UNIT-IV

Cloud Security

Cloud computing security refers to the set of procedures, processes and standards designed to provide information security assurance in a cloud computing environment.

Cloud computing security addresses both physical and logical security issues across all the different service models of software, platform and infrastructure. It also addresses how these services are delivered (public, private or hybrid delivery model).

CLOUD INFORMATION SECURITY OBJECTIVES

Developing secure software is based on applying the secure software design principles that form the fundamental basis for software assurance. software assurance as "the basis for gaining justifiable confidence that software will consistently exhibit all properties required to ensure that the software, in operation, will continue to operate dependably despite the presence of sponsored (intentional) faults.

The Data and Analysis Center for Software (DACS) requires that software must exhibit the following three properties to be considered secure:

Dependability — Software that executes predictably and operates correctly under a variety of conditions, including when under attack or running on a malicious host.

Trustworthiness — Software that contains a minimum number of vulnerabilities or no vulnerabilities or weaknesses that could sabotage the software's dependability.

Survivability (Resilience) — Software that is resistant to or tolerant of attacks and has the ability to recover as quickly as possible with as little harm as possible.

Confidentiality, Integrity, and Availability

Confidentiality, integrity, and availability are sometimes known as the CIA triad of information system security, and are important pillars of cloud software assurance.

1) **Confidentiality**

   Confidentiality refers to the prevention of intentional or unintentional unauthorized disclosure of information. Confidentiality in cloud systems is related to the areas of intellectual property rights, covert channels, traffic analysis, encryption, and inference:

- **Intellectual property rights —** Intellectual property (IP) includes inventions, designs, and artistic, musical, and literary works. Rights to intellectual property are covered by copyright laws, which protect creations of the mind, and patents, which are granted for new inventions.

- **Covert channels —** A covert channel is an unauthorized and unintended communication path that enables the exchange of information. Covert channels can be accomplished through timing of messages or inappropriate use of storage mechanisms.

- **Traffic analysis —** Traffic analysis is a form of confidentiality breach that can be accomplished by analyzing the volume, rate, source, and destination of message traffic, even if it is encrypted. Increased message activity and high bursts of traffic can indicate a major event is occurring.

- **Encryption —** Encryption involves scrambling messages so that they cannot be read by an unauthorized entity, even if they are intercepted. The amount of effort (work factor) required to decrypt the message is a function of the strength of the encryption key and the robustness and quality of the encryption algorithm.

- **Inference —** Inference is usually associated with database security. Inference is the ability of an entity to use and correlate information protected at one level of security to uncover information that is protected at a higher security level.

2) **Integrity**

The concept of cloud information integrity requires that the following three principles are met:

- Modifications are not made to data by unauthorized personnel or processes.
- Unauthorized modifications are not made to data by authorized personnel or processes.
- The data is internally and externally consistent — in other words, the internal information is consistent both among all sub-entities and with the real-world, external situation.

3) **Availability**

Availability ensures the reliable and timely access to cloud data or cloud computing resources by the appropriate personnel. Availability guarantees that the systems are functioning properly when needed. In addition, this concept guarantees that the security services of the cloud system are in working order. A denial-of-service (DOS) attack is an example of a threat against availability.

CLOUD SECURITY SERVICES

Additional factors that directly affect cloud software assurance include authentication, authorization, auditing, and Accountability.

Authentication

Authentication is the testing or reconciliation of evidence of a user's identity. It establishes the user's identity and ensures that users are who they claim to be. For example, a user presents an identity (user ID) to a computer login screen and then has to provide a password. The computer system authenticates the user by verifying that the password corresponds to the individual presenting the ID.

Authorization

Authorization refers to rights and privileges granted to an individual or process that enable access to computer resources and information assets. Once a user's identity and authentication are established, authorization levels determine the extent of system rights a user can hold.

Auditing

To maintain operational assurance, organizations use two basic methods: system audits and monitoring. These methods can be employed by the cloud customer, the cloud provider, or both, depending on asset architecture and deployment. In addition, IT auditors might recommend improvements to controls, and they often participate in a system's development process to help an organization avoid costly reengineering after the system's implementation.

- A system audit is a one-time or periodic event to evaluate security.
- Monitoring refers to an ongoing activity that examines either the system or the users, such as intrusion detection.

Information technology (IT) auditors are often divided into two types: internal and external.

- ➢ Internal auditors typically work for a given organization, whereas external auditors do not. Internal auditors usually have a much broader mandate than external auditors, such as checking for compliance and standards of due care, auditing operational cost efficiencies, and recommending the appropriate controls.
- ➢ External auditors are often certified public accountants (CPAs) or other audit professionals who are hired to perform an independent audit of an organization's financial statements.

IT auditors typically audit the following functions:

- System and transaction controls
- Systems development standards
- Backup controls
- Data library procedures
- Data center security
- Contingency plans.

An audit trail or log is a set of records that collectively provide documentary evidence of processing, used to aid in tracing from original transactions forward to related records and reports, and/or backward from records and reports to their component source transactions. Audit trails may be limited to specific events or they may encompass all of the activities on a system.

Audit logs should record the following:

- The transaction's date and time
- Who processed the transaction
- At which terminal the transaction was processed
- Various security events relating to the transaction

Accountability

- ➢ Accountability is the ability to determine the actions and behaviors of a single individual within a cloud system and to identify that particular individual.
- ➢ Audit trails and logs support accountability and can be used to conduct later studies in order to analyze historical events and the individuals or processes associated with those events.
- ➢ Accountability is related to the concept of nonrepudiation, wherein an individual cannot successfully deny the performance of an action.

CLOUD SECURITY DESIGN PRINCIPLES

Computer software was not written with security in mind; but because of the increasing frequency and sophistication of malicious attacks against information systems, modern software design methodologies include security as a primary objective. The goal is to have a system that is secure enough for everyday use while exhibiting reasonable performance and reliability.

11 security design principles:

- Least privilege
- Separation of duties
- Defense in depth
- Fail safe
- Economy of mechanism
- Complete mediation
- Open design
- Least common mechanism
- Psychological acceptability
- Weakest link
- Leveraging existing components

Least Privilege

- ➢ The principle of least privilege maintains that an individual, process, or other type of entity should be given the minimum privileges and resources for the minimum period of time required to complete a task.
- ➢ This approach reduces the opportunity for unauthorized access to sensitive information.

Separation of Duties

- ➢ Separation of duties requires that completion of a specified sensitive activity or access to sensitive objects is dependent on the satisfaction of a plurality of conditions.
- ➢ For example, an authorization would require signatures of more than one individual, or the arming of a weapons system would require two individuals with different keys.

Defense in Depth

- ➢ Defense in depth is the application of multiple layers of protection wherein a subsequent layer will provide protection if a previous layer is breached.

The defense-in-depth strategy as defined in IATF (Information Assurance Technical Framework) promotes application of the following information assurance principles:

- **Defense in multiple places —** Information protection mechanisms placed in a number of locations to protect against internal and external threats.
- **Layered defenses —** A plurality of information protection and detection mechanisms employed so that an adversary or threat must negotiate a series of barriers to gain access to critical information.
- **Security robustness —** An estimate of the robustness of information assurance elements based on the value of the information system component to be protected and the anticipated threats.
- **Deploy KMI/PKI —** Use of robust key management infrastructures (KMI) and public key infrastructures (PKI).
- **Deploy intrusion detection systems —** Application of intrusion detection mechanisms to detect intrusions, evaluate information, examine results, and, if necessary, take action

Fail Safe

- ➢ Fail safe means that if a cloud system fails it should fail to a state in which the security of the system and its data are not compromised.
- ➢ One implementation would be to make a system default to a state in which a user or process is denied access to the system. A complementary rule would be to ensure that when the system recovers, it should recover to a secure state and not permit unauthorized access to sensitive information.
- ➢ In the situation where system recovery is not done automatically, the failed system should permit access only by the system administrator and not by other users, until security controls are reestablished.

Economy of Mechanism

- ➢ Economy of mechanism promotes simple and comprehensible design and implementation of protection mechanisms, So that unintended access paths do not exist or can be readily identified and eliminated.

Complete Mediation

- ➢ In complete meditation, every request by a subject to access an object in a computer system must undergo a valid and effective authorization procedure.
- ➢ This mediation must not be suspended or become capable of being bypassed, even when the information system is being initialized, undergoing shutdown, being restarted, or is in maintenance mode.

