A Survey on the Methods for Security, Integrity and Privacy Preserving For Data Cloud Storage

L.Sara Anantha Kumari[1], Suma Sira Jacob[2]

[1]PG Scholar, Department of Computer Science and Engineering, Christian College of Engineering and Technology, Dindigul, Tamilnadu - 624619, India.
saralawrence117@gmail.com

[2]Assistant Professor, Department of Computer Science and Engineering, Christian College of Engineering and Technology, Dindigul, Tamilnadu - 624619, India.
sumasarajacob@gmail.com

Abstract— **Data Storage in cloud via an internet connection is easy which provides easy access of the stored data without any underlying architectures. But the data stored in the cloud must be highly secured so that only the data owner or the member of the group can view, update and save back the data. To ensure the security of the stored data, two properties must be satisfied viz. integrity and privacy. Here a survey of the available methods for providing data integrity and privacy preservation is presented so as to find a possible way to improve the cloud data security.**

Keywords— **Cloud computing, security, auditing, privacy.**

I. INTRODUCTION

Cloud computing is the concept of using remote services through a network using various resources. At present, Security and Privacy issues are major concern to cloud service providers who are responsible for hosting the services. In such cases the cloud provider must guarantee that their infrastructure is secure and clients' applications and data are safe by implementing security policies and mechanisms. The customers must be satisfied; that the cloud provider has taken proper security measures to protect their information. Cloud security issues such as physical security transmission security, storage security, access security, application security and data security The ultimate challenge in cloud is, not the cloud computing provider rather data-level security, and sensitive data in the enterprise domain. Security needs to move in to data level so that enterprises would be sure that its data is protected wherever it goes. It can force encryption of particular type of data and permit only specified users to access those data. Although data stored in the cloud server are encrypted those encrypted data are very likely to be vulnerable to attacks and business interests become compromised once the server is invaded. [1].

The data's are stored at a remote location on servers and maintained by a cloud provider. This approach is to store encrypted data on storage provider's site and data must be encrypted and decrypted by some other service provider. The storage provider's administrator is not aware of the decryption key and the data will be secured. [3].

II. CLOUD COMPUTING SERVICE MODELS

The three most widely used service models of cloud computing are described below.

Software-as-a-Service (SaaS)

SaaS is a model of software deployment in which applications are licensed for use as a service provided on demand of customers. On-demand licensing and use relieves the customer on the burden of equipping a device with every application to be used. It is a software delivery for business applications like content delivery, accounting, Human resource management (HRM), Enterprise resource planning (ERP) etc on demand on pay-as-you go model Mohit Marwaha et al. [2].

Platform-as-a-Services (PaaS)

PaaS is an outgrowth of the SaaS application delivery model. With the PaaS model, all the facilities that are required to support the complete lifecycle which includes development and delivery of web applications and services are available to developers. PaaS is also known as "cloud ware". It provides the facilities for application development, design, testing, and deployment and application services such as web service integration, database integration, security, scalability, storage, persistence, and application versioning and developer community facilitation. These services are provided as an combined solution over the web Mohit Marwaha et al. [2].

Infrastructure-as-a-Services (IaaS)

IaaS is the delivery of computer infrastructure as a service. These virtual infrastructure stacks comes under the category of everything-as-a-service trend and share many of the characteristics. Rather than purchasing hardware infrastructures, clients buy these resources as a fully outsourced service. The service is billed individually on the basis of utility and the quantity of resources consumed Mohit Marwaha et al. [2].

III. CLOUD DEPLOYMENT MODELS

A cloud infrastructure may be operated in one of the following deployment models: public cloud, private cloud, community cloud, or hybrid cloud. The differences are based on how exclusive the computing resources are made to a cloud customer [4].

Public Cloud

A public cloud is the one in which the cloud infrastructure and resources for computing are made available over a public network. A public cloud is owned by an organization who will be selling cloud services to various groups of clients.

Private Cloud

A private cloud gives the exclusive access, usage of the computational resources and infrastructure to every individual cloud consumer's organization, which may be hosted on the organization's premises or outsourced to the hosting company

Community Cloud

A community cloud provides service to a group of cloud consumers who have shared their data such as objectives on mission, security, privacy and compliance on policy. Alike private clouds, a community cloud may be a managed organizations or third party which can be implemented on customer premise or outsourced to hosting company.

