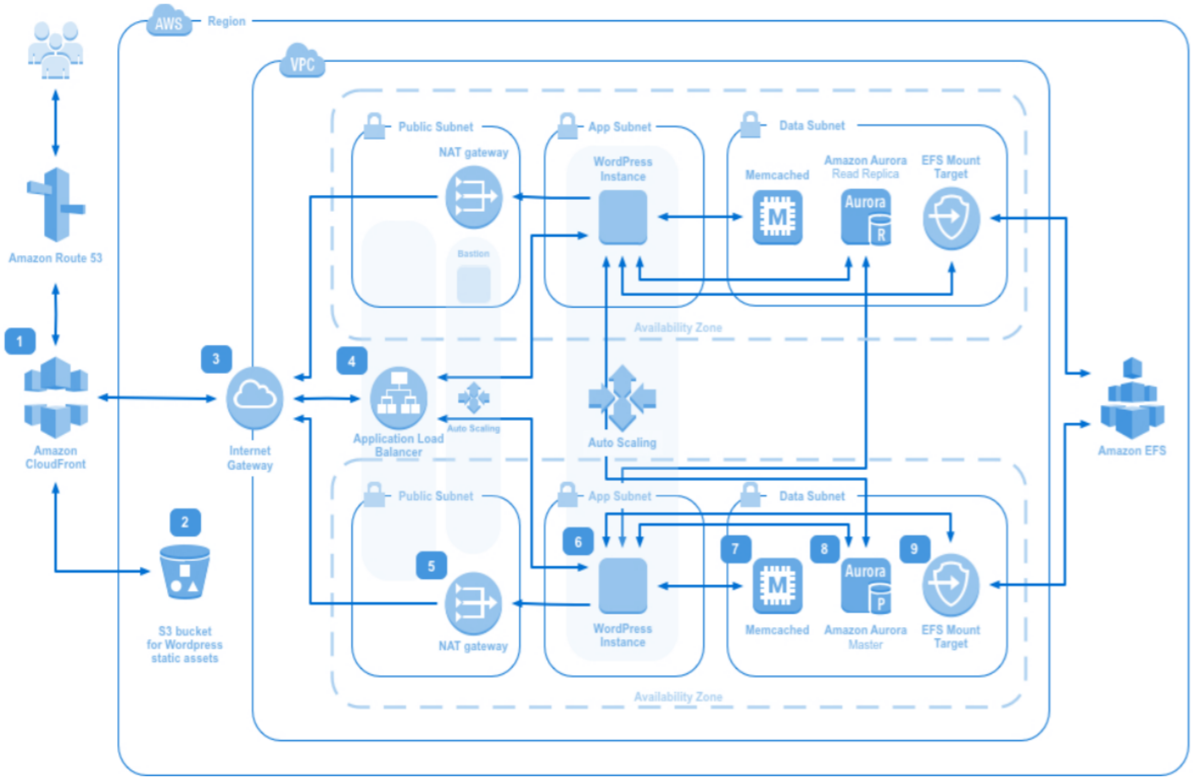
Application Note

A Blueprint for Cloud Observability

A Guide for Designing and Deploying Observability to Reduce Service Outages, Strengthen Security, and Accelerate Incident Response



Highlights

- Learn about different types of network flows for Multi-Cloud Observability and Security
- Learn how to capture, store, and analyze network packets in the cloud
- Learn about Observability deployment use cases and the trade-offs

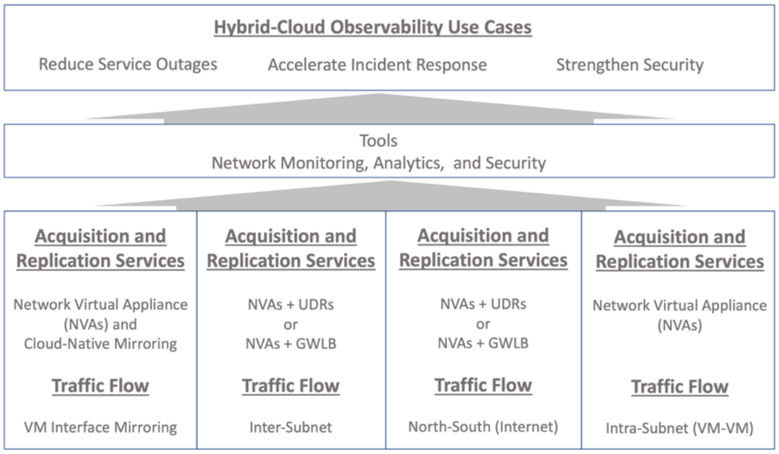
Audience

This document is intended for solution architects, sales engineers, consultants, network, and security implementors to assist in Cloud Observability architectures and use case deployments. The scope of this document covers planning and deployment to scope and design monitoring and security architectures. It assumes that the reader has a fundamental understanding of public cloud infrastructures such as AWS, Azure, Google Cloud, and others.

Introduction

Gaining access to network insights without adding agents to the virtual machines is a challenge. cPacket Networks provides a suite of agentless cloud-ready services to collect, forward and store network packets. These capabilities are critical to reducing service outages, strengthening security, and accelerating incident response. The monitoring services are added without impairing the production application workloads and virtual machine resources (vCPU, MEM, IO). This is achieved by redirecting the network flows via network load balancers to virtual appliances that offload the replication service to be forwarded to security and network monitoring tools. Also, the cloud service providers support additional cloud-native mirroring and gateway load-balancer services to access network interface and flow traffic.

