www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 4 Issue 9 Sep 2015, Page No. 14336-14340

Survey On Verification Of Storage Correctness In Cloud Computing

Jeevitha M[1], Chandrasekar A[2], Karthik S[3]

[1]Department of CSE, SNS College of Technology,
Coimbatore, India
mail2jeevitha@gmail.com

[2]Department of CSE, SNS College of Technology,
Coimbatore, India
chandru.as76@gmail.com

[3]Department of CSE, SNS College of Technology,
Coimbatore, India
profskarthik@gmail.com

Abstract: ***In the recent era of the world of computing, cloud has become a trend-setter by providing the ease of storing and sharing the data in enormous volumes. Cloud computing has evolved as a boon to its users by not only relieving them from the burden of storing and maintaining huge data but also give them the feel that their data is straightforwardly accessible from anywhere across the globe. The users will no more have the physical possession of their data, which will mount the security concerns of storing the data in cloud, which is a third party server whose location is not known to the user. Therefore verification of data to ensure the storage correctness becomes a significant task in cloud. As the users of cloud are free from the burden of storing and maintaining their data, they must also be free to use their data without worrying about the need for any integrity checking. So cloud uses a Third Party Auditor (TPA) who has the capability for verifying the data for its consistency. In this paper we will discuss about the various techniques that can be adopted by the TPA in ensuring the storage correctness of dynamic data in cloud.***

Keywords: Cloud Computing, Security, Auditing, Provable Data Possession, Proofs of Retrievability, Third Party Storage Auditing Services

1. Introduction

Cloud Computing is a recently emerging hopeful technology that provides countless computing services over the Internet with an assured Quality of Service, reduces the cost of implementing high cost infrastructures and provides on demand access to the shared pool of resources such as, Infrastructure, Platform, Software, harmonizing to the needs of the users. The most prominent feature that cloud gives its users is storage, which makes the data accessible from any location and from any device via the Internet. It is this characteristic of cloud that has attracted millions of users towards it. By moving our data into the cloud, it can be operated on whenever needed without any saddle of maintaining our own hardware and software. But that introduces certain new security threats since the user does not have substantial control over the data or we can say that there is no storage dependence. The cloud users will develop a state of disbelief that their data might be traced out by some unauthorized person, or the data might be lost. In order to deal with the security issues we need to implement strict auditing techniques which will ensure the storage correctness at regular periods and also detect the misbehaving servers immediately and report to the users. In order to enforce reliability of cloud storage, persuasive auditing mechanisms must be employed in cloud. The Third Party Auditor must do all these reliability measures in a causal manner without the knowledge of the cloud user, whose foremost concern is the ease of data openness and the assurance of secure storage. Let us discuss about the essential characteristics of auditing mechanisms and the various protocols that are available for better verification of data integrity in cloud storage in the following sections.

2. The Cloud Model

The There are three Service Models in cloud. They are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

i) Software as a Service (SaaS): Through this service model Cloud Computing offers software applications to its end users which they can make use of through the internet instead of installing and maintaining their own software. SaaS applications are also known as On-demand software or hosted software [1].

ii) ii) Platform as a Service: This service model provides the platform for deployment and hosting

DOI: 10.18535/ijecs/v4i9.46

the software applications over the internet. It provides tools and Integrated Development Environment (IDE) to do the coding and testing of software applications over the web [1].

iii) iii) Infrastructure as a Service: This model provides infrastructure services such as servers, storage systems, network, and operating systems. Instead of purchasing own software, servers, datacenter space or network equipment, clients can outsource those resources as a service on demand.

3. The Architecture of Cloud

The network architecture for cloud storage service model is illustrated in the figure (Figure 1.). It consists of the following entities [5]:

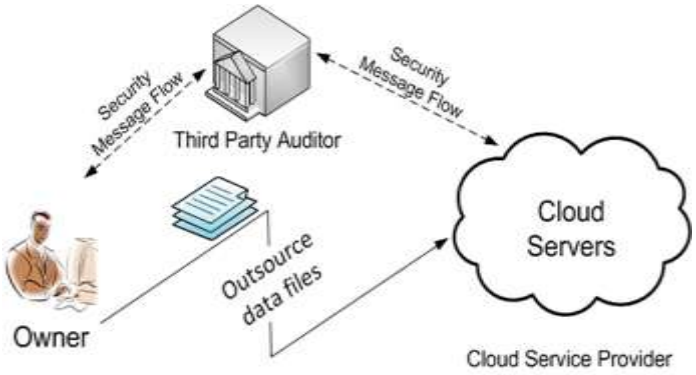

Figure 1: Data Storage in cloud

Owner: Individual users and different organizations or enterprises might have their data to be stored in the cloud and they solely There are a number of security issues that can be classified into two categories: security issues faced by cloud service providers(organizations providing software-, platform-, or infrastructure-as-a-services through cloud) and security issues faced by the cloud customers (organizations who host their applications or store data on the cloud) [2].depend upon the cloud for data storage and computation.

