A REVIEW ON VARIOUS DATA SECURITY ISSUES IN CLOUD COMPUTING ENVIRONMENT AND ITS SOLUTIONS

Balasubramanian V.¹ and Mala T.²
¹Department of Computer Science and Engineering, SSN College of Engineering, Chennai, India
²Department of Information Science and Technology, College of Engineering, Anna University, Chennai, India
Email: balasubramanv@ssn.edu.in

ABSTRACT
Cloud Computing is a relatively new computing model that provides on demand business and IT services over the Internet. One of the main concerns in adapting Cloud Computing is its security. When outsourcing the data and business application to a third party cloud causes the security and privacy issues critical. Cloud service users need to understand the risk of data breaches in the cloud environment. In this paper, a survey of the various cloud computing models, different security risks that affects the cloud environment in the area of confidentiality, integrity and computing on data is thoroughly investigated. This paper also provides the solutions for the different security issues due to the cloud service delivery models.

Keywords: cloud computing, security, confidentiality, integrity, cloud delivery model.

1. INTRODUCTION
Cloud computing is the newest term for computing as a utility. It has become a hot topic in both industry and academia. It represents a new business model and computing paradigm. It enables convenient, on-demand provisioning of computational and storage resources. The resources can be rapidly deployed with great efficiency (I. Foster et al., 2008).

Since the adaption of cloud computing is increasing, there is an explicit and constant effort to evaluate the current trends in security for such technology. It considers problems already identified with possible solutions. Also concerns are being raised about the security issues through the adoption of this new model.

1.1. Cloud characteristics
The cloud security alliance has summarized five essential characteristics and illustrates the relations and differences from traditional computing (P. Mell et al., 2011; Z. Xiao et al., 2013).

1.1.1 On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as and when needed automatically without requiring human interaction with the cloud service provider.

1.1.2 Broad network access: Services are available over the Internet and accessed through standard mechanisms by heterogeneous thin or thick client platforms like mobile phones, tablets, laptops, and workstations.

1.1.3 Resource pooling: The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model. The different physical and virtual resources are dynamically assigned and reassigned according to consumer demand. Examples of resources include: storage, processing, memory, network bandwidth and virtual machines.

1.1.4 Rapid elasticity: Capabilities can be elastically provisioned and released automatically, to scale rapidly out-ward and inward based on the demand.

1.1.5 Measured service: The service purchased by customers can be quantified and measured. For both the provider and customers, resource usage will be monitored, controlled, metered and reported.

1.2. Cloud supporting techniques
Cloud computing has leveraged a collection of existing techniques, such as Data Center Networking (DCN), Virtualization, distributed storage, MapReduce, web applications and services, etc.

1.2.1 Modern data center has been practically employed as an effective carrier of cloud environments (J. Dean et al., 2008). It provides massive computation and storage capability by composing thousands of machines with DCN techniques.

1.2.2 Virtualization technology has been widely used in cloud computing to provider dynamic resource allocation and service provisioning, especially in IaaS (J. Dean et al., 2008). With virtualization, multiple OSs can co-reside on the same physical machine without interfering each other.

1.2.3 MapReduce is a programming framework that supports distributed computing on mass data sets. This breaks large data sets down into small blocks that are distributed to cloud servers for parallel computing (J. Dean et al., 2008; D. Zissis et al., 2012). It speeds up the batch processing on massive data, which makes this become the preference of computation model for cloud vendors.

1.3. Cloud vulnerabilities
Although cloud computing’s benefits are tremendous, security and privacy concerns are the primary obstacles to wide adoption. Because cloud service providers (CSPs) are separate administrative entities, moving to the commercial public cloud deprives users of direct control over the systems that manage their data and applications.
Even if CSPs’ infrastructure and management capabilities are much more powerful and reliable than those of personal computing devices, the cloud platform still faces both internal and external security and privacy threats, including media failures, software bugs, malware, administrator errors and malicious insiders. Noteworthy outages and security breaches to cloud services appear from time to time: Apple’s iPad subscriber privacy leak (Techcrunch, 2010) Amazon S3’s recent downtime (Amazon, 2008), and Gmail’s mass email deletions (Techcrunch, 2006) are all such examples.