The objective to have actionable observability in the cloud environment is driven by service health, security monitoring requirements, available network services, and the required traffic flows. Figure 1 shows some of the questions in the early design stages.



Hybrid-Cloud Observability Use Cases

Reduce Service Outages | Accelerate Incident Response | Strengthen Security

Tools
Network Monitoring, Analytics, and Security

Acquisition and Replication Services	**Acquisition and Replication Services**	**Acquisition and Replication Services**	**Acquisition and Replication Services**
Network Virtual Appliance (NVAs) and Cloud-Native Mirroring	NVAs + UDRs or NVAs + GWLB	NVAs + UDRs or NVAs + GWLB	Network Virtual Appliance (NVAs)
Traffic Flow	**Traffic Flow**	**Traffic Flow**	**Traffic Flow**
VM Interface Mirroring	Inter-Subnet	North-South (Internet)	Intra-Subnet (VM-VM)

What's the observability objective?

What tools are being used?

What network services are needed for the environment?

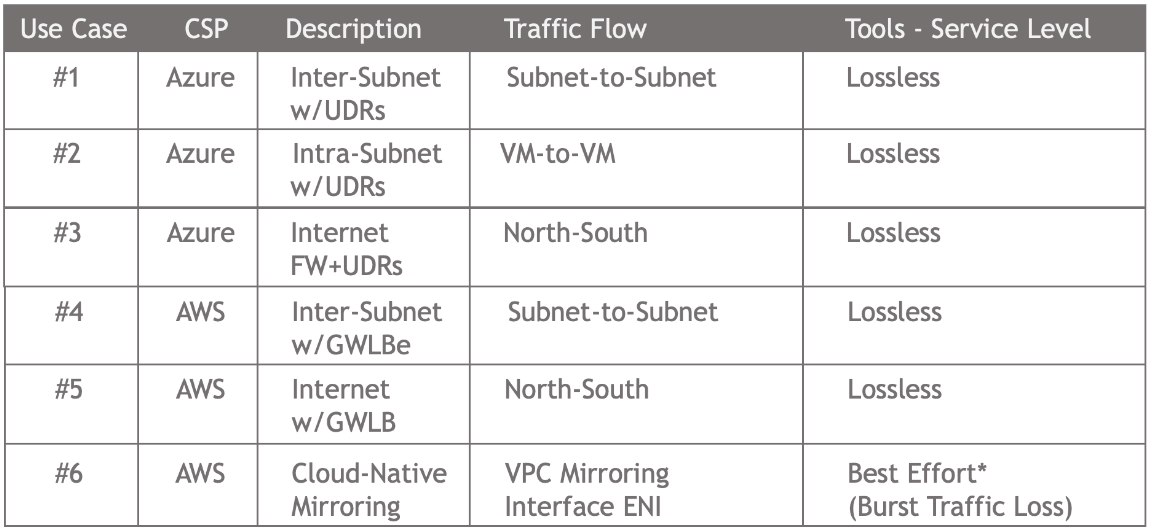
What are the traffic flows?

Figure 1 - Cloud Traffic Flows and Objectives

Cloud Deployment Use Cases

The observability deployment use cases are defined from the perspective of the traffic flow types i.e., North-South traffic, Inter-Subnet traffic, and VM interface mirroring. This allows us to focus on the type of instrumentation required in a public cloud environment and the level of service required, including an always-on service. It's essential to consider the entire service chain performance as any device in the chain may cause a bottleneck and limit throughput. Traffic replication and forwarding depend on the tool ingestion requirements, total replicated traffic, and packet capture storage~~.~~

Table 1 shows example use cases, including traffic flow, tool service, and the network packets delivery service level. The cloud-native mirroring service replicates network traffic directly from the instance's virtual network interface or Elastic Network Interface (ENI). Traffic mirroring is a service that runs on VM and creates resource contention within the instance, so when there is traffic congestion, production traffic is given a higher priority than mirrored traffic. Mirrored traffic is therefore dropped when there is congestion, especially during traffic bursts. Dropped mirrored packets cause visibility gaps that make troubleshooting problems difficult and create an exploitable security risk. Cloud-native mirroring consumes additional bandwidth on the VMs. Best effort is made to forward replicated traffic to the mirror targets, collector destinations, or endpoints. Sizing is dependent on the capabilities of the instance.



Use Case	CSP	Description	Traffic Flow	Tools - Service Level
#1	Azure	Inter-Subnet w/UDRs	Subnet-to-Subnet	Lossless
#2	Azure	Intra-Subnet w/UDRs	VM-to-VM	Lossless
#3	Azure	Internet FW+UDRs	North-South	Lossless
#4	AWS	Inter-Subnet w/GWLBe	Subnet-to-Subnet	Lossless
#5	AWS	Internet w/GWLB	North-South	Lossless
#6	AWS	Cloud-Native Mirroring	VPC Mirroring Interface ENI	Best Effort* (Burst Traffic Loss)

*AWS p_roduction traffic has higher priority than mirrored traffic. As a result, mirrored traffic is dropped when there is congestion._
https://docs.aws.amazon.com/vpc/latest/mirroring/traffic-mirroring-considerations.html

Table 1 – Public Cloud Observability Use Cases

Cloud Inter-Subnet, North-South, and Internet traffic flow using the Gateway Load Balancer services are processed by the network and do not consume additional bandwidth on the VMs. The VM, application, and users are unaware of the network replication and cannot turn the service on or off, so the service is always on. In Azure, the traffic is redirected to the NLB using User Defined Routes (UDRs) to enable the ability to monitor and analyze traffic. All components in the service chain may impact the throughput capabilities of the service.