Cloud Server: It is an entity, which is administered by Cloud Service Provider (CSP) to provide data storage service which has significant storage space and computation resources.

Third-Party Auditor (TPA): The TPA has proficiency and expertise that users may not possess. He is the trusted person to assess and expose the risk of cloud services on behalf of the users upon request.

4. Cloud Security

Cloud Security is an up-coming sub-domain of computer security which refers to a broad set of policies, technologies, and controls established in cloud in order to preserve the stored information, applications, and the associated infrastructure of cloud computing. Since users store their data at multiple locations in cloud ensuring storage correctness is a very important and chaotic task. And so Data Integrity Verification is the most significant security issue in cloud. Such a security assurance is significantly required for communications between cloud users and cloud servers, and also for data at rest on cloud servers. In particular, cloud users may have great concerns on data integrity when outsourcing valuable data assets in the cloud for storage. Verifying data integrity is a challenging task because users should not be known that such verification is taking place and they must be free to use the data stored in cloud. Auditing is necessary to check the storage correctness of the data which is stored in some third party cloud server. Without affecting the client's original data, auditing is performed[3].The cloud service providers should be able to provide the required security services to meet individual security requirements of the cloud users' at the same time should also confine to the regulations/compliances. In non-sensitive applications, it is also important to protect the critical data of the users and help them verify the services provided by the cloud. Secure auditing mechanisms are usually necessary for this purpose [4]. Data storage seems to be more cumbersome because of the following reasons:

i. Cloud users may not be willing to fully rely on cloud service providers for providing data integrity protection. The reason behind this is that cloud services are usually provided by third-party providers who cannot be expected to be in the same trust domain of the cloud users.

ii. Data integrity service should be provided in the timely manner, because in practical applications the cloud users find out data corruption too late, ie., when they are actually retrieving the data.

iii. The"self-served" data integrity check requires active involvement of cloud users, necessary expertise and computing power of the users. The cloud users vary greatly in their available resources and expertise. Most cloud users may not have the ability to perform data integrity check by themselves. A sensible solution to this problem is to assign the task of data verifiability check to a third professional party of their trust (i.e.,a third party auditor (TPA)) which has the necessary resources and expertise.

iv. As data stored on cloud servers may subject to modification by cloud users, the data verification mechanism should efficiently support such data dynamics. Modification of one block of data should not have its effect on other data blocks in terms of data integrity protection.

v. Preferably, a data integrity protection mechanism should address all these above mentioned issues. While performing data integrity verification there should not be any degradation in the performance of the cloud usage and it should not be a hindrance to the cloud users. Therefore we need to choose a well organized auditing mechanism that will serve the purpose effectively.

5. A General Idea of Existing Audit Mechanisms

In this paper we are going to survey on some of the existing auditing schemes and we do not propose any new scheme for auditing.

There are three approaches for ensuring data integrity. They are:

5.1 Provable Data Possession (PDP) Schemes

Provable Data Possession (PDP) is a cryptographic technique that permits the cloud users to store their data at some untrusted server and have probabilistic guarantees that the server possesses the original data. The client stores only his cryptographic keys and never has to retrieve the file.

The PDP protocol is the following:

i. The first step is the pre-computation of tags for each block of a file by the user.
ii. The user transmits the file and tags to the storage server.
iii. To verify possession, the user produces challenge and sends it to the server.
iv. The server generates a proof based on the challenge and replies to the client.
v. The client verifies the proof.

PDP uses homomorphic verifiable tags that reduces the overhead of server computation, network traffic and block accesses while achieving a strong guarantee of data possession[6].

5.1.1 Sampling PDP and Efficient PDP

These PDP Schemes are based on Homomorphic Verifiable Tags (HVT) and Homomorphic Linear Authenticators (HLA). The files are stored in blocks in storage server and Method Tags are allocated to the corresponding File Blocks. In order to verify the Integrity of the Data Stored we make use of the Cumulative Sum of all the Blocks. 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 auditing of data takes place. Data Possession is guaranteed only as a whole block and therefore it incurs more cost.

