CPSC 436C
Cloud Computing for Data Science

Cloud Security – Part 1

Maryam R.Aliabadi

mraiyata@cs.ubc.ca

Spring 2024

Last Class's Review

- Resource management Systems
  - Mesos
    - Offered-based
    - Max-Min fairness: DRF
  - YARN
    - Request-based
    - RM, AM, NM

Motivation

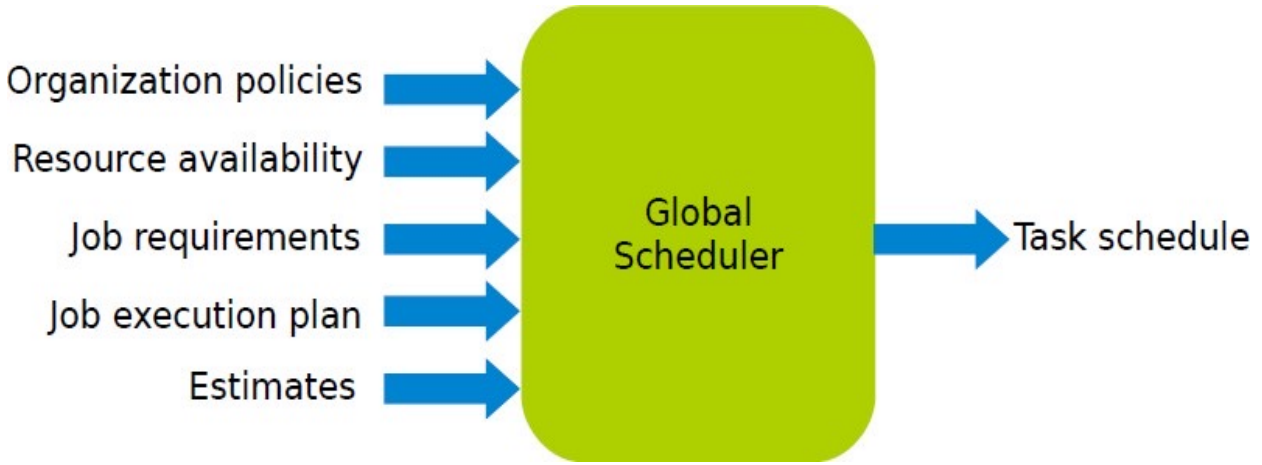
- Rapid innovation in cloud computing.
- No single framework optimal for all applications.
- Running each framework on its dedicated cluster:
  - Expensive
  - Hard to share data

Proposed Solution

- Running multiple frameworks on a single cluster.
- Maximize utilization and share data between frameworks.
- Top resource management systems:
  - Mesos
  - YARN

