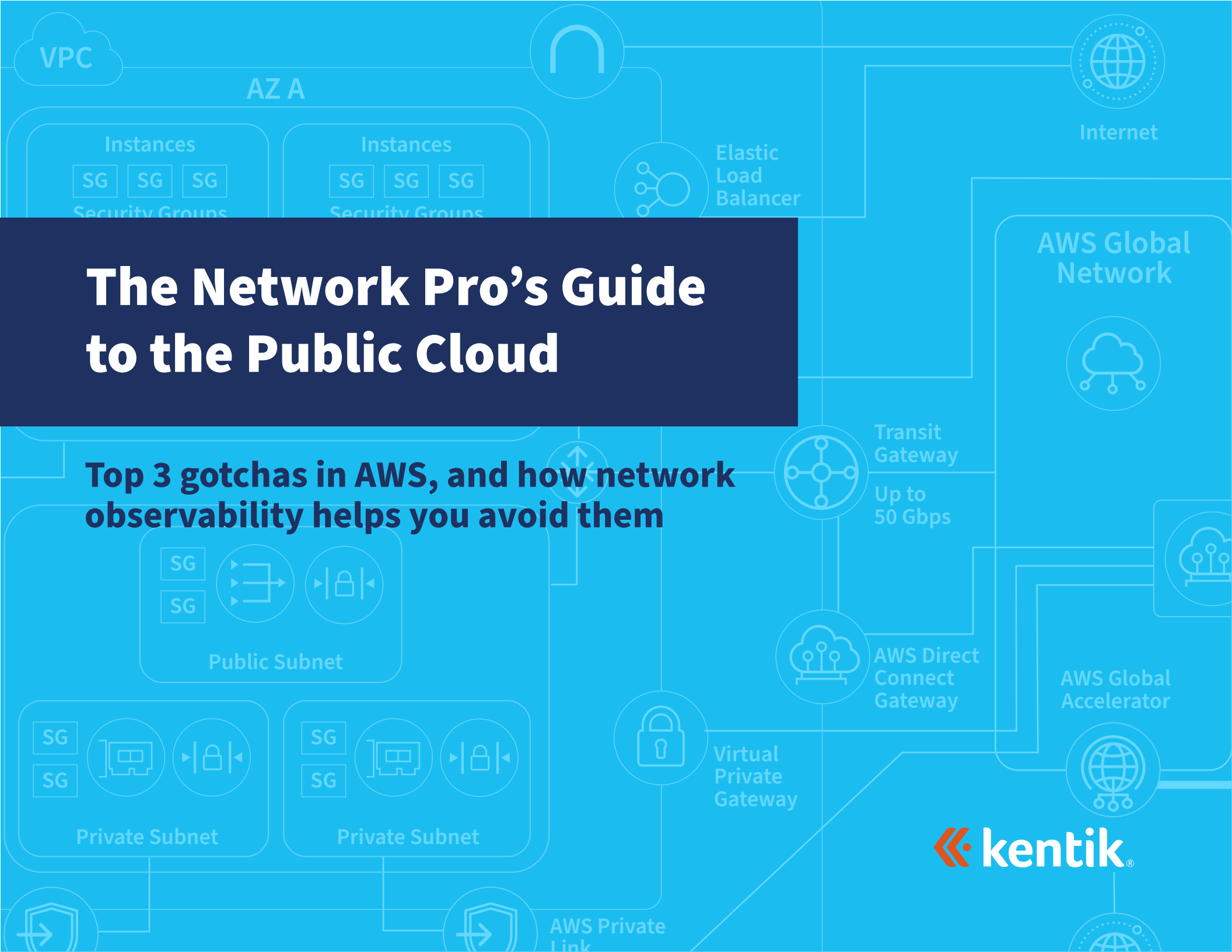
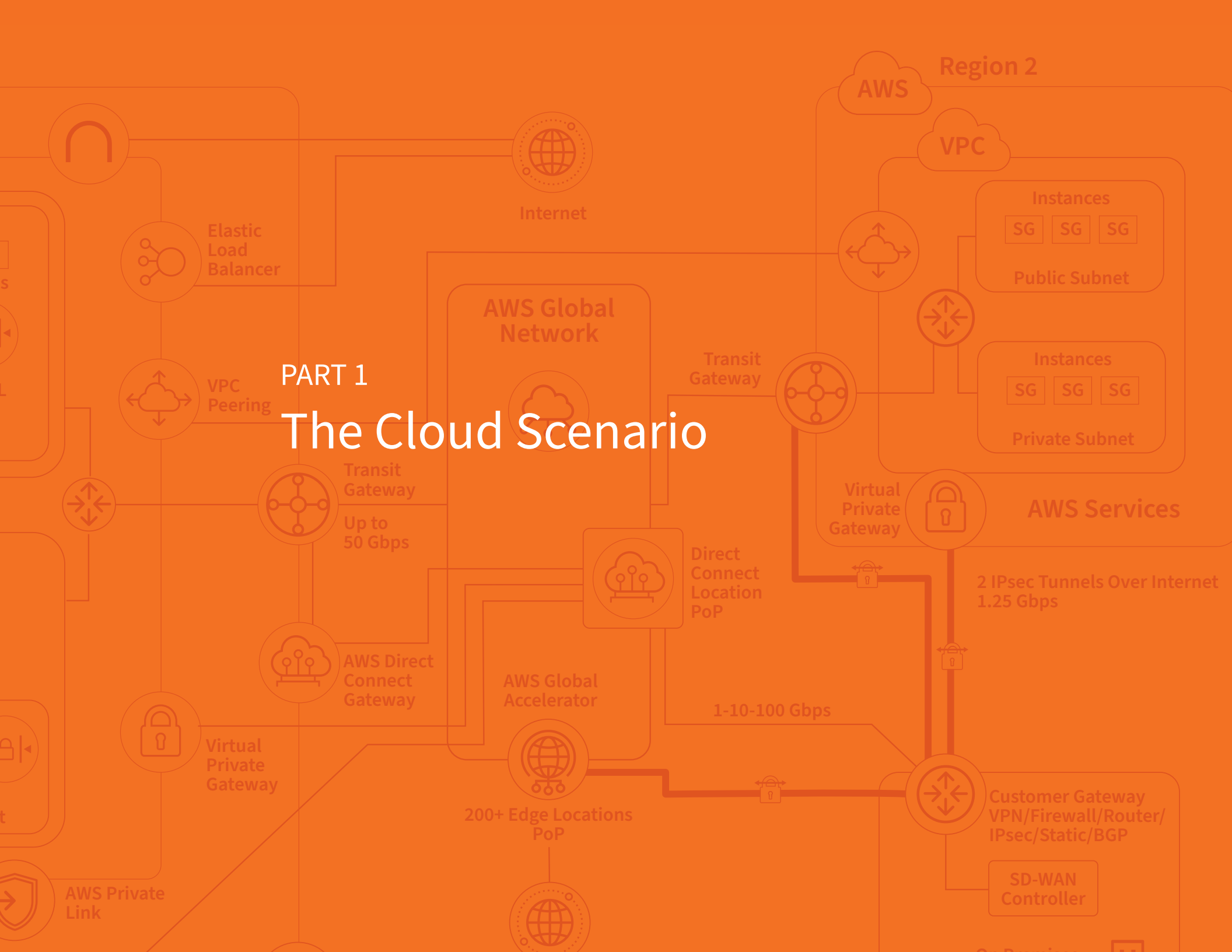


The Network Pro's Guide to the Public Cloud

Top 3 gotchas in AWS, and how network observability helps you avoid them

Kentik provides hybrid and multi-cloud network observability covering all major public clouds (AWS, Google Cloud, Microsoft Azure, IBM Cloud, Alibaba Cloud).

This ebook focuses on AWS networking constructs to, from, and across VPCs, AZs, regions, and on-premises connectivity.



PART 1

The Cloud Scenario

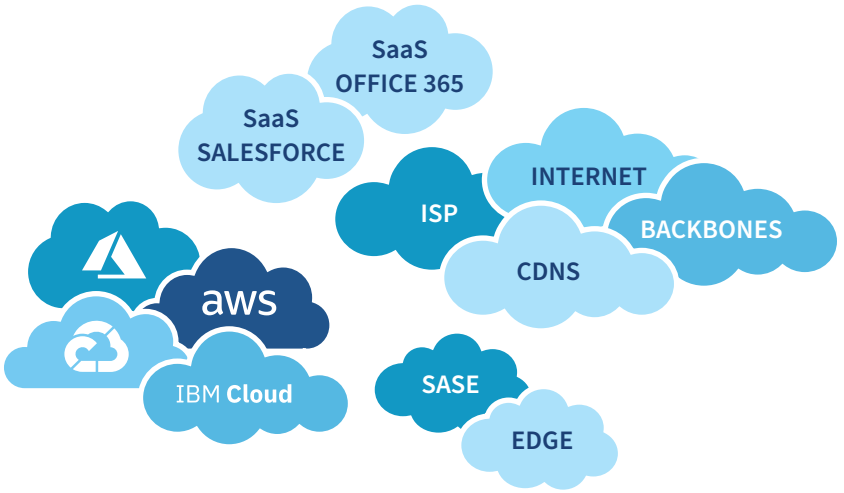
Cloud strategy requires cloud expertise

It's no surprise workloads are moving to the cloud in record numbers. From reliable global reach on day one to elastic on-demand resources, migrating or starting a business in the cloud is a no-brainer.