The main goal of this paper is to give a broad outline of various critical security challenges, to point out their importance, and to motivate further investigation for security solutions. We have also identified, classified, organized and quantified the main security concerns and their solutions.

2. CLOUD COMPUTING MODEL

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable resources like networks, servers, storage, applications, and services (J. Geelan, 2008; R. Buyya, 2009). The resources can be rapidly provisioned and released with great efficiency and minimal management.

2.1. Service models

This Cloud computing utilizes three delivery models by which different types of services are delivered to the end user. The delivery models are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) (Cloud Security Alliance, 2009; Youseff et al., 2005).

2.1.1 Infrastructure as a Service (IaaS): In this model, the cloud provider provides the physical processing, storage, networking, and other fundamental computing resources. Also manages the hosting environment and cloud infrastructure. It offers the consumer the capability to provision processing, storage, networks, and other computing resources, and allow to deploy and run arbitrary software, like operating systems and application software. The consumer has control over operating systems, storage, deployed applications, and select networking components. This model completely abstracts the hardware beneath it (Cloud Security Alliance, 2009).

2.1.2 Platform as a service (PaaS): The cloud provider manages the cloud infrastructure for the platform, and provisions tools and execution resources for the platform consumers to develop, test, deploy, and administer applications. Consumers have control over the applications and possibly the hosting environment settings, but cannot access the infrastructure. It provides an integrated set of developer environment that a developer can tap to build their applications without knowing what is going on underneath the service. It affords the consumer with the capability to deploy onto the cloud infrastructure; consumer created or acquired applications, produced using programming languages and tools supported by the provider. The consumer does not control the underlying cloud infrastructure including network, servers, operating systems, or storage. Consumer has control over the deployed applications and hosting environment configurations. (Cloud Security Alliance, 2009).

2.1.3 Software as a service (SaaS)

Here the cloud provider deploys, configures, maintains, and updates the operation of the software applications on a cloud infrastructure so that the services are provisioned at the expected service levels to cloud consumers. The cloud consumers have limited administrative control of the applications. It is a software deployment model where applications are remotely hosted by the service provider and made available to customers on demand over the internet. It offers the consumer with the capability to use the applications running on a cloud infrastructure. The applications are accessible from various client devices, through web browser. The consumer does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, storage, etc. (Cloud Security Alliance, 2009).

2.2. Cloud deployment models

To differentiate between cloud implementations considerably different in nature, a set of deployment models have been defined - similar to how the service models have been defined based on National Institute of Standards and Technology (NIST) (Mell and Grance, 2011).

2.2.1 Public cloud: This cloud infrastructure is accessible by the general public or a cluster of organizations. This system is hosted, managed and owned by an organization selling cloud services (D. Zissis et al, 2012).

2.2.2. Private cloud: This cloud infrastructure is operated solely for one organization. There are two types of private clouds (D. Zissis et al, 2012):

- The private internal cloud: The organization acquires the necessary hardware and maintains it for itself.
- The private external cloud: The organization pays a cloud provider to provide this as a service.

2.2.3 Community cloud: This cloud infrastructure functions for multiple organizations with a set of shared concerns - typically cooperative, mission- or domain specific concerns. The cloud environment is managed or hosted by the organizations, or any third party vendors (D. Zissis et al, 2012).

2.2.4 Hybrid cloud: This cloud infrastructure is a combination of two or more of the above models. The clouds composing a hybrid cloud remain unique entities and are bound together by technology allowing communication between the clouds (D. Zissis et al, 2012).
3. CLOUD SECURITY ISSUES

Security, in general related to important aspects of Confidentiality, Integrity and availability. These are the basic building blocks for secure systems. Several critical security challenges and its importance are discussed in this session. The security solutions for various data security challenges are discussed (K. Ren et al., 2012; Z. Xiao et al. 2013)

Figure-1 illustrates the potential security threats in cloud security and possible solutions to mitigate the threats. Figure-2 classifies the cloud data security into three important types: Confidentiality, Integrity and Computational cloud security. Figure-1 also suggests the available solutions to address these classified challenges.

