IJCAT International Journal of Computing and Technology, Volume 1, Issue 6, July 2014
ISSN : 2348 - 6090
www.IJCAT.org

# Survey on Data Integrity and Security Issues in Cloud Storage

**[1] Lethija J, [2] Shyja N K**

[1] Department of Computer Science and Engineering
Malabar College of Engineering and Technology, Thrissur

[2] Department of Computer Science and Engineering
Malabar College of Engineering and Technology, Thrissur

**Abstract -** Cloud computing is a novel internet based network architecture that focuses on sharing of computing and resources. It is defines as enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. In a cloud system, for data owners, computing and data are not in the places they used to be in. It relieves the user from the burden of local storage and maintanence. With cloud storage data is stored to a set of concurrently running distributed servers. Since the data owner no more possess the data it may have to suffer both internal and external threats at cloud service provider.The risks of compromised security and privacy may be lower overall, however, with cloud computing than they would be if the data were to be stored on individual machines instead of in a so called "cloud" (the network of computers used for remote storage and maintenance).This paper reviews various security threats and data integrity issues and briefs some solutions for them**.**

***Keywords*** **- Cloud computing, Integrity, Security**

## 1. Introduction

Cloud computing is set of resources and services offered through the Internet. Cloud services are delivered from data centres located throughout the world. Cloud computing facilitates its consumers by providing virtual resources via internet. Possession of data and resource implies control in the sense that if data is in customers' computer they ultimately control when it gets deleted and have some ability to move it around, copies it and operate on it. If a computation happens in the cloud instead of customers' computers, the cloud has the ability to stop it, the ability up to the point to reprogram and put some change on it. Control what happens in the world of computing implies power. Data and computations on the data are outsourced to the cloud means the data owner loses control over his data and subsequently gave the power to perform any activity on it to the third party (cloud service provider).

Cloud computing refers to both the applications delivered as services over the Internet and the hardware and software in the data centres that provide those services. Cloud providers use virtualization technologies combined with self service abilities for computing resources via network infrastructure. In cloud environments, several kinds of virtual machines are hosted on the same physical server as infrastructure. In cloud, customers must only pay for what they use and have not to pay for local resources which they need to such as storage or infrastructure. Nowadays, we have three types of cloud environments: Public, Private, and Hybrid clouds. A public cloud is standard model which providers make several resources, such as applications and storage, available to the public. Public cloud services may be free or not. In public clouds which they are running applications externally by large service providers and offers some benefits over private clouds.

Private Cloud refers to internal services of a business that is not available for ordinary people. Essentially Private clouds are a marketing term for an architecture that provides hosted services to particular group of people behind a firewall. Hybrid cloud is an environment that a company provides and controls some resources internally and has some others for public use. Combination of private and public clouds is called Hybrid cloud. In this type, cloud provider has a service that has private cloud part which only accessible by certified staff and protected by firewalls from outside accessing and a public cloud environment which external users can access to it. There are three major types of service in the cloud environment: SaaS, PaaS, and laaS [1]. Also there is In Cloud computing technology there are a set of important policy issues, which include issues of privacy, security, anonymity, telecommunications capacity, government surveillance, reliability, and liability, among others [I]. But the most important between them is security and how cloud provider assures it.

## 2. Cloud Security Issues

Using Cloud results applications and data will move under third-party control. Predominantly, the first question is an information security officer must answer

to that whether he has adequate transparency from cloud services to manage the governance (shared responsibilities) and implementation of security management processes such as detection and prevention solutions to assure the costumers that the data in the cloud is appropriately protected. Actually, the answer to this question has two parts: what security controls must the customer provide over and above the controls inherent in the cloud platform, and how must an enterprise's security management tools and processes adapt to manage security in the cloud.