Hybrid-Cloud Observability Architecture

The observability objective and tool requirements will determine the architecture and design of the type of packet acquisition in the cloud environment. Cloud service providers continually add new network services, such as mirroring, load-balancing, and forwarding services which provide improved ability to observe traffic.

The cPacket Cloud™ Visibility Suite is designed for cloud observability; it provides agentless network packet brokering that includes filtering, replication, and forwarding to multiple tools. It also can capture network packets for storage and analyze both the streaming and stored network packet data to facilitate observability.

The cPacket cCloud™ Visibility Suite consists of:

- **cClear®-V** – Analytics, data visualizations via customizable dashboards, and fabric management all from a single-pane-of-glass
- **cStor®-V** – Virtualized Packet Capture that provides forensic evidence that can be searched, retrieved, exported to PCAP files, and enables replaying traffic linked to events for stateful and low-latency analysis
- **cVu®-V** – Virtualized Network Packet Broker (vNPB) that performs packet acquisition, replication, filtering, and forwarding

The Virtualized NPB provides packet acquisition, replication, filtering, and forwarding from the mirroring service. Due to the mirrored service using shared resources, it's essential to offload further replication and distribution services to the Virtualized NPB appliance fleet to minimize the impact on your production application. The replication requirements will be determined by the number of security and monitoring tools and the number of forwarding flows.

cCloud Visibility Suite Sizing

Figure 2 shows the minimum cCloud Visibility Suite configuration comprising of three instances of the virtualized packet brokers providing offload filtering, replication, and forwarding to multiple tools, Packet Capture Storage, and Observability Analytics. Adjust the number of cVu-V instances in the back-end pool to allocate capacity for failover and scale-out scenarios depending on the needed ingress traffic and replication traffic.

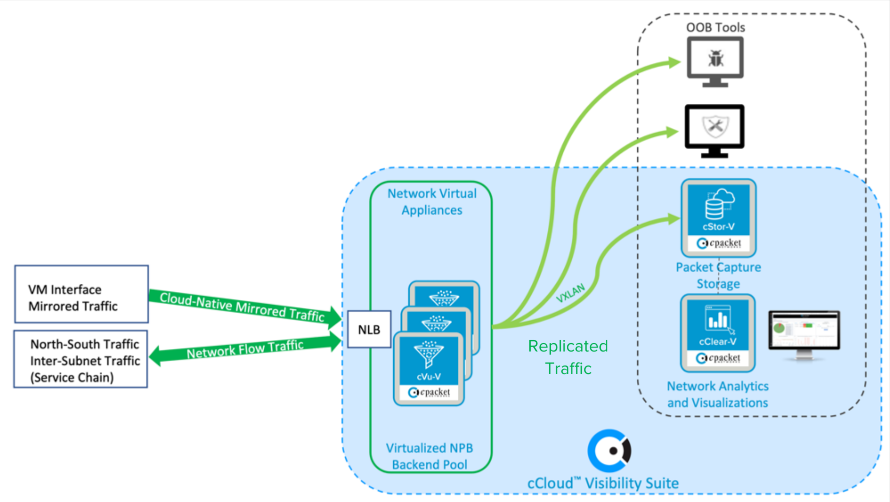

OOB Tools
Network Virtual Appliances
VM Interface Mirrored Traffic
Cloud-Native Mirrored Traffic
North-South Traffic Inter-Subnet Traffic (Service Chain)
Network Flow Traffic
NLB
cVu-V
cpacket
VXLAN
Replicated Traffic
cStor-V
Packet Capture Storage
cClear-V
Network Analytics and Visualizations
Virtualized NPB Backend Pool
cCloud™ Visibility Suite

Figure 2 - cCloud Minimum Deployment

Sizing the components of the cCloud Visibility Suite depends on your architecture, traffic, and service requirements. The following details are a recommendation for sizing the VM or instance types. These can be adjusted to support higher network bandwidth. For further performance details and supported options, please check cPacket cCloud Visibility Suite datasheet:
https://www.cpacket.com/wp-content/uploads/2020/09/cPacket-cCloud-DataSheet.pdf

Recommended minimum AWS instances for 30Gbps* monitoring throughput (depends on region availability):

-	**cVu®-V**	4x vCPUs, 16 GiB mem, 40GB	m5a.xlarge (x3)
-	**cStor®-V**	4x vCPUs, 16 GiB mem, 40GB +1TB	m5n.xlarge
-	**cClear®-V**	8x vCPUs, 32 GiB mem, 50GB+250GB	m5a.2xlarge

Recommended minimum Azure VMs for 30Gbps* monitoring throughput (depends on region availability):

-	**cVu®-V**	4x vCPUs, 16 GiB mem, 40GB	Standard D4_v5 (x3)
-	**cStor®-V**	4x vCPUs, 16 GiB mem, 40GB +1TB	Standard D4s_v5
-	**cClear®-V**	8x vCPUs, 32 GiB mem, 50GB+250GB	Standard D8s_v5

Examples in this document show a theoretical throughput traffic 30Gbps* (cVu®-V back-end pool) and three destination tools for the replicated traffic. The aggregated monitoring capacity is dependent on the instance size of each Virtualized Packet Broker cVu®-V.