Figure-1. Cloud Security issues and mitigation strategies.

Figure-2. Data security vulnerabilities and solutions.

3.1. Confidentiality

Confidentiality refers to any authorized parties having access to protected data. It implies that customer’s data and computational task are to be kept confidential from both cloud provider and other customers (Z. Xiao et al. 2013).

Threat 1: Data is moved to the cloud owing to its flexibility and cost efficiency. But, since the user no longer physically possesses the data, hence its confidentiality and integrity of the data is at risk.

Solution: Cryptography is the most employed practice to secure sensitive data, so data encryption technique is used on
the data before outsourcing. It protects data privacy and also prevents unsolicited access.

**Threat 2:** Cryptography makes deploying traditional data utilization service, such as plaintext keyword search over textual / non textual data difficult.

**Solution a:** The trivial solution is to download all the data and decrypt it locally. But it seems to be impractical, due to huge bandwidth cost.

**Solution b:** Searchable encryption technique, which has gained recent attention allows search on encrypted data. Here at a higher level, a searchable encryption scheme employs a prebuilt encrypted search index that lets users with appropriate tokens securely search over the encrypted data via keywords without first decrypting it. It is particularly challenging because to meet performance and scalability (Ateniese et al., 2007).

**Threat 3:** Cross Virtual Machine via Side Channel Attack Side-channel attacks have played a major role in the history of cryptography. Usually, these attacks occur when a machine leaks details of its internal operation through some unexpected vector-for example, computation time or electromagnetic emissions. The cloud environment offers a bonanza of potential side channels because different VMs share physical resources-for example, processor, instruction cache, or disk-on a single computer. If an attacking program can carefully monitor those resources’ behavior, it can theoretically determine what another program is doing with them. The existence of Cross-VM attacks in an Amazon EC2 platform has been reported. A Cross-VM attack exploits the nature of multi-tenancy, which enables that VMs belonging to different customers may co-reside on the same physical machine. Data confidentiality can also be breached unintentionally, due to data remanence. It is the residual representation of data that have been deleted or removed (Z. Xiao et al., 2013).

**Solution:** The problem mentioned can be solved by using the strategies given below: Since the attack is so elaborate, cloud users need not get panic of it (Z. Xiao et al., 2013; Ristenpart et al., 2011; Aviram et al., 2010)

a) Placement prevention intends to reduce the success rate of placement. Also to reduce the success rate of placement, cloud providers might let the users decide where to put their VMs.

b) Physical isolation enforcement, which can be incorporated in Service Level Agreements (SLA’s). Also the infrastructure be shared only with “friendly” VMs which are owned by the same customer or other trustworthy customers.

c) New cache designs, which can overcome the attack, which works by measuring the behavior of the shared instruction cache.

### 3.2. Data integrity

Another key aspect of Information Security is integrity. Integrity refers that data can be modified only by authorized parties or in authorized ways. It refers to protecting data from unauthorized deletion, modification or manipulation. By preventing unauthorized access, organizations can achieve greater data confidentiality and system integrity (Ateniese et al., 2007; Ateniese et al., 2008; C. Wang et al., 2011). Additionally, such mechanisms offer the greater visibility into determining who or what may have altered data or system information, potentially affecting their integrity. It implies that data should be honestly stored on cloud servers, and any violations should be detected. That is, data lost, altered, or compromised should be detected.

**Threat 1:** Data loss: When data service is outsourced to the cloud, then protecting its integrity and long-term storage correctness arises. Outsourcing data to the cloud is economically attractive for long term, large scale storage. There is no guarantee for data integrity. Since users no longer locally possess their data, they can’t utilize traditional cryptographic primitives to protect its correctness. In cloud storage, applications deliver storage as a service. Servers keep large amounts of data, and some of it might be accessed on rare occasions. There is a possible threat that data may be lost or modified maliciously or accidentally (Ateniese et al., 2007; Ateniese et al., 2008). It can happen because of error during regular data backup and restore or data migration.