5.1.2 Scalable PDP

This PDP Scheme is based on the cryptographic Technique which uses symmetric key. 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 [8].

5.1.3 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.

5.1.4 Dynamic PDP

The original PDP scheme applies only to static (or append-only) files. We present a definitional framework and efficient constructions for dynamic provable data possession (DPDP), which extends the PDP model to support provable updates to stored data. A new version of authenticated dictionaries is used based on rank information. Thus we are able to support efficient authenticated operations on files at the block level, such as authenticated insert and delete. We prove the security of our constructions using standard assumptions [7].

5.1.5 Multi -Replica Provable Data Possession MR-PDP

MR-PDP extends previous work on data possession proofs for a single copy of a file in a client/server storage system. Using MR-PDP to store t replicas is computationally much more efficient than using a single-replica PDP scheme to store t separate, unrelated files (e.g., by encrypting each file separately prior to storing it). Another advantage of MR-PDP is that it can generate further replicas on demand, at little expense, when some of the existing replicas fail[9].

5.1.6 Cooperative Provable Data Possession

A cooperative PDP (CPDP) scheme is based on homomorphic verifiable response and hash index hierarchy. An effective construction of CPDP scheme is based on modern cryptographic techniques, such as interactive proof system (IPS), which does not compromise on data privacy. This construction is proved to be a multi-prover zero knowledge proof system (MP-ZKPS), which has completeness, knowledge soundness, and zero-knowledge properties. These properties ensure that CPDP scheme can implement the security against data leakage attack and tag forgery attack [10].

5.2 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 undamaged and not modified by any attack. 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 gives the proof of Existence. They consume less bandwidth than the file itself and hence can be used in remote environment. The important feature of this model is that they can correct any Data Corruptions that is found by using Error Correction codes.

5.2.1 Compact Proofs of Retrievability

It uses Homomorphic Property to reduce the size of authenticator value and hence there is considerable reduction in 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. Two Types of CPOR are Compact Proofs of Retrievability-I, Compact Proofs of Retrievability-II, where the former supports Public Verifiability where anyone not only the data owner can verify the integrity of the Uploaded Data and the latter supports Private Verifiability, where the verifiability option is restricted to the Data Owner alone.

5.3 Third-Party Storage Auditing Service (TSAS)

The TSA is an efficient, privacy preserving auditing protocol for cloud storage that supports dynamic data operations and

batch auditing for multiple Data Owners and multi-cloud servers. This scheme uses Bilinearity property to verify the data integrity[11].

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.

The TSAS should possess the following characteristics:

i. An Efficient and Privacy Preserving Auditing Protocol: The auditing protocol should protect the data privacy against the auditor himself. (a) For public data, the auditor may obtain the data information by recovering the data blocks from the data proof. (b) For encrypted data, the auditor may obtain content keys somehow through any special channels and with that key he is able to decrypt the data. To solve the data privacy problem, TSAS generates an encrypted proof with the challenge stamp and it makes use of the Bilinearity property of the bilinear pairing, such that the auditor cannot decrypt it. But the auditor can verify the correctness of the proof without decrypting it.

ii. Supports Secure Dynamic Auditing:The Data Owners will dynamically update their data stored in cloud. Therefore the auditing protocol must be designed such that it supports data dynamics and also secure the static archived data. The server faces two attacks: (a) Replay Attack The server may not update correctly the owner's data on the server and may use the previous version of the data to pass the auditing. (b) Forge Attack: When the data owner updates the data to the current version, the server may get enough information from the dynamic operations to forge the data tag. If the server could forge the data tag, it can use any data and its forged data tag to pass the auditing. To prevent the replay attack, an Index Table (ITable) is introduced to record the abstract information of the data. To deal with the forge attack, it can modify the tag generation algorithm TagGen [11].

iii. Batch Auditing for Multi-Owner and Multi-Cloud: Data storage auditing is a significant service in cloud computing which helps the owners check the data integrity on the cloud servers. Due to the large number of data owners, the auditor may receive many auditing requests from multiple data owners. In this scenario, it would greatly improve the system performance, if the auditor could combine these auditing requests together and only conduct the batch auditing for multiple owners simultaneously.