### 2.1 Data Leakage

Innately, when moving to a cloud there is two changes for customer's data. First, the data will store away from the customer's local machine. Second, the data is moving from a single-tenant to a multi-tenant environment. These changes can raise an important concern that called data leakage. Because of them, Data leakage has become one of the greatest organizational risks from security standpoint [2]. Nowadays, for mitigate effects of such problem there has been interested in the use of data leakage prevention (DLP) applications to protect sensitive data. But if data stored in a public cloud because of nature of it, using DLP products is value less to protect the confidentiality of that data in all types of cloud. Inherently, in SaaS and PaaS discovery of client's data with DLP agents is impossible except when the provider put ability of it to its service. In private clouds, Costumer has direct control over the whole infrastructure; it is not a policy issue whether DLP agents are deployed in connection with SaaS, PaaS, or IaaS services. However, it may well be a technical issue whether DLP agents interoperate with your SaaS or PaaS services as architected [3]. In hybrid cloud, if service is IaaS, client could set in DLP agents for some control over data.

### 2.2 Cloud Security Issues

Internet is communication infrastructure for cloud providers that use well-known TCP/IP protocol which users' IP addresses to identify them in the Internet. Similar to physical computer in the Internet that have IP address, a virtual machine in the Internet has an IP address as well. A malicious user, whether internal or external, like a legal user can find this IP addresses as well. In this case, malicious user can find out which physical servers the victim is using then by implanting a malicious virtual machine at that location to launch an attack. Because all of users who use same virtual machine as infrastructure, if a hacker steals a virtual machine or take control over it, he will be able to access to all users' data within it. Therefore, The hacker can copy them into his local machine before cloud provider detect that virtual machine is in out of control then the hacker with analysis the data may be find valuable data afterward .

#### 2.2.1 Attacks in Cloud

As the cloud can give service to legal users it can also provide service to users that have malicious purposes. A hacker can use a cloud to host a malicious application for achieve his object which may be a DDoS attacks against cloud itself or arranging another user in the cloud. For example, assume an attacker knew that his victim is using typical cloud provider, now attacker by using same cloud provider can sketch an attack against his victim. This situation is similar to this scenario that both attacker and victim are in same network but with this difference that they use virtual machines instead of physical network.

#### 2.2.1 a) DDoS Attacks against Cloud

Distributed Denial of Service (DDoS) attacks typically focus high quantity of IP packets at specific network entry elements; usually any form of hardware that operates on a Blacklist pattern is quickly overrun and will become in out of- service situation. In cloud computing where infrastructure is shared by large number of clients, DDoS attacks make have the potential of having much greater impact than against single tenanted architectures [4]. If cloud has not plenty resource to provide services to its costumers then this is may be cause undesirable DDoS attacks. Solution for this event is a traditional solution that is increase number of such critical resources. But serious problem is when a malicious user deliberately done a DDoS attack.

Most network counter measures cannot protect against DDoS attacks as they cannot stop the deluge of traffic and typically cannot distinguish good traffic from bad traffic. Intrusion Prevention Systems (IPS) are effective if the attacks are identified and have pre-existing signatures but are ineffectual if there is legitimate content with bad intentions [2]. Unfortunately, similar to IPS solutions, firewalls are vulnerable and ineffective against DDoS attacks because attacker can easily bypass firewalls and also IPSs since they are designed to transmit legitimate traffic and attacks generate so much traffic from so many distinct hosts that a server, or for cloud its Internet connection, cannot handle the traffic [6]. It may be more accurate to say that DDoS protection is part of the Network Virtualization layer rather than Server Virtualization. For example, cloud systems use virtual machines can be overcome by ARP spoofing at the network layer and it is really about how to layer security across multivendor networks, firewalls and load balances.

#### 2.2.1 b) Cloud against DDoS Attacks

When a DDoS attack is launched, it sends a heavy flood of packets to a Web server from multiple sources. In this situation, the cloud may be part of the solution. It’s interesting to consider that websites experiencing DDoS

attacks which have limitation in server resources can take advantage of using cloud that provides more resource to tolerate such attacks. In the other hand, cloud technology offers the benefit of flexibility, with the ability to provide resources almost instantaneously as necessary to avoid site shutdown.