*For expected VM or instance network bandwidth, refer to the following information for details:
AWS https://aws.amazon.com/ec2/instance-types/
Azure https://docs.microsoft.com/en-us/azure/virtual-machines/sizes

Designing for Scale and Availability

Figure 3 shows that the network virtual appliance back-end pool has an added cVu®-V instance as the service scales up to accommodate bandwidth growth or provide availability for a possible single virtual machine or instance failure.

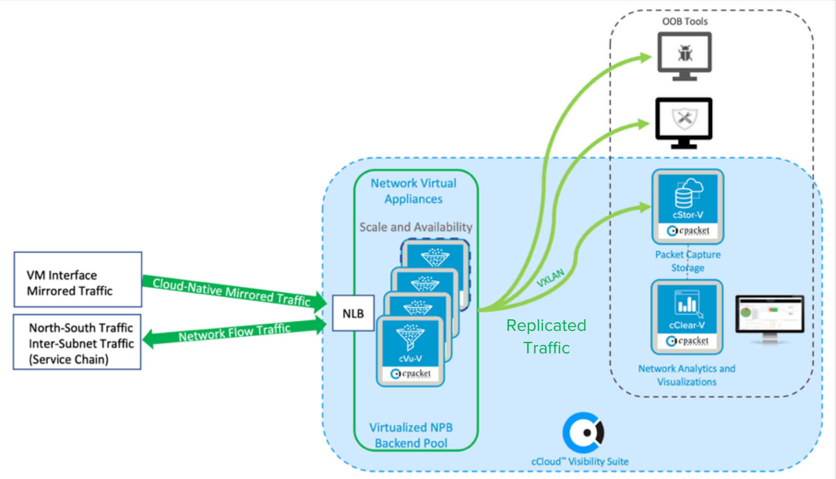


Figure 3 - cCloud Availability and Scale-Out

Azure Cloud Deployment Use Cases

When designing and sizing, the CSP supported quotas are subject to change. Please refer to the Azure Service Limits documentation using this link:
https://docs.microsoft.com/en-us/azure/azure-resource-manager/management/azure-subscription-service-limits

Azure Inter-Subnet Monitoring

Inter-Subnet monitoring in Azure is processed by the network and is redirected via UDRs to the Network Virtual Appliance (NVA) Virtualized Network Packet Broker back-end pool. Sizing is dependent on the Azure NLB performance and the cVu-V size used to support the traffic throughput.

Deployment Use Case #1 – Azure Inter-Subnet with UDRs

Figure 4 shows an example Inter-Subnet (bi-directional) monitoring deployment with UDRs, including the NVAs replicating traffic to three destination tools.

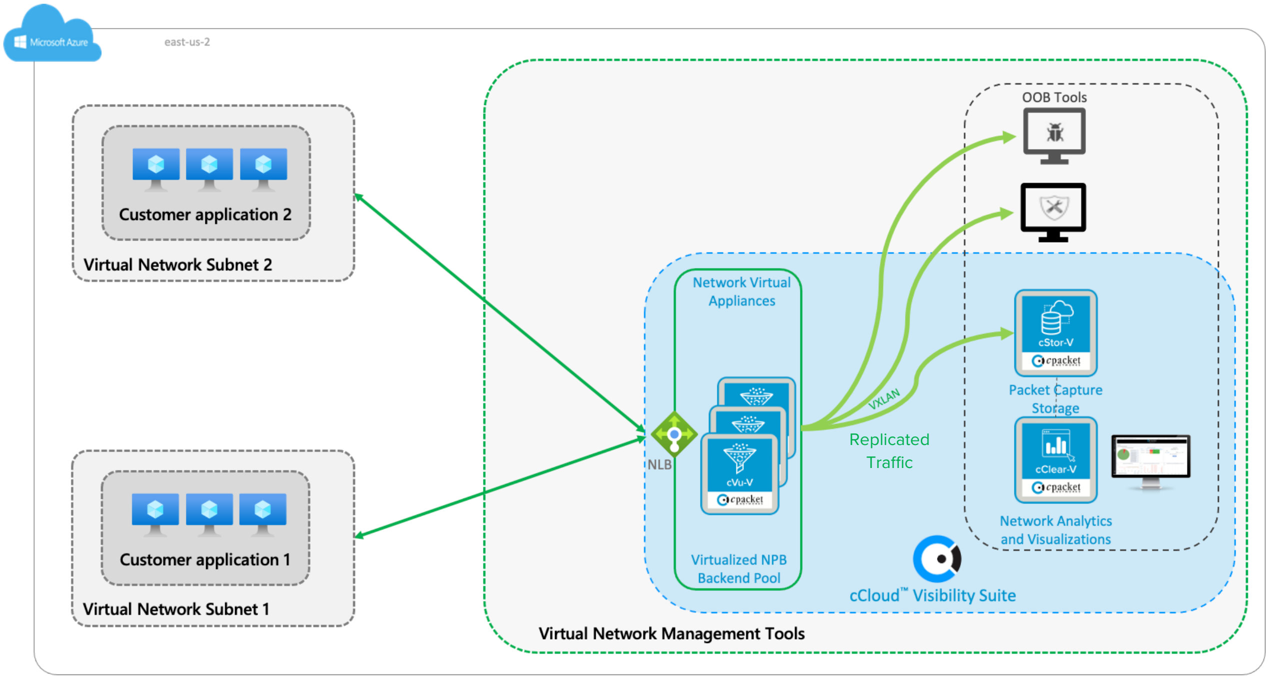


Figure 4 - Microsoft Azure Inter-Subnet example

Expected replication throughput service:

	Total NLB Ingress	+	Total NVA Replicated Traffic	+	Subnet Destination Traffic
Example (Figure 4),	5Gbps	+	(3x 5Gbps)	+	5Gbps

Recommend for Inter-Subnet traffic, critical always-on production environments, and lossless traffic requirements.