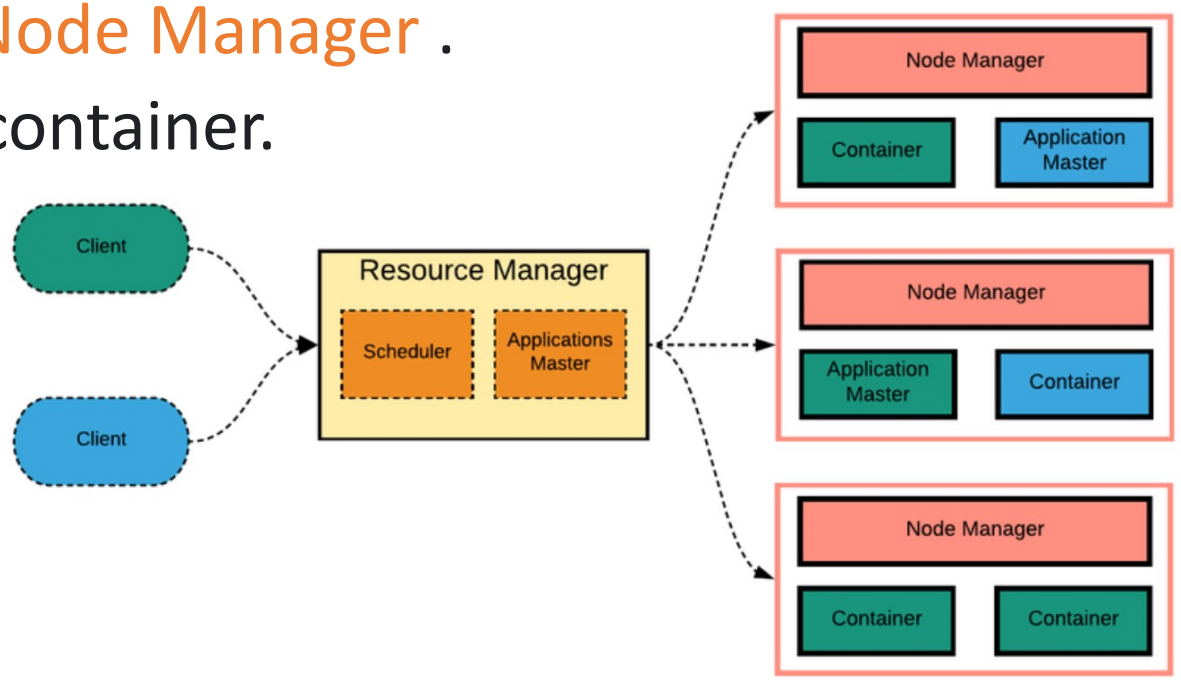
Global Scheduler

- Job requirements
  - Response time
  - Throughput
  - Availability
- Job execution plan
  - Task DAG
  - Inputs/outputs
- Estimates
  - Task duration
  - Input sizes
  - Transfer sizes

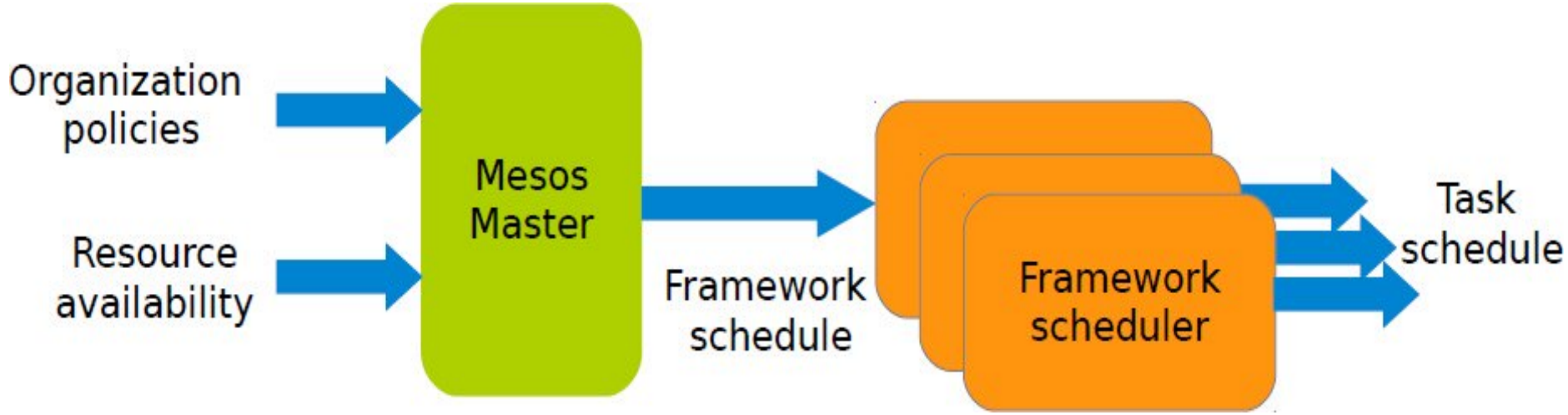


Global Scheduler : Yarn

- The client submits an application to the Resource Manager.
- The Resource Manager allocates a container (Application Master).
- Application Mater negotiates resources with the Resource Manager.
- The Application Master contacts Node Manager .
- The Node Manager launches the container.
- The container executes the task.

