International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 05, May-2015

A Literature Survey on Integrity Verification Techniques

[1.] Anusha Priya.G,
[1.] Post-GraduateStudent,
Department of Information and Technology,
Karunya University,India

[2.] Mrs. Esther Daniel
[2.] Assistant Professor,
Department of Information Technology (IT),
Karunya University, India

Abstract:- **Increase in the use of cloud has lead to number of data integrity (correctness of data) and security issues. User data's integrity in the cloud servers is the most important concerns of users now a days. The data integrity proofs the consistency regularity and validity of the data. It is a secure way for writing the data in the persistent data storage which can be retrieved in the same layout as it was stored later. Cloud storage therefore becoming popular for day-to-day management of outsourcing data.So integrity monitoring of the data in the cloud is also very important to escape all possibilities of data corruption and data crash.In this paper we are going to analyze different methodologies and protocols, which the users can use to check the correctness of their data with the simplest possible way and less overhead at the customer side and to overcome the challenges faced by cloud servers for the security and integrity of users data.**

Keywords: Cloud Computing, Data Integrity, Cloud storage.

1. INTRODUCTION

Cloud computing is an advanced technology in which it uses high speed internet-based computing by which user can access their resource from the remote site. Providers ensure that all critical data are encrypted and only authorized users have access to data in its entirety. The digital identities and credentials must be protected as should any data that the provider collects or produces about customer activity in the cloud. With data storage and sharing services (such as Dropbox and Google Drive) provided by the cloud, people can easily work together as a group by sharing data with each other. Once a user creates shared data in the cloud, every user in the group is able to not only access and to modify the shared data, but also share the latest version of the shared data with the rest of the group. But there is no guarantee that data stored in the cloud is secured and not altered by the cloud or Third Party Auditor (TPA). In order to overcome the threat of integrity of data, the user must be able to use the assist of a TPA.

Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life cycle and it is a difficult aspect to the implementation, design and the usage of any system which stores the data, processes, or retrieves data. Data Integrity is given much more importance among the other cloud storage issues because only data integrity ensures that data is of high and good quality, correct, consistent and accessible. After moving the data to the cloud the owner of the data hopes that their data and applications are in secured manner. But that hope may fail some times (i.e.) the owner's data may be altered or deleted. In that case, it is important to verify if one's data has been modified or deleted. To validate data, often a user must download the data. If the outsourced data is very large files, downloading to determine data integrity may become prohibitive or costly in terms of increased cost of bandwidth and time, especially if the data is to be checked frequently. If the owner wants to check the data integrity, he need to access the entire file so it's expensive to the cloud server. Also transmitting the file across a network may consume high bandwidth. It's further complicated for the owner of the data whose devices like Personnel Digital Assist and mobile phones. Because these devices can have only a limited amount of battery power, CPU power, storage capacity and communication bandwidth. Owner can check over the data integrity by enabling a new role which is TPA because it possesses experience capabilities that the customer does not.

The need for security in cloud is to provide security for the shared data. Third Party Auditors can understand the threats and they know how to deal with the identified threats. The TPA will be able to verify over any threats in online storage services that are represented in the cloud server. Thus, the user who owns the data can depend on the TPA to verify the data in the cloud without involving with the procedure. Encryption can also be used for privacy of the data in which the basic idea is based on scrambling the information that only the one who have the secret key can expose it by decryption. The encryption only will not be enough to ensure the data integrity over the cloud. Sometimes TPA may modify file and upload it in cloud again.