Complete mediation entails the following:

1. Identification of the entity making the access request
2. Verification that the request has not changed since its initiation
3. Application of the appropriate authorization procedures
4. Reexamination of previously authorized requests by the same entity

Open Design

- ➢ There has always been an ongoing discussion about the merits and strengths of security designs that are kept secret versus designs that are open to scrutiny and evaluation by the community at large.
- ➢ A good example is an encryption system. Some feel that keeping the encryption algorithm secret makes it more difficult to break. The opposing philosophy believes that exposing the algorithm to review and study by experts at large while keeping the encryption key secret leads to a stronger algorithm because the experts have a higher probability of discovering weaknesses in it.

For most purposes, an open-access cloud system design that has been evaluated and tested by a myriad of experts provides a more secure authentication method than one that has not been widely assessed.

Least Common Mechanism

- ➢ This principle states that a minimum number of protection mechanisms should be common to multiple users, as shared access paths can be sources of unauthorized information exchange.
- ➢ Shared access paths that provide unintentional data transfers are known as covert channels. Thus, the least common mechanism promotes the least possible sharing of common security mechanisms.

Psychological Acceptability

- Psychological acceptability refers to the ease of use and intuitiveness of the user interface that controls and interacts with the cloud access control mechanisms.
- Users must be able to understand the user interface and use it without having to interpret complex instructions.

Weakest Link

- “A chain is only as strong as its weakest link,” the security of a cloud system is only as good as its weakest component. Thus, it is important to identify the weakest mechanisms in the security chain and layers of defense, and improve them so that risks to the system are mitigated to an acceptable level.

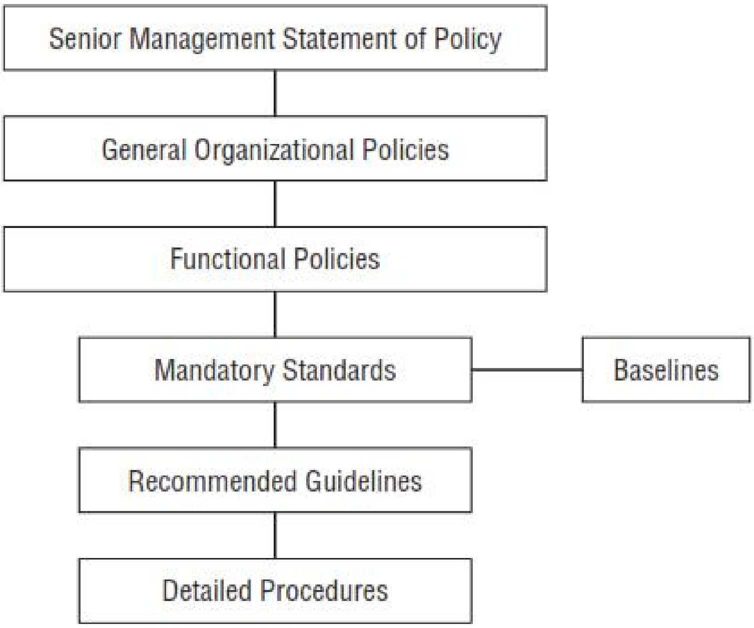
Leveraging Existing Components

- Another approach that can be used to increase cloud system security by leveraging existing components is to partition the system into defended subunits. Then, if a security mechanism is penetrated for one sub-unit, it will not affect the other sub-units, and damage to the computing resources will be minimized.

CLOUD SECURITY POLICY IMPLEMENTATION

Security policies are the foundation of a sound security implementation. Often organizations will implement technical security solutions without first creating this foundation of policies, standards, guidelines, and procedures, unintentionally creating unfocused and ineffective security controls.

A policy is one of those terms that can mean several things. For example, there are security policies on firewalls, which refer to the access control and routing list information. Standards, procedures, and guidelines are also referred to as policies in the larger sense of a global information security policy.



Implementation Issues

Cloud software security requirements are a function of policies such as system security policies, software policies, and information system policies.

Important areas addressed by a software system’s cloud security policy include the following:

- Access controls
- Data protection
- Confidentiality
- Integrity
- Identification and authentication
- Communication security
- Accountability

Security policy functional requirement:

- Derive the detailed functional requirements, e.g., “The server should return public-access Web pages to any browser that requests those pages.”
- Identify the related constraint requirements, e.g., “The server should return restricted Web pages only to browsers that are acting as proxies for users with authorized privileges sufficient to access those Web pages.”
- Derive the functional security requirements, e.g., “The server must authenticate every browser that requests access to a restricted Web page.”
- Identify the related negative requirements, e.g., “The server must not return a restricted Web page to any browser that it cannot authenticate.”

Source of inputs to secure software policies which specifies the following items:

System and Services Acquisition — “Organizations must (i) employ system development life cycle processes that incorporate information security considerations; (ii) employ software usage and installation restrictions; and (iii) ensure that third-party providers employ adequate security measures to protect information, applications, and/or services outsourced from the organization.”

System and Communications Protection — “Organizations must . . . (i) employ architectural designs, software development techniques, and systems engineering principles that promote effective information security within organizational information systems.”

System and Information Integrity — “Organizations must: (i) identify, report, and correct information and information system flaws in a timely manner; (ii) provide protection from malicious code at appropriate locations within organizational information systems.”

Three main objectives common to all system security policies

- They must allow authorized access and connections to the system while preventing unauthorized access or connections, especially by unknown or suspicious actors.
- They must enable allowable reading, modification, destruction, and deletion of data while preventing unauthorized reading (data leakage), modification (data tampering), destruction (denial of service), or deletion (denial of service).
- They must block the entry of content (user input, executable code, system commands, etc.) suspected of containing attack patterns or malicious logic that could threaten the system’s ability to operate according to its security policy and its ability to protect the information.

Policy Types

1. Senior Management statement of policy
2. Regulatory policy
3. Advisory Policy
4. Informative policy

Decomposing Critical Security Issues into Secure Cloud Software Requirements

An information system security policy addresses the critical issues of confidentiality, integrity, availability, identification, authentication, authorization, and auditing; and decomposes their elements into the following secure software requirements.

Confidentiality

Confidentiality in a cloud system policy is concerned with protecting data during transfers between entities. A policy defines the requirements for ensuring the confidentiality of data by preventing the unauthorized disclosure of information being sent between two end points. The policy should specify who can exchange information and what type of data can be exchanged. These policy statements should translate into requirements that address the following:

- Mechanisms that should be applied to enforce authorization
- What form of information is provided to the user and what the user can view
- The means of identity establishment
- What other types of confidentiality utilities should be used

Integrity

A cloud policy has to provide the requirements for ensuring the integrity of data both in transit and in storage. It should also specify means to recover from detectable errors, such as deletions, insertions, and modifications. The means to protect the integrity of information include access control policies and decisions regarding who can transmit and receive data and which information can be exchanged. Derived requirements for integrity should address the following:

- Validating the data origin
- Detecting the alteration of data
- Determining whether the data origin has changed

Availability

Cloud policy requirements for availability are concerned with denying illegitimate access to computing resources and preventing external attacks such as denial-of-service attacks. Additional issues to address include attempts by malicious entities to control, destroy, or damage computing resources and deny legitimate access to systems. While availability is being preserved, confidentiality and integrity have to be maintained. Requirements for this category should address how to ensure that computing resources are available to authorized users when needed.

Authentication and Identification

A cloud system policy should specify the means of authenticating a user when the user is requesting service on a cloud resource and presenting his or her identity. The authentication must be performed in a secure manner. Strong authentication using a public key certificate should be employed to bind a user to an identity. Some corresponding requirements include the following:

- Mechanisms for determining identity
- Binding of a resource to an identity
- Identification of communication origins
- Management of out-of-band authentication means
- Reaffirmations of identities

Authorization

After authentication, the cloud system policy must address authorization to allow access to resources, including the following areas:

- A user requesting that specified services not be applied to his or her message traffic
- Bases for negative or positive responses
- Specifying responses to requests for services in a simple and clear manner
- Including the type of service and the identity of the user in an authorization to access services
- Identification of entities that have the authority to set authorization rules between users and services

Auditing

The auditing of a cloud system has characteristics similar to auditing in the software development life cycle (SDLC) in that the auditing plan must address the following:

- Determination of the audit’s scope
- Determination of the audit’s objectives
- Validation of the audit plan
- Identification of necessary resources
- Conduct of the audit
- Documentation of the audit
- Validation of the audit results
- Report of final results

CLOUD COMPUTING SECURITY CHALLENGES

1) Security Policy Implementation

Security policies are the foundation of a sound security implementation. Often organizations will implement technical security solutions without first creating this foundation of policies, standards, guidelines, and procedures.