Clouds make the deployment of horizontally scaled applications significantly easier, creating an expectation for high reliability. So, while cloud technology is the enabler of modern, fast-paced, global internet business, it's also the reason we no longer tolerate glitches, breaches, and inefficiencies.

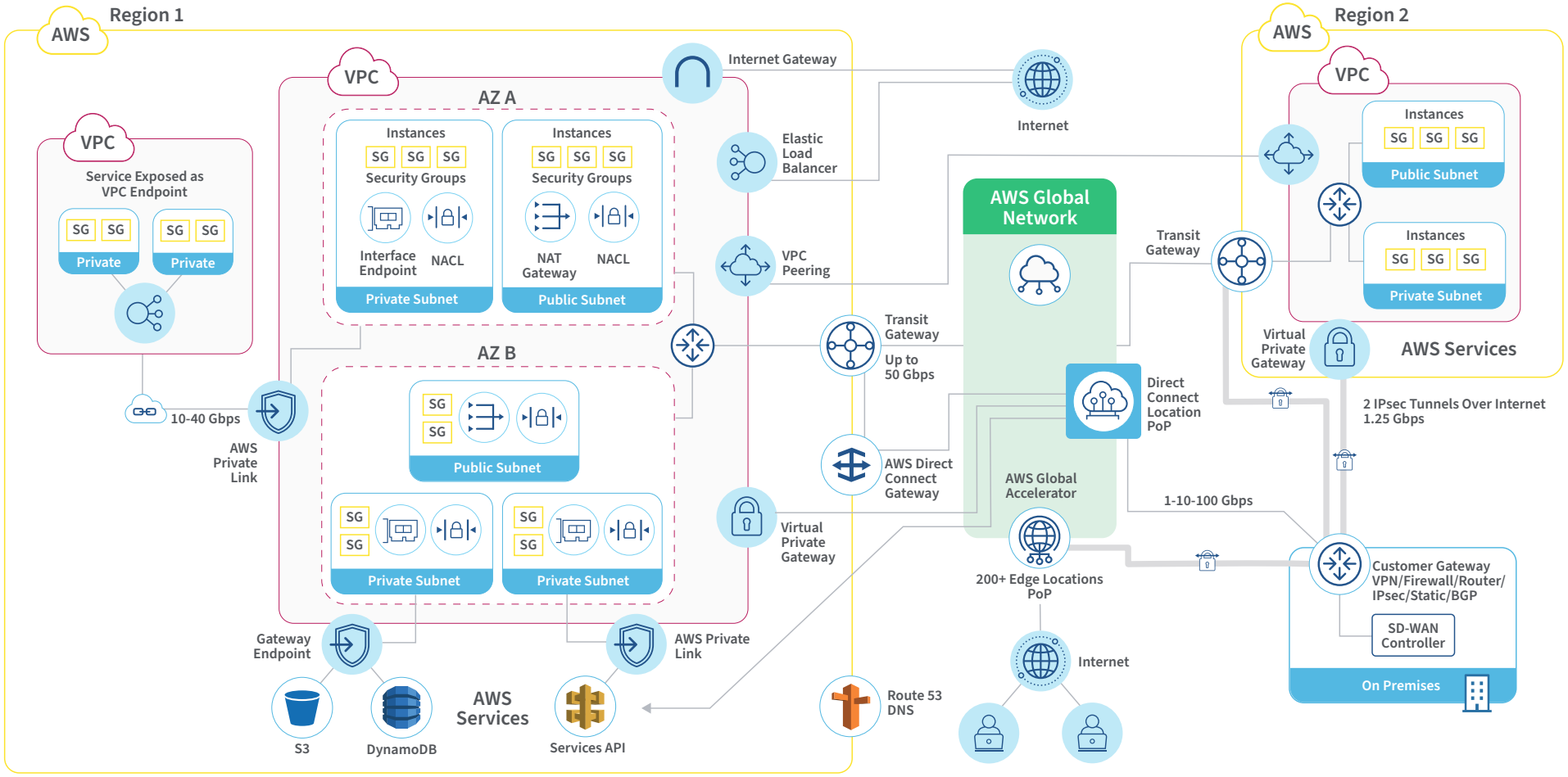
Poor performance is a non-starter, security incidents are a death knell, and cloud costs eat into profit margins as you grow. Therefore, observing, understanding, and optimizing public cloud and hybrid networking is fundamental to thriving in the cloud.

When it comes to cloud ways of doing things, networking feels strange and overwhelming. Cloud networking is not a forklift migration or even a translation from data center networking. It requires new know-how to be effective while taking full advantage of what public clouds enable.



> **"Traditional networking engineers entering the world of Amazon Web Services for the first time often feel a bit like Alice in Wonderland. Everything looks and sounds familiar, and yet it all feels a bit different."**
>
> **— Ivan Pepelnjak, Independent Network Architect, Author, Blogger**

Cloud networking gets complex fast



Manage complexity with smart networking choices

Embracing flexibility and speed is great, but it demands visibility and the ability to get answers fast from what inevitably will become a complex cloud network environment.

As applications are broken into services and deployed in multi-region networks of thousands (or even hundreds of thousands) of VPCs, cloud teams will have to make smart network decisions. Considerations about optimized internet and third-party service access, content pushed to the edge, and connection to on-premises workloads will continue to raise the bar on complexity. Given the direction that modern business is taking, your cloud networking diagram will eventually look like the one on the previous page.

Cloud network complexity can grow to unmanageable levels for various reasons: lack of planning, team boundaries, regulatory requirements, or ad-hoc and PoC deployments that move to production with suboptimal networking choices.

You can manage complexity with well-designed cloud networking. Take the first step in the right direction: learn the cloud ways.

AWS offers a broad set of cloud networking services. You need to build and manage reliable, high-performance, secure, and cost-effective public and hybrid cloud environments. With this objective, it is necessary to understand cloud networking constructs, connectivity types and their respective trade-offs.

In the pre-cloud world, adding or changing the network was an important decision, involving physical hardware and static configuration changes that were difficult to reverse.

Now, with IaaS, the ease with which changes can be made causes a negative side-effect. In many cases, poor configuration choices lead to inefficiencies because “we can always go back and fix things” if needed.

