Carnegie Mellon University Software Engineering Institute

Working with Cloud Flow Data

JANUARY 2024

Tim Shimeall
Principal Engineer

Document Markings

DM24-0001

Overview

- Why this still isn’t a vendor issue
- Usage Scenario: Software as a Service
- Usage Scenario: Platform as a Service
- Usage Scenario: Infrastructure as a Service
- Tradeoff: shared security and network visibility

Why This Still Isn't Just a Vendor Issue

Presentation updates talk from FloCon22

Cloud hosting services are dedicated to provision

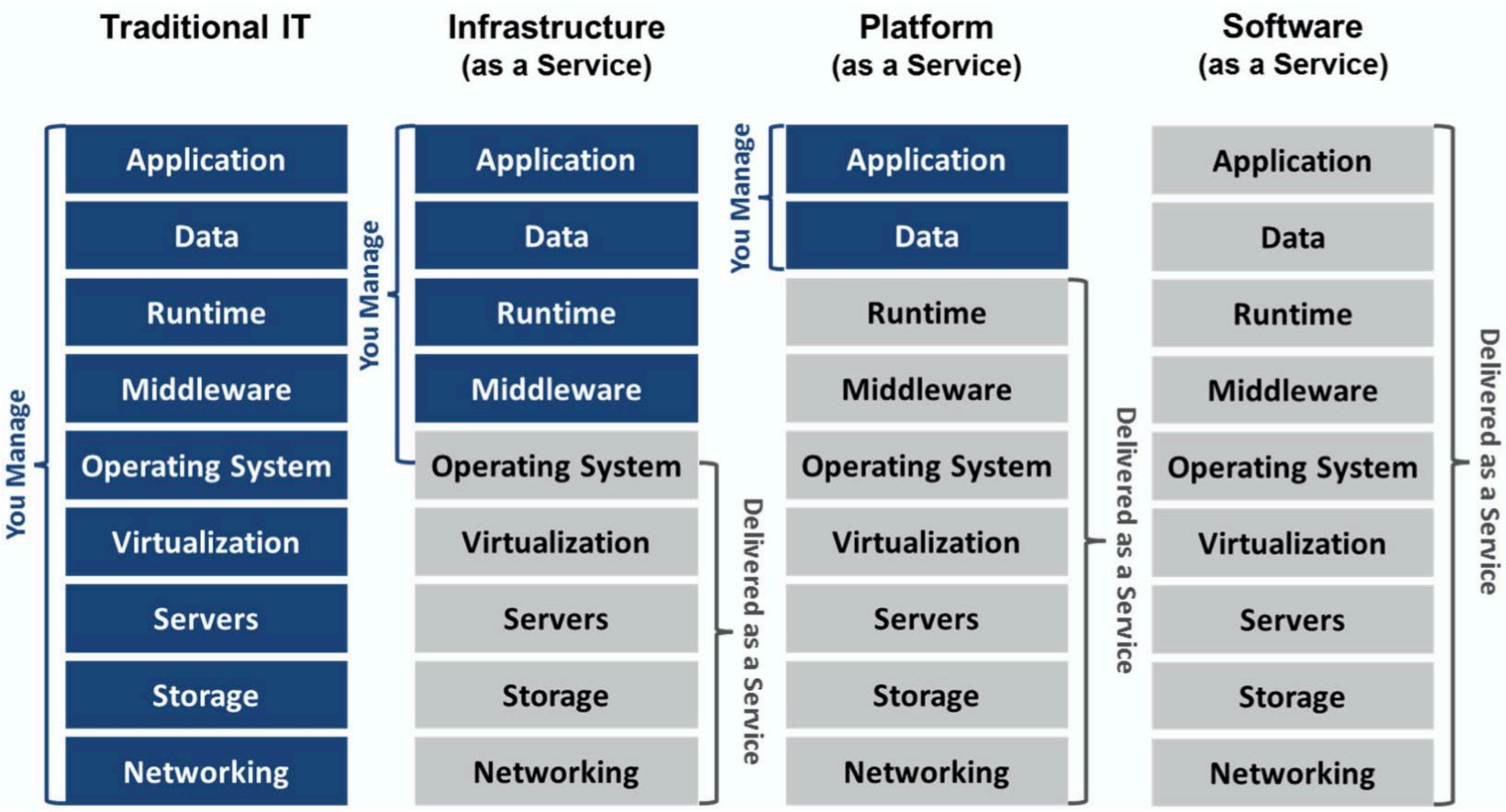
The organization that uses cloud services is responsible for security