Senior Management Statement of Policy

- This is a general, high-level policy that acknowledges the importance of the computing resources to the business model.
- States support for information security throughout the enterprise; and commits to authorizing and managing the definition of the lower-level standards, procedures, and guidelines.

Regulatory Policies

- Regulatory policies are security policies that an organization must implement due to compliance, regulation, or other legal requirements.
- These companies might be financial institutions, public utilities, or some other type of organization

Advisory Policies

- Advisory policies are security policies that are not mandated but strongly suggested, perhaps with serious consequences defined for failure to follow them.
- A company with such policies wants most employees to consider these policies mandatory.

Informative Policies

- Informative policies are policies that exist simply to inform the reader. There are not implied or specified requirements, and the audience for this information could be certain internal (within the organization) or external parties.
- This does not mean that the policies are authorized for public consumption but that they are general enough to be distributed to external parties (vendors accessing an extranet, for example) without a loss of confidentiality.

2) Computer Security Incident Response Team (CSIRT)

CSIRT is a structured incident-handling program of intrusion detection and response, a Computer Emergency Response Team (CERT) or computer security incident response team (CSIRT) is commonly created. The main tasks of a CSIRT are as follows:

- Analysis of an event notification
- Response to an incident if the analysis warrants it
- Escalation path procedures
- Resolution, post-incident follow-up, and reporting to the appropriate parties

The prime directive of every CIRT is incident response management, which reflects a company's response to events that pose a risk to its computing environment.

- CIRT consist of following:
- Coordinating the notification and distribution of information pertaining to the incident to the appropriate parties through a predefined escalation path.
- Mitigating risk to the enterprise by minimizing the disruptions to normal business activities and the costs associated with remediating the incident.
- Assembling teams of technical personnel to investigate the potential vulnerabilities and resolve specific intrusions.

Additional examples of CIRT activities are:

- Management of the network logs, including collection, retention, review, and analysis of data
- Management of an incident's resolution, management of a vulnerability's remediation, and post-event reporting to the appropriate parties.
- Response includes notifying the appropriate parties to take action in order to determine the extent of an incident's severity and to remediate the incident's effects.
- Provide the ability to respond quickly and effectively

- Contain and repair the damage from incidents. When left unchecked, malicious software can significantly harm an organization's computing

3) Virtualization Security Management

The global adoption of virtualization is a relatively recent event, threats to the virtualized infrastructure are evolving just as quickly. Historically, the development and implementation of new technology has preceded the full understanding of its inherent security risks, and virtualized systems are no different.

Virtual Threats

Some threats to virtualized systems are general in nature, as they are inherent threats to all computerized systems (such as denial-of-service, or DoS, attacks).

Various organizations are currently conducting security analysis and proof of-concept (PoC) attacks against virtualized systems, and security in virtual environments highlights some of the vulnerabilities exposed to any malicious-minded individuals:

- **Shared clipboard** Shared clipboard technology allows data to be transferred between VMs and the host, providing a means of moving data between malicious programs in VMs of different security realms.
- **Keystroke logging** Some VM technologies enable the logging of keystrokes and screen updates to be passed across virtual terminals in the virtual machine, writing to host files and permitting the monitoring of encrypted terminal connections inside the VM.
- **VM monitoring from the host** Because all network packets coming from or going to a VM pass through the host, the host may be able to affect the VM by the following:
  - Starting, stopping, pausing, and restart VMs
  - Monitoring and configuring resources available to the VMs, including CPU, memory, disk, and network usage of VMs
  - Adjusting the number of CPUs, amount of memory, amount and number of virtual disks, and number of virtual network interfaces available to a VM
  - Monitoring the applications running inside the VM
  - Viewing, copying, and modifying data stored on the VM's virtual disks
- **Virtual machine monitoring from another VM** Usually, VMs should not be able to directly access one another's virtual disks on the host. However, if the VM platform uses a virtual hub or switch to connect the VMs to the host, then intruders may be able to use a hacker technique known as "ARP poisoning" to redirect packets going to or from the other VM for sniffing.
- **Virtual machine backdoors** A backdoor, covert communications channel between the guest and host could allow intruders to perform potentially dangerous operations.

Virtual threat is classified into three levels of compromise:

- **Abnormally terminated** Availability to the virtual machine is compromised, as the VM is placed into an infinite loop that prevents the VM administrator from accessing the VM's monitor.
- **Partially compromised** The virtual machine allows a hostile process to interfere with the virtualization manager, contaminating stet checkpoints or over-allocating resources.
- **Totally compromised** The virtual machine is completely overtaken and directed to execute unauthorized commands on its host with elevated privileges.

Hypervisor Risks

The hypervisor is the part of a virtual machine that allows host resource sharing and enables VM/host isolation. Therefore, the ability of the hypervisor to provide the necessary isolation during intentional attack greatly determines how well the virtual machine can survive risk.

Vulnerabilities in Hypervisor

Rogue Hypervisor: Rootkits that target virtualization, and in particular the hypervisor, have been gaining traction in the hacker community. VM-based rootkits can hide from normal malware detection systems by initiating a "rogue" hypervisor and creating a cover channel to dump unauthorized code into the system.

External Modification of the Hypervisor

In additional to the execution of the rootkit payload, a poorly protected or designed hypervisor can also create an attack vector. Therefore, a self-protected virtual machine may allow direct modification of its hypervisor by an external intruder. This can occur in virtualized systems that don't validate the hypervisor as a regular process.

VM Escape

Due to the host machine's fundamentally privileged position in relationship to the VM, an improperly configured VM could allow code to completely bypass the virtual environment, and obtain full root or kernel access to the physical host. This would result in a complete failure of the security mechanisms of the system, and is called *VM escape*. Virtual machine escape refers to the attacker's ability to execute arbitrary code on the VM's physical host, by "escaping" the hypervisor.

LEGAL ISSUES IN CLOUD COMPUTING

The legal issues that arise in cloud computing are wide ranging. Significant issues regarding privacy of data and data security exist, specifically as they relate to protecting personally identifiable information of individuals, but also as they relate to protection of sensitive and potentially confidential business information either directly accessible through or gleaned from the cloud systems (e.g., identification of a company's customer by evaluating traffic across the network).

1) Data Privacy And Security Issues

Data Breach Notification Requirements

- Data breach is a loss of unencrypted electronically stored personal information. This information is usually some combination of name and financial information (e.g., credit card number, Social Security Number).
- Avoidance of a data breach is important to both cloud providers and users of cloud services because of the significant harm, both to the user and to the provider, when a breach occurs. From the user's viewpoint, if personal Information is compromised, there is a risk of identity theft and of credit or debit card fraud. From the provider's viewpoint, financial harm, potential for lawsuits, Federal Trade Commission (FTC) investigations, loss of customers, and damage to reputation are all likely results of when a data breach occurs.
- For purposes of data breach law, data in the cloud are treated no differently than any other electronically stored information. Cloud providers that have had their systems compromised will be required to notify affected persons and will have to coordinate with the cloud users who provided the data in order to do so.

2) Cloud Contracting Models

Licensing Agreements versus Services Agreements

- A traditional software license agreement is used when a licensor is providing a copy of software to a licensee for its use (which is usually non-exclusive). This copy is not being sold or transferred to the licensee, but a physical copy is being conveyed to the licensee.
- The software license is important because it sets forth the terms under which the software may be used by the licensee. The license protects the licensor against the inadvertent transfer of ownership of the software to the person or company that holds the copy.

3) Jurisdictional Issues Raised By Virtualization And Data Location

The geographical location of the data in a cloud computing environment will have a significant impact on the legal requirements for protection and handling of the data.

The Issues Associated with the Flexibility of Data-Location

- One of the benefits of cloud computing from the cloud provider's perspective is the ability of the cloud provider to move data among its available data center resources as necessary to maximize the efficiencies of it overall system. From a technology perspective, this ability to move data is a reasonably good solution to the problem of underutilized machines.
- In the cloud environment it is possible that the same data may be stored in multiple locations at the same time. For example, real time-transaction data may be in one geographic location while the backup or disaster recovery systems may be elsewhere. In fact, a few cloud providers (of which Amazon.com is one) are

allowing cloud customers of certain service offerings to choose whether their data are kept in a U.S. or European data center.

4) **Confidentiality and Government Access to Data.**

- Each jurisdiction has its own regime to protect the confidentiality of information. In the cloud environment, given the potential movement of data among multiple jurisdictions, the data housed in a jurisdiction is subject to the laws of that jurisdiction, even if its owner resides elsewhere.
- Given the inconsistency of confidentiality protection in various jurisdictions, a cloud user may find that its sensitive data are not entitled to the protection with which the cloud user may be familiar, or that to which it contractually agreed.
- A government's ability to access data is also directly connected to the jurisdiction in which the data reside. If the jurisdiction has laws that permit its government to get access to data that data may be subject to interception by the government.