2. CLOUD STORAGE SECURITY ISSUES AND DATA INTEGRITY TYPES

There are several number of security issues for cloud asit deals with many technologies. Data security explains encryption ofthe data as well as ensuring the appropriate policies for enforced data sharing. Data storage in the cloud also requires particular consideration. Many industries dealing with cloud may operate and generate particular issues with cloud-based data storage. Strong cryptographic protection is very much essential, no matter the

information is at rest or in transit. Some of the challenges faced in cloud storage are:

a. Data Leakage

Because of insufficient authentication, authorization, and audit controls, such as deletion or alteration of records without a backup of the original content many ways had compromised. Effective destruction may cause because of the Loss of an encoding key. This results to the gain of access to sensitive data by the Unauthorized parties. The data might be deletedby the hacker.

b. Account Hijacking

If an attacker gains access to credentials, he can eavesdrop on user activities and transactions, manipulate data, falsify information, and redirect your clients to illegal sites.

c. Malicious insiders

A provider may not reveal how it allows employee's access to physical and virtual assets, how it monitors these employees, or how it analyzes. In cloud computing, the organization doesn't need to know the technical details of how the services are delivered. In situations, the risk is great. Without full knowledge and control, your organization may be at risk. In situations, the risk is great. Without full knowledge and control, your organization may be at risk.

d. Unknown risk Profile

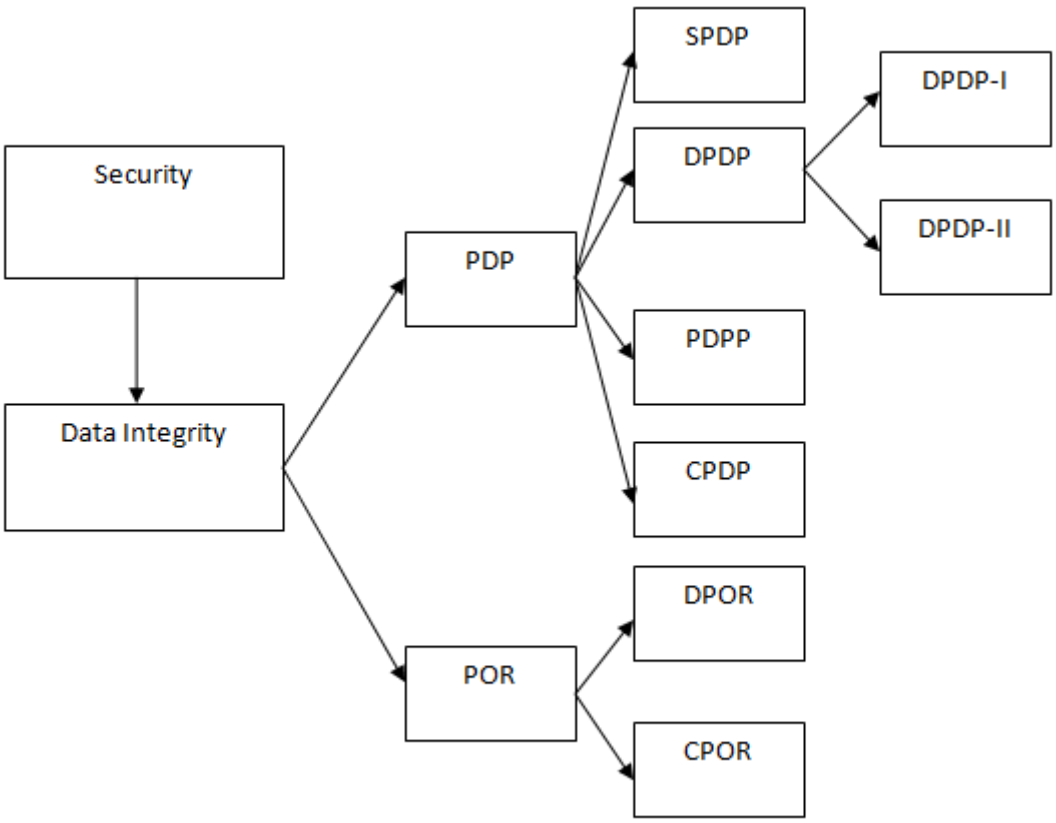
Versions of software, code modifications, security policies and applications, vulnerability reports, interference attempts, and security design, are all important factors for estimating company's security status. Information about who is sharing your infrastructure may be relevant.

e. Data breaches

The most important thing is to prevent any data violation. The challenge addressing the threats of data loss and data leakage is that "the measures you put in place to improve one can worsen the other". Data is encrypted to reduce the impact of a violation, but if the encryption key is lost, then data will be lost. However, if offline backups of data are chosen to reduce data loss, exposure data breaches are increased.