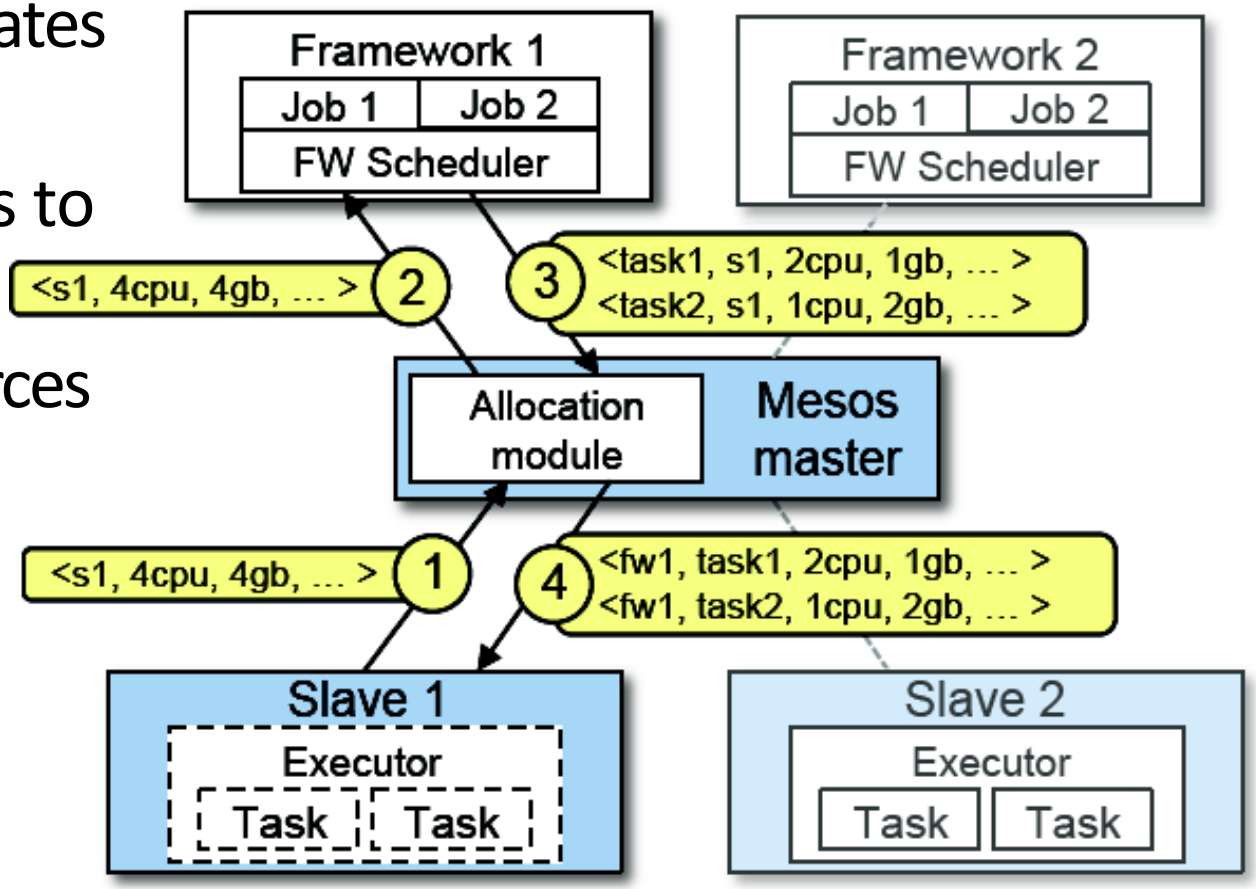


Distributed Scheduler



Distributed Scheduler : Mesos

- Slaves continuously send status updates about resources to the Master.
- Mesos Master sends resource offers to frameworks.
- Framework scheduler selects resources and provides tasks.
- Framework executors launch tasks.