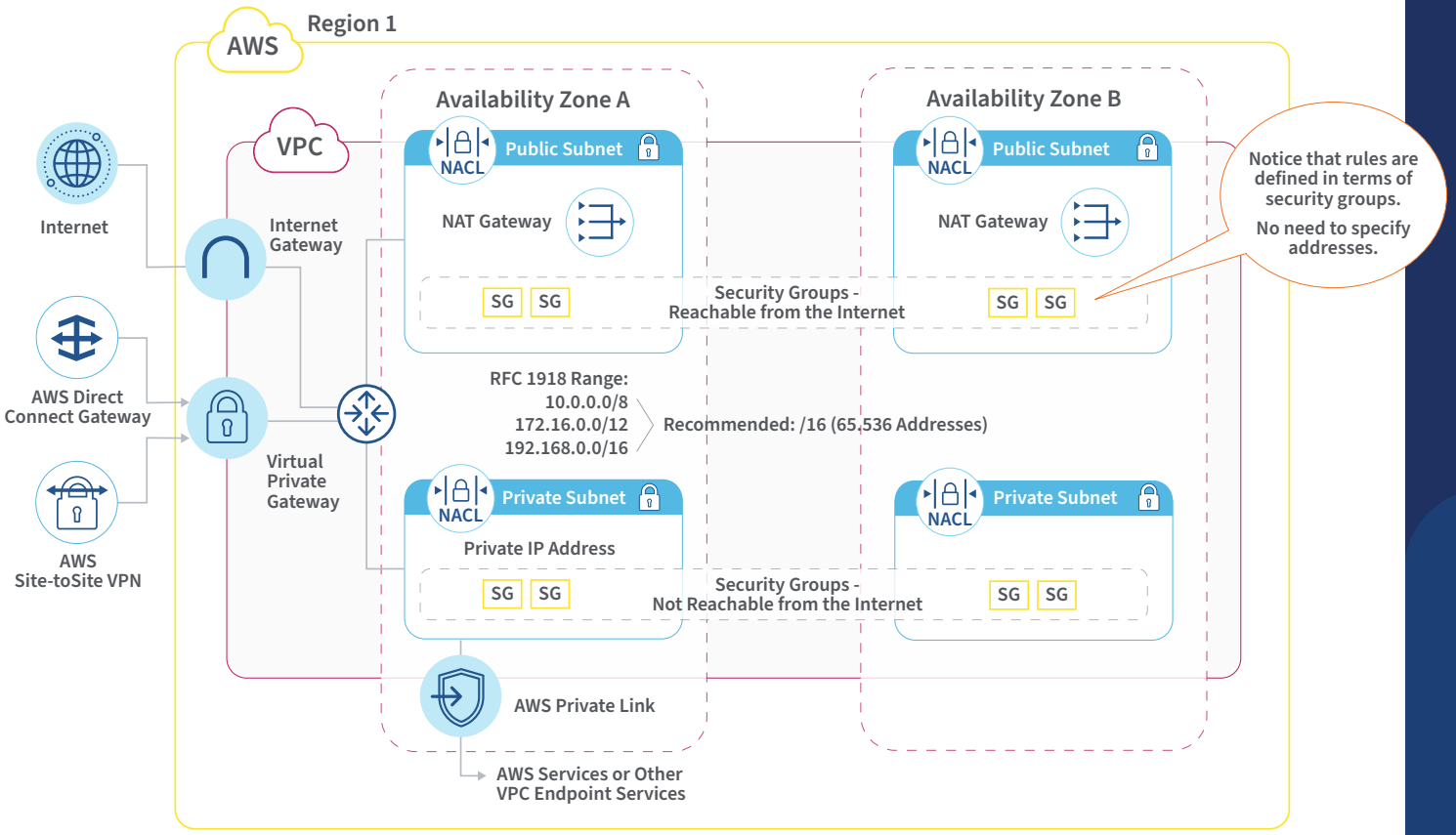
Unfortunately it is also harder to see suboptimal choices after the fact.

PART 2

Cloud Network Intricacies

First, let's look at how AWS VPCs work

Virtual Private Clouds (VPCs) are isolated networks on AWS infrastructure that act as your virtual data center networks in the cloud. VPCs can span multiple availability zones in the same region. They support Network Access Control Lists (NACLs) at subnet level and Security Groups (SG) at the instance level.



KEY VPC NETWORKING CONCEPTS

Public and Private Subnets:
Public (reachable over the internet) and private (not routable outside the VPC) IP addresses are assigned, respectively

Route Tables:
Set of rules to direct network traffic from/to subnets or gateways

Internet Gateway (IGW):
For internet access from public subnets

NAT Gateway:
For internet access from private subnets

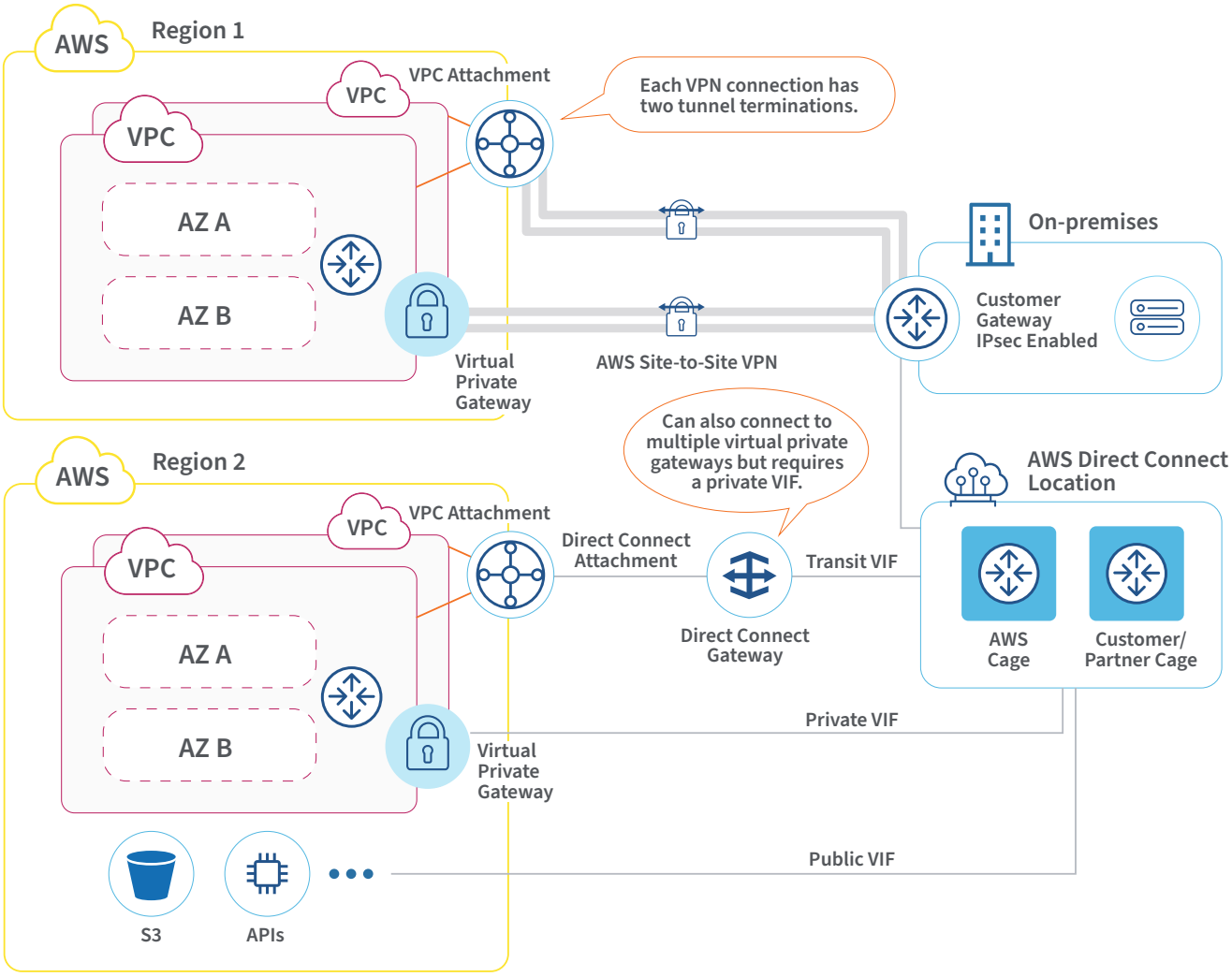
Virtual Private Gateway (VGW):
For on-premises communications via VPN or Direct Connect

VPC Endpoint:
For private connections to supported AWS services and VPC endpoint services via PrivateLinks (no public address required)

Network ACLs (NACLs):
Stateless firewalls applied at the subnet level ("allow" and "deny" rules)

Security Group (SG):
Stateful distributed firewalls applied at the instance level (only "allow" rules)

Next, let's see how VPCs work in hybrid environments



It is common for companies to need connectivity between their AWS VPCs and on-premises networks, such as data centers, offices, branches, and campuses.

AWS offers AWS Site-to-Site VPN and AWS Direct Connect to extend VPC access to on-premises networks.