Hybrid Cloud

A hybrid cloud is the composition or combination of two or more clouds. That remains as distinct entities but still they are bound together by standardized technology that enables data and application portability.

IV. SECURITY ISSUES IN CLOUD

A. Data Security

Security for the data in the cloud is mandatory to prevent the access of data by unauthorized users who tries to bypass security check or gain access through some other ways.

For the first layer of the security, RSA algorithm is used to store the data in the cloud. RSA is a public key algorithm that works with a secret key and a shared key. RSA algorithm can also be used for signing the data, when a particular user writes back an available data.

Apart from security given to the individual data, more concentration is given to the adversary models. Security measures for adversary models differ depending on the nature whether it is weak or strong.

For a weak adversary model, the probability for data modification and probability for misbehaving servers are found and counter measures are taken to minimize or nullify the effect of adversary nodes [8].

For a strong adversary model, the security measure has to be taken by knowing the pattern in which the data is stored in the cloud. Data in the cloud is stored by a method called file distribution preparation which uses the file matrix(F) and the parity matrix (P). Hackers try to find P so that they can find F [8].

To overcome this problem, the encoded file matrix is added with random perturbations and thereby hiding P. By this method actually a known noise is added to the resulting data matrix after encoding with the parity matrix which makes the hacker tough to get the parity matrix and also the data matrix [8].

A cloud storage is a place where there will be plenty of users each having their own data. Some of the misbehaving users try to steal or delete or modify the available data without the knowledge of the data owner who really stores the data. So as to overcome this problem, security has to be implemented in several ways. All the security measures done are basically divided as two major categories.

i. Security for the cloud

Whenever the service is provided, there will be hackers trying to get access of the services without proper permission. To avoid such users, the cloud storage must be protected against attacks. This can be done by providing security to the cloud storage. There are many ways to secure the cloud storage space. Some of them are

RSA algorithm
Onetime password
MD5
AES(Advanced Encryption Standard)

ii. Security for the data in cloud

Security for the cloud environment is provided in several ways as mentioned in previous section. There are certain situation in which the members those who have access to the cloud storage but does not have permission to specific data. There are chances that these members will try to access data without permission and can use, modify or delete them. To overcome this situation, the data stored in the cloud should be provided with security. There are many ways to provide security for the data that is stored in the cloud.

B. Data integrity for cloud data

When a user stores the data in a public cloud, there might be some other users willing to use the same data. Similarly there will be a group of members willing to share their data and use others data with proper permission.

When a particular data is used by one user and saved back, there should be a confidence that the saved data is correct and it is modified by an authenticated user. This property of the data to ensure its correctness is known as data integrity. To ensure integrity there must be a verifier who verifies the correctness of the data. The process of verifying the correctness of the data is known as auditing.

By employing the Third Party Auditor (TPA), the correctness of data can be ensured. While auditing the integrity of the shared data that are stored in the cloud storage are verified [12]. The auditing mechanism that can be performed simultaneously is known as Batch auditing. (i.e) multiple auditing tasks performed concurrently. This batch auditing reduces the overhead on communication resources and computation [6].

The data stored in the cloud may easily get corrupted because of the unavoidable hardware or software failures. The correctness of the shared data is defined as the detection of corrupted data that has been identified by the TPA.

There are three parties involved in cloud service storage [9]. They are (i) Users, (ii) Cloud Server, (iii) Third Party Auditor.

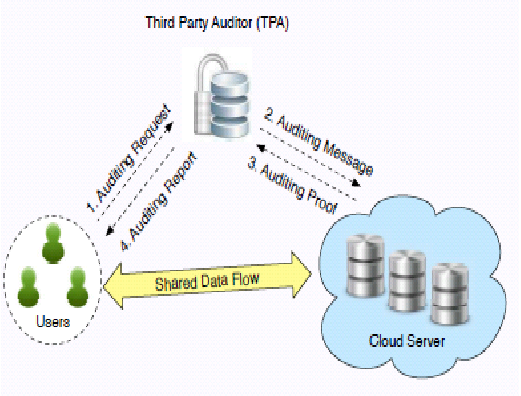

Fig. 1 Cloud Service Storage

To check the integrity of shared data, the user sends an auditing request to the Third Party Auditor (TPA). After getting the auditing request, an auditing message will be generated by the TPA to the cloud server. Then the cloud server will be responded to the TPA by sending the auditing proof of shared data. After the retrieval of auditing proof, the TPA verifies the correctness of data and sends an auditing report to the user as the feedback.

