructure. They need to focus on their business process to increase profitability [13]. Therefore, the importance of cloud computing is increasing, becoming a huge market and receiving much attention from the academic and industrial community's Schematic definition of cloud computing can be illustrated in fig 1.


***Fig. 1: Schematic Definition of Cloud Computing***

Cloud model provides three service models namely Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) and four deployment models namely Private Cloud, Public Cloud, Community Cloud and Hybrid Cloud.

## II. SECURITY ISSUES IN THE CLOUD

Security is the biggest concern when it comes to cloud computing. It is a mix of technologies, controls to safeguard the data, and policies to protect the data, services, and infrastructure.

A company essentially gives away private data and information to cloud that might be sensitive and confidential. It is then up to the cloud service provider to manage, protect and retain them, thus the provider's reliability is very critical. By outsourcing the data, users lose their physical control over data when it is stored in a remote server. As a result, users are at the mercy of their cloud service providers for the availability and integrity of their data. Recent downtime of Amazon's S3 is such an example [2]. Therefore there are new security requirements in the cloud compared to traditional environments. Traditional security architecture is broken because the customer does not own the infrastructure any more.

Despite powerful and reliable server compared to client processing power and reliability, there are many threats facing the cloud not only from an outsider but also from an insider which can utilize cloud vulnerabilities to do harm [3]. Some untrusted providers could hide data breaches to save their reputations or free some space by deleting the less used or accessed data [4]. These threats

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 9, September 2016

may harm data confidentiality, data integrity, and data availability.

One of the important concerns in the cloud computing that need to be addressed is to assure cloud data integrity. Accordingly in the next section we will discuss about data integrity issues and solutions to protect data integrity.

## III. DATA INTEGRITY ISSUES

Data integrity means data should be kept from unauthorized modification [5]. Any modification to the data should be detected. Data that is stored in the could suffer from the damage on transmitting to/from cloud data storage. Since the data and computation are outsourced to a remote server, the data integrity should be maintained and checked constantly in order to prove that data and computation are intact.

Data integrity help in getting lost data or notifying if there is data manipulation. The following are the two examples of how the data integrity could be violated [14].

### *A. Data loss or Manipulation*

Basic types of data loss include data destruction, data corruption and unauthorized data access. The reason for these types of loss includes infrastructure malfunctions, software errors and security breaches. Users have a huge number of user files. Therefore, cloud providers provide Storage as services (SasS). Those files can be accessed every day or sometimes rarely. Therefore, there is a strong need to keep them correct.

In many cases, data could be altered intentionally or accidentally. Also, there are many administrative errors that could cause losing data such as getting or restoring incorrect backups. The attacker could utilize the user's outsourced data since they have lost the control over it.

### *B. Computation of Untrusted Remote Server*

Cloud computing is not just about storage. There is some intensive computing that need cloud processing power in order to perform their tasks. Therefore, users or clients outsource their computation. Since the cloud provider is not in the security boundary and is not transparent to the owner of the tasks, no one will prove whether the computation integrity is intact or not.

Sometimes, the cloud provider behaves in such a way that no one will discover a deviation of computation from normal execution. Because the resources have a value to the cloud provider, the cloud provider could not execute the task in a proper manner. Even if the cloud provider is considered more secure, there are many issues such as those coming from the cloud provider's underlying systems, vulnerable code or misconfiguration.

## IV. SOLUTIONS TO DATA INTEGRITY ISSUES

There are two traditional ways of proving the integrity of data outsourced in a remote server [14].

1) By using Message Authentication Code (MAC) algorithm.
2) Computing the Hash Value in the cloud by using a hash tree.

MAC algorithms take two inputs, a secret key and variable length of data, which produce one output, a MAC (tag). In this way this algorithm is run on the client side. After getting a MAC, the data owner outsources those data to the cloud. For checking its integrity, the data owner downloads the outsourced data and then calculates the MAC for it and compares it with the one calculated before outsourcing that data. By using this method accidental and intentional changes will be detected. This method consumes more bandwidth.

The second one is to compute that hash value in the cloud by using a hash tree. In this technique, the hash tree is built from bottom to top where the leaves are the data and parents are also hashed together until the root is reached. The owner of data only stores the root. When the owner needs to check his data, he asks for just root value and compares it with the one he has.

To some extent it is not practical because computing the hash value of a huge number of values consume more computation. Therefore, there is a need to find a way to check data integrity while saving bandwidth and computation power.