VPCs in hybrid environments CONTINUED

AWS SITE-TO-SITE VPN

Securely connect AWS VPC to remote networks and users via VPN tunnels

Main Use Cases

- One-to-one connection: Customer VPN to a VPC Virtual Private Gateway
- One-to-many connections: Customer VPN to a Transit Gateway

Considerations and Limitations

- 1.25Gbps IPSec tunnel
- Two tunnels terminating in different AZs for failover (not combined)
- Static and dynamic (BGP) routing

AWS DIRECT CONNECT

Provides more reliable direct-to-fiber connectivity between AWS global network and on-premises infrastructure bypassing internet providers

Main Use Cases

- Private virtual network interface (VIF) to connect to VPCs
- Transit VIF to connect to Transit Gateways (via Direct Connect Gateway)
- Public VIF to access AWS public services

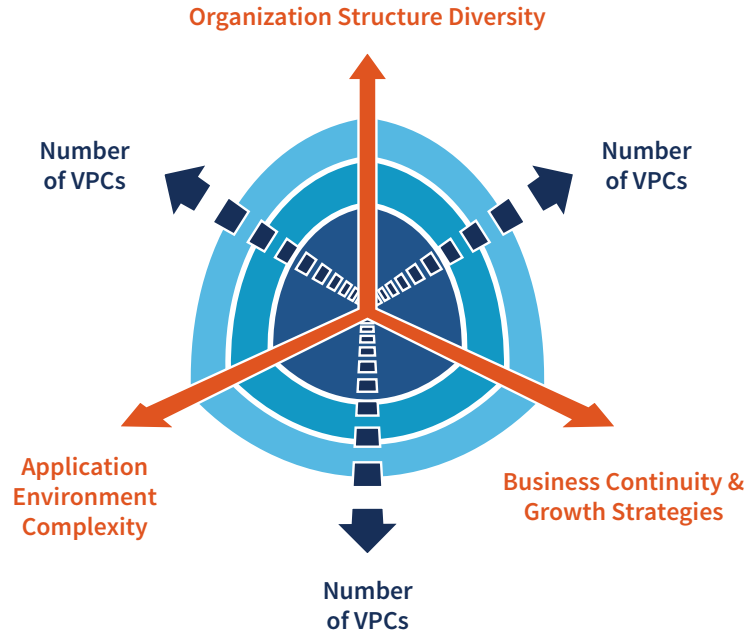
Considerations and Limitations

- 10 VGWs per region (no overlapping CIDRs) per Direct Connect Gateway
- 1/10/100 Gbps port speed (LAG supported)
- 50 VIFs (private or public) and 1 Transit VIF (1 VIF of any kind per hosted connection)
- 10 active connections per region per account
- Priced per port hours and outbound data transfer

Then, scale to multiple VPCs in different regions

There are many reasons for deploying multiple VPCs. That's because VPCs provide, in addition to network isolation, accountability, manageability, and security layers for:

- **Organization structure:** Departments, business units
- **Workload types:** Application components, shared services
- **User types:** Employee, customers, partners
- **Environment types:** Development, testing, production
- **Business continuity strategies:** Geo-coverage, HA, DR
- **Growth strategies:** M&A, offices, branches, campuses



The steep growth of public cloud adoption shows that VPCs will grow quickly in numbers requiring different networking constructs and connectivity methods.

As organizations increase their cloud presence in multiple regions with a growing number of VPCs and workloads, they will need VPC networking capabilities.

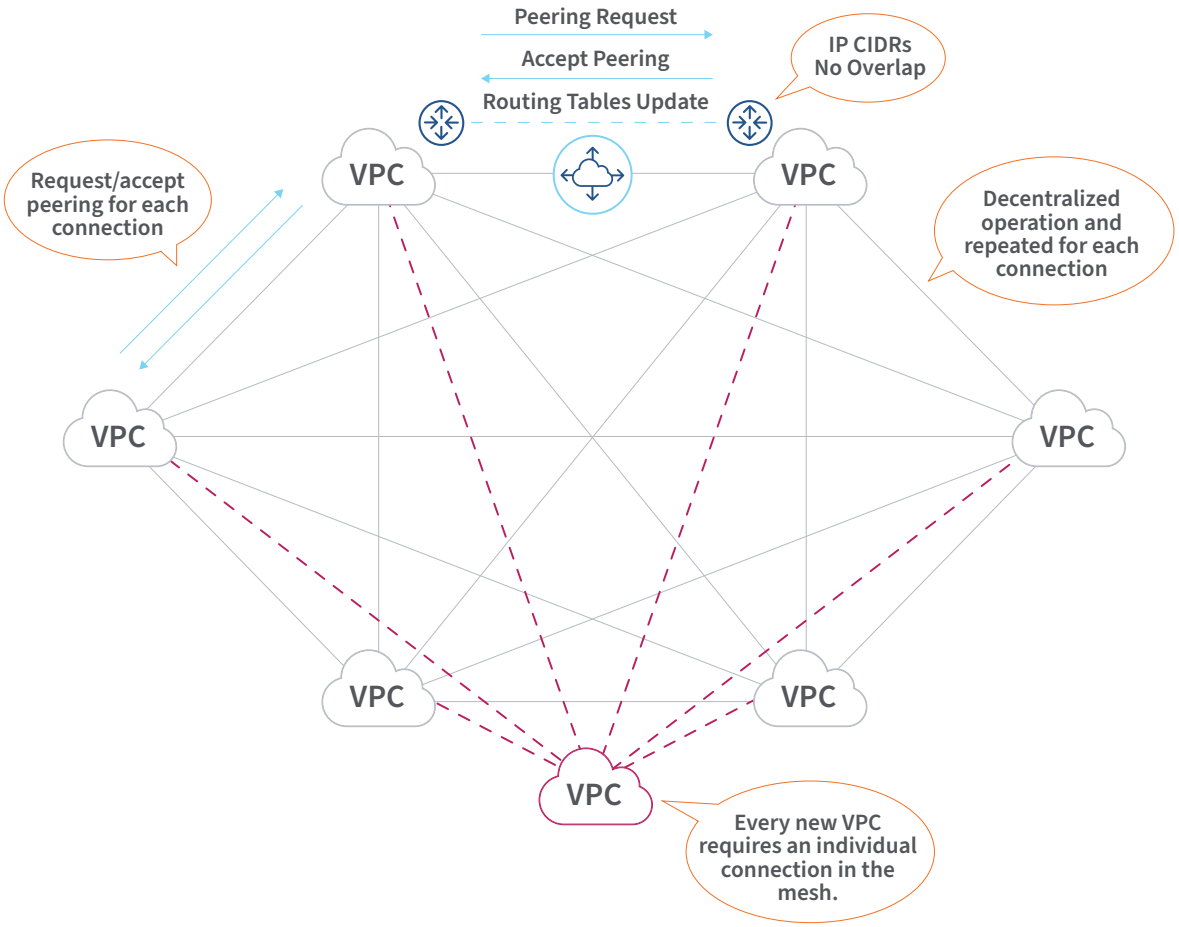
Let's take a look into key cloud networking constructs:

- VPC Peering
- Transit Gateways
- VPC Endpoints
- AWS Global Accelerator

KEY CLOUD NETWORKING CONSTRUCT

AWS VPC Peering: One-to-one connections

Peering VPCs in AWS is like connecting two data centers with a dedicated line between them, but instead of a physical connection, VPC peering uses a logical private IP connection.



VPC PEERING

Main Use Cases

- Interconnect VPCs of any account in any region
- Decentralized networking: request/accept transaction between account owners

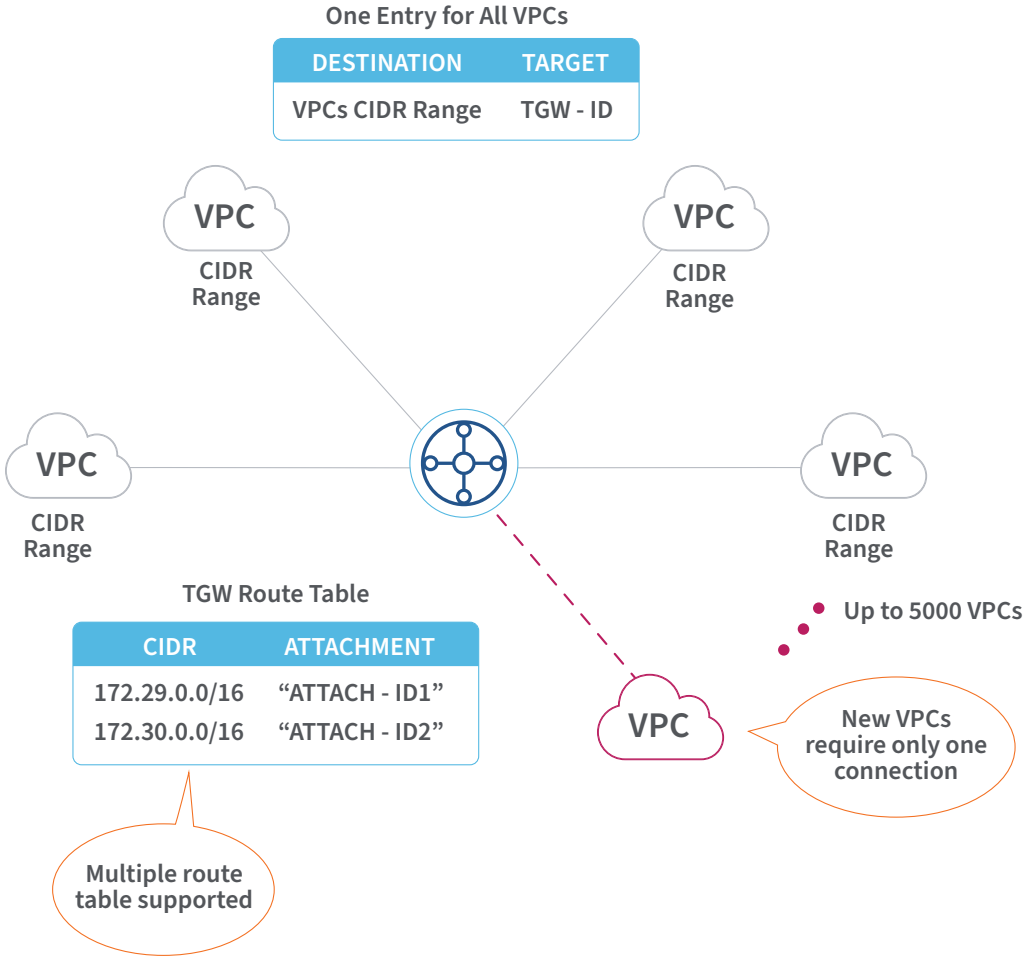
Considerations and Limitations

- Maximum of 125 peering connections per VPC. But one-to-one architecture is not recommended beyond 20 VPCs in a mesh.
- VPC CIDR ranges must not overlap
- No connectivity charges, but charges based on data transfer