- Provisioning and monitoring is done jointly with cloud service provider (CSP)
- Identify requirements and expectations, compare with contract statements
- Using organization: content, not infrastructure
- Using organization: differentiate abuse and activity

A using organization may host services on more than one vendor

Understand trade-offs and risks

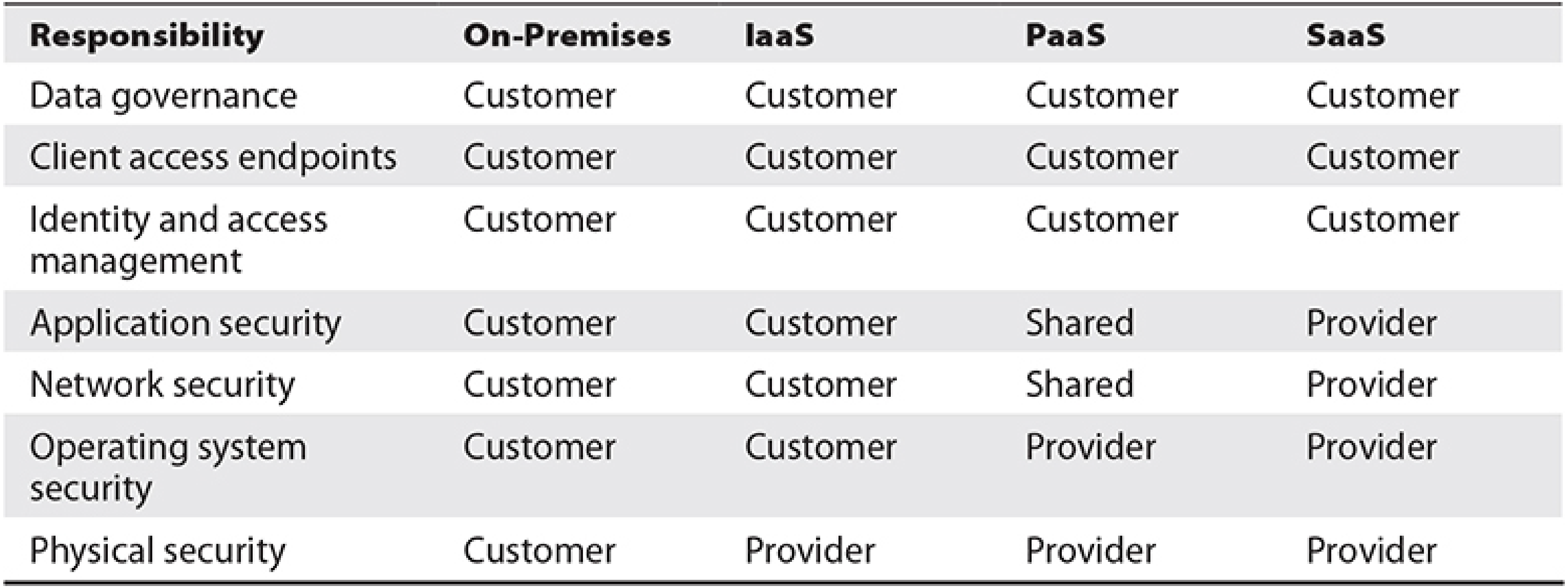
Shared Responsibility Model – 1



Traditional IT	Infrastructure (as a Service)	Platform (as a Service)	Software (as a Service)
Application	Application	Application	Application
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
Operating System	Operating System	Operating System	Operating System
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.]

