A Survey on Data Integrity Verification Schemes in Cloud Computing

John Abinash Paul
PG Student of Information and Technology,
Department of Computer Science and Technology,
Karunya University, Coimbatore, India.

Mrs. Esther Daniel
Assistant Professor,
Department of Computer Science and Technology,
Karunya University, Coimbatore, India

***Abstract*--Data Integrity has become one of the serious Concerns that has to be addressed in Cloud Environment due to increase in the Amount of distribution of data storage across Internet. Due to the Increase in data storage it becomes necessary to outsource the data storage service of a Cloud Service Provider to a third Party thus creating doubts over loss of Integrity in the outsourced Data. This paper discusses briefly on several Data Integrity Techniques namely, Provable Data Possession in Cloud Computing, Proofs of Retrievability and Third Party Storage Auditing Service.**

Keywords-Cloud,Data Integrity

I. INTRODUCTION

Cloud computing is one of the fastest emerging technologies that has redefined the way of computing and the types of services offered across the internet. One of the advantage is the option of getting Services as required anywhere at any time especially that of data Storage. Hence there is possibility of Storage of Sensitive Information and Hence Data Integrity becomes a Necessity.

There are various challenges which occur in Cloud Storage Service of cloud computing. Some of which are Data Security and billing, reliability, service delivery, performance interoperability and bandwidth. Along with the above challenges mentioned, there is another major Issue that cloud has to deal with. This major Issue comes under Security of Storage namely Data Integrity. Data Integrity enables the Cloud Service Providers of cloud services to Offer Assurance to the users that the Data that is being stored is secure enough from various Integrity Attacks.

There Exist an Environment which integrates Multiple Clouds from Various Services hence named Multi Cloud and managed by Integrated Database system and Virtualization Technologies and tools such as VMWare. Since the Environment becomes more complex there is a possibility for Loss of Data Integrity as the Data is spread across multiple Clouds.

Hence Data Integrity Verification is carried out in Data Level of Cloud Environment.

The End Users who store the Data have no Idea where their Data is being stored. Only thing they can be made aware of is the mechanism that is being used to preserve the Data Integrity of the Data that they store in the Cloud and hence it is mandatory to use a reliable and trusted Mechanism to acquire their trust.

There are several attacks that are carried out on data that is being stored in cloud especially like tampering, Data Forgery and there are threats like tag Forgery in circumstances where even a service Provider in cloud Environment tries to cheat the end users when they just try to verify the data instead of downloading and checking it. It is almost impossible to find out the attack in the latter case. Since Data Integrity becoming such a big threat it is mandatory for an efficient data verification Scheme in the Cloud Scenario and also while implementing such schemes it is important to check the amount of communication and computation cost and other overheads that occurs during implementation of Verification Process.

This paper consists of six sections. Section 2 of this paper deals with the related work which gives a brief introduction of various Integrity Verification techniques in Cloud. Section 3 deals with Data Integrity Verification techniques in which, Provable Data Possession in Cloud, Proofs of Retrievability and Third Party Storage Auditing Service are discussed in detail. Section 4 provides the advantages and disadvantages of Data Integrity Verification techniques. Section 5 provides the Conclusion of the paper. Section 6 provides the References.

II. RELATED WORK

Verifying the Data Integrity is a challenging process as it involves optimization of several metrics. There should not be any performance degradation and there should not be any loss of Usability. Several Researches have been carried out in this Area especially to optimize the Verification Process and also improve Security.

In [1] a model is proposed called as Provable data possession which is a publicly verifiable mechanism which allows not only the Owner but anyone to Challenge the Server as it is Challenge-Response Algorithm and it utilises the Homomorphic Properties.There has been many improvisation provided based on this Mechanism and to support the evolving Multi-Cloud Environment.But it provide a communication cost of order O(1).In [2] a model is proposed called as Proofs of Retrievability which depends on preprocessing the Data by the client before sending the Data or Uploading it.

Some Issues with Updating were overcome in Compact version but it can only optimise to Communication Cost O(t) .

In [3] a model is proposed called as Third-Party Storage Auditing Service which uses Data Fragment Technique and Homomorphic Verifiable Tags to reduce Communication Cost as well as to improve Performance.

The following section gives the detailed explanation on Data Integrity Verification Schemes. The Mechanisms that are included are Provable data possession in Cloud and some of its Variants, Proofs of Retrievability and Third-Party Storage Auditing Service

III. DATA INTEGRITY VERIFICATION SCHEMES

This Paper does not propose any new Schemes or any Architecture. It just surveys on Existing Schemes on Data Integrity in Cloud Environment. This Section provides the working of Several Integrity Verification Schemes proposed based on Researches carried out.

The Major three approaches of data Integrity that are included in this study are

- Provable data possession (PDP)
- Proofs of Retrievability (POR)
- Third-Party Storage Auditing Service (TSAS)

A. PROVABLE DATA POSSESSION (PDP) SCHEMES:

Provable Data possession is the mechanism of ensuring the Integrity of the Data when it is being outsourced to a third party as Data Storage Service. It enables the option of checking the Integrity of Uploaded data without obtaining the Entire Stored Data from the Server which is useful on the circumstance that there has been a large Data stored on the server. It was introduced as an alternate for the traditional signatures and Hash functions .It is also based on the challenge – Response concept and hence challenges are mandatory for Server to [1].