Remote data auditing, by which the data integrity or correctness of remotely stored data is investigated, has

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 9, September 2016

been given more attention recently. There has been a lot of development in this field and lots of algorithms have been proposed by various researchers. In our survey we found that there are four techniques proposed by researchers for ensuring Integrity of cloud Data.

***A. Third Party Auditor (TPA)***

TPA is the third party auditor who will audit the data of data owner or client. The TPA has expertise and capabilities that users do not. TPA eliminates the involvement of the client through the auditing process and checks whether his data stored in the cloud are indeed intact.

Balusamy et al. [5] proposed a framework, which involves the data owner in checking the integrity of their outsourced data. This scheme involves the data owner in the auditing process. First, TPA uses normal auditing processes. Once they discover any modification to the data, the owner is notified about those changes. The owner checks the logs of the auditing process to validate those changes. If the owner suspects that unusual actions have happened to his data, he can check his data by himself or by another auditor assigned by him.

Therefore, the owner is always tracking any modification to his own data. There is an assigned threshold value that a response from the third party auditor should not exceed. The data owner validates all modifications lesser than or equal to this threshold. If the time exceeds this threshold, the data owner is supposed to do surprise auditing.

Cong Wang et al. [6] proposed a secure cloud storage system supporting privacy-preserving public auditing and further extended to enable the TPA to perform audits for multiple users simultaneously and efficiently.

They considered a cloud data storage service involving three different entities: the cloud user (U), who has large amount of data files to be stored in the cloud; the cloud server (CS), which is managed by the cloud service provider (CSP) to provide data storage service and has significant storage space and computation resources; the third party auditor (TPA), who has expertise and capabilities that cloud users do not have and is trusted to assess the cloud storage service reliability on behalf of the user upon request. To achieve privacy-preserving public auditing, they integrated the homomorphic linear authenticator with random masking technique.

In this proposed system, 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.

***B. Proof of Retrievability***

Proof of Retrievability (POR) mechanism tries to obtain and verify a proof that the data that is stored by a user at remote data storage in the cloud (called cloud storage archives or simply archives) is not modified by the archive and thereby the integrity of the data is assured [13].

The simplest Proof of retrievability (POR) scheme can be made using a keyed hash function. In this scheme the verifier, before archiving the data file F in the cloud storage, computes the cryptographic hash of F and stores this hash as well as the secret key K. To check the integrity of the file F, the verifier releases the secret key K to the cloud archive and asks it to compute and return the value of hash key of file F.

A. Juels et al. [7] proposed an error correction code and spot checking to prove the possession and retrievability of the data. The verifier hides some sentinels among file blocks before sending them to the remote server. When the verifier wants to check retrievability of the data, it only asks the server for those sentinels. In order to keep those sentinels indistinguishable for the remote server, the data owner encrypts the file after adding sentinels. It uses one key regardless of the size of the file. Hence the entire file is not processed, it accesses only parts of file. Therefore the I/O operations are less.

***C. Provable Data Possession***

A Provable Data Possession scheme makes sure that the files in the cloud server are retained. The owner of that data processes his files to store its metadata locally. The file is then uploaded by the owner to cloud and deletes its local copy. The owner checks that his data is contained in the cloud by using the challenge response protocol. In this way, the clients use this integrity proving

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 9, September 2016

technique to verify the integrity of their data and to periodically check its availability.

Ateniese et al. [8] proposed the Provable Data Possession (PDP) scheme to investigate statically the correctness of the data outsourced to cloud storage without retrieving the data.

The proposed model is to check that data stored in a remote server are still in its possession and that the server has the original data without retrieving it. This model is based on probabilistic proofs by randomly choosing a set of blocks from the server to prove the possession. They used a RSA-based homomorphic verifiable tag, which is combines tags in order to provide a message that the client can use to prove that the server has specific block regardless of whether the client has access to this specific block or not. In the case of a prover that is untrusted or has malicious intent, this scheme fails in proofing data possession.

G. Ateniese et.al. [9] overcome the limitation of [8] by using symmetric cryptography. They proposed a PDP scheme that supports partial and dynamic verification. The limitation of this proposition is that it does not support auditability.

Wang et al. [10] proposed a new dynamic PDP for auditing remote dynamic data. They use the Merkle Hash Function (MHT) and the bilinear aggregate signature. They modify Merkle Hash Function structure by sorting leaf nodes of MHK to be from left to right. This sorting will help in identifying the location of the update. However, this method incurs more computation overhead when the file is large.