5) **International Conflicts of Laws**

- The body of law known as "conflict of laws" acknowledges that the laws of different countries may operate in opposition to each other, even as those laws relate to the same subject matter.
- Every nation is sovereign within its own territory. That means that the laws of that nation affect all property and people within it, including all contracts made and actions carried out within its borders. When there is either (1) no statement of the law that governs a contract, (2) no discussion of the rules regarding conflicts of laws in the agreement, or (3) a public policy in the jurisdiction which mandates that the governing law in the agreement will be ignored.
- A cloud environment, the conflicts of laws issues make the cloud provider's decisions regarding cross-geography virtualization and multi-tenancy, the cloud user's lack of information regarding data location, and the potential issues with geographically diverse subcontractors highly relevant.

6) **The Physical Location of Your Data:** It is important to know that where your data is stored physically. Your data cloud be stored in any country and may not even you know that. The physical location raises the issues of physical governance and legal operation or storage of the data. The Customer must be aware of provisions of prevailing laws in that particular county. Data storage in a country with fewer laws or without laws for data protection. If a dispute arises, what will be the place of jurisdiction? In case of a conflict arises between customer and cloud vendor, which country's laws and court will settle the dispute.

7) **Liability of Cloud Vendor/Broker:** Liability issues include in odd situations, including the questions below:

- What if the data-center is hit by a disaster?
- Is there any liability coverage for breach of privacy?
- What can be done if data-center gets hacked?
- Who will take responsibility and liability for unexpected shut-down of services?

8) **Intellectual Property Rights:** Includes questions below:

- Is your data protected under intellectual property right laws?
- How secure are trade secrets?
- What could be done if intellectual property rights are compromised?
- Which third party can access the data?
- Who is responsible for providing protection of the privacy?

9) **Contractual Issues with Vendors:** To protect the interests of your business, it is extremely essential that you read the terms and conditions deeply and set on understanding of contract before signing up for cloud services.

If the cloud service provider provides a standard form of contract (which is a general practice), then you must be fully aware of all the terms and conditions written in it. This will save you from tricky surprises and you will be financially, mentally and legally prepared to save your business from unfavorable consequences of cloud computing.

Some issues with contract are mentioned below:

- Unclear software warranties
- Varied intellectual property rights in world
- Warranties and clause to protect customers
- Unclear jurisdiction/legal governance of data-center
- Restrictive data export regulations.
- Conflict in inter-country laws.
- Data storage in a country with fewer laws or without laws for data protection.

10) **Service-Level Agreements (SLAs): SLAs** are important in any cloud service contract, We have to give attention on these points:

- How is the availability calculated by the provider?
- What will be independent measurement of performance?
- How, much downtime should be acceptable?
- What procedure to be followed if service provider fails to provide satisfactory delivery of services?
- What procedure will be followed at the end-of-service for destruction of data from data-centers?

11) **Lack of Laws Dedicated to Cloud Computing:** Most of the countries don't have specialized rule, regulations and laws for dealing with legal issues in cloud computing. Most disputes and claims are settled by companies according to their own guidelines, most of which are clearly not favorable for customers. If we take the examples of developed countries like U.S.A. or E.U. countries, there some rule have been developed for internet and data protection. In the case of jurisdiction lied under the service provider's nation customer may feel helpless to get legal help in issues. In recent times new regulation and laws are being developed world wide and still there is the need of more specialized laws and legal regulatory institutory actions to develop a strong legal base to handle legal issues in cloud computing.

CLOUD COMPUTING SECURITY ARCHITECTURE

Architectural Considerations

A variety of factors affect the implementation and performance of cloud security architecture. There are general issues involving regulatory requirements, adherence to standards, security management, information classification, and security awareness. Then there are more specific architecturally related areas, including trusted hardware and software, providing for a secure execution environment, establishing secure communications, and hardware augmentation through microarchitectures.

General Issues

A variety of topics influence and directly affect the cloud security architecture. They include such factors as compliance, security management, administrative issues, controls, and security awareness. Compliance with legal regulations should be supported by the cloud security architecture.

Compliance

In a public cloud environment, the provider does not normally inform the clients of the storage location of their data. In fact, the distribution of processing and data storage is one of the cloud's fundamental characteristics. However, the cloud provider should cooperate to consider the client's data location requirements. In addition, the cloud vendor should provide transparency to the client by supplying information about storage used, processing characteristics, and other relevant account information. Another compliance issue is the accessibility of a client's data by the provider's system engineers and certain other employees.

This factor is a necessary part of providing and maintaining cloud services, but the act of acquiring sensitive information should be monitored, controlled, and protected by safeguards such as separation of duties.

Security Management

Security architecture involves effective security management to realize the benefits of cloud computation. Proper cloud security management and administration should identify management issues in critical areas such as access control, vulnerability analysis, change control, incident response, fault tolerance, and disaster recovery and business continuity planning.

Controls

The objective of cloud security controls is to reduce vulnerabilities to a tolerable level and minimize the effects of an attack. To achieve this, an organization must determine what impact an attack might have, and the likelihood of loss. Examples of loss are compromise of sensitive information, financial embezzlement, loss of reputation, and physical destruction of resources. The process of analyzing various threat scenarios and producing a representative value for the estimated potential loss is known as a risk analysis (RA). Controls function as countermeasures for vulnerabilities. There are many kinds of controls, but they are generally categorized into one of the following four types:

- **Deterrent controls:** Reduce the likelihood of a deliberate attack.
- **Preventative controls:** Protect vulnerabilities and make an attack unsuccessful or reduce its impact. Preventative controls inhibit attempts to violate security policy.
- **Corrective controls:** Reduce the effect of an attack.
- **Detective controls:** Discover attacks and trigger preventative or corrective controls. Detective controls warn of violations or attempted violations of security policy and include such controls as intrusion detection systems, organizational policies, video cameras, and motion detectors.

Information Classification

Major area that relates to compliance and can affect the cloud security architecture is information classification. The information classification process also supports disaster recovery planning and business continuity planning.

Information Classification Benefits

Employing information classification has several clear benefits to an organization engaged in cloud computing. Some of these benefits are as follows:

- It demonstrates an organization’s commitment to security protections.
- It helps identify which information is the most sensitive or vital to an organization.
- It supports the tenets of confidentiality, integrity, and availability as it pertains to data.
- It helps identify which protections apply to which information.
- It might be required for regulatory, compliance, or legal reasons.

Public data: Information that is similar to unclassified information; all of a company’s information that does not fit into any of the next categories can be considered public. While its unauthorized disclosure may be against policy, it is not expected to impact seriously or adversely the organization, its employees, and/or its customers.

Sensitive data: This information is protected from a loss of confidentiality as well as from a loss of integrity due to an unauthorized alteration. This classification applies to information that requires special precautions to ensure its integrity by protecting it from unauthorized modification or deletion. It is information that requires a higher-than-normal assurance of accuracy and completeness.

Private data: This classification applies to personal information that is intended for use within the organization. Its unauthorized disclosure could seriously and adversely impact the organization and/or its employees. For example, salary levels and medical information are considered private.

Confidential data: This classification applies to the most sensitive business information that is intended strictly for use within the organization. Its unauthorized disclosure could seriously and adversely impact the organization, its stockholders, its business partners, and/or its customers.

For example, information about new product development, trade secrets, and merger negotiations is considered confidential.

Security Awareness

Security awareness is often overlooked as an element affecting cloud security architecture because most of a security practitioner’s time is spent on controls, intrusion detection, risk assessment, and proactively or reactively administering security. Employees of both the cloud client and the cloud provider must be aware of the need to secure information and protect the information assets of an enterprise. An effective computer security awareness and training program requires proper planning, implementation, maintenance, and periodic evaluation.

The purpose of computer security awareness, training, and education is to enhance security by doing the following:

- Improving awareness of the need to protect system resources
- Developing skills and knowledge so computer users can perform their jobs more securely
- Building in-depth knowledge, as needed, to design, implement, or operate security programs for organizations and systems

A computer security awareness and training program should encompass the following seven steps

- Identify program scope, goals, and objectives.
- Identify training staff.
- Identify target audiences.
- Motivate management and employees.
- Administer the program.
- Maintain the program.
- Evaluate the program.