→ Sampling PDP and Efficient PDP:

These PDP Schemes are based on Homomorphic Verifiable Tags (HVT) and Homomorphic Linear Authenticators (HLA) [4].
In this Method Tags are allocated to File Blocks as the Files are stored in blocks in Storage Server. The Cumulative Sum of all the Blocks is used as to verify the Integrity of the Data Stored.

Sampling PDP and Efficient PDP Schemes carries out Verification based on the KEA1 assumption (Knowledge of Exponent Assumption) where if an untrusted Server stores the data uploaded by a Client, Error is reported while Audition of data takes place. Data Possession is guaranteed only as a whole block which incurs more Cost.

→ Scalable PDP:

This PDP Scheme is based on the cryptographic Technique which uses symmetric key [5].
The Data Owner Pre-Computes several tokens for a set of blocks of data before handing the Data over to the Server. Verification Scheme is done based on the challenge issued by server for random blocks of Data and the Server in turn should carry out the check just on the specified data blocks only hence minimizing time and cost for large Data Blocks

→ Dynamic PDP:

This PDP Scheme is used to address the problem when a user tries to update a data block like modification of Data or any other operations carried out on the Data [6]. When the User inserts some difference in data then the previous approach causes some inconvenience.
Dynamic PDP uses Authenticated Dictionary and also utilizes a slight Variant where rank information is used for organizing the entries which is essential for authenticated Insert or Delete.
Another Extension of Dynamic PDP known as DPDP-II uses RSA tree for authentication hence increased chance of Detection but there is more time consumption for update Operation.

→ Basic Multi Copy PDP:

This PDP Scheme has a different Concept of making copies of the Data and generating Different Keys for each copy of Data and the generated keys are kept as secret from the cloud Service Provider [7]. Hence it disables the possibility of any forgery that a cloud Service Provider can cause.
In This Scheme the Client can verify the possibility of any Integrity breach by challenging each copy of the Data that was created using any existing PDP Schemes .Hence it can be used as an extension for any PDP Scheme

→ Distributed and Replicated PDP:

This PDP Scheme is used to attain data replication over more than one server and hence distribution. This is achieved by means of using one of the Cloud Service Providers as Organizer which co-ordinates the communication between servers. Hence it supports multi-cloud storage [8].The Organizer performs only load balancing and no group or disk operations which can prove to be expensive. Hence it is feasible to use such an option .Also Replication of Organizer is made possible and essential to avoid failures due to excess Loads. This PDP does not allow Servers to communicate with each other in the absence of an Organizer.
The Computation done by Organizer is larger when compared to Server hence it might create some extra Cost when verification is done in a Multi-Cloud Environment.

→ Co-Operative PDP:

This PDP Scheme uses Hash Index Hierarchy and Homomorphic Verifiable Response for implementation of integrity verification in Hybrid Clouds. It hides the location of Data Storage but this method reduces communication bandwidth [9].
The Previous approaches did not support fully multi cloud environment which provided added advantage for this PDP Scheme.

Hash Index Hierarchy uses a collision resistant Hash Function which combines multiple responses from various Cloud Service Providers and Homomorphic Verifiable Response provides the necessary response for the Data Owner from the Client. Hence this is also considered to have Multi Prover Zero Knowledge Proof System (MP-ZKPS), a type of Interactive Proof System (IPS).

The Properties that this PDP Scheme satisfies include completeness, soundness, and zero-knowledge. This Scheme is Fool-Proof against data leakage and Tag Forgery attack.

→ Pairing based Provable Multi Copy Data Possession (PB-PMDP):

This PDP Scheme uses Verification method that can be any one not only the Data Owner. Users can access the copies anytime, anywhere without any constraint [10]. But the Copies that are created should be made different from each other .It uses the diffusion property of the PDP Schemes in order to implement the above, thus disabling the Cloud Service Provider of Cheating the Data Owner which might end up showing that it stored multiple copies when only one copy exists

B. PROOFS OF RETRIEVABILITY (POR) SCHEMES:

A Proof of retrievability is similar scheme to that of Provable Data Possession ,It provides the proof that a file is Intact and not modified by any attack [11].This helps more in defining the existence of data than that of Integrity (i.e.) Helps more in Checking the full Existence of Data .Hence it is gives the proof of Existence.

They consume less bandwidth than the file itself and hence can be used in remote environment.

The main feature that occurs in this Scheme is that they can correct any Data Corruptions that is found by using Error Correction codes

→ Compact Proofs of Retrievability:

Compact Proofs of Retrievability uses Homomorphic Property to reduce the size of authenticator value and hence reduces the computational cost.

This Scheme makes use of blocks called as Sentinel Blocks which are randomly inserted to detect data corruption in the Data that was uploaded by Client.