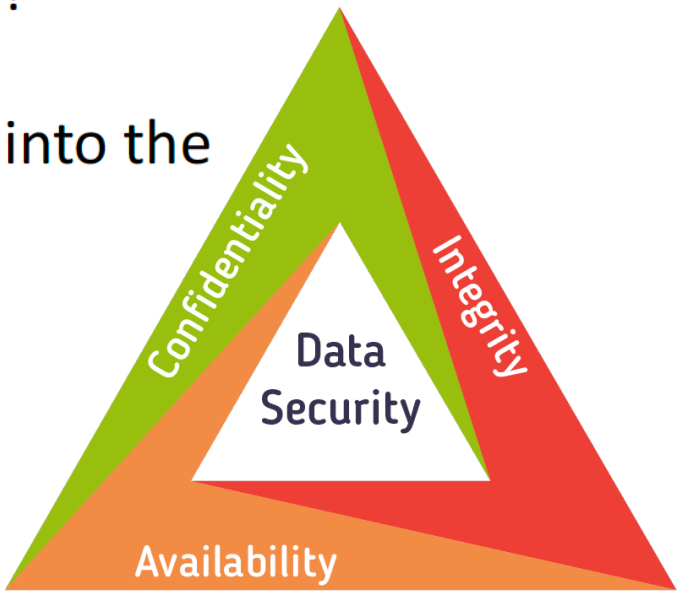
Today's Topic : Cloud Security

Data: The Main Asset In The Cloud

- Data is the most important component of the data center and Cloud
- Data is unique for the organization: everything else we saw above (building, software, hardware) can be replaced, rebought, rebuilt except data.
- Data should be protected from not only the risk of loss, but also from the risk of unauthorized access.

Cloud Security Issues in Cloud

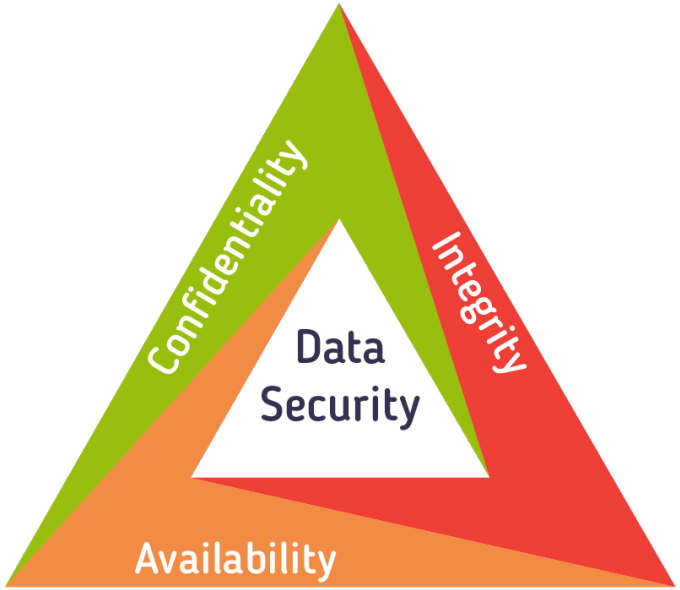
- Confidentiality
  - Fear of loss of control over data
    - Will the sensitive data stored on a cloud remain confidential?
    - Will cloud compromises leak confidential client data
  - Will the cloud provider itself be honest and won't peek into the data?



- Integrity
  - How do I know that the cloud provider is doing the computations correctly?
  - How do I ensure that the cloud provider really stored my data without tampering with it?

Cloud Security Issues (cont.)

- Availability
  - Will critical systems go down at the client, if the provider is attacked in a Denial of Service attack?
  - What happens if cloud provider goes out of business?
  - Would cloud scale well-enough?
  - Often-voiced concern
    - Although cloud providers argue their downtime compares well with cloud user's own data centers



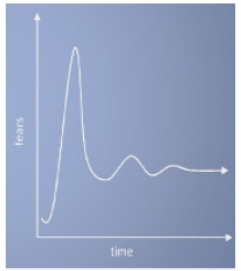
Cloud Security Issues (cont.)