Identity Management and Access Control

Identity management and access control are fundamental functions required for secure cloud computing. The simplest form of identity management is logging on to a computer system with a user ID and password. However, true identity management, such as is required for cloud computing, requires more robust authentication, authorization, and access control.

Identity Management

Identification and authentication are the keystones of most access control systems. Identification is the act of a user professing an identity to a system, usually in the form of a username or user logon ID to the system. Identification establishes user accountability for the actions on the system. User IDs should be unique and not shared among different individuals. In many large organizations, user IDs follow set standards, such as first initial followed by last name, and so on. Authentication is verification that the user's claimed identity is valid, and it is usually implemented through a user password at logon. Authentication is based on the following three factor types:
Type 1 Something you know, such as a personal identification number (PIN) or password
Type 2 Something you have, such as an ATM card or smart card
Type 3 Something you are (physically), such as a fingerprint or retina scan

Passwords

Because passwords can be compromised, they must be protected. In the ideal case, a password should be used only once. This "one-time password," or OTP, provides maximum security because a new password is required for each new logon. A password that is the same for each logon is called a static password. A password that changes with each logon is termed a dynamic password.

Passwords can be provided by a number of devices, including tokens, memory cards, and smart cards.

- **Tokens** Tokens, in the form of small, hand-held devices, are used to provide passwords. The following are the four basic types of tokens:
- Static password tokens
- Synchronous dynamic password tokens, clock-based
- Synchronous dynamic password tokens, counter-based
- Asynchronous tokens, challenge-response

- **Memory Cards** Memory cards provide nonvolatile storage of information, but they do not have any processing capability. A memory card stores encrypted passwords and other related identifying information. A telephone calling card and an ATM card are examples of memory cards.

- **Smart Cards** Smart cards provide even more capability than memory cards by incorporating additional processing power on the cards. These credit-card-size devices comprise microprocessor and memory and are used to store digital signatures, private keys, passwords, and other personal information.

- **Biometrics** An alternative to using passwords for authentication in logical or technical access control is biometrics.

Access Control

Access control is intrinsically tied to identity management and is necessary to preserve the confidentiality, integrity, and availability of cloud data. These and other related objectives flow from the organizational security policy. This policy is a high-level statement of management intent regarding the control of access to information and the personnel who are authorized to receive that information. Three things that must be considered for the planning and implementation of access control mechanisms are threats to the system, the system's vulnerability to these threats, and the risk that the threats might materialize.

These concepts are defined as follows:

- **Threat** — An event or activity that has the potential to cause harm to the information systems or networks
- **Vulnerability** — A weakness or lack of a safeguard that can be exploited by a threat, causing harm to the information systems or networks
- **Risk** — The potential for harm or loss to an information system or network; the probability that a threat will materialize

Autonomic Security

Autonomic security refers to security techniques based on autonomic computing which is self-managed, reconfigurable according to changing conditions and self-healing. It offers capabilities that can improve the security of information system and cloud computing.

The ability of autonomic security to collect and interpret data and recommend or implement solutions can enhance security and provide recovery from harmful events.

Autonomic security system is self-managing, monitors changes 'that affect the system and maintains internal balances of processes associated with security. It has:

- Sensory input
- Decision making capabilities
- Ability to implement remedial actions
- Ability to maintain an equilibrium state of normal operations.

Examples of events that can be handled autonomously by system include the following:

- Malicious attacks
- Hardware or software faults
- Power failures
- Organizational policies
- Software updates
- Interactions among systems
- Unintentional operator errors

Characteristics of autonomic computing systems introduced by IBM are given:

- Self-awareness
- Self-configuring
- Self-optimizing
- Self-healing
- Self-protecting
- Context aware
- Open
- Anticipatory

- Autonomic security and protection techniques involve detection of harmful situation and taking actions that will mitigate the situation. These systems will be designed to predict problems from analysis of sensoiy inputs and initiate corrective actions.
- An autonomous system security response is based on network knowledge, capabilities of connected resources, information and complexity of situation as well as impact on affected application/component.
- The decision making element of autonomic computing can take actions such as changing the strength of required authentication or modifying encryption keys. According to current security position and context, the state of system can be changed and level of authorization can be modified immediately.

Guidelines for autonomous protection systems

- Minimize overhead requirements
- Be consistent with security policies
- Optimize security-related parameters.
- Minimize impact on performance.
- Minimize potential vulnerabilities.
- Conduct regression analysis
- Ensure that reconfiguration processes are secure

DATA SECURITY IN CLOUD COMPUTING

The concept of data security in cloud is taking information and making it secure, so that only certain user, designated for or allowed for access, can read, write and modify data. Application of data security is viewed as a series of options for easy and efficient access controls assuring the security of system, data and networks. Data security does not mean to build any impassable barriers for user, it is just deployed as the process of minimizing the effect of malicious attacks on data, by prevention and risk mitigation.

Business Continuity Planning/Disaster Recovery Planning

Business continuity planning (BCP) and disaster recovery planning (DRP) involve the preparation, testing, and updating of the actions required to protect critical business processes from the effects of major system and network failures.

From the cloud perspective, these important business processes are heavily dependent on cloud- based applications and software robustness and security. BCP comprises scoping and initiating the planning, conducting a business impact assessment (BIA), and developing the plan.

Designing, developing, and implementing a quality and effective BCP and DRP is a major undertaking, involving many person-hours and, in many instances, high hardware or software costs. These efforts and costs are worthwhile and necessary, but they impact a large number of organizational resources.

Disaster

A disaster is a rapidly occurring or unstoppable event that can cause suffering, loss of life, or damage.

- ➢ In many instances, the aftermath of a disaster can impact social or natural conditions for a long period of time.
- ➢ A DRP is a comprehensive statement of consistent actions to be taken before, during, and after a disruptive event that causes a significant loss of information systems resources.
- ➢ The number one priority of DRP is personnel safety and evacuation, followed by the recovery of data center operations and business operations and processes.

Specific areas that can be addressed by cloud providers include the following:

- Protecting an organization from a major computer services failure
- Providing extended backup operations during an interruption
- Providing the capability to implement critical processes at an alternate site
- Guaranteeing the reliability of standby systems through testing and simulations
- Returning to the primary site and normal processing within a time frame that minimizes business loss by executing rapid recovery procedures.
- Minimizing the decision-making required by personnel during a disaster
- Proving an organized way to make decisions if a disruptive event occurs
- Minimizing the risk to the organization from delays in providing service

A business continuity plan addresses the means for a business to recover from disruptions and continue support for critical business functions. It is designed to protect key business processes from natural or man-made failures or disasters and the resultant loss of capital due to the unavailability of normal business processes. A BCP includes a business impact assessment (BIA), which, in turn, contains a vulnerability assessment.

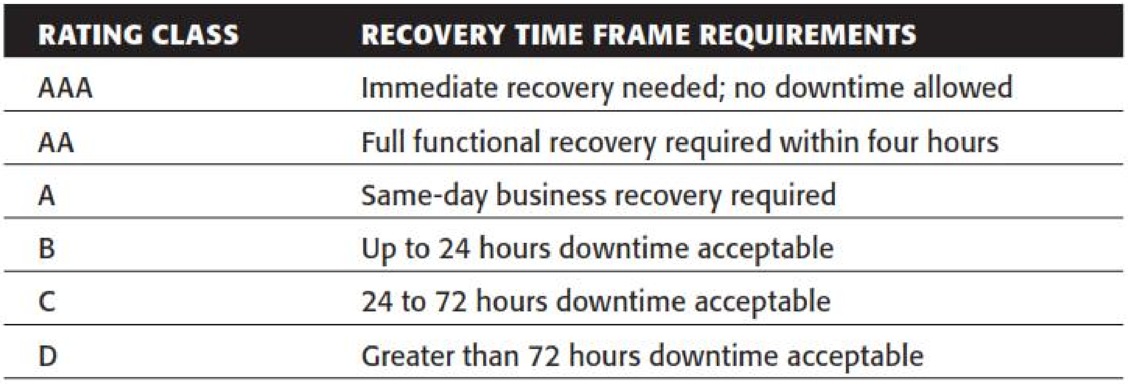
A BIA is a process used to help business units understand the impact of a disruptive event. A vulnerability assessment is similar to a risk assessment in that it contains both a quantitative (financial) section and a qualitative (operational) section.

DISASTER RECOVERY PLANNING

The primary objective of a disaster recovery plan is to provide the capability to implement critical processes at an alternate site and return to the primary site and normal processing within a time frame that minimizes loss to the organization by executing rapid recovery procedures. In many scenarios, the cloud platforms already in use by a customer are extant alternate sites. Disasters primarily affect availability, which impacts the ability of staff to access the data and systems, but it can also affect the other two tenets, confidentiality and integrity. In the recovery plan, a classification scheme such as the one shown in Table can be used to classify the recovery time-frame needs of each business function.