After that the Data Error codes can be used to recover the data from corruption. Encrypted File Verification is being carried out in this Scheme and queries are limited to certain extent. Two Types of CPOR are

- Compact Proofs of Retrievability-I
- Compact Proofs of Retrievability-II

The Difference that differentiates the two schemes is that the construction to support Verification option, the former offers Public Verifiability where anyone not only the data owner can verify the integrity of the Uploaded Data and the latter provides private Verifiability which grants verifiability option only to the Data Owner.

C. THIRD-PARTY STORAGE AUDITING SERVICE (TSAS):

This Scheme is also used for checking data Integrity in the cloud Environment. The property used in this mechanism is Bilinearity Property. This Property is used to create a proof system that is encrypted along with a stamp that helps the data owner to challenge the Cloud Server. Hence Data Privacy is managed in such a way that a Third Party Auditor cannot decrypt the message while auditing. Also it does not require any Organizer while doing auditing on Multi-Cloud environment .Also This Scheme enables the server to check the value sometimes so that can be used by Auditor thus reducing the load on the Auditor itself thus load balancing is achieved so that more efficiency is achieved [3].

TSAS consists of five Algorithms clubbed together that is Generation of Key, Generation of tags, Challenge issued by Owner, Proof given by the Cloud Server, Verification Algorithm is carried out in order to verify the integrity of the data.

In this Scheme, the Data Fragment Technique and Homomorphic Verifiable Tags are utilized in order to improve the performance. Using the Homomorphic verifiable tags, enables the server only to respond to the auditor, the sum of data blocks and product of tags irrespective of number of Data Blocks challenged. It results in considerable reduction of the cost incurred in communication.

The data fragment Technique is capable of reducing number of data tags, hence it results in minimizing the overhead that occurs in storage and the system performance is greatly improved.

IV. ADVANTAGES AND LIMITATIONS:

Advantages:

- PDP Schemes reduces the Server and client computation cost to greater Extent.
- Sampling PDP gives stronger guarantee of Data Possession.
- Flexibility to Data Changes allowed in Dynamic PDP
- Proofs of Retrievability assures the Data Retrieval
- Proofs of Retrievability corrects data corruptions in the case of occurrence in Client Data.
- Compact Proofs of Retrievability use shorter queries and response time.
- Multi Copy PDP Schemes increases Data Possession.
- TSAS reduces load on the Cloud Service Provider
- TSAS supports both Multi-Owner and Multi-Cloud Environment.

Limitations:

- Sampling PDP can support only Static Data.
- Except for TSAS and CPDP others does not support Multi-cloud Storage.
- There is always tradeoff between computation cost and communication cost.

- The best normalized efficiency achieved with all features enabled is in TSAS in order O(t)

V. CONCLUSION

Cloud computing has emerged as a technology that has redefined the way of computing and the types of services offered across the internet. There are several major concerns that occur in cloud Environment especially Data Integrity, This Paper carried a study on various data integrity Techniques in Cloud Environment. Also Several Variants of these Techniques have been studied Hence this Paper Surveyed existing Techniques to initiate Future Research in this Area of Data Integrity to enhance Cloud Security.

REFERENCES

[1] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner,Z. N. J. Peterson, and D. X. Song, "Provable data possession at untrusted stores," in ACM Conference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F.Syverson, Eds. ACM, 2007, pp. 598–609.

[2] A. Juels and B. S. K. Jr., "Pors: proofs of retrievability for large files," in ACMConference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds.ACM, 2007, pp. 584–597.

[3] Kan Yang and Xiaohua Jia.,"TSAS: Third-Party Storage Auditing Service" in Security for Cloud Storage Systems,SpringerBriefs in Computer Science 2014, pp 7-37

[4] G. Ateniese, R. Burns, R. Curtmola, J.Herring, L. Kissner, Z. Peterson, and D.Song, "Provable data possession at untrusted stores," in CCS '07: Proceedings of the 14th ACM Conference on Computer and Communications Security, New York, NY,USA, 2007, pp. 598–609.

[5] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Secure Comm '08: Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, New York, NY, USA, 2008, pp. 1–10.

[6] Q. Wang, C. Wang, J. Li, K. Ren, and W.Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in ESORICS'09: Proceedings of the 14th European Conference on Research in Computer Security, Berlin, Heidelberg, 2009, pp. 355–370.

[7] Ayad F.Barsoum and M.Anwar Hasan Provable Possession and Replication of Data over Cloud Servers.

[8] Mohammad Etemad and Alptekin Koc University,Istanbul,Turkey."Transparent,Distributed, and Replicated Dynamic Provable Data Possession".

[9] Zhu,Y.,Hu, H.,Ahn, G.,Yu, M.: Cooperative provable data possession for integrity verification in multi-cloud storage. IEEE Trans. Parallel Distrib. Syst. 23(12) 2231–2244 (2012).

[10] Ayad F. Barsoum, M. Anwar Hasan. "Integrity Verification of Multiple Data Copies over Untrusted Cloud Servers".

[11] A. Juels and B. S. K. Jr., "Pors: proofs of retrievability for large files," in ACMConference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds.ACM, 2007, pp. 584–597.