f. Types of Data Integrity

Data integrity means completeness or wholeness and it is basic requirement of information technology. Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle**.**Data integrity ensures the data is the same as it was when it was originally recorded.Data integrity can be roughly divided into two overlapping categories Physical integrity and logical integrity Physical integrity deals with challenges related to storing and fetching of the data. Challenges for the physical integrity may include electromechanical faults, design flaws, material fatigue, corrosion, power outages, natural disasters, acts of war and terrorism. Physical integrity makes use of error detecting algorithms known as error correcting codes. Logical integrity is related with the correctness or rationality of a piece of data.



Types of Security Measures

3. OVERVIEW OF INTEGRITY VERIFICATION TECHNIQUES

Clients make the Integrity verification to make sure that their data has been properly stored and maintained in the third party auditor. The overview of various integrity verification techniques are summarized as follows:

3.1 Provable Data Possession Techniques

Proving the integrity of data stored at untrusted servers in resource-sharing networks is more important. In PDP [1], explains to check the correctness of the outsourced data statically in the cloud storage without having to retrieve the data. They have used homomorphic verifiable tags based on RSA to combine and to build a proof message that permits the client to check whether the server has specific blocks, even if the client has no access to the blocks. In SPDP [2] It provides secured data in encrypted form by using symmetric cryptographic key and also allows public verifiability. It provides efficient PDP by encryption and it is light weight PDP scheme to support homomorphic hash function but it lacks in randomness hence by using the previous challenges, client can cheat the server very easily.

In DPDP (I) [3] this scheme supports dynamic updates in each blocks which allows block modification.It use authenticated dictionaries based on rank list. Block modification and updation of block is allowed and efficient integrity verification is made by querying and updating DPDP scenario though it provides efficient verification but construction of rank based scheme is difficult .In DPDP (II)[3] It also supports dynamic updation of data with blockless verification scheme in which entire data need not to be download. Blockless verification where particular block can be queried for integrity verification and RSA trees use homomorphic tag where tag are small and easy to use these schemes with RSA tree construction is efficient with dynamic option but it cannot be adapted to the multi-cloud.

In Flex DPDP[4] it uses homomorphic verifiable tags as in DPDP [3] multiple tags can be combined to obtain a single tag that corresponds to combined blocks [1].These tags are small compared to data blocks, enabling storage in memory. Authenticity of the skip list makes sure of the tags integrity, and the integrity of the data blocks is protected by the tags. In [5] cooperative PDP for integrity verification in multi-cloud storage uses homomorphic verifiable response and hash index hierarchy and supports dynamic scalability on multiple storage servers. It can resist various attacks even if it is deployed as a public audit service in clouds and only a small amount of computation and communication overheads is introduced.In [6] trapdoor commitment scheme is used fordata auditing in cloud. This approach greatly reduces the security related issues and a key is generated using RSA algorithm which can be obtained by the Third Party Auditor (TPA) only by using a trapdoor commitment.

In [7] Dynamic audit services for integrity verification of outsourced storages in clouds used for verifying the integrity of untrusted and outsourced storage it supports dynamic data operations and timely abnormal detection with the help of several effective techniques, such as random sampling,fragment structure, and index-hash table. It has lower computation overhead , as well as a shorter extra storage for integrity verification. In this paper [8] Towards Efficient Provable Data Possession in Cloud Storage a PDP scheme EPOS is proposed which is very efficient in communication, storage and computation. Compared to [1], POS is much more efficient in computation (400 times faster in setup), and equally efficient in communication and storage.

In [9] Provable data possession (PDP) schemes provide data format independence, which is a similar kind of feature in the practical deployments also, and have no restriction on the number of times the client can challenge the server to prove data possession. Also, a variant of our main PDP scheme offers public verifiability. In [10] Based on the bilinear pairing technique proxy provable data possession in public clouds is designed and it plays a vital role when the client cannot perform the remote data possession checking and it secure and efficient. In Privacy-Preserving public auditing mechanism [11] it supports public auditing on shared data stored in the cloud and performs multiple auditing tasks simultaneously instead of verifying them one by one.