Azure Intra-Subnet Monitoring (VM-to-VM)

Cloud Intra-Subnet Monitoring is processed by the network and does not consume additional bandwidth on the VMs. Connectivity traffic from a VM is redirected to the NLB via UDRs and forwarded for replication and to the destination target virtual machine.

Deployment Use Case #2 – Azure Intra-Subnet Traffic with UDRs

Figure 5 shows an example Intra-Subnet monitoring deployment with User Defined Routes for VM-to-VM traffic, including replication traffic to three destination tools.

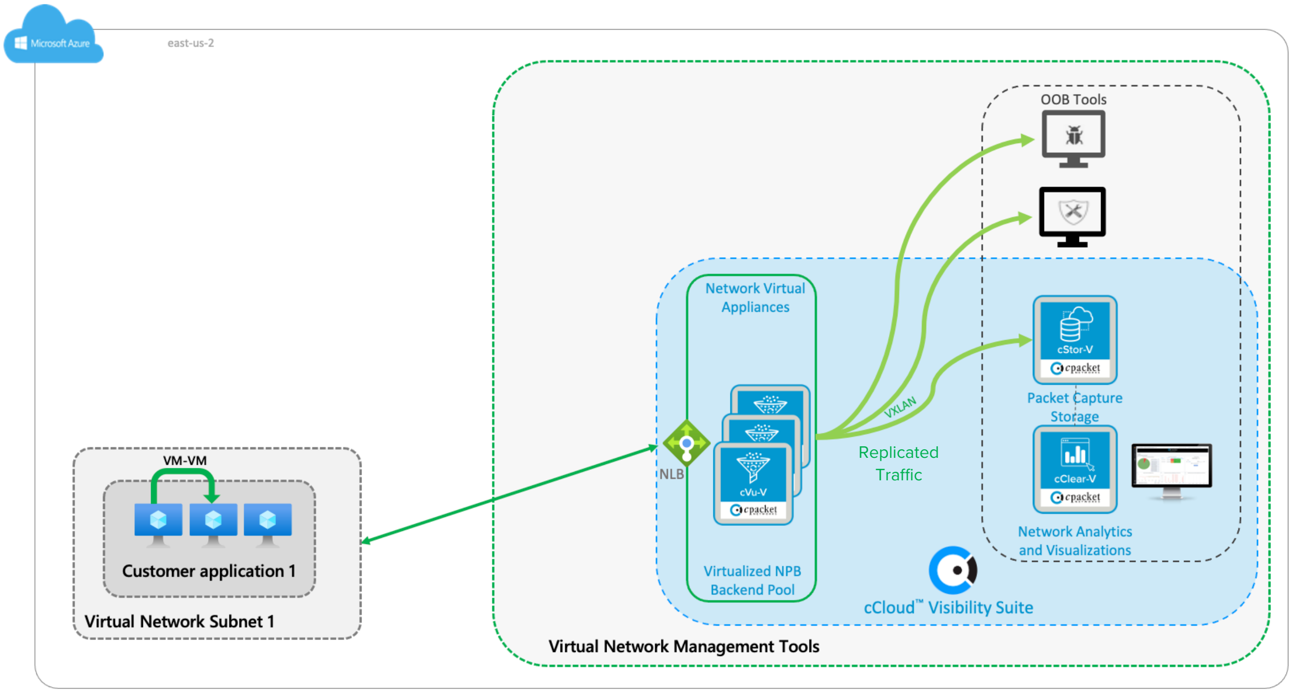


Figure 5 - AWS Azure Intra-Subnet Traffic example

Expected replication throughput service:

	Total NLB Ingress	+	Total NVA Replicated Traffic	+	VM Destination Traffic
Example (Figure 5),	5Gbps	+	(3x 5Gbps)	+	5Gbps

Recommend virtual machine traffic for critical always-on production environments and lossless traffic requirements.

Azure North to South (Internet) Monitoring

Azure North-South traffic is redirected from the firewall to the NLB using User Defined Routes to enable the replication service for the filtered firewall traffic. All components in the service chain may impact the throughput capabilities of the service.

Deployment Use Case Use #3 - Azure Internet Traffic with Firewall and UDRs

Figure 6 shows an example of monitoring North-South traffic, including a Firewall and the NVAs replicating traffic to three destination tools.