KEY CLOUD NETWORKING CONSTRUCT

AWS Transit Gateway routing

Transit Gateway (TGW) is an AWS routing construct that you can think of as a cloud router. It is meant to scale the management of VPC connectivity in AWS cloud and to on-premises.



AWS TRANSIT GATEWAY

Main Use Cases

- Interconnect multiple VPCs in any-to-any hub and spoke design
- Centralized inter-VPC networking
- Static and dynamic (BGP) routing for VPN and Direct Connect attachments
- Multicast routing concepts
- Entry point for on-premises connections (see diagram on following page)

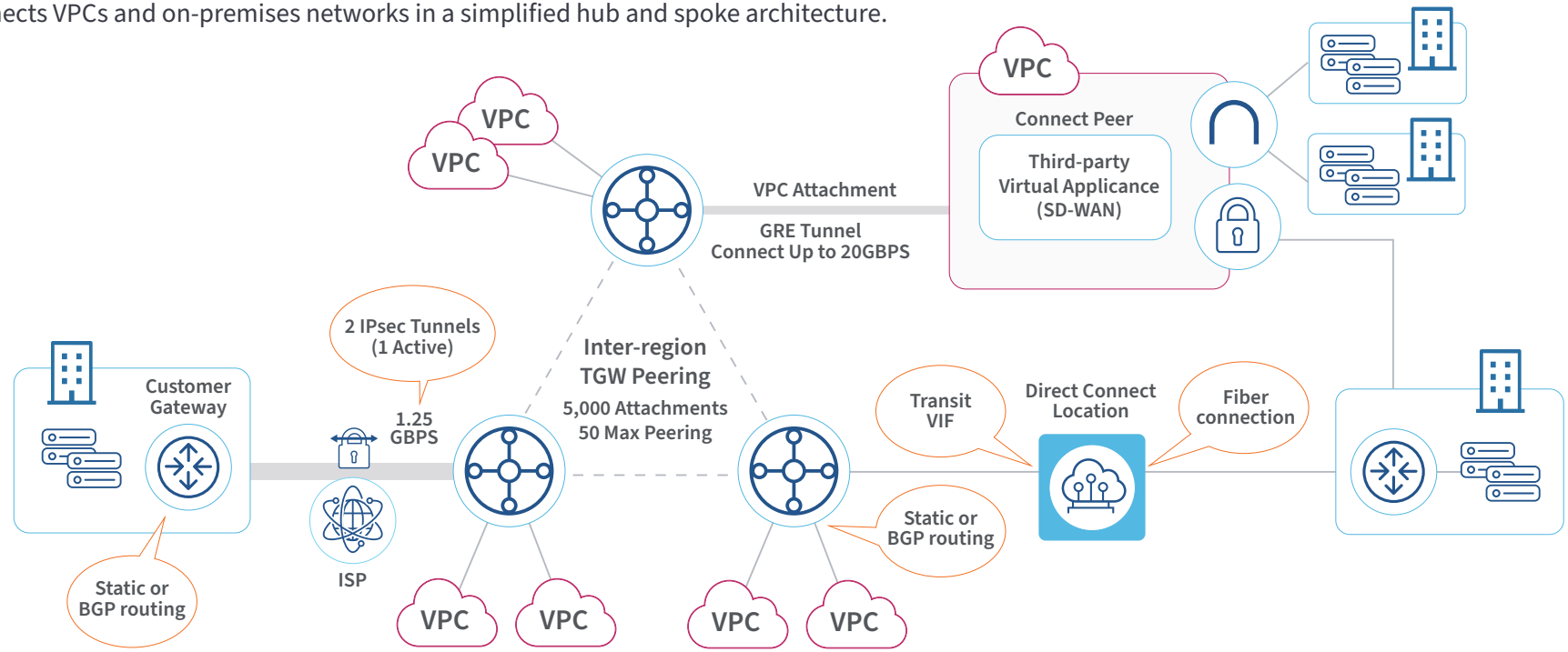
Considerations and Limitations

- Regional constructs. Inter-region connectivity requires TGW peering (up to 50 per TGW)
- Up to 5 TGW attachments per VPC; up to 5 TGW per region per account
- Up to 5000 attachments; up to 50 Gbps per VPC attachment
- Up to 20 route tables supported per TGW (soft limit)
- Charges apply to data transfer and attachment connection

KEY CLOUD NETWORKING CONSTRUCT

AWS Transit Gateway Global Hybrid Network

Transit Gateways can work as peered regional hubs to build a global hybrid network that connects VPCs and on-premises networks in a simplified hub and spoke architecture.

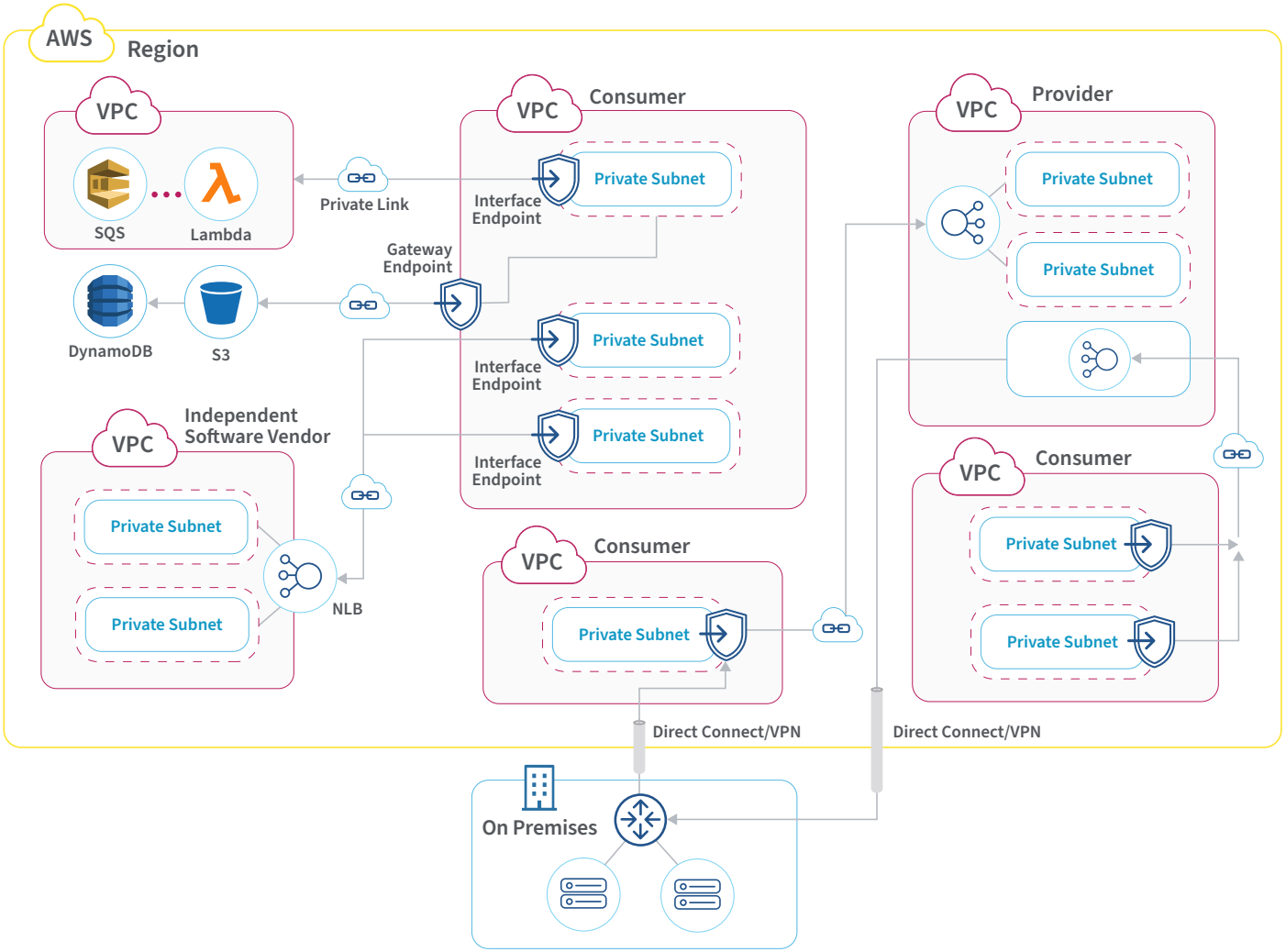


TGWs support the following attachments types:

- VPCs
- VPN
- Direct Connect
- Transit Gateway peering
- Transit Gateway Connect (third-party appliances/SD-WAN)
  - Requires an existing VPC or Direct Connect attachment as the transport attachment. Supported in some AWS regions.