- Privacy issues raised via massive data mining
  - Cloud now stores data from a lot of clients, and can run data mining algorithms to get large amounts of information on clients
- Increased attack surface
  - Entity outside the organization now stores and computes data, and so
  - Attackers can now target the communication link between cloud provider and client
  - Cloud provider employees can be phished

Cloud Security Issues (cont.)

Cloud Computing is **a security nightmare** and it can't be handled in traditional ways.

John Chambers
CISCO CEO



- Security is one of the most difficult task to implement in cloud computing.
  - Different forms of attacks in the application side and in the hardware components
- Attacks with catastrophic effects only needs one security flaw

Who is the attacker?

Insider?

- Malicious employees at client
- Malicious employees at Cloud provider
- Cloud provider itself

Outsider?

- Intruders
- Network attackers?

Malicious insider

- At client
  - Learn passwords/authentication information
  - Gain control of the VMs
- At cloud provider
  - Log client communication

Cloud Provider

- What?
  - Can read unencrypted data
  - Can possibly peek into VMs, or make copies of VMs
  - Can monitor network communication, application patterns

Outside attacker

- What?
  - Listen to network traffic (passive)
  - Insert malicious traffic (active)
  - Probe cloud structure (active)
  - Launch DoS

Traditional systems security Vs. Cloud Computing Security

Analogy

Securing a house

Owner and user are often the same entity

Securing a motel

Owner and users are almost invariably distinct entities

Traditional systems security Vs. Cloud Computing Security

Securing a house

Biggest user concerns
Securing perimeter
Checking for intruders
Securing assets

Securing a motel

Biggest user concern
Securing room against (the bad guy in next room | hotel owner)

Cloud Computing brings new threats

- Traditional system security mostly means keeping bad guys out. The attacker needs to either compromise the auth/access control system, or impersonate existing users.

- But clouds allow co-tenancy: Multiple independent users share the same physical infrastructure. An attacker can legitimately be in the same physical machine as the target

Challenges for the attacker

How to find out **where** the target is located

How to be **co-located** with the target in the same (physical) machine

How to gather information about the target

More on attacks…

1- Can one determine where in the cloud infrastructure an instance is located?

More on attacks…

1- Can one determine where in the cloud infrastructure an instance is located?

2- Can one easily determine if two instances are co-resident on the same physical machine?

More on attacks...

1- Can one determine where in the cloud infrastructure an instance is located?

2- Can one easily determine if two instances are co-resident on the same physical machine?

3- Can an adversary launch instances that will be co-resident with the other instances?

More on attacks…

1- Can one determine where in the cloud infrastructure an instance is located?

2- Can one easily determine if two instances are co-resident on the same physical machine?

3- Can an adversary launch instances that will be co-resident with the other instances?

4- Can an adversary exploit cross VM information leakage once co-resident?

More on attacks…

1- Can one determine where in the cloud infrastructure an instance is located?

2- Can one easily determine if two instances are co-resident on the same physical machine?

3- Can an adversary launch instances that will be co-resident with the other instances?

4- Can an adversary exploit cross VM information leakage once co-resident?

Answer: Yes to all

Finding the Victim's Location

- IP Address Geolocation: The IP address assigned to an instance might offer an approximate location. IP geolocation databases can provide details like the city or region associated with an IP address. However, this isn't precise and might only indicate the general area of the data center.

- Network Tracing and Latency: Tracing network routes and measuring latency between instances might give clues about their physical proximity. However, this won't precisely identify the physical location but could hint at the network's topology.