***D. Proof of Ownership***

Halevi et al. [11] proposed the concept of Proofs of Ownership in remote storage systems and introduced the proof of ownership idea. The ideas behind proving the ownership are the Collision Resistant Hash functions and Merkle Hash Tree. In this proposed scheme, the owner of a file creates a Merkle Hash Tree (MHT) and sends the file to the cloud, called verifier. Once it is received by cloud, the file is divided into bits using pair wise independent hash and then the verifier creates a Merkle Hash Tree for this file.

Once the prover asks for the ownership of the file, the verifier sends a challenge, which is the root and the number of leaves. The prover calculates the sibling path and returns it to verifier as proof of ownership of this file. The verifier after receiving the sibling path checks this path against what the merkle tree has and validates the prover.

However, this violates the privacy of users since their sensitive data is leaked to the remote server. Therefore, there has to be a way to prevent that remote server from accessing outsourced data.

Mainul Mizan et.al [12] proposed the concept of accountable proof of ownership for data using timing element in cloud services. This system uses original data, or data hash as input. The proposed scheme allows a service provider to incorporate timing accountability of data generated at the provider, by requesting proofs from accountability servers in the cloud.

The unique feature of this system design is that the size of the proof is independent of the data size. The scalability of the system has been evaluated using the Amazon EC2.

In the proposed system, clients and servers both keep their own local history for future use which helps keeping load off the servers and at the same time facilitate audit when client looses part of the proof. The limitation of this system is that it doesn't discuss about the, Denial of Service (DoS) attack from the client.

## V. CONCLUSION

Though Cloud computing offers great potential to improve productivity and reduces costs. It also imposes many new security risks which are related to cloud storage. As cloud is mainly used for the storage of the data, data integrity is the main issue to deal with security of cloud data. We mentioned the security issues related to data integrity on clouds and presented the solutions to cloud data integrity issues that protect data seen by the cloud service provider. The study done in this paper will help the researchers to take these issues and solutions into detailed discussion giving them a new view point.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 9, September 2016

## REFERENCES

[1] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 1 – 11, 2011.

[2] Amazon.com, "Amazon Web Services (AWS)," Online at http://aws. amazon.com, 2008.

[3] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing,". In INFOCOM, Proceedings IEEE, pp. 1–9, 2010.

[4] K. Yang and X. Jia, "Data storage auditing service in cloud computing: challenges, methods and opportunities," World Wide Web, vol. 15, no. 4, pp. 409_428, 2012.

[5] B. Balusamy, P. Venkatakrishna, A. Vaidhyanathan, M. Ravikumar, and N. Devi Munisamy, "Enhanced security framework for data integrity using third-party auditing in the cloud system," Advances in Intelligent Systems and Computing. Springer India, vol. 325, pp. 25–31, 2015.

[6] C.Wang, Q.Wang, K. Ren, andW. Lou, "Ensuring data storage security in cloud computing,". In Proceedings of IWQoS'09, pp. 1–9, July 2009.

[7] Juels Jr., A., Kaliski, B.S, "Pors: proofs of retrievability for large files". In Proceedings of the ACM Conference on Computer and Communications Security, pp. 584–597. 2007.

[8] Ateniese, G., Burns, R.C., Curtmola, R., Herring, J., Kissner, L., Peterson, Z.N.J., Song, D.X. "Provable data possession at untrusted stores". In Proceedings of the ACM Conference on Computer and Communications Security, pp. 598–609, 2007.

[9] Ateniese, G., Pietro, R.D., Mancini, L.V., Tsudik, G, "Scalable and efficient provable data possession". In Proceedings of the 4th International Conference on Security and Privacy in Communication Networks, pp. 1–10, 2008.

[10] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing". In Proceedings of ESORICS'09, volume 5789 of LNCS. Springer-Verlag, pp. 355–370, Sep. 2009.

[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems." In Proceedings of the 18th ACM conference on Computer and communications security. ACM, pp. 491–500, 2011.

[12] Mainul Mizan, Md Lutfor Rahman, Rasib Khan, Munirul Haque, Ragib Hasan, "Accountable Proof of Ownership for Data using Timing Element in Cloud Services". In Proceedings of International Conference on High Performance Computing and Simulation (HPCS), 2013

[13] Harshest. K. Raj, "A Survey on Cloud Computing", Internal Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issues 9, July 2014.

[14] W. A. Sultan Aldossary, "Data Security, Privacy, Availability and Integrity in cloud computing". International Journal of Advanced Computer Science and Applications,Vol. 7, No. 4, 2016.