KEY CLOUD NETWORKING CONSTRUCT

Now, let's learn how to connect to services using VPC endpoints



AWS service connectivity is provided via VPC endpoints powered by AWS PrivateLink connections.

This type of connectivity is very versatile because it provides secure service connectivity (including to third-party services from independent software vendors) to a large number of VPCs across different accounts without requiring networking or firewall rules changes. Endpoint policy and security group association can be used to limit access.

KEY CLOUD NETWORKING CONSTRUCT

VPC endpoints CONTINUED

VPC Endpoint and AWS PrivateLinks

One-way entry point (only service consumers can start a connection) to privately connect instances on VPCs to services in AWS without requiring internet and NAT gateways or VPC peering

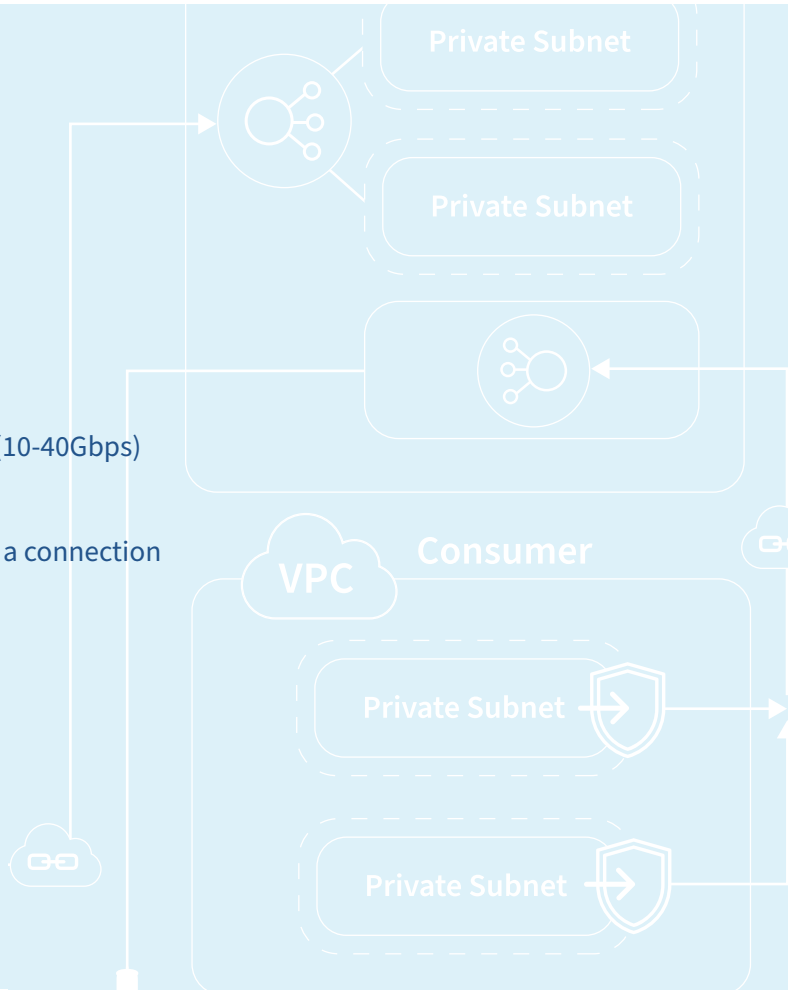
Main Use Cases

- Endpoint Gateway*, interface endpoint and Gateway Load Balancer endpoint to access, via private address, a supported AWS service or a VPC endpoint service (AWS PrivateLink-powered service)
- Abstract networking to service endpoints for agile app development
- Cloud services accessed via private IP address without IPAM or concerns over IP CIDR overlap issues

**Endpoint Gateway supports only S3 and DynamoDB*

Considerations and Limitations

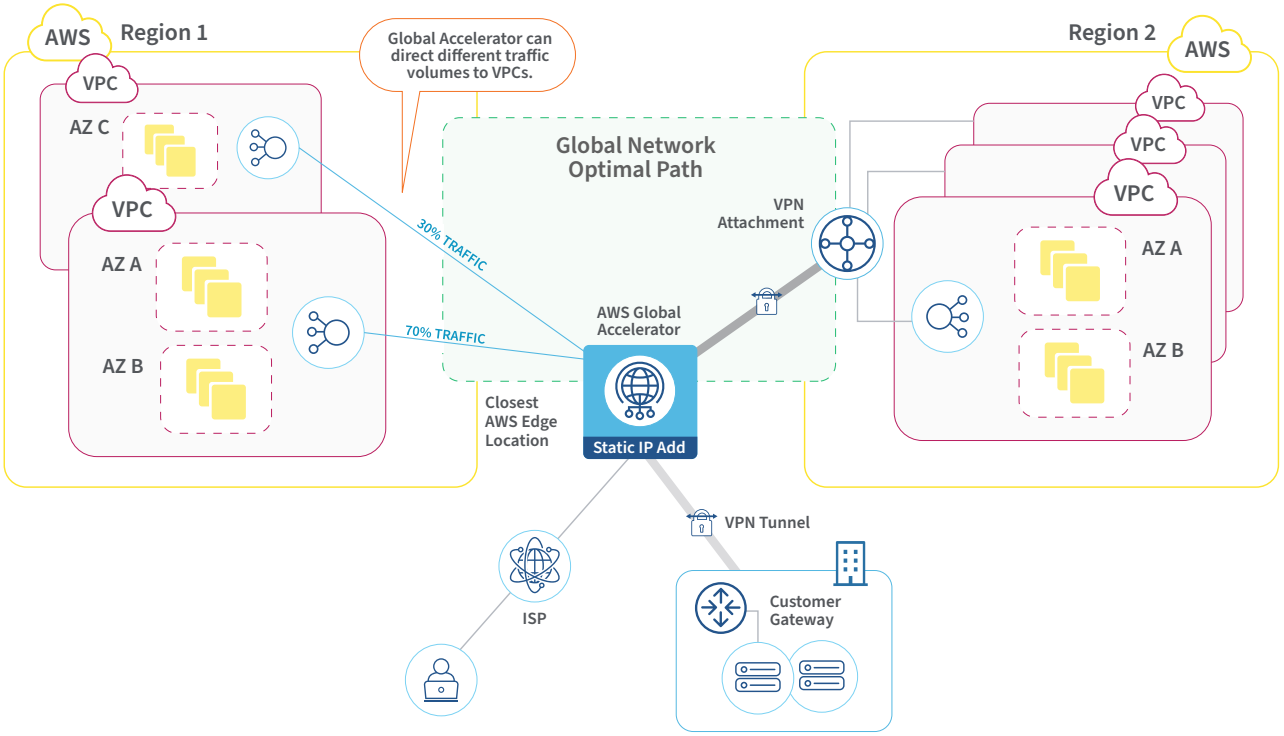
- Regional construct
- 50 AWS PrivateLink connections (10-40Gbps)
- TCP traffic only
- Only the consumer side can start a connection
- No CloudWatch metrics



KEY CLOUD NETWORKING CONSTRUCT

Add Global Accelerator to improve experience

Global Accelerator is a network service offered at AWS edge locations to provide an optimal or custom-defined networking path on the AWS network to applications deployed on AWS cloud.



Global Accelerator supports intelligent routing based on health checks, user locations and custom policies. It provides a fixed entry point (static IP addresses) to application endpoints to which specified traffic volume dial can be applied.