Being the Victim's Co-Resident

- Choosing the same Region

- Brute-force placement (launch many instances over a relatively long period of time)
  - Of 1686 target victims co-residence achieved with 141 victim servers (8.4% coverage)

- Leveraging placement locality
  - Target recently launched instances (take advantage of the tendency for EC2 to assign fresh instances to the same small set of machines)

Exploiting Co-Residence

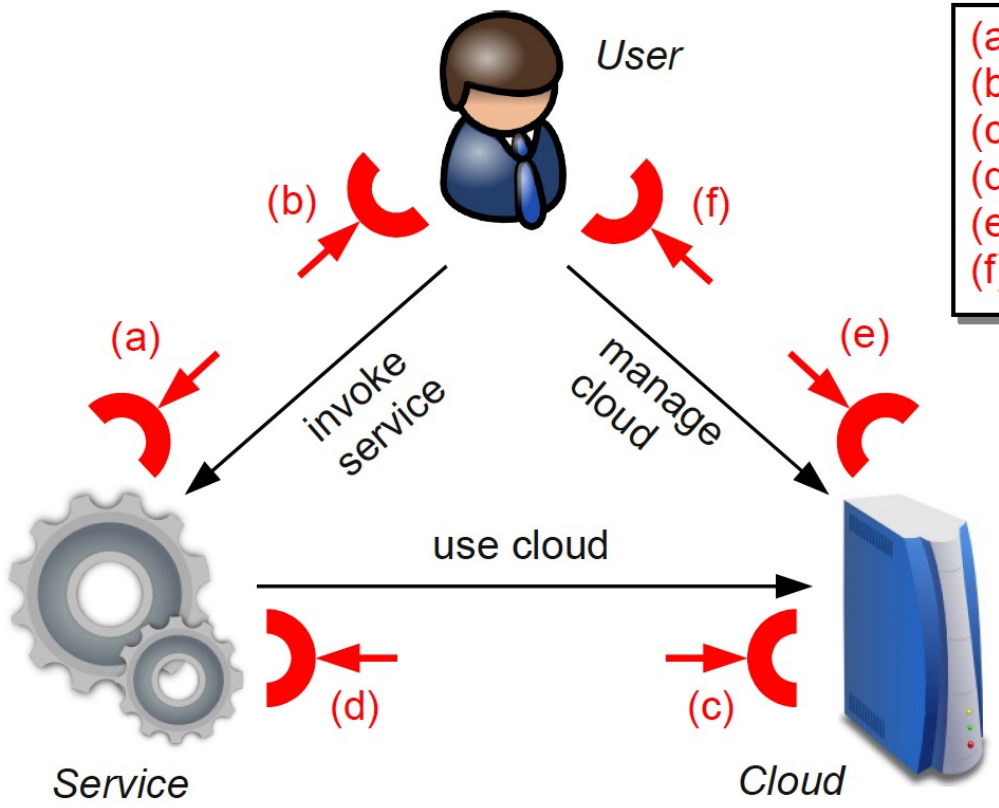
- Cross-VM attacks can allow for information leakage
- How can we exploit the shared infrastructure?
  - Gain information about the resource usage of other instances
  - Create and use covert channels to intentionally leak information from one instance to another

- Side/Covert Channel is a passive attack in which attacker gains information about target through indirect observations.
  - Time required to access a file or resource
  - Power consumption data of a computational process
  - Measuring cache usage

Clouds extend the attack surface

- An Attack Surface is the set of locations in the hardware or software that are reachable by attacker.
- The vulnerabilities that are located on the attack surface are potentially exploitable.
- How?
  - By requiring users to communicate with the cloud over a public/insecure network
  - By sharing the infrastructure among multiple users (Multi-Tenancy)

Analyzing Attack Surfaces in Clouds



(a) service-to-user
(b) user-to-service
(c) cloud-to-service
(d) service-to-cloud
(e) cloud-to-user
(f) user-to-cloud

Cloud attack surfaces can be modeled using a 3 entity model (user, service, cloud)

Figure from: Gruschka et al., Attack Surfaces: A Taxonomy for Attacks on Cloud Services.