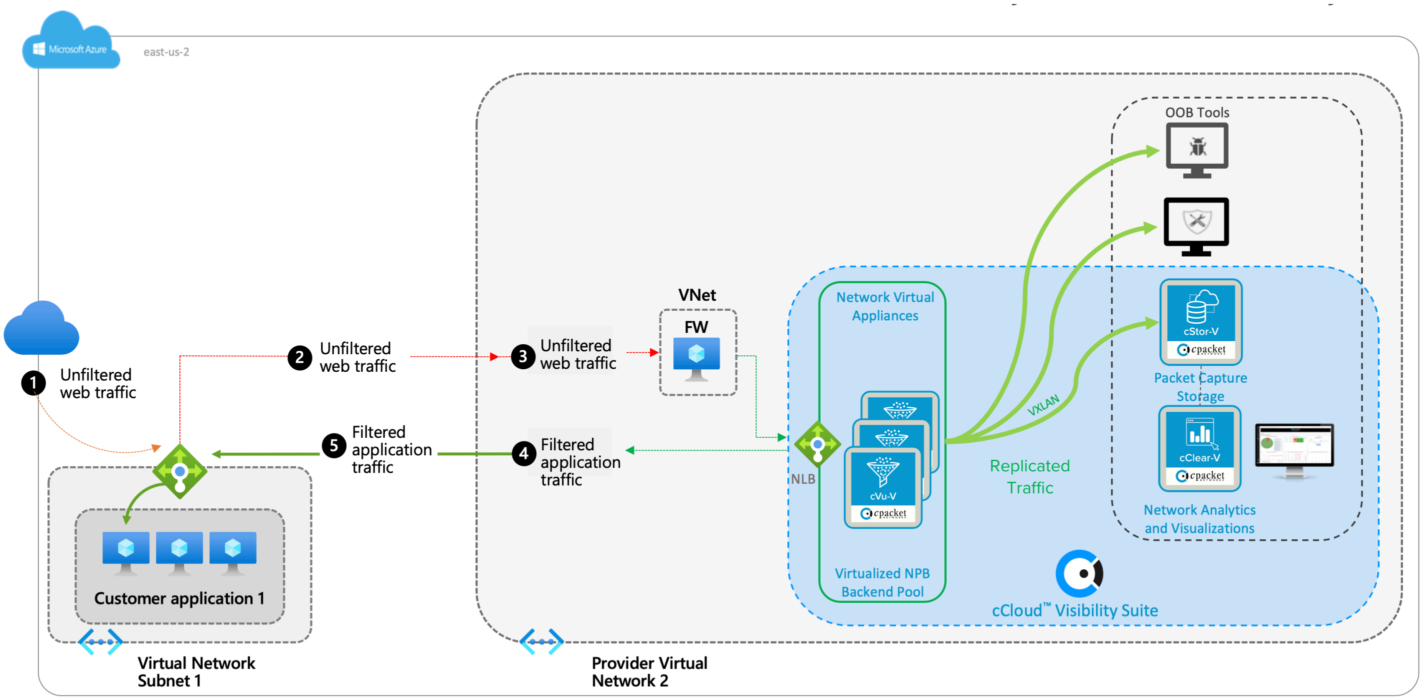


Figure 6 – Azure North-South Internet Traffic example

Azure Internet Monitoring throughput service:

	NLB Ingress /Egress	+ Total NVA Replicated Traffic	+ Subnet 1 Destination Traffic
Example (Figure 6),	2Gbps	+ (3x 2Gbps)	+ 2Gbps

NB: Azure Firewall Data Throughput limit is 30Gbps, and the standard NLB up to 1,000,000 flows (see Azure Service Limits)

Recommend for North-South with Firewall traffic for critical always-on production environments and lossless traffic requirements.

AWS Cloud Deployment Use Cases

When designing and sizing, the CSP supported quotas are subject to change. Please refer to the AWS Gateway Load Balancer (GWLB) limits documentation using this link:
https://docs.aws.amazon.com/elasticloadbalancing/latest/gateway/quotas-limits.html

AWS Inter-Subnet Monitoring

Inter-Subnet monitoring is processed by the network and does not consume additional bandwidth on the VMs. Connectivity traffic from a VM is redirected to the GWLBe and forward for replication.

Deployment Use Case Use #4 – AWS Inter-Subnet Monitoring with GWLB

Figure 7 shows an example Inter-Subnet (bi-directional) monitoring deployment with GWLB and Gateway Load Balancer Endpoint (GWLBe) and the NVAs replicating traffic to three destination tools.