GLOBAL ACCELERATOR

Application traffic (TCP & UDP) is directed to the closest AWS edge location and is optimally and custom routed to regional resources and endpoints (NLB, ALB, EC2 instances, etc.)

Main Use Cases

- Improve experience of performance-sensitive applications (e.g., gaming, VoIP, video conferencing, mobile, etc.) by reducing latency and jitter
- Blue/green deployment, A/B testing, DR/multi-region resiliency
- AWS Accelerated Site-to-Site VPN

Considerations and Limitations

- 20 per account
- 10 listeners (TCP, UDP ports)
- Endpoint types: NLB, ALB, EC2, Elastic IP address, VPC subnet

PART 3

Challenges in Cloud Networking

The overarching challenge with cloud networking: You don't see it!

Cloud networking is not only complex, it is like stepping into a black box. Cloud tools don't know anything about *how* traffic is flowing from A to B. But it is *in the how* that you can make a difference in performance, cost, and security.

There are many networking constructs, such as gateways, load balancers, connectivity types, routing and tunneling options to connect VPCs across regions and to on-premises networks. Chances are that something will need fixing and optimizing over time. Network teams will be under pressure collating data from multiple sources and tools in an attempt to assemble the network paths and trace cascading events leading to identify root cause.

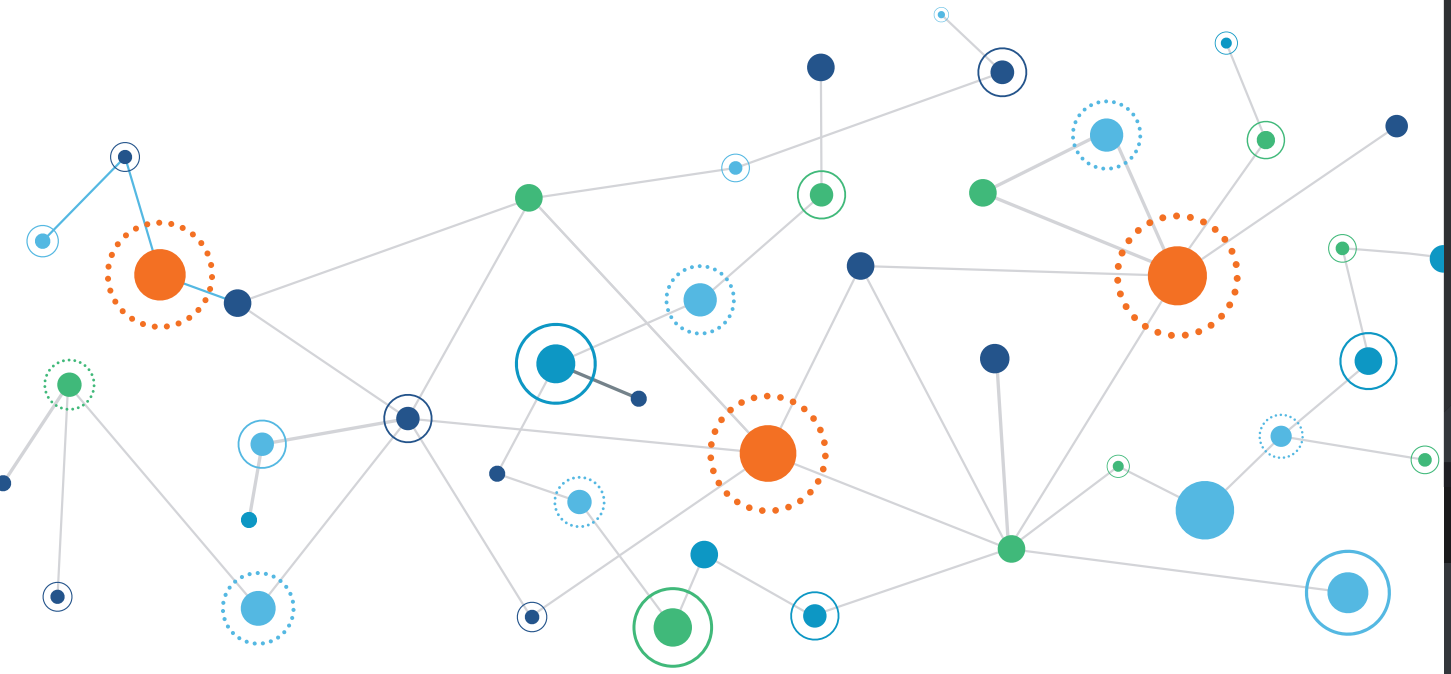
With legacy and siloed monitoring tools:

- **You can't see the network paths nor the traffic on them.**
- **You can't see how they are performing.**
- **You can't see the cost they are incurring.**
- **You can't see potential security and policy violations.**

What you don't see and don't know leads to the top 3 cloud networking gotchas

The black box effect has multiple downsides:

- You can't visualize, for instance, how flows are routed through gateways, load balancers, endpoints, or unexpected paths composing your end-to-end network
- You can't easily monitor performance across intra- and inter-cloud connections and hybrid networks (data centers, branches, campuses) from the users' perspective
- You can't easily govern cloud networking cost



UNDERSTANDING THE BLACK BOX: DEFINING OBSERVABILITY

Like so many other terms in software engineering, "observability" is a term borrowed from an older physical discipline: in this case, control systems engineering. Observability is the mathematical dual of controllability.

"Less formally, this means that one can determine the behavior of the entire system from the system's outputs. If a system is not observable, this means that the current values of some of its state variables cannot be determined through output sensors."

— Charity Majors, Co-founder and CTO, Honeycomb.io

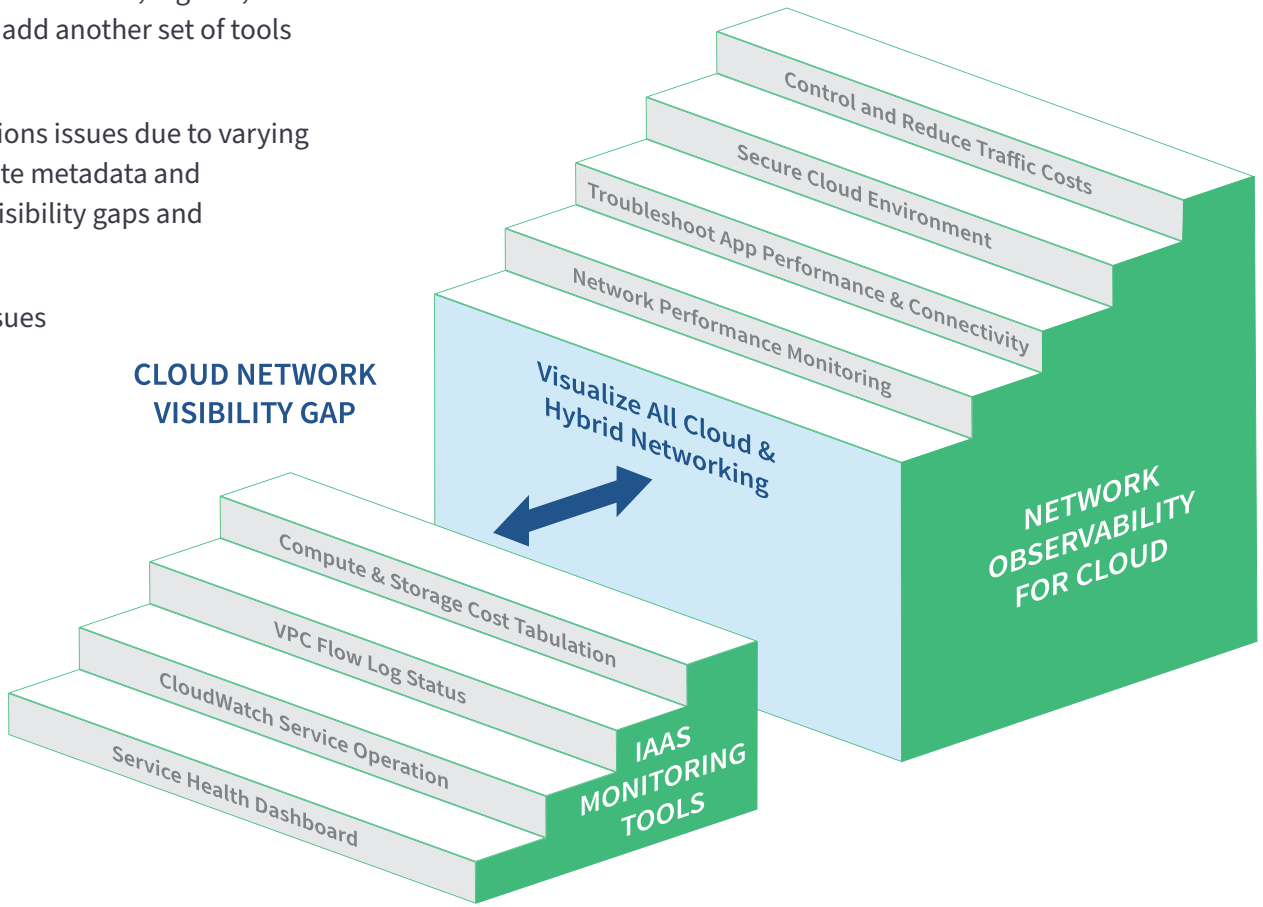
Gotcha #1

Visibility gaps create fragmented views and hinder operation

Network, DevOps, and development teams use multiple tools to monitor the networks and traffic to and from workloads in and across clouds, regions, and VPCs. In the case of hybrid deployment, they have to add another set of tools for data centers, WANs, and SD-WANs.