C. Auditing on cloud data

Auditing is the process of verifying the correctness of the data that has been used in the cloud by a group of users [9]. Auditing can be done in two ways [5].

Internal auditing is the process of verifying the correctness of data by the group manager who is present inside the cloud group. By doing this the cost is low and the time taken is also less. But the confidentiality among others nodes of the audit report is not up to the mark.

To improve the confidentiality of the audit report, the audit must be done by a third party auditor who is common to all the users in the cloud. The process of checking the integrity of the data by an external third party checker is known as Third Party Audit(TPA).

Data in the cloud is uploaded by a particular user who is the owner for the data. The data owner has the rights to form a group consisting of members who would like to get access of the data. The data owner will provide access permissions to the group members to view, modify and save back [6].

The data that is uploaded by the data owner and the same data that has been used by the group member must satisfy the general characteristics of data such as security, confidentiality, availability, privacy and integrity [8]. Of all these characters security and integrity is the main aspect has to be maintained when a particular data is used by the group.

To ensure security and integrity the data must be checked each and every time it’s been used. The process of checking them is known as auditing. Auditing can be done in two ways [11].

First is the local audit, done by data owner or the group manager. Second is the external audit, done by the Third Party Auditor (TPA).

D. Privacy Preserving

Preserving the identity of the user to improve security is the main concern when comes to public audit. When internal audit is done there is no fear of privacy leakage. But confidentiality cannot be assured. This is because the group members will have a doubt on audit performed by group manager or the data owner, so to improve data confidentiality for the users, Third party audit is done.

When a user accesses a data and saves it back, the user's signature is added. So when TPA is done there is a chance that the list of users those who accessed a particular file is disclosed, thereby decreasing the security [6].

To overcome this problem TPA should be done with privacy preservation. There are several methods by which privacy could be preserved during TPA. Some of them are listed below [7].

(i) Encryption methods
(ii) Access control mechanisms
(iii) Query integrity/Keyword searches
(iv) Auditability schemes
  a. Remote Data Possession at Untrusted Host
  b. Public Verifiability for Storage Security

c. Remote Data Checking Using Provable Data Possession
d. Privacy Preserving Data Integrity Checking
e. Privacy Preserving Public Auditability for Storage Security.

i) Encryption Methods

These are the approaches that are specific for the encryption techniques to attain privacy in cloud. To resolve the problem of privacy preserving encryption method, it comprises the data in an organization's design structure, the creation and the access control management of symmetric keys communication among participants and managing the modifications on providing the user's access and allows the user to perform dynamic operations such as delete, append and update [6].It uses an interactive protocol and an extirpation based key derivation algorithm. It ensures data confidentiality solve ineffectiveness of key derivation, reduces the burden of encryption and decryption, can be able to manage numerous keys, saves owners storage space, reduce run-time overheads of the system, gives excellent privacy security and can apply to multiple users, data owners and service providers But it needs to have techniques to reduce owner's encryption burden and to work on cipher text.

A method for improving user privacy with secret key recovery in cloud storage that allows users? To encrypt their files in the cloud storage has been proposed. A Secret sharing Algorithm to Key Recovery Mechanism is used. AES-128 to encrypt user's file, the key length is set to 128 bits is used Key Recovery scheme partially trusted because no one has the full information about the encryption key except the user himself. The compression algorithm used here is ZIP. The user's privacy is protected and it decreases the risk of encryption key lose. But it puts a big computation burden for users. It has concerns about transforming speed. Renewing user's key is a challenge here, users can't search words and there is dispersal of information. The works that involves [7] provides a privacy-preserving cloud storage framework supporting cipher text retrieval, it is to solve the problems while operating on an encrypted data and to reduce the data owner's workload on management of data and support data sharing. Interaction protocol, Key derivation Algorithm, combination of symmetric and asymmetric encryption and Bloom Filter is used here. It can operate on encrypted data; reduce data owner's workload on managing the data and storage space, reduce communication, computation and storage overhead. It can manage numerous keys and is efficient, safe and economic. But it supports only owner-write-user-read and lacks in technique that support cipher text-based computing.