There are two main schemes used by the TPA. They are as follows:

5.3.1 HMAC - Hash based Message Authentication Code

It is a cryptographic hash function where the message and key are hash them together. In HMAC, the generation of authentication code uses secret and hash based algorithm. This code ensures the usage of hash function is expanded in many places. It conserves the performance of hash function. By using secret key, we can calculate the hash function of message authentication code. SHA is a hash algorithm with is used to generate the authentication code for the message. It is used to check the message authentication by using secret key and verify the data storage correctness. The strength of HMAC is determined based on the strength of hash function, hash output size and key size. HMAC supports for has algorithms like MD5, SHA-1, SHA-256, etc [3].

5.3.2 Homomorphic Encryption

It is an encryption technique in which it is possible to do computations on cipher text. It generates an encrypted result which, when decrypted, matches the result of operations performed on the plain text. A Homomorphic Encryption system is used to perform operations only on encrypted data not on decryption data. While performing operations, it does not know the users private key. It knows only the users secret key.

5.4 Advantages of the Auditing Schemes

i. PDP supports remote data checking of large data sets in widely distributed storage systems. It minimizes the number of file block accesses, the computation and communication cost incurred on the server

ii. POR supports dynamic operations such as insertion,deletion and modification.Itassures the Data Retrieval

iii. Proofs of Retrievability corrects data corruptions in the case of occurrence in Client Data

iv. Compact Proofs of Retrievability use shorter queries and response time

v. Multi Copy PDP Schemes increases Data Possession

vi. TSAS reduces load on the Cloud Service Provider

vii. TSAS supports both Multi-Owner and Multi-Cloud Environment

6. Conclusion

Cloud computing has given a new definition for the field of computing and the way internet can be used for storage services. Several major concerns are to be taken into account when we decide to utilize cloud services. This paper carried out a survey on various storage verification techniques in Cloud Environment. Also several methods of these Techniques have been discussed. Hence we have analyzed the existing techniques to initiate further research in ensuring storage verification thereby enhance Cloud Security.

References

[1] Ben Kepes, Understanding the Cloud Computing Stack: SaaS, PaaS, IaaS, *Rackspace Support*, Article ID 35, published on Oct 22,2013

[2] Sandeepraja Batchu, J.N. Chaitanya, A study on Security Issues Associated with Public Clouds in Cloud Computing, Sanketika Vidyaparishad Engineering College, *International Journal of Advanced Computer*

Technology, volume 2, number 2, 2012

[3] G.Mahinder , G.Sudhakar , Distributed Storage and Integrity Auditing Mechanism for Secure Cloud Storage , Aurora's Scientific Technological and Research Academy, Hyderabad, India, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 2, Issue 10, October 2013

[4] Ramya.D, Dr.Raja.K, An Analysis of Third Party Auditing Techniques in Cloud Computing, Alpha College of Engineering, Chennai, Tamil Nadu, India. *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 3, Issue 11, November 2014

[5] Chapter 5.3: Data Security in Cloud Computing Shucheng Yu, University of Arkansas at Little Rock, AR, USA , Wenjing Lou, Virginia Polytechnic Institute and State University, VA, USA , and Kui Ren, Illinois Institute of Technology, IL, USA

[6] Chaoling Li, Yue Chen ,Towards comprehensive provable data possession in cloud computing,*Wuhan University Journal of Natural Sciences*, June 2013, Volume 18, Issue 3, pp 265-271

[7] C. Chris Erway , Alptekin Kupcu , Charalampos Papamanthou , Roberto Tamassia, Dynamic Provable Data Possession , Brown University,Providence RI November 29, 2009

[8] John Abinash Paul, Mrs. Esther Daniel,A Survey on Data Integrity Verification Schemes in Cloud Computing, *International Journal of Engineering Research & Technology (IJERT)* IJERT ISSN: 2278-0181 Karunya University, Coimbatore, India, Vol. 3 Issue 4, April - 2014

[9] Reza Curtmola, Purdue University, Osama Khan, Randal Burns, Giuseppe Ateniese, Johns Hopkins University, MR-PDP: Multiple-Replica Provable Data Possession, *Distributed Computing Systems, ICDCS '08*. The 28th International Conference, 2008

[10] Yan Zhu, Hongxin Hu, Gail-Joon Ahn,Senior Member, IEEE , Mengyang Yu, Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage,*IEEE Transactions on Parallel and Distributed Sy stems*, 2012

[11] Chapter 2 TSAS: Third-Party Storage Auditing Service, K. Yang and X. Jia, Security for Cloud Storage Systems, *Springer Briefs 7 in Computer Science*, DOI: 10.1007/978-1-4614-7873-7_2, © The Author(s) 2014