## 3. Solutions for Against Cloud Security Problems

There are several traditional solutions to mitigate security problems that exist in the Internet environment, as a cloud infrastructure, but nature of cloud causes some security problem that they are especially exist in cloud environment [7]. In the other hand, there is also traditional counter measure against popular Internet security problems that may be usable in cloud but some of them must be improved or changed to work effectively in it.

### 3.1 Access Control

Access control mechanisms are tools to ensure authorized user can access and to prevent unauthorized access to information systems. Therewith, formal procedures should be in place to control the allocation of access rights to information systems and services. Such mechanisms should cover all stages in the lifecycle of user access, from the initial registration of new users to the final de-registration of users who no longer require access to information systems and services. Special attention should be given, where appropriate, to the need to control the allocation of privileged access rights, which allow users to override system controls. Generally, in the SaaS model the cloud provider is responsible for managing all aspects of the network, server, and application infrastructure. In this model, since the application is delivered as a service to end users, usually via a web browser, network-based controls are becoming less relevant and are augmented or superseded by user access controls, e.g., authentication using a one-time password [4, 7]. Hence, customers should focus on user access controls (authentication, federation, privilege management, provisioning, etc.) to protect the information hosted by SaaS.

In the PaaS delivery model, the Cloud provider is responsible for managing access control to the network, servers, and application platform infrastructure. However, the customer is responsible for access control to the applications placed on a PaaS platform. Access control to applications manifests as end user access management, which includes provisioning and authentication of users. IaaS customers are entirely responsible for managing all aspects of access control to their resources in the cloud. Access to the virtual servers, virtual network, virtual storage, and applications hosted on an IaaS platform will have to be designed and managed by the customer. In an IaaS delivery model, access control management falls into one of the following two categories. Access control management to the host, network, and management applications that are owned and managed by the Cloud provider and user must manage access control to his/her virtual server, virtual storage, virtual networks, and applications hosted on virtual servers[8].

## 4. Data Integrity in Cloud Computing

Cloud computing provides universal data access and avoids capital expenditure on hardware software and personnel maintenance While Cloud Computing makes these advantages more appealing than ever, it also brings new a challenging security threats towards users' outsourced data. One of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. For example, the storage service provider, which experiences Byzantine failure occasionally, may decide to hide the data errors from the clients for the benefit of their own. How to efficiently verify the correctness of outsourced cloud data without the local copy of data files becomes a big challenge for data storage security in Cloud Computing. Note that simply downloading the data for its integrity verification is not a practical solution due to the expensiveness in cost and transmitting the file across the network. Considering the large size of the outsourced data and the users constrained resource capability, the ability to audit the correctness of the data in a cloud environment can be formidable and expensive for the cloud users. Besides, it is often insufficient to detect the data corruption when accessing the data, as it might be too late for recover the data loss or damage. In order to solve the problem of data integrity checking, many schemes are proposed under different systems and security models.

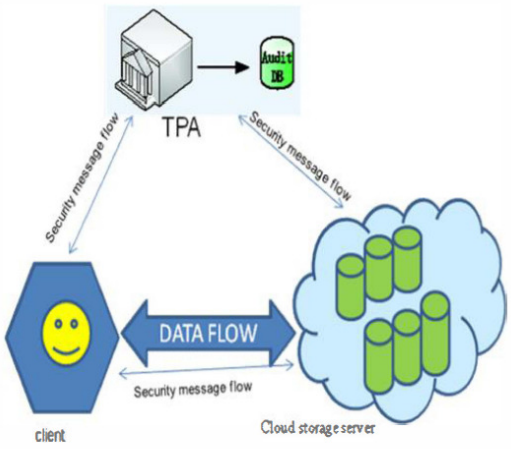


Fig. 1 Cloud data storage architecture

### 4.1 Requirements for Remote Data Integrity Checking Protocol

Remote data possession checking protocols are available. Using one of such protocols, a prover can

convince a verifier that the prover has access to a complete and uncorrupted version of a data file.