The works listed in [7] also discusses about controllable privacy preserving search functionalities which include revocable delegated search and decryptable delegated search that are based on symmetric predicate encryption in the cloud storage. Thus the Owner of cloud can easily control lifetime and search privileges of data which is suitable for delegation-based business applications. But it cannot support complex access control and search privileges. A method using discretion algorithm for preserving privacy through data control in a cloud computing architecture, which provides security solution that requires more than user authentication and digital certificate are discussed. Here the SP can directly use data without any key and is more flexible and safe to protect individual’s privacy. But the use of Encryption limits data usage and needs communication and compatibility with heterogeneous host.

The main problem in using encryption based technique is that it limits the data usage and puts into an additional burden. The access control mechanisms are available which will overcome the burden of the above overheads.

ii) Access Control Mechanisms

Access security is important for prohibiting the unwanted users from accessing resources and sending unauthorized queries to servers. Normally, this is accomplished through the use of firewall that prevents unwanted traffics from communicating with the business applications. Verification should be done whether the cloud provider has firewall protection to prevent intruders and denial of service on attacks. [3].

A privacy preserving user authenticated access control mechanism on securing data in clouds which verifies the user's authenticity without knowing the user's identity before stored data gets displayed. The valid user can be able to decrypt the already stored data. It provides security from the reply attacks and provisioning of the data privacy [10]. This is a decentralized approach that allows several read and writes operations and it robust in nature.

iii) Query Integrity/Keyword Searches

There are approaches that make use of queries and keyword search scheme to check the privacy in cloud. The works listed in [7] proposed a privacy preserving keyword search scheme in cloud computing that allows a service provider to participate in partial decipherment and enables them to search the keywords on encrypted files. It makes use of an efficient privacy preserving keyword search scheme (EPPKS). It provides protection of user data privacy, queries privacy and support key word search on encrypted data. It is found efficient, practical and provably and semantically secure but the computation on encrypted data was a challenge. A privacy preserving approach for data outsourcing in cloud environment which make use of fragmentation and heuristic algorithm is used by Sayi et.al [13]. It proves to be efficient and effective but confidentiality is not achieved.

iv) Auditability Schemes

Auditing reduces risk for customers as well as give incentives to providers to improve their services. Auditability falls under two categories as follows when we consider the available schemes in auditability: private auditability and public auditability. Even though schemes with private auditability can attain higher scheme efficiency, public auditability permits anyone, not just the client (data owner), to deal with the cloud server for correctness of data storage while keeping no private information. Then, clients are able to pass on the evaluation of the service performance to an independent third party auditor (TPA), without giving their computation resources. So we can denote the types of auditing protocols as Data Owner Auditing and Third Party Auditing.

The methods of data storage auditing methods can be categorized into three: Message Authentication Code (MAC) - based methods, RSA- based methods and Boneh-Lynn-Shacham signature (BLS) - based Homomorphic methods. The challenging issues of data storage auditing include Dynamic Auditing Collaborative Auditing and Batch Auditing. We need to meet the three performance criteria when comes to designing of auditing protocols as: low storage overhead, low communication cost and low computational complexity.

a. Remote Data Possession at Untrusted Host

A RDPC scheme is proposed, which is efficient in terms of computation and communication; it allows verification without the need for the challenger to compare against the original data; it uses only small challenges and responses, and users need to store only two secret keys and several random numbers. Finally, a challenge updating method is proposed based on Euler's theorem.

b. Public Verifiability for Storage Security

A study of the problem in ensuring the data integrity in cloud storage is done. To ensure the correctness of data they allow a third party auditor to work on behalf of the cloud consumer, to check the integrity of the stored data in the cloud. This scheme assures that the storage at the client side is minimal which will be helpful for thin clients.

c. Remote Data Checking Using Provable Data Possession

A model for provable data possession which can be used for remote data checking is proposed. By having a sampling random set of blocks from the server, this model produces probabilistic proofs of possession which will significantly reduce I/O costs. In order to minimize network communication the challenge/response protocol transmits a small and constant amount of data. The model incorporates some mechanisms for pacifying random data corruption and it is long-lasting. It offers two efficient secure PDP schemes and the overhead at the server is low. To add durability to any remote data checking scheme based on spot checking, it proposes a generic transformation.