The DRP should address all information processing areas of the company:

- Cloud resources being utilized
- LANs, WANs, and servers
- Telecommunications and data communication links
- Workstations and workspaces
- Applications, software, and data
- Media and records storage
- Staff duties and production processes



RATING CLASS	RECOVERY TIME FRAME REQUIREMENTS
AAA	Immediate recovery needed; no downtime allowed
AA	Full functional recovery required within four hours
A	Same-day business recovery required
B	Up to 24 hours downtime acceptable
C	24 to 72 hours downtime acceptable
D	Greater than 72 hours downtime acceptable

Recovery Time frame Classification

Backup services are important elements in the disaster recovery plan. The typically used alternative services are as follows:

Mutual aid agreements: An arrangement with another company that might have similar computing needs. The other company may have similar hardware or software configurations or may require the same network data communications or Internet access.

Subscription services: Third-party commercial services that provide alternate backup and processing facilities. An organization can move its IT processing to the alternate site in the event of a disaster.

Multiple centers: Processing is spread over several operations centers, creating a distributed approach to redundancy and sharing of available resources. These multiple centers could be owned and managed by the same organization (in-house sites) or used in conjunction with a reciprocal agreement.

Service bureaus: Setting up a contract with a service bureau to fully provide all alternate backup-processing services. The disadvantages of this arrangement are primarily the expense and resource contention during a large emergency.

Disaster Recovery Plan Testing

The major reasons to test a disaster recovery plan are summarized as follows:

- To inform management of the recovery capabilities of the enterprise
- To verify the accuracy of the recovery procedures and identify deficiencies
- To prepare and train personnel to execute their emergency duties
- To verify the processing capability of the alternate backup site or cloud provider

BUSINESS CONTINUITY PLANNING

A BCP is designed to keep a business running, reduce the risk of financial loss, and enhance a company's capability to recover promptly following a disruptive event. The four principle components of a BCP are as follows:

- **Scope and plan initiation**: Creating the scope and other elements needed to define the plan's parameters.
- **Business impact assessment (BIA):** Assisting the business units in understanding the impact of a disruptive event. This phase includes the execution of a vulnerability assessment.
- **Business continuity plan development:** Using information collected in the BIA to develop the actual business continuity plan. This process includes the areas of plan implementation, plan testing, and ongoing plan maintenance.
- **Plan approval and implementation:** Obtaining the final senior management sign-off, creating enterprise wide awareness of the plan, and implementing a maintenance procedure for updating the Plan as needed.

The BIA

A key element of the BCP process is conducting a BIA. The purpose of a BIA is to create a document that outlines what impact a disruptive event would have on the business. The impact might be financial (quantitative) or operational (qualitative), such as the inability to respond to customer complaints.

A BIA has three primary goals:

- **Criticality prioritization:** Every critical business unit process must be identified and prioritized, and the impact of a disruptive event must be evaluated.
- **Downtime estimation:** The BIA is used to help estimate the maximum tolerable downtime (MTD) that the business can withstand and still remain viable; that is, what is the longest period of time a critical process can remain interrupted before the company can never recover? The BIA process often determines that this time period is much shorter than expected.
- **Resource requirements:** The resource requirements for the critical processes are also identified at this time, with the most time-sensitive processes receiving the most resource allocation.

A BIA generally involves four steps:

1. Gathering the needed assessment materials
2. Performing the vulnerability assessment
3. Analyzing the information compiled
4. Documenting the results and presenting recommendations

The Vulnerability Assessment

The vulnerability assessment is often part of a BIA. It is similar to a risk assessment but it is smaller than a full risk assessment and is focused on providing information that is used solely for the business continuity plan or disaster recovery plan.

The function of a vulnerability assessment is to conduct a loss impact analysis. Because there are two parts to the assessment, a financial assessment and an operational assessment, it is necessary to define loss criteria both quantitatively and qualitatively.

Quantitative loss criteria can be defined as follows:

- Incurring financial losses from loss of revenue, capital expenditure, or personal liability resolution
- Incurring additional operational expenses due to the disruptive event
- Incurring financial loss resulting from the resolution of violating contract agreements
- Incurring financial loss resulting from the resolution of violating regulatory or compliance requirements

Qualitative loss criteria can consist of the following:

- The loss of competitive advantage or market share
- The loss of public confidence or credibility, or incurring public embarrassment

Secure Remote Access

In order for cloud-based BCP/DRP to be effective, the cloud applications and data must be securely accessible from all parts of the globe. One solution is for the cloud vendor to establish a global traffic management system that provides the following customer services:

- Meets service-level agreements for availability and performance
- Regulates and controls traffic among virtual machines located at multiple data centers
- Maximizes speed and performance by directing traffic to the closest and most logical cloud data center.

These services have to be implemented and conducted in a secure environment to protect both the cloud consumer and cloud provider from compromises and attacks.

Integration into Normal Business Processes

Services provided by a cloud vendor at a remote location are, in almost all cases, isolated geographically from the customer's facilities. The cloud enterprise is strongly protected both physically and technically. At the consumer's site, if cloud processing and data storage are integrated into the daily routine of the business, recovery from a disruptive event at the user organization can be more rapid and involve less time and personnel.

RISK MITIGATION

- ➢ Risk mitigation is a systematic approach to reduce the extent of exposure to a risk and the probability of its occurrence.
- ➢ In cloud, risk mitigation is process of the selection and implementation of security controls to reduce the risk to a level acceptable to the cloud provider and customer.
- ➢ It is identification of ways to minimize or eliminate expected and conquered risks. Depending upon impact of risk and the level of effort for the mitigation strategies, it may be appropriate to initiate several mitigation activities.
- ➢ Mitigation strategy reflects an organizational perspective on what mitigations are employed and where the mitigations are applied to reduce risks to organizational operations and resources. Risk mitigation strategies are the 'primary links between organizational risk management process and security policies. Effective risk mitigation strategies consider the general placement and allocation of mitigations, the degree of intended mitigation, and cover mitigations at each level of organization.
- ➢ Risk mitigation is the final step of risk management process, it includes prioritization of risks, risk evaluation and implementation of appropriate risk-reducing controls recommended from the risk assessment process.

Risk mitigation has a major phase of potential risk treatment or potential risk mitigation after risks have been identified and assessed, all techniques to manage the risk fall into one of these four major categories:

- Risk avoidance (eliminate, withdraw from or not become involved)
- Risk reduction (optimize mitigate)
- Risk sharing (transfer, outsource or insure)
- Risk retention (accept and adjust)

The process of risk mitigation in cloud environment include these major goals:

- Preparing the system technically and managerially to face the threats.
- Prepare the proper risk management plan that include solutions for risk treatments.
- Reconfigurability of system is designed as per requirements.
- Minimize the effect of an intentional or unintentional disruptive event on the system under threat.
- Providing recommended solution for handling consequences.
- Prepare proper backup plan and alternative services options.

UNDERSTANDING AND IDENTIFICATION OF THREATS IN CLOUD

The security practices in cloud computing consists of answers to some critical questions, which arise from the basics of security concerns and principles. At start of the developing any security, it must be considered or asked that why and from whom we need security in cloud environment. This leads us to the process of understanding and identification of threats in clouds.