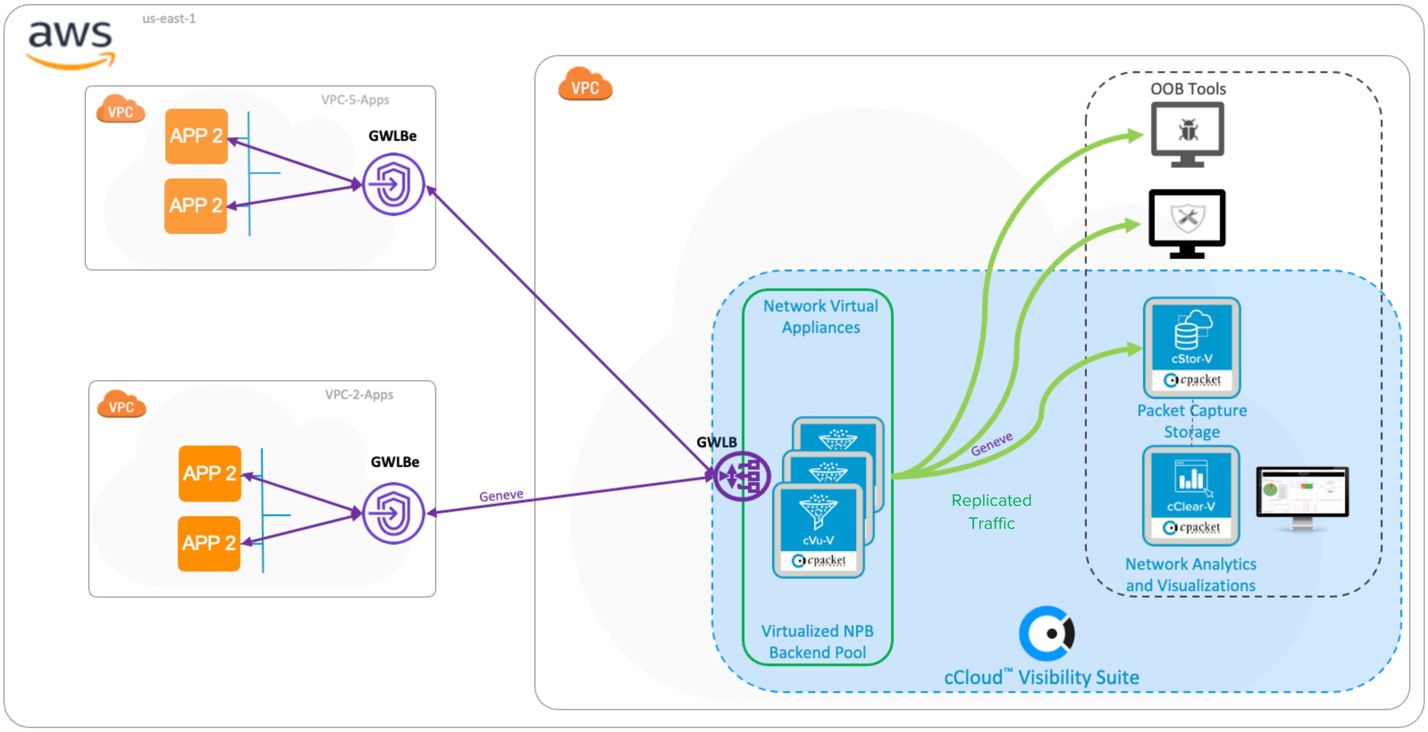


Figure 7 - AWS Inter-Subnet with GWLB example

Expected replication service depends on:

	Total GWLB Ingress/Egress Traffic	+ Total NVA Replicated Traffic
Example (Figure 7),	7Gbps	+ (3x 7Gbps)

For AWS GWLB Quota limits:
https://docs.aws.amazon.com/elasticloadbalancing/latest/gateway/quotas-limits.html

Recommend Inter-Subnet traffic for critical always-on production environments and lossless traffic requirements.

AWS North-South Monitoring (Internet)

AWS North-to-South monitoring is processed by the network and does not consume additional bandwidth on the VMs. Traffic flow from the external network is directed to the GWLB via the GWLBe for replication.

Deployment Use Case Use #5 - AWS Internet Monitoring with GWLB

Figure 8 shows an example AWS North-South (bi-directional) monitoring deployment with Gateway Load Balancer (GWLB) and the NVAs replicating traffic to three destination tools.

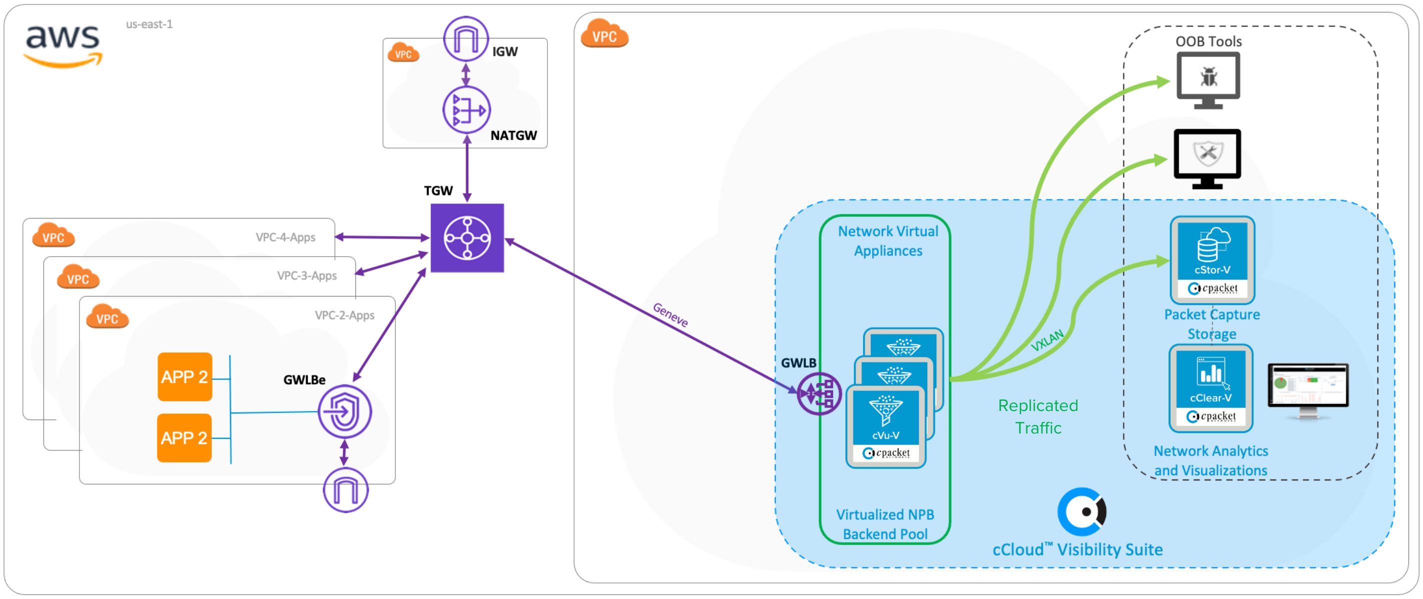


Figure 8 - AWS North-South with GWLB Traffic example

Expected Replication Throughput Service:

	GWLB Ingress/ Egress Traffic	+ Total NVA Replicated Traffic	+ Subnet Destination Traffic
Example (Figure 8),	2Gbps	+ (3x 2Gbps)	+ 2Gbps

For AWS GWLB Quota limits:
https://docs.aws.amazon.com/elasticloadbalancing/latest/gateway/quotas-limits.html

Recommend North to South traffic for critical always-on production environments and lossless traffic requirements.