d. Privacy Preserving Data Integrity Checking

A privacy preserving technique for checking the data integrity at remote locations with data dynamics and public verifiability make use of a Remote Data Integrity Checking Protocol. The protocol provides public verifiability without the help of a third party auditor. It doesn't leak any privacy information to third party, which provides good performance without the support of the trusted third party and provides a method for independent arbitration of data retention contracts. But it gives unnecessary computation and communication cost.

e. Privacy Preserving Public Auditability for Storage Security

The study about the problem that ensures integrity of the data storage in Cloud Computing has been analyzed. It allows a third party auditor to confirm the integrity of dynamic data stored in the cloud. This scheme achieves both public auditability and dynamic data operations. A public auditing scheme with privacy preserving for secure cloud storage is proposed [7]. The protocol design to achieve the security and performance guarantees Public Auditability, Storage Correctness, Privacy-Preserving, Batch auditing, and to be Lightweight. The method is found to be scalable and efficient which provides complete outsourcing solution, integrity checking and thus saves amount of auditing time. It relies on third party auditors and has the use of expensive modular exponentiation operations which leads to storage overhead on server and extra communication cost. An efficient auditing service outsourcing for data integrity in clouds is proposed based on the creation of an interactive PDP protocol to inhibit the dishonesty of proven soundness property and the leakage of verified data (zero-knowledge property). It describes the periodic verification for improving the performance of audit services. Here the approach adopts a way of sampling verification. The scheme not only prevents the forgery and deception of cloud storage providers, but also prevents the leakage of data that are outsourced in the process of verification. It supports an adaptive parameter selection. The system shows only lower computation cost as well as a shorter extra storage and the scheme is less complex due to fragment structure. It achieves Audit without-downloading, Verification-correctness, Privacy-preserving and High performance.

V. CONCLUSION

Cloud computing provides Storage support, Service support and infrastructure support. The usage and the application can be extended when cloud support is given with the discussed properties. Improving data security and integrity with preserving the privacy of shared data provides trustworthy to users. In this survey several methods are studied which can be combined to create a novel method for securing stored cloud data.

VI. REFERENCES

[1] http://en.wikipedia.org/wiki/Cloud_computing

[2] Mohit Marwaha, Rajeev Bedi, “Applying Encryption algorithm for Data Security and Privacy in Cloud Computing”, IJCSI International Journal of Computer Science Issues, Vol 10, Issue 1, No 1, January 2013, pp. 367-370.

[3] Vishakha Lokhande et al., “Efficient Encryption and Decryption Services for Cloud Computing”, International Journal of Soceital Applications of Computer Science, Vol 1 Issue 2 December 2012, pp. 71-75.

[4] Fang Liu et al., “NIST Cloud Computing Reference Architecture”, National Institute of Standards and Technology U.S Department of Commerce, Special Publication 500-292, pp. 10-12.

[5] “Auditing to Keep Online Storage Services Honest” HOTOS'07 Proceedings of the 11th USENIX workshop on Hot topics in operating systems

[6] B. Wang, B. Li, and H. Li, “Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud,” in Proceedings of IEEE Cloud 2012, 2012, pp.295–302.

[7] Neethu Mariam Joseph et al, “Survey on Privacy-Preserving Methods for Storage in Cloud Computing”, Amrita International Conference of Women in Computing (AICWIC’13) Proceedings published by International Journal of Computer Applications® (IJCA).

[8] Cong Wang et al, “Ensuring Data Storage Security in Cloud Computing”, US National Science Foundation under grant CNS-0831963, CNS-0626601, CNS-0716306, and CNS-0831628.

[9] Miss. M.Sowparnika1, Prof. R. Dheenadayalu2, “Improving data integrity on cloud storage services”, International Journal of Engineering Science Invention ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org Volume 2 Issue 2 ǁ February. 2013 ǁ PP.49-55.

[10] Ranjita Mishra et al, “A Privacy Preserving Repository for Securing Data across the Cloud”.

[11] Shingare Vidya Marshal, “Secure Audit Service by Using TPA for Data Integrity in Cloud System”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue-4, September 2013.

[12] Imran Ahmad et al, “Privacy-Preserving Public Auditing & Data Integrity for Secure Cloud Storage”, International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV, PP.100-104.

[13] Sayi, T. J. V. R. K., Krishna, R. S., Mukkamala, R., & Baruah, P. K. 2012. Data Outsourcing in Cloud Environments: A Privacy Preserving Approach. In Information Technology: New Generations (ITNG), 2012 Ninth International Conference on (pp. 361-366).