According to Cloud Security Alliance (CSA) recommendations we will identify the cloud threats in seven kind of different possible categories:

- **Abuse and nefarious use of cloud computing:** Cloud providers offer their customers the illusion of unlimited compute, storage and network facilities, where anyone can register and immediately can use cloud services. Some of the providers offer free of cost services based on some terms and conditions. By abusing this usage model, spammers, malicious code authors and other criminal mind peoples have been able to conduct their activities here. This leads to misuse and wastage of resources, and limited or compromised availability.
- **Insecure APIs:** Cloud providers expose a set of APIs that customers use to manage and interact with cloud services. The security and availability of general cloud services are depending upon security features of these APIs. These APIs must implement the basic access controls and authentications properly, otherwise it may lead to risk for system.
- **Malicious insiders:** Malicious insiders are hackers, spying employees or untrained unsophisticated users practicing malicious activities in cloud intentionally or unintentionally. The level of access granted to these can enable more threats to confidential data and parts of system.
- **Vulnerabilities of shared technology:** Cloud services provide scalability by sharing the infrastructure, but this infrastructure generally not designed to follow strong isolation properties for a multi-tenant architecture. To manage this gap, a virtualization hypervisor mediates access between guest OS and the physical resource. These hypervisors exhibit flaws that have enabled guest OS to gain in appropriate levels of controls or influence on underlying platform. These drawbacks become vulnerabilities for a shared technology.
- **Data loss of leakage:** There are many ways to comprise data security. Deletion or alteration of records without a backup is an example. Unlinking of a record, unreliable storage, unrecoverable media are also threats to data security. Loss of an encryption/ encoding key many result in effective destruction. Finally, unauthorized parties must be prevented from gaining access to sensitive data.
- **Account or Service hijacking:** Attack methods such as phishing, fraud and exploitation of software vulnerabilities still achieve results, not favorable for cloud. Credentials and passwords are often reused, which amplifies the impact of such attacks. If an attacker gains access to your credentials, he can monitor and misuse your transactions, information and return modified information or redirect your clients to the illegitimate sites. This also may affect your reputation.
- **Unknown risk profile:** Using cloud services provide financial and operational benefits, but on the other hand the security concerns are complicated by the fact that cloud deployments are driven by underlying benefits, so this leads to loss of security. Compromise in security must not be the cost of these benefits. Versions of software, code updates, security practices, vulnerabilities, intrusion attempts and security design are all important factors for estimating your companies' security posture. Security at low attention can result in unknown risk exposure. It may also impair the in-depth analysis required highly controlled and regulated operational.

SERVICE LEVEL AGREEMENT (SLA)

Service-level agreement provides a framework within which both seller and buyer of a service can pursue a profitable service business relationship. It outlines the broad understanding between the service provider and the service consumer for conducting business and forms the basis for maintaining a mutually beneficial relationship.

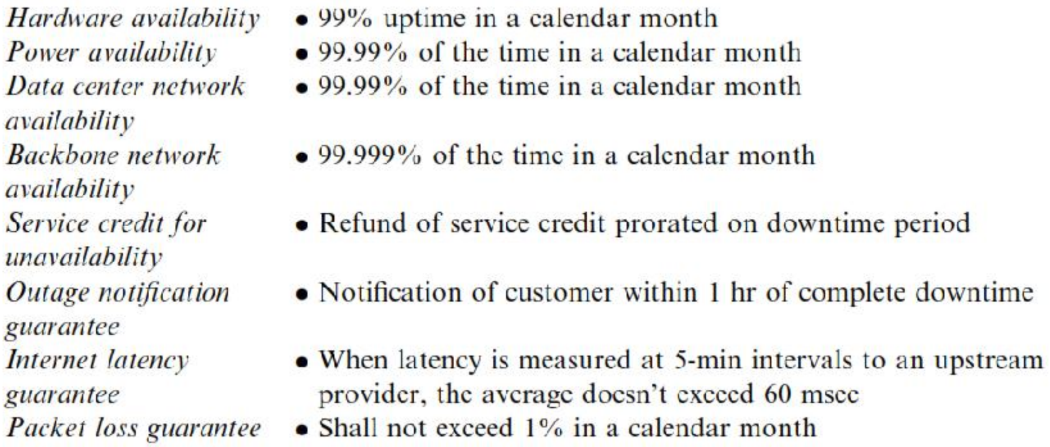
From a legal perspective, the necessary terms and conditions that bind the service provider to provide services continually to the service consumer are formally defined in SLA.

TYPES OF SLA

There are two types of SLAs from the perspective of application hosting.

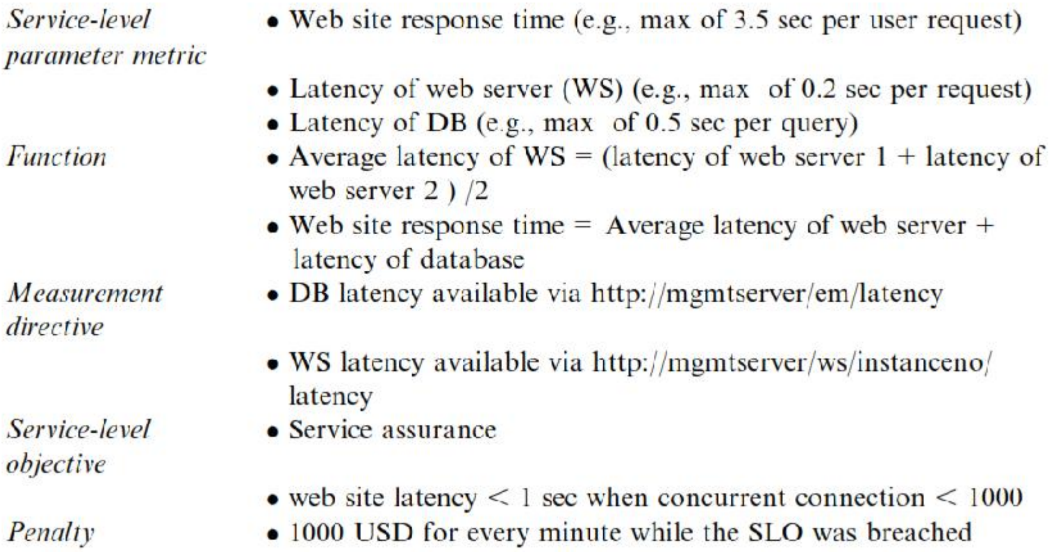
Infrastructure SLA. The infrastructure provider manages and offers guarantees on availability of the infrastructure, namely, server machine, power, network connectivity, and so on. Enterprises manage themselves, their applications that are deployed on these server machines. The machines are leased to the customers and are isolated from machines of other customers.
Service-level guarantees offered by infrastructure providers is shown in Table 1.

Application SLA. In the application co-location hosting model, the server capacity is available to the applications based solely on their resource demands. Hence, the service providers are flexible in allocating and de-allocating computing resources among the co-located applications.



Hardware availability	• 99% uptime in a calendar month
Power availability	• 99.99% of the time in a calendar month
Data center network availability	• 99.99% of the time in a calendar month
Backbone network availability	• 99.999% of the time in a calendar month
Service credit for unavailability	• Refund of service credit prorated on downtime period
Outage notification guarantee	• Notification of customer within 1 hr of complete downtime
Internet latency guarantee	• When latency is measured at 5-min intervals to an upstream provider, the average doesn't exceed 60 msec
Packet loss guarantee	• Shall not exceed 1% in a calendar month

Table 1. Infrastructural SLA



Service-level parameter metric	• Web site response time (e.g., max of 3.5 sec per user request)
	• Latency of web server (WS) (e.g., max of 0.2 sec per request)
	• Latency of DB (e.g., max of 0.5 sec per query)
Function	• Average latency of WS = (latency of web server 1 + latency of web server 2) /2
	• Web site response time = Average latency of web server + latency of database
Measurement directive	• DB latency available via http://mgmtserver/em/latency
	• WS latency available via http://mgmtserver/ws/instanceno/latency
Service-level objective	• Service assurance
	• web site latency < 1 sec when concurrent connection < 1000
Penalty	• 1000 USD for every minute while the SLO was breached

Table 2. Application SLA

- ➢ It is also possible for a customer and the service provider to mutually agree upon a set of SLAs with different performance and cost structure rather than a single SLA.
- ➢ The customer has the flexibility to choose any of the agreed SLAs from the available offerings. At runtime, the customer can switch between the different SLAs.

From SLA perspective there are multiple **challenges for provisioning infrastructure.**

1) The application is a black box to the MSP (Managed service provider) and the MSP has virtually no knowledge about the application runtime characteristics. Therefore, the MSP needs to determine the right amount of computing resources required for different components of an application at various workloads.

2) The MSP needs to understand the performance bottlenecks and the scalability of the application.
3) The MSP analyzes the application before it goes on-live. However, subsequent operations/enhancements by the customer's to their applications or auto updates beside others can impact the performance of the applications, thereby making the application SLA at risk.
4) The risk of capacity planning is with the service provider instead of the customer. If every customer decides to select the highest grade of SLA simultaneously, there may not be a sufficient number of servers for provisioning and meeting the SLA obligations of all the customers.