The following requirements should to be satisfied for a remote data possession checking protocol to be of practical use:

- the verifier should not be required to keep an entire copy of the file( s) to be checked. It would be impractical for a verifier to replicate the content of all provers to be verified. Storing a reduced-size digest of the data at the verifier should be enough.
- The protocol has to stay secure even if the prover is malicious. A malicious prover is interested in proving knowledge of some data that she does not entirely know; security means that such a prover ought to fail in.
- The amount of communication required by the protocol should be low.
- The protocol should be efficient in terms of communication
- It can be possible to run the verification an unlimited number of times.
- Data dynamics. Data dynamics means after clients store their data at the remote server, they can dynamically update their data at later times. The main operations supported by data dynamics are data insertion, data modification and data deletion. Moreover, when data is updated. The updating overhead should be made as small as possible.
- Unforgeability: to ensure that there exists no dishonest cloud server that can pass the audit from the user/TPA without indeed storing users' data intact;
- Public verifiability. Public verifiability allows anyone (not just a client) to perform the integrity checking operation. This function avoids disputes between the client and the server regarding data integrity. Whenever there is such a dispute, a third party authority can easily judge whether the data integrity is maintained using the verification protocol.
- Privacy. When the verification is performed by a third party verifier (not by a client),the protocol must ensure that no private information contained in the data is leaked.
- Batch auditing. Multiple delegated auditing tasks from different users can be performed simultaneously by the third party auditor. Cloud server may concurrently handle multiple verification sessions from different client.
- Fast localization of data error. To effectively locate the mal functioning server when data corruptions has been detected.
- Dependability. To enhance data availability against Byzantine failure, malicious data modification and server colluding attacks, i.e. minimizing the effect brought by the data errors or servers failures.
- Lightweight. To enable users to perform storage correctness checks with minimum overhead.

## 5. Integrity Proof Techniques

The following sections deal with existing integrity proof techniques

.

### 5.1 Provable Data Possession (PDP) at untrusted stores [8]

A PDP protocol checks that an outsourced storage site retains a file, which consists of a collection of n blocks. The data owner pre-processes the file, generating a piece of metadata that is stored locally, transmits the file to the server and deletes its local copy. Therefore correct possession of data is verified in a challenge-response protocol between the data owner and the server that stores the file. Ateniese et al.[8] are the first to construct and formally define public verifiability PDP model which is provably secure for remote data checking. They introduce the concept of RSA-based homomorphic verifiability tags (HVTs) which are a building block for PDP scheme and using HVTs they also alleviate the lack of block less verification in techniques based on aggregate signatures [9] and condensed-RSA [10] to mention some. It is easy to tell that schemes which fail block less verification are not suitable for third party auditing.

### 5.2 Proof of Retrievability (POR) for Large Files

A POR is a protocol in which a server proves to a data owner that a target file is intact, in the sense that the client can retrieve the entire file from the server with high probability. Hence, POR guarantees not only correct data possession but it also assures retrievability upon some data corruptions.
A POR system by Juels and Kaliski [11] includes six functions *keygen, encode, challenge , respond, verify, extract*

- The verifier performs the pre-processing phase and storage as follows. A $l$ – 1bit block is employed as a basic unit in this scheme. First, using the function *keygen* a secret key is generated. Then the verifier carves the file A into *m*-1-block chunks, calls the function *encode* and:

- Makes an (n, k, d)1-error correcting code *C* on each chunk over Galois Field GF($2^1$ ).This operation expands each chunk into n blocks and yields a first layer encoded file F′ with b' = bn/k blocks.