Cloud-Native Interface Mirroring

Interface mirroring happens on the virtual machine (VM) instance, giving production traffic a higher priority than mirrored traffic when there is congestion, causing mirrored traffic to be dropped.

Deployment Use Case Use #6 – AWS VPC Mirroring

Figure 9 shows an example architecture using AWS VPC Traffic Mirroring with a theoretical ingress traffic 30Gbps* (cVu-V back-end pool) and three destination tools for the replicated traffic.

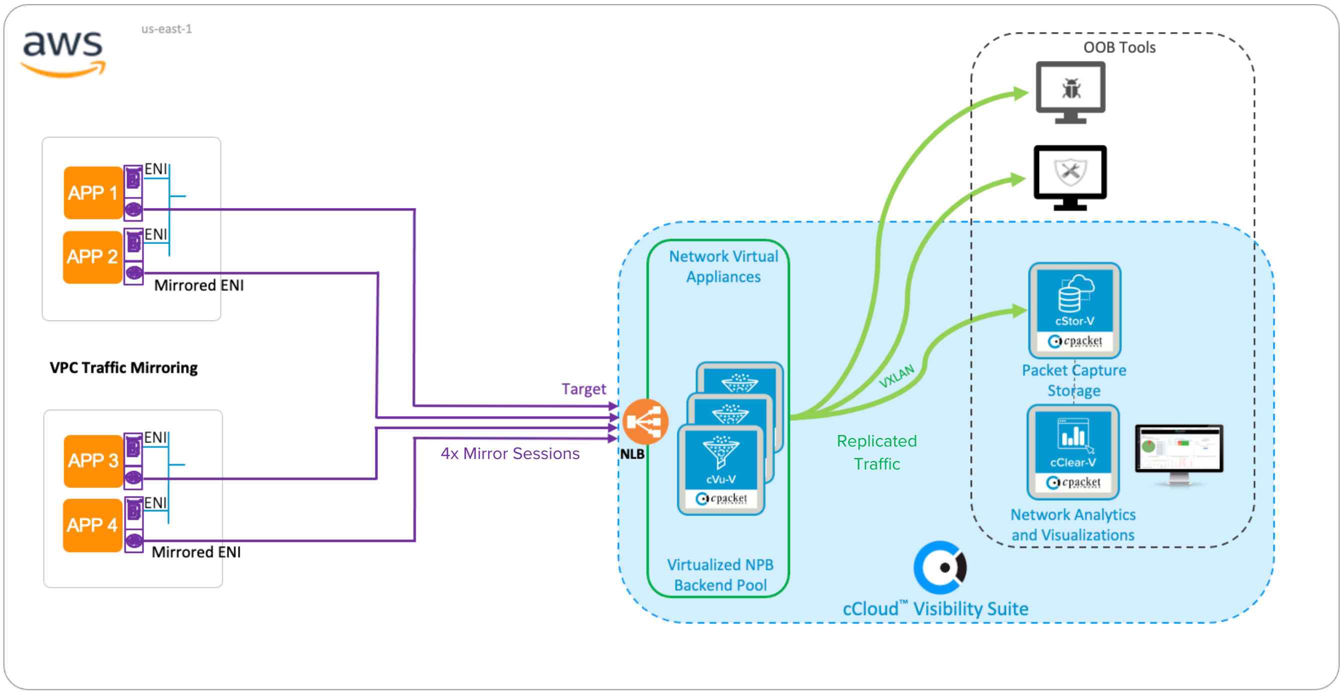


Figure 9 - AWS VPC Mirroring Traffic example

Expected replication throughput service:

	Total NLB Ingress Traffic	Total NVA + Replicated Traffic
Example (Figure 9),	7.5Gbps (4x 1.875)	+ (3x 7.5Gbps)

For AWS NLB Performance (up to 3Mpps ~30Gbps):
https://aws.amazon.com/blogs/aws/new-network-load-balancer-effortless-scaling-to-millions-of-requests-per-second/

Recommend VPC Mirrored Traffic for non-critical, non-burst traffic, and ad-hoc troubleshooting

Summary of Hybrid-Cloud Observability Deployment Use Cases

This document shows several common use cases to gain access to traffic flows so the IT team and the tools they use can benefit from observability to reduce service outages, strengthen the security posture, and accelerate incident response. It is necessary to understand the available services for each public cloud service provider. This guide provides recommended architectures to meet the specific requirements for the use cases presented. The sizing of the monitoring fabric and the required CSP and cCloud Visibility Suite services vary depending on traffic flows, and consideration must be given to bandwidth growth and availability. Examples provide guidance for design purposes, and it's recommended to check the CSPs limits and quotas as they are subject to change.