**Solution 1:** Provable Data Possession [PDP]: Integrity checking on data is a long-term research topic. Provable data possession accidentally (Ateniese et al., 2007; Ateniese et al., 2008; C. Wang et al., 2011) is a light weight remote data integrity checking model. This idea consists of the client computing a hash value for file \( F \) with a key \( k \) (i.e., \( h(k, F) \)) and subsequently sending \( F \) to the server. Once the client finds a necessity to check the file, it releases \( k \) and sends \( k \) to the server, which is subsequently asked to recompute the hash value, based on the \( F \) and \( k \); after this, the server replies to the client with the hash result for comparison. The client can initiate multiple checks by keeping different keys and hash values. This approach provides strong proof that the server still retains \( F \). It is a challenge-response scheme.

The client can have a meta data about the file i.e., hash value for the file, message digest or homomorphic tags on the client side, which can be used for verification purposes subsequently, for the data to be sent to the cloud server. Once the client feels a necessity to check the data integrity at a later time, he/she sends a challenge to the cloud server, which will respond with a message based on the data content. After comparing the reply and the local meta-data, the client is able to prove whether the integrity of the data is intact or violated. Normally PDP is applicable only to static files.

**Solution 2:** Proof of Retrievability [POR]: Proof of Re-trievability (PoR) is also a light weight protocol and it attempts to minimize the storage in client and server side (Juels and Kaliski, 2007; Dodis et al., 2009). The user stores only a key, which is used to encode the file \( F \) in order to get the encrypted file \( F' \). The task is that a set of sentinel values are embedded into \( F' \), and the server only stores \( F' \) without knowing where the sentinels may be. The sentinels are indistinguishable from regular data blocks. In the challenge and response protocol, the server is asked to return a certain subset of sentinels in \( F' \). If the server has tampered with or deleted \( F' \), there is high probability that certain sentinels are also corrupted or lost; this causes the server to be unable to
generate a complete proof for the original file. Therefore, a client has evidence to prove that the server has corrupted the file.

**Solution 3:** Scalable PDP: Scalable PDP adopts symmetric key encryption instead of public-key to reduce computation overhead. Scalable PDP has added dynamic operations on remote data.

**Solution 4:** Dynamic PDP: It support full dynamic operations (e.g., append, insert, modify, and delete). The purpose of dynamic operations is to enable authenticated insert and delete functions with rank-based authenticated directories that are built on a skip list.

### 3.3. Computational security

Another fundamental service enabled within the cloud paradigm is computation outsourcing. By outsourcing workloads to the cloud, users’ computational power is no longer limited by their resource constrained devices. Users can make use of unlimited computing resources in a pay-per-use model.

**Threat 1:** Computation outsourcing security: However, current outsourcing practice operates in plaintext that is, it reveals both data and computation results to the commercial public cloud (Z. Xiao *et al.*, 2013). This can raise big security concerns, especially when the out sourced computation work load’s contain sensitive information, such as personal health information. Furthermore, the cloud’s operational details aren’t transparent enough to user’s and also the cloud can behave unfaithfully and return incorrect results.

**Solution:** A recent breakthrough in fully homomorphic encryption (FHE) has shown the general results of secure computation outsourcing to be viable in theory atleast now. Homomorphic encryption (P. Paillier, 1999) is a type of encryption that allows computations to take place on the cipher text to get the cipher text and it is the same result as the computations carried out on the plain text. Usually the homomorphic function supports either addition or multiplication. If a particular homomorphic function supports both of them, then it is called fully homomorphic encryption. The goal of homomorphic function is that the data is encrypted and is sent to the receiver without decryption of the text. When the information is stored in the cloud as c(m), that is the cipher text of m, calculations on the encrypted is done as f(c(m)), which means the calculations are done on c(m) and the information is not revealed by decrypting it on the cloud. Let f be a function. Let a and b be two values. The function f is said to be homomorphic over multiplication if it satisfies the following property: 

\[ f(a \cdot b) = f(a) \cdot f(b) \]

Similarly f is said to be homomorphic over addition if it satisfies the following property: 

\[ f(a + b) = f(a) + f(b) \]

But applying this general mechanism to everyday computing tasks is still far from practical due to FHE operations extremely high complexity.