Attack Surface 1 : Service to User

- Service interface exposed toward clients

- Possible attacks: common attacks in client-server architecture
  - Buffer overflow
  - SQL Injection
  - Privilege escalation

Attack Surface 2: User to Service

- User exposed to the service

- Common attacks
  - SSL certificate spoofing
  - Phishing

Attack Surface 3: Cloud to Service

- Cloud resources/interfaces exposed to service

- Attacks run by service on the cloud infrastructure
  - Resource exhaustion
  - Denial of Service

Attack Surface 4: Service to Cloud

- Service interface exposed to the cloud

- Privacy attacks
- Data integrity attacks
- Data confidentiality attacks

Attack Surface 5: Cloud to User

- Cloud interface exposed to users
- Attacks on cloud control

Attack Surface 6: User to Cloud

- User exposed to the cloud

- How much the cloud can learn about a user?

OWASP Top Ten In The Cloud

OWASP Top Ten In Web Applications

A1: Injection
A2: Broken Authentication
A3: Sensitive Data Exposure
A4: XML External Entities (XEE)
A5: Broken Access Control
A6: Security Misconfiguration
A7: Cross-Site Scripting
A8: Insecure Deserialization
A9: Using Components with Known Vulnerabilities
A10: Insufficient Logging and Monitoring

OWASP Top Ten In The Cloud

R1. Accountability & Data Ownership (Loss of control)
R2. User Identity Federation
R3. Legal & Regulatory Compliance
R4. Business Continuity & Resiliency
R5. User Privacy & Secondary Usage of Data
R6. Service & Data Integration
R7. Multi-tenancy & Physical Security
R8. Incidence Analysis & Forensics
R9. Infrastructure Security
R10. Non-production Environment Exposure

OWASP Top Ten In The Cloud

R1. Accountability & Data Ownership (Loss of control)
R2. User Identity Federation
R3. Legal & Regulatory Compliance
R4. Business Continuity & Resiliency
R5. User Privacy & Secondary Usage of Data
R6. Service & Data Integration
R7. Multi-tenancy & Physical Security
R8. Incidence Analysis & Forensics
R9. Infrastructure Security
R10. Non-production Environment Exposure

Multi-Tenancy

- Tenants share a pool of resources and have opposing goals

- Cloud computing naturally bring new threats:
  - Multiple independent users share the same physical infrastructure

  - Thus an attacker can legitimately be in the same physical machine as the target

Minimize Multi-Tenancy Risks

- Can try to increase isolation between tenants
  - Strong isolation techniques (VPC to some degree)
  - QoS requirements need to be met
  - Policy specification

- Can try to increase trust in the tenants
  - Who's the insider, where's the security boundary? Who can I trust?
  - Use SLAs to enforce trusted behavior

Loss of control in the cloud

- Data, applications, resources are located with provider
- User identity management is handled by the cloud
- User access control rules, security policies and enforcement are managed by the cloud provider
- Consumer relies on provider to ensure Data security and privacy
- Resource availability
- Monitoring and repairing of services/resources

Minimizing the loss of control

- Many possible layers of access control
  - E.g. access to the cloud, access to servers, access to services, access to databases (direct and queries via web services), access to VMs, and access to objects within a VM
  - Depending on the deployment model used, some of these will be controlled by the provider and others by the consumer

Minimizing the loss of control

- User Identity Federation
  - Keep control over user identities as they move services applications to the different cloud providers.
  - Digital identity is a key part of cybersecurity.

- Implement a modern identity service or platform to provide robust, persistent, verified identity controls.
  - Security Assertion Markup Language (SAML) as the underlying identity protocol to federate across Cloud apps and providers.
  - OAuth and OpenID Connect as mechanisms for user identity federation.

Recap

- Cloud Security Issues
- Cloud Attack Surface
- OWASP Top Ten in the Cloud

Next Class:

Practical Cloud Security (Guest Speaker from AWS)