3.2 Proof of Retrievability Techniques

A malicious storage provider might even choose to delete rarely accessed files to save money. To assure such concerns, a simple auditing procedure for clients to verify that their data is stored correctly. Such audits, called Proofs of Retrievability. It describes the preprocessing steps client should do before uploading the data to provider server by using Message Authentication Code and it is the simple way to audit the server.In Compact Proofs of Retrievability [12] improves the efficiency and security of the original POR based on the data fragmentation concept this technique uses homomorphic property to aggregate a proof into authenticator. It gives dynamic cost by integrating the Compact POR scheme and Merkle Hash Tree (MHT) into the DPDP. Homomorphic Verifiable Response from provider gives proof of stored data intact and it provides authenticated proof value but their solution is also static and could not prevent the leakage of data blocks during the verification process.

In [13] POR for large files an auditor has the capacity to recover and mitigate corruption of data by using forward error-correcting codes when data is stored in untrusted cloud. In order to achieve this goal, the data owner needs using a one-way function creates a set of sentinel blocks and inserts the sentinels randomly on the data blocks before uploading to the server. If the server wants to modify even a small amount of the data in the file, the verifier can easily

find it and check the integrity of a file due to the effect of file modification on the sentinels. However, the number of queries in these kind of method depends on the number of inserted sentinel blocks. Moreover, the Proof of retrievability method incurs high computation overhead on the client side because of the error recovery and data encryption processes.

In [14] Towards secure and dependable storage services in cloud computing a flexible distributed storage integrity auditing mechanism is used. It allows users to audit the cloud storage with very lightweight communication and computation cost and ensures strong cloud storage correctness guarantee, and achieves fast data error localization. It is highly efficient and resilient against byzantine failure, malicious data modification attack, and even server colluding attacks. In [15] an efficient and secure dynamic auditing protocol for data storage in cloud computing privacy-preserving auditing protocol is used. It ensures the data privacy by using cryptography method and the bilinearity property of the bilinear pairing, instead of using the mask technique but the data content may sometimes get leaked to the auditor.

The majority of POR methods failed to efficiently support dynamic data update because the server is unable to realize the relation between the data blocks and encrypted code words DPOR via oblivious RAM [16] is the first technique to overcome this issue the client can execute an efficient audit protocol to ensure that the server maintains the latest version of the client data. The main advantage is it incurs high computation overhead on the client and server side.In POR with public verifiability and constant communication cost in cloud [17] allows public verification andreleases the data owners from the burden of staying online and it doesen't have a thirdparty auditor. This scheme achieves constant communication size, efficient computation performance as well as low storage overhead by supporting the public verifiability as well.

In Proofs of Retrievability via Hardness Amplification[18] it allows the client to store a file on an untrusted server, and after some time it runs an efficient audit protocol in which the server proves that it still possesses the client's data. Constructions of Proof of retrievability schemes[12] attempt to minimize the storage of client and server ,audits communication complexity, and the number of file-blocks accessed by the server during the audit. In [19] Towards efficient proofs of retrievability in cloud storage is based on Strong diffie-hellman assumption where efficient POR scheme with private verifiability is designed. The proposed scheme needs only linear communication bits with respect to the security parameter per verification. In Table [1] we provide a comparison of several schemes different factors included are methodology and the protocol used,description of the scheme, merits and demerits etc.

Table 1

S.NO	Title	Methodology and protocol Name	Description	Merits	Demerits
1	Provable Data Possession at Untrusted Stores [1]	PDP scheme is used.	• It allows client to verify the server that possess their data without downloading the actual data by using homomorphic Verifiable Tag.	• It provides security to data based on RSA scheme. • It allow public verifiability in which access privilege can be set in cloud..	• It is more efficient scheme but can applicable only for static files. • It is insecure against dynamic block of data
2	Scalable and Efficient Provable Data Possession. [2]	PDP,MHT	• It provides secured data in encrypted form by using symmetric cryptographic key and also allows public verifiability.	• It provides efficient PDP by encryption. • It is light weight PDP scheme to support Homomorphic hash function.	• It lacks in randomness hence by using the previous challenge, client can easily deceive the server.
3	Dynamic Provable Data Possession. (DPDP-I)[3]	Authenticate d Skip List	• It supports dynamic updates in each blocks which allows block modification .It use authenticated dictionaries based on rank list.	• Block modification and updation of block is allowed. • Efficient integrity verification is made by querying and updating DPDP scenario	• It provides efficient verification but construction of rank based scheme is difficult.