Shared Responsibility Model - 2



Responsibility	On-Premises	IaaS	PaaS	SaaS
Data governance	Customer	Customer	Customer	Customer
Client access endpoints	Customer	Customer	Customer	Customer
Identity and access management	Customer	Customer	Customer	Customer
Application security	Customer	Customer	Shared	Provider
Network security	Customer	Customer	Shared	Provider
Operating system security	Customer	Customer	Provider	Provider
Physical security	Customer	Provider	Provider	Provider

Cloud flow logs vary in content and format

AWS VPC Flow logs:

- Cloud-specific fields: version, account-id, interface-id, action, log-status, vpc-id, subnet-id, instance-id, region, az-id, sublocation-type, sublocation-id
- Conventional: srcaddr, pkt-srcaddr, start, end, dstaddr, pkt-dstaddr, srcport, dstport, protocol, packets, bytes, tcp-flags, type, flow-direction
- Other: pkt-src-aws-service, pkt-dst-aws-service, traffic-path

Google VPC Flow logs:

- Cloud-specific fields: src_instance, dest_instance, src_location, dest_location, src_vpc, dest_vpc
- Conventional: connection (5-tuple), bytes_sent, packets_sent, start_time, end_time
- Other: reporter, rtt_msec, src_gke_details, dst_gke_details

Azure Flow logs:

- Cloud-specific fields: systemId, resourceId,
- Conventional: mac (sensor), time stamp, source IP, destination IP, source port, destination port, protocol, traffic flow (direction), packets sent, bytes sent, packets received, bytes received
- Other: time, version, rule, traffic decision, flow state

Cloud Flow Log limitations

AWS:

- Collection options impact data capture
- Default max aggregation of 10 minutes
- Start, end in seconds, +/- 60 seconds
- TCP flags only SYN, RST, FIN, SYN-ACK

Google:

- Collection inside VPC firewalls (before egress, after ingress)
- Flow records stored for 30 days in logs (can export)
- Flow records generated by sampling 1 of 30 or less (subscriber cannot change)
- Aggregation: 5 seconds (default), 30 seconds, 1 minute, 5 minutes, 10 minutes, 15 minutes
- Logs are sampled before storage (50% by default, can be 0%-100%)

Azure: (NSG – VNET records offer more cloud info with 1 minute aggregation)

- Flow records are events – no duration (Begin, Continue, End flow state)
- Aggregation: 5 minutes
- Protocol: T (TCP) or U (UDP)
- Traffic flow: I (inbound), O (outbound)
- No TCP flags

Cloud Flow Log Advantages

Observe the provider viewpoint

Access to instance identifiers beyond IP address

Access to hosting specifics

Access to provider-hosted visualization and analysis dashboards

Usage Scenario: Software as a Service

- Cloud, but not virtual infrastructure:
  - Organization contracts cloud services as needed, logged
  - Supports access through personal devices or thin clients
  - On premises network provides identification and authorization, local devices, monitoring
- No cloud flow
- Integration of service logs, on-premises flow logs, access logs, possibly via SIEM

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.]

Usage Scenario: Platform as a Service

- Cloud usage as individual hosts
  - Organization contracts services as needed
  - Organization contracts hosting for shared computing and storage resources
  - Supports access through personal devices or thin clients
  - Mixed provision of identification and authorization
  - On-premises network provides local devices, monitoring
  - Platform-based flow collection (adapted local sensor)
- No vendor-supplied cloud flow logs (platform-based flow may be exported)

Usage Scenario: Infrastructure as a Service

- Full infrastructure in the cloud, interacting with local infrastructure
- Organization has multiple options for traffic mirroring, and flow generation
- Need to balance cost, flexibility in hosting logs, log volume
- Cloud usage as shared structure
  - Interface through fixed gateway from variety of clients
  - Instances of infrastructure vary by load
  - Logging tied to infrastructure
  - Storage dedicated to logging
  - Analysis by mix of provider analytics, custom analytics

Tradeoffs

- Provider charges vs. customer tool provisioning
- Cloud storage vs. export costs
- Flexibility vs. visibility
- Logs for provision tracking vs. logs for security tracking
- Logs as resources vs. logs as attack targets

Contact Info

Tim Shimeall
Principal Engineer

SEI Contact Info
Netsa-help@cert.org
+1 412-268-5800

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.]