LIFE CYCLE OF SLA

Each SLA goes through a sequence of steps starting from identification of terms and conditions, activation and monitoring of the stated terms and conditions, and eventual termination of contract once the hosting relationship ceases to exist. Such a sequence of steps is called SLA life cycle and consists of the following five phases:

1. Contract definition
2. Publishing and discovery
3. Negotiation
4. Operationalization
5. De-commissioning

Contract Definition. Generally, service provider's define a set of service offerings and corresponding SLAs using standard templates. These service offerings form a catalog. Individual SLAs for enterprises can be derived by customizing these base SLA templates.
Publication and Discovery. Service provider advertises these base service offerings through standard publication media, and the customers should be able to locate the service provider by searching the catalog. The customers can search different competitive offerings and shortlist a few that fulfill their requirements for further negotiation.
Negotiation. Once the customer has discovered a service provider who can meet their application hosting need, the SLA terms and conditions needs to be mutually agreed upon before signing the agreement for hosting the application. This phase could be automated. For customized applications that are hosted on cloud platforms, this phase is manual. The service provider needs to analyze the application's behaviour with respect to scalability and performance before agreeing on the specification of SLA. At the end of this phase, the SLA is mutually agreed by both customer and provider and is eventually signed off.
Operationalization. SLA operation consists of SLA monitoring, SLA accounting, and SLA enforcement. SLA monitoring involves measuring parameter values and calculating the metrics defined as a part of SLA and determining the deviations. On identifying the deviations, the concerned parties are notified. SLA accounting involves capturing and archiving the SLA adherence for compliance. As part of accounting, the application's actual performance and the performance guaranteed as a part of SLA is reported. Apart from the frequency and the duration of the SLA breach, it should also provide the penalties paid for each SLA violation. SLA enforcement involves taking appropriate action when the runtime monitoring detects a SLA violation. Such actions could be notifying the concerned parties, charging the penalties besides other things.
De-commissioning. SLA decommissioning involves termination of all activities performed under a particular SLA when the hosting relationship between the service provider and the service consumer has ended. SLA specifies the terms and conditions of contract termination and specifies situations under which the relationship between a service provider and a service consumer can be considered to be legally ended.

SLA MANAGEMENT IN CLOUD

SLA management of applications hosted on cloud platforms involves five phases.
1. Feasibility
2. On-boarding
3. Pre-production
4. Production
5. Termination

1) Feasibility Analysis

MSP conducts the feasibility study of hosting an application on their cloud platforms. This study involves three kinds of feasibility: (1) technical feasibility, (2) infrastructure feasibility, and (3) financial feasibility.

The technical feasibility of an application implies determining the following:
1. Ability of an application to scale out.
2. Compatibility of the application with the cloud platform being used within the MSP's data center.
3. The need and availability of a specific hardware and software required for hosting and running of the application.
4. Preliminary information about the application performance and whether they can be met by the MSP.

The Infrastructure feasibility involves determining the availability of infrastructural resources in sufficient quantity so that the projected demands of the application can be met.

The financial feasibility study involves determining the approximate cost to be incurred by the MSP and the price the MSP charges the customer so that the hosting activity is profitable to both of them.

A feasibility report consists of the results of the above three feasibility studies. The report forms the basis for further communication with the customer. Once the provider and customer agree upon the findings of the report, the outsourcing of the application hosting activity proceeds to the next phase, called "onboarding" of application.

2) On-Boarding of Application

Once the customer and the MSP agree in principle to host the application based on the findings of the feasibility study, the application is moved from the customer servers to the hosting platform. Moving an application to the MSP's hosting platform is called on-boarding. As part of the on-boarding activity, the MSP understands the application runtime characteristics using runtime profilers. This helps the MSP to identify the possible SLAs that can be offered to the customer for that application. This also helps in creation of the necessary policies (also called rule sets) required to guarantee the SLOs (Service level objective) mentioned in the application SLA. The application is accessible to its end users only after the onboarding activity is completed.

On-boarding activity consists of the following steps:

a. Packing of the application for deploying on physical or virtual environments. Application packaging is the process of creating deployable components on the hosting platform (could be physical or virtual). Open Virtualization Format (OVF) standard is used for packaging the application for cloud platform.
b. The packaged application is executed directly on the physical servers to capture and analyze the application performance characteristics. It allows the functional validation of customer's application. Besides, it provides a baseline performance value for the application in nonvirtual environment. This can be used as one of the data points for customer's performance expectation and for application SLA. Additionally, it helps to identify the nature of application—that is, whether it is CPU-intensive or I/O intensive or network-intensive and the potential performance bottlenecks.
c. The application is executed on a virtualized platform and the application performance characteristics are noted again. Important performance characteristics like the application's ability to scale (out and up) and performance bounds (minimum and maximum performance) are noted.
d. Based on the measured performance characteristics, different possible SLAs are identified. The resources required and the costs involved for each SLA are also computed.
e. Once the customer agrees to the set of SLOs and the cost, the MSP starts creating different policies required by the data center for automated management of the application. This implies that the management system should automatically infer the amount of system resources that should be allocated/de-allocated to/from appropriate components of the application when the load on the system increases/decreases. These policies are of three types: (1) business, (2) operational, and (3) provisioning. Business policies help prioritize access to the resources in case of contentions. Business policies are in the form of weights for different customers or group of customers. Operational policies are the actions to be taken when different thresholds/conditions are reached. Also, the actions when thresholds/conditions/triggers on service-level parameters are breached or about to be breached are defined. The corrective action could be different types of provisioning such as scale-up, scale-down, scale-out, scale-in, and so on, of a particular tier of an application.

3) Preproduction

Once the determination of policies is completed as discussed in previous phase, the application is hosted in a simulated production environment. It facilitates the customer to verify and validate the MSP's findings on application's runtime characteristics and agree on the defined SLA. Once both parties agree on the cost and the terms and conditions of the SLA, the customer sign-off is obtained. On successful completion of this phase the MSP allows the application to go on-live.

4) Production

In this phase, the application is made accessible to its end users under the agreed SLA. However, there could be situations when the managed application tends to behave differently in a production environment compared to the preproduction environment. This in turn may cause sustained breach of the terms and conditions mentioned in the SLA. Additionally, customer may request the MSP for inclusion of new terms and conditions in the SLA.

If the application SLA is breached frequently or if the customer requests for a new non-agreed SLA, the on-boarding process is performed again. In the case of the former, on-boarding activity is repeated to analyze the application and its policies with respect to SLA fulfillment. In case of the latter, a new set of policies are formulated to meet the fresh terms and conditions of the SLA.

5) Termination

When the customer wishes to withdraw the hosted application and does not wish to continue to avail the services of the MSP for managing the hosting of its application, the termination activity is initiated. On initiation of termination, all data related to the application are transferred to the customer and only the essential information is retained for legal compliance. This ends the hosting relationship between the two parties for that application, and the customer sign-off is obtained.

SLAs are also categorized at different levels:

Customer-based SLA: An agreement with an individual customer or group, covering all services they use.

Service-based SLA: An agreement with all customers using the service being delivered similarly by the service provider.

Multilevel SLA: The SLA is split into the different levels, each addressing different set of customers for the same services, in the same SLA. For example:

- Corporate-level SLA
- Customer-level SLA
- Service-level SLA

TRUST MANAGEMENT

Probably the most critical issue to address before cloud computing can become the preferred computing paradigm is that of establishing trust. Mechanisms to build and maintain trust between cloud computing consumers and cloud computing providers, as well as between cloud computing providers among themselves, are essential for the success of any cloud computing.

With the popularity and growth of cloud computing, service providers make new services available on clouds. All these service and service providers have varying levels of quality and also due to anonymous nature of the cloud computing, some dishonest or unprincipled or crooked service providers may tend to cheat unaware, unsuspecting clients. Hence it becomes necessary to identify the quality of services and service providers who would meet the trust requirements of customers.

- ➢ Trust management is a key issue- that needs special attention and it is an important component of cloud security.
- ➢ Trust management is an abstract system that processes symbolic representations of social trust, usually to aid automated decision making process. It increases and establishes the trust for the cloud computing systems among the users. Such representations like cryptographic credentials, can link the abstract system of trust management with results of trust assessment. Trust management is popular in implementing information security, specifically access control policies.
- ➢ Trust management system provide are assurance to users- that their data is secure and confidential with particular cloud service provider.

The concept of trust management has been introduced to help and assist the automated verification of actions against security policies. The definition of trust covers honesty, truthfulness, competence and reliability.

Trust management in cloud computing involves these aspects:

- Data integrity and privacy protection
- Trusted cloud computing over data-centers
- Security Aware cloud architecture
- Virtual network security and trust negotiation
- Defence of virtualized resources
- Guarantee of confidentiality, integrity and availability.

Various techniques used for trust management include following:

- Data coloring and water marking
- Encryption or cryptography
- Hardware security in data-centers
- Replication of data
- Administrative record security
- Tightly secured access controls
- VPN technology
- Compliance with world-class security and operational standard certifications.
- Standard trust and reputation management practices.
- Third party cloud security and backup services
- Formal accreditation, audit and standards