4	DPDP-II[3]	RSA Trees	• Italso supports dynamic updation of data with blockless verification scheme in which entire data need not to be download.	• Blockless verification where particular block can be queried for integrity verification. • RSA trees use Homomorphic tag where tag are small and easy to use.	• DPDP scheme with RSA tree construction is efficient with dynamic option but it cannot be adapt to the multi-cloud.
5	An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing[15]	privacy-preserving auditing protocol is used.	• Ensures the data privacy by using cryptography method and the Bilinearity property of the bilinear pairing, instead of using the mask technique.	• Less communication cost between the auditor and the server. • Reduces the computing loads of the auditor by moving it to the server.	• Data content may sometimes get leaked to the auditor.
6	POR (Proof of retrievability) for large files.[13]	MAC	• It describes the preprocessing steps client should do before uploading the data to provider server by using Message Authentication Code.	• Preprocessing steps can be made by client before storing their data. • It is the simple way to audit the server.	• It is difficult to build the system for e-client probably with secured data during the audit.
7	CPOR (Compact POR)[12]	HVR	• This technique which uses homomorphic property to aggregate a proof into authenticator. • It gives dynamic cost by integrating the Compact POR scheme and Merkle Hash Tree (MHT) into the DPDP.	• Dynamic cost of data provides more flexibility to user. • Homomorphic Verifiable Response from provider gives proof of stored data intact.	• It provide authenticated proof value but their solution is also static and could not prevent the leakage of data blocks during the verification process.
8	Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds.[7]	Dynamic audit services are used for verifying the integrity of untrusted and outsourced storage.	• Supports dynamic data operations and timely abnormal detection . • With the help of several effective techniques, such as fragment structure, random sampling, and index-hash table.	• Lower computation overhead, as well as a shorter extra storage for integrity verification.	• Less frequent activities may not detect anomalies in a timely manner.
9	Towards Secure and Dependable Storage Services in Cloud Computing.[14]	A flexible distributed storage integrity auditing mechanismis used.	• Ensures strong cloud storage correctness guarantee, and achieves fast data error localization.	• Highly efficient and resilient against Byzantine failure, malicious data modification attack, and even server colluding attacks.	• Though the benefits are clear new security risks may arise towards the correctness of the data in cloud.

10	POR with Public Verifiability and Constant Communication Cost in Cloud.[17]	POR scheme is used.	• Allows public verification and releases the data owners from the burden of staying online and it doesen't have a thirdparty auditor.	• This scheme achieves constant communication size, efficient computation performance as well as low storageoverhead is achieved by supporting the public verifiability as well.	• Low scalability is the main drawback.

In table [2] we provide a comparison of several schemes. Different factors are considered like single servers, static/dynamic, supports batch auditing, public auditing, etc.

Table [2]

S.NO	Scheme	Static/dynamic	Supports batch auditing	Public Auditing	Single Server
1	Scalable PDP[2]	Dynamic	No	No	Yes
2	PDP(Provable Data Possession)[1]	Static	No	Yes	Yes
3	DPDP-I (Dynamic PDP –I)[3]	Dynamic	No	No	Yes
4	DPDP-II (Dynamic PDP – II)[3]	Dynamic	No	No	Yes
5	POR (Proof of retrievability)[13]	Static	No	No	No
6	CPOR (Compact POR) [12]	Dynamic	No	No	No
7	Flex DPDP[4]	Dynamic	No	No	Yes

Table[1] and Table [2] explains different schemes (which includes brief explanation of the protocol its working, advantages and disadvantages) and whether this particular schemes support batch auditing, public auditing, it is static or dynamic and whether it has single or multiple server is discussed above.