- Apply a per-block asymmetric key cipher *E* to *F'*. to yield a second layer encoded file *F''*. A per-block encryption is made independently on plaintext blocks in order to recover A even when the server deletes or corrupts blocks.
- Creates data check blocks called *sentinels* using a suitable one-way function. Individual *sentinel* is used for one time verification.
- Finally, to make the *sentinels* indistinguishable from the data blocks, a pseudorandom permutation is applied on *F''*, yielding the output file *F~*.. Then this output is stored on the server.
- In verification phase The verifier generates q-sentinel positions and sends them to the server as a challenge.
- The server returns corresponding sentinel values to the client. The verifier calls a function verify and determines whether the response from the server is valid or not. The function outputs '1' if verification succeeds and'0' otherwise.

### 5.3 Other Integrity Proof Schemes

In this section we summarized some other POR and PDP protocols assuming that in one way or another the schemes by Juels et al., Ateniese et al. and Shacham et al. are the basis for their research idea. Ateniese et al.[12] proposed dynamic version of the previous PDP scheme constructed based on symmetric key cryptography, while not requiring any bulk encryption. However, this scheme also imposes a prior bound on the number of queries and does not support fully dynamic data operations, i.e., it only allows very basic block operations with limited functionality i.e., for instance block insertion cannot be supported. Erway et al. [13] introduce a formal framework for dynamic PDP and provide the first fully dynamic PDP solution to support provable updates on the stored data. They extend the PDP model in [6] to support provable updates to stored data files using rank based authenticated skip lists. Bowers et al. [14] proposed HAIL (High-Availability and Integrity Layer) that improves PORs by providing efficiently computable proofs with servers and it can also verify and reallocate file shares leaving the design of more efficient *redistribute* algorithms for future work.

The file system availability in HAIL protocol is enhanced through a careful interleaving of different types of error-correcting layers. Bowers et al. [15] state their theoretical POR design by employing error-correcting codes in two phases (i.e., *inner* and *outer* ) to offer an improvement over the protocols. They propose a new variant so that it can achieve lower storage overhead and tolerate higher error rates but just like the protocol in [11] it gives no security guarantee against fully byzantine adversarial servers stated as in [16]. In these schemes data dynamics and public verifiability are not treated and hence they are not suitable for third party auditing. Zheng et al.[17] proposed the first dynamic POR with a new property called *fairness* which is necessary to prevent dishonest clients from accusing an honest server about modifying their own stored data. As Stefanov et al. [18] understands, Zheng el al.'s scheme is only a PDP and not considered as a POR since their error correcting code is applied at the block level instead of the full file level, and looses the amplification property required to extract the full file during corruption, which is detected through the challenge-response protocol. Wang et al. [19] claim to improve the proof of retrievability model and achieve a dynamic PoR solution, but in fact only provide a dynamic PDP scheme. In their subsequent work wang et al.[13] proposed a privacy preserving public auditing system with a random mask technique which was good and fascinating in addressing data leakage problems of the protocol .

## 6. Conclusion

Doubtless, Cloud computing helps IT enterprises use various techniques to optimize and secure application performance in a cost-effective manner. But this suggestive way for decentralized application and access every time and everywhere to data, occasion and introduce new set of challenges and security problems that must consider before transfer data to a cloud environment. Additionally, just because the software can run in a Virtual machine does not mean that it performs well in cloud environment necessarily. Thereupon, in cloud there are risks and hidden costs in managing cloud compliance. The key to successful cloud computing initiatives is achieving a balance between the business benefits and the hidden potential risks which can impact efficiency. Cloud providers often have several powerful servers and resources in order to provide appropriate services for their users but cloud is at risk similar to other Internet-based technology.