### 3.4 Security issues in service delivery models

A different security issue that has emanated due to the service delivery models has been presented in this section. Cloud computing utilizes three delivery models by which different types of services are delivered to the end user (N. Gonzalez *et al.*, 2011). The three delivery models are the SaaS, PaaS and IaaS which provide infrastructure resources, application platform and software as services to the consumer. Figure-3 illustrates the security issues in different cloud service delivery models.

![Figure-3. Security issues in service delivery models.](image-url)
3.4.1 Security issues in SaaS

It is a software deployment model where applications are remotely hosted by the service provider and made available to customers on demand, over the Internet. It offers the customers with significant benefits, such as improved operational efficiency and reduced costs. SaaS is rapidly emerging as the dominant delivery model. Here the client has to depend on the provider for proper security measures. The provider must do the work to keep multiple users’ from seeing each other’s data, since it employs multitenant. So it becomes difficult to the user to ensure that right security measures are in place. Also in SaaS, the cloud service provider will be substituting new software applications for old ones and hence achieving a successful data migration is also needs to be taken care.

The enterprise data needs to be secured; it is done by using strong encryption techniques and by fine grain based access control. Cross VM placement can be avoided by proper VM placement and co-residence detection. Data flow from the enterprise to the SaaS application and SaaS storage should be secured to protect Man-in-the-Middle attack, IP spoofing, packet sniffing. It is achieved by using strong network traffic encryption techniques like Secure Socket Layer (SSL) and Transport Layer Security (TLS).

Data locality can be provided reliably to the SaaS consumer by Service level agreement (SLA) to ensure where the data is securely stored. As a result of multi-tenancy, multiple users’ data are stored at the same location. Intrusion needs to be prevented and it can be achieved by proper VM placement and identifying co-resident VM.

Data access risk can be addressed by incorporating specific access policies in the SaaS application itself. Web application security addresses the programming interface design through which virtualized system and resources are accessed. ‘Verizon Business data breach investigation report’ reported 59% of breaches involves API hacking. The report reveals that external criminals pose the greatest threat (73%), but they can achieve least impact (30,000 records compromised), whereas insiders pose least risk (18%), but achieved 3, 75, 000 records compromised. (N. Gonzalez et al., 2011)

3.4.2 Security issues in PaaS

PaaS facilitates deployment of cloud-based applications as a service, without the cost of buying and maintaining the underlying hardware and software (N. Gonzalez et al., 2011). It depends on a secure and reliable network and secure web browser. PaaS security comprises of two types: Security of the PaaS platform itself, normally provided by cloud service provider; and Security of customer applications deployed on a PaaS platform.

3.4.3 Security issues in IaaS

IaaS provides resources such as servers, storage, networks, and other computing resources in the form of virtualized systems, which can be accessed through the Internet. Users are entitled to run any software with full control on the resources allocated to them. With IaaS, cloud users have better control over the security Compared to the other models. Cloud service provider should ensure there is no security hole in the virtual machine monitor (N. Gonzalez et al., 2011). User controls the software running in their virtual machines, and they are capable to configure security policies accordingly. But the underlying compute, net-work, and storage infrastructure is controlled by cloud service provider. IaaS providers must take efforts to secure their systems to minimize the threats that arise from creation, communication, monitoring, modification, and mobility.

Here is some of the security issues associated to IaaS.

Virtualization: It allows users to create, copy, share, mi-grate, and roll back virtual machines, which may allow them to run a variety of applications. Virtual machine security becomes as important as physical machine security, and any flaw in either one may affect the other. The Virtual Machine Monitor (VMM) or hypervisor is responsible for virtual machines isolation; therefore, if the VMM is compromised, its virtual machines may potentially be compromised as well.

4. CONCLUSIONS

Security challenges and the privacy of data are the major obstacles for the success of cloud computing. We have performed a systematic review of security issues for cloud environments where we enumerated the main cloud threats and vulnerabilities. This article is intended as a call for action to motivate further investigation of many challenging security issues. The security issues should be well understood and the solutions suggested in this paper should be implemented in a proper way to achieve cloud environment more secure and safe. We hope this review will be helpful in shaping the future research works in the area of cloud data security.

REFERENCES