4. CONCLUSION

In this paper we tried to cover some of the important schemes of integrity verification techniques for secure storing of data in the cloud servers. Byzantine failure is one of the main reasons of corrupting users data. Due to this failure the servers begin to behave improperly. So, if the data is distributed on multiple servers in the cloud it will have good availability and reliability. So that if one server fails to respond then data is available on other servers to respond to users queries/requests. This achieves constant communication size, efficient computation performance as well as low storage overhead and supports public verifiability schemes.

REFERENCES

[1]Ateniese.G and Randal.B"Provable Data Possession at Untrusted Stores", in proceedings of the 14th IEEE/ACM conference on Computer and communications security, pp.3,November 2007.

[2]Ateniese.G and Di Pietro.R"Scalable and Efficient Provable Data Possession", in Proceedings of the 4thIEEE/ACM conference on Security and privacy in communication netowrks,pp.3,September 2008.

[3]Alptekin.K andChris Erway.C"Dynamic Provable Data Possession", in Proceedings of the 16th ACM conference on Computer and communications security, pp.3, November 2009.

[4]Adilet.K andErtem.E , "FlexDPDP: FlexList-based Optimized Dynamic Provable Data Possession",In proceedings of International journal of Engineering Reserch and technology pp.3,March 2013.

[5] Gail-Joon.A and Yan.Z"Cooperative PDP for integrity verification in multi-cloud storage",IEEE transaction, in proceedings of IEEE transactions on parallel and distributed systems, pp.3, December 2012.

[6]Vijini.M andRoshniThanka.M "Trapdoor commitment scheme fordata auditing in cloud",In proceedings of International journal of Engineering Reserch and technology pp.4,March 2013.

[7]Yan.Z and Wang.H"Dynamic audit services for integrity verification of outsourced storages in cloud",in proceedings of 2011 ACM conference on applied computing, pp. 4, March 2011.

[8]Chien.Eand Jia.X"Towards Efficient Provable Data Possession in Cloud Storage", in proceedings of ACM conference on Computer and communications security,pp.4,2011.

[9]Ateniese.G and Randal.B"Remote Data Checking Using Provable Data Possession",in proceedings of the ACM Transactions on Information and System security, pp.4,May 2011.

[10]HuaqunWang "Proxy provable data possession in public clouds",in proceedings of theIEEE Transactions on services computing ,Vol. 6,pp.4,2013.

[11]Li.B and Wang.B"Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud", in proceedings of the IEEE transactions on Cloud computing,Vol.2,pp.4, January-March 2014.

[12]Shacham.H and Waters.B"Compact Proofs of Retrievability",inProceedings of the 14thInternational Conference on the Theory and Application of Cryptology and Information Security: Advances in CryptologyACM conference, pp.4, February 2008.

[13] Ari Juels and Burton S"Proof of Retrievability for large files",in Proceedings of the 14th ACM conference on Computer and communications security, pp.4,2007.

[14] Wang.C and Ren.K "Towards secure and dependable storage services in cloud computing", In proceedings of the IEEE Transactions on service computing, Vol. 5,pp.4,April-June 2012.

[15] Jia.X and Kan Yang "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing",Inproceedins of IEEE Transactionson parallel and distributed systems,Vol. 24,pp.4, September 2013.

[16]Cash.D and Daniel.W "DPOR via Oblivious RAM", in proceedings of32nd Annual Springer International Conference on the Theory and Applications of Cryptographic Techniques, pp.5, 2013.

[17] Jiawei.Y "POR with Public Verifiability and Constant Communication Cost in Cloud",in proceedings of the ACM 2013 international workshop on Security in cloud computing, pp.5, May 2013.

[18] Dodis.Y and Daniel.Y"Proofs of Retrievability via Hardness Amplification",in Proceedings of the 6th Theory of Cryptography ACM Conference on Theory of Cryptography,pp.5, 2009.

[19]Chien.Eand Jia.X "Towards Efficient Proofs of Retrievability" in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security,pp.5,2012.

.