In this survey, we analyzed a large list of protocols for remote data integrity verification and compared them. The majority of the existing verification protocols are built with data integrity verification as primary objective and also other security primitives like guarantying data confidentiality or owner privacy [6] addressed. Moreover most of the protocols support dynamic operations in cloud storage. We have studied different data integrity proving schemes and understand the basic requirements to design a complete protocol such as security (unforgeability), unbounded use of queries, dynamicity, retrievability of data (in the sense that data recoverability during corruption), block less verification, and public verifiability. Most of previous works deal with static data which leads to a security flaw when it is tried to apply on a dynamic environment. Therefore we can see that designing efficient, secure and fully dynamic remote data integrity scheme with

recoverability feature is quit challenging and remain an open problem for the community in this area.

## References

[1] S. Roschke, et aI., "Intrusion Detection in the Cloud," presented at the Eighth IEEE International Conference on Dependable, AutonomIc and Secure Computing, Chengdu, China, 2009.

[2] C. Almond, "A Practical Guide to Cloud Computing Security," 27 August 2009

[3] http://cloudsecurity.trendmicro.com/

[4] T. Mather. (2011). Data Leakage Prevention and Cloud Computing.Available: http://www.kpmg.com/Globa1/Pages/default.aspx

[5] N. Mead, et ai, "Security quality requirements engineering(SQUARE) methodolgy," Carnegie Mellon Software Engineering Institute.

[6] Security Management in the Cloud. Availablehttp://mscerts.net/programming/Security%20Management%20in%20the%20Cloud.aspx

[7] z. Zorz, "Top 7 threats to cloud computing," 2010.

[8] G. Ateniese, *et al.*, "Provable data possession at untrusted stores," presented at the Proceedings of the 14th ACM conference on Computer and communications security, Alexandria, Virginia, USA, 2007.

[9] D. Boneh, *et al.*, "Aggregate and VerifiablyEncrypted Signatures from Bilinear MapsAdvances in Cryptology — EUROCRYPT 2003." vol.2656, E. Biham, Ed., ed: Springer Berlin /Heidelberg, 2003, pp. 641-641.

[10] E. Mykletun, *et al.*, "Authentication and integrity in outsourced databases," *Trans. Storage,* vol. 2, pp. 107-138, 2006.

[11] Juels and J. Burton S. Kaliski, "Pors: proofs of retrievability for large files," presented at the Proceedings of the 14th ACM conference on Computer and communications security, Alexandria, Virginia, USA, 2007.

[12] G. Ateniese, *et al.*, "Scalable and efficient provable data possession," presented at the Proceedings of the 4th international conference on Security and privacy in communication netowrks, Istanbul, Turkey, 2008.

[13] Erway, *et al.*, "Dynamic provable data possession," presented at the Proceedings of the 16th ACM conference on Computer and communications security, Chicago, Illinois, USA, 2009.

[14] K. D. Bowers, *et al.*, "HAIL: a high-availability and integrity layer for cloud storage," presented at the Proceedings of the 16th ACM conference on Computer and communications security, Chicago, Illinois, USA, 2009

[15] K. D. Bowers, *et al.*, "Proofs of retrievability: theory and implementation," presented at the Proceedings of the 2009 ACM workshop on Cloud computing security, Chicago, Illinois, USA, 2009.

[16] Y. Dodis, *et al.*, "Proofs of Retrievability via Hardness Amplification, Theory of Cryptography, ." vol. 5444, O. Reingold, Ed., ed: Springer Berlin / Heidelberg, 2009, pp. 109-127.

[17] Q. Zheng and S. Xu, "Fair and dynamic proofs of retrievability," presented at the Proceedings of the first ACM conference on Data and application security and privacy, San Antonio, TX, USA, 2011.

[18] E. Stefanov, *et al.*, "Iris: A Scalable Cloud File System with Efficient Integrity Checks."

[19] Q. Wang, *et al.*, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing; Computer Security – ESORICS 2009." vol. 5789, M. Backes and P. Ning, Eds., ed: Springer Berlin / Heidelberg, 2009, pp. 355-370.