Different tools also create data collation and correlations issues due to varying granularity, lack of real-time monitoring, and disparate metadata and metrics mapping, among other things. The result is visibility gaps and data islands.

Tool siloes create operational inefficiencies due to issues brewing in the dark. The negative results are more incidents, reduced uptime, and higher MTTRs.



Gotcha #2

Cloud network complexity makes it hard to early-detect and troubleshoot performance issues

The cloud networking black box effect becomes even more prominent when trying to troubleshoot performance in distributed environments, with multiple types of connectivity in the cloud and to on-premises sites. Besides routing and traffic dynamics, functionality and resource constraints can cause propagating degradations that are hard to isolate.

When network complexity meets speed of change and cascading effects, cloud teams, SREs, developers, or a network team can easily be overwhelmed and become:

- The target for all blame
- Reactive to problems reported by applications teams or users
- Unable to quickly identify root cause

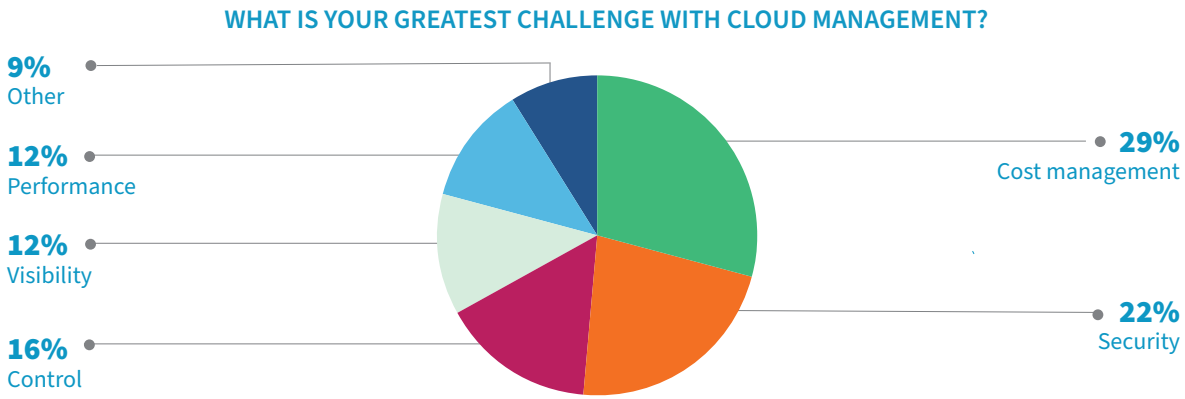
In complex environments, it is critical to be proactive in detecting early signs of performance degradation in each cloud and hybrid network path.

Gotcha #3

Inability to monitor cloud network costs leads to poor (and costly) network architectures

Not being able to centrally manage and control networking cost is a shared pain by all organizations using cloud networking. It is a direct consequence of the lack of cost visibility into VPC traffic and accounts. Existing cloud cost management tools don't provide the granularity and details needed for various types of cloud gateways and networking services. This has significant implications to costs.

Cloud networking cost management requires visibility into traffic volume (data being transferred), route connectivity, and networking services utilized.



In a recent survey of 310 AWS users, we found that cost management was the number one challenge reported. [SOURCE]

A COMMON SCENARIO LEADING TO UNNECESSARY COSTS

Intra-region traffic goes through an Internet Gateway — with incurring cost starting at $0.09/GB (after the first 1GB) — instead of using lower cost and more secure intra-region networking constructs, such as AWS PrivateLink endpoints.

PrivateLink endpoints charges have an hourly component on provisioned VPC endpoints and a data processing component based on data volume processed through the VPC endpoint. Despite two cost components, **the overall cost is substantially lower than via an Internet Gateway.**

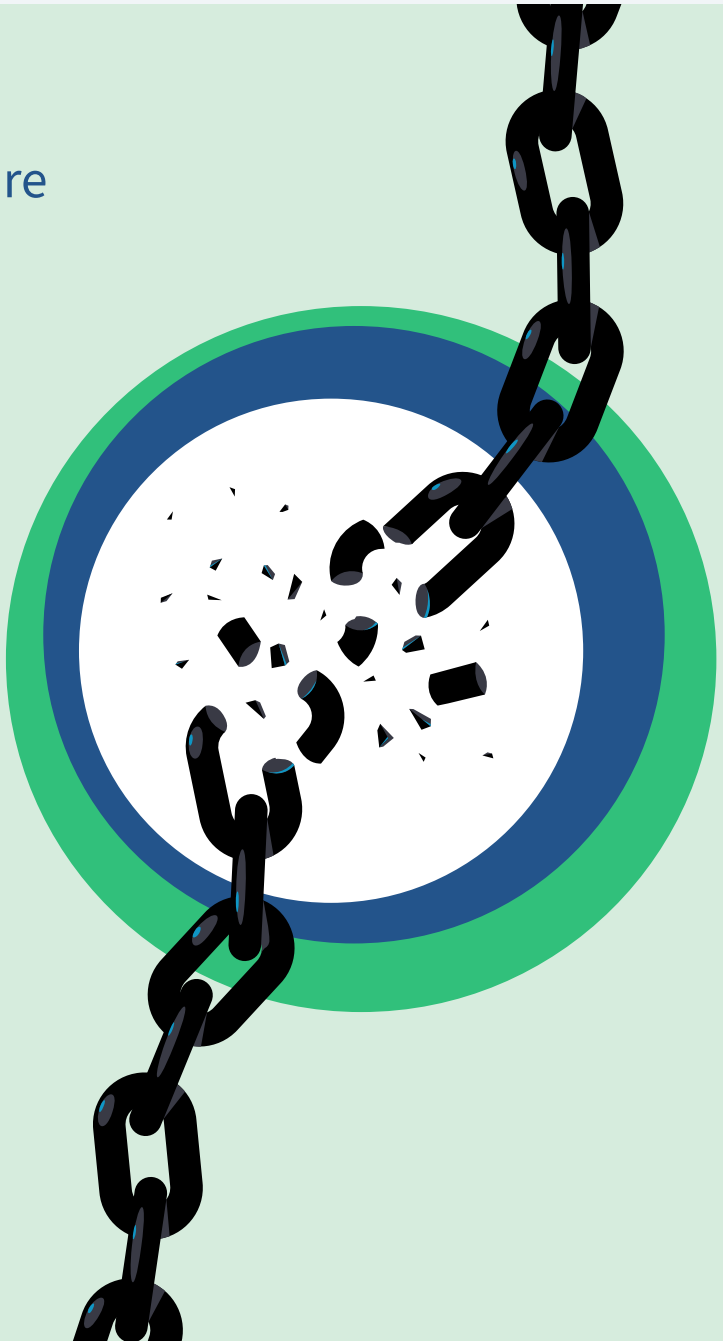
For example, for the US West (Oregon) region, it's $0.01/hour of provisioned endpoint and $0.0035/GB of data processed.

Consider a monthly 1TB intra-region data transfer using VPC endpoints **could represent an 80% cost reduction** as compared to using an Internet Gateway. If a gateway endpoint is used (for Amazon S3 or DynamoDB services), greater savings could be achieved since there is no charge for using gateway endpoints.

It is also difficult to maintain a strong security posture

There is no true translation from data center networking, connectivity types, firewall, and access control rules to the cloud without defeating the purpose of modern cloud frameworks. Cloud security maturity requires alignment of security, network, DevOps, and development teams.

A chain is as strong as its weakest link. Security misconfigurations in cloud networking can open doors to violations, breaches, and attacks. Networking teams need to stay on top of cloud traffic to ensure that routing, access control, and load balancing policies are functioning as intended. They also need to get real-time alerts when there are anomalies. Flexibility must be provided for accountability and visibility.



PART 4

What’s Needed

Cloud networking is hard without network observability.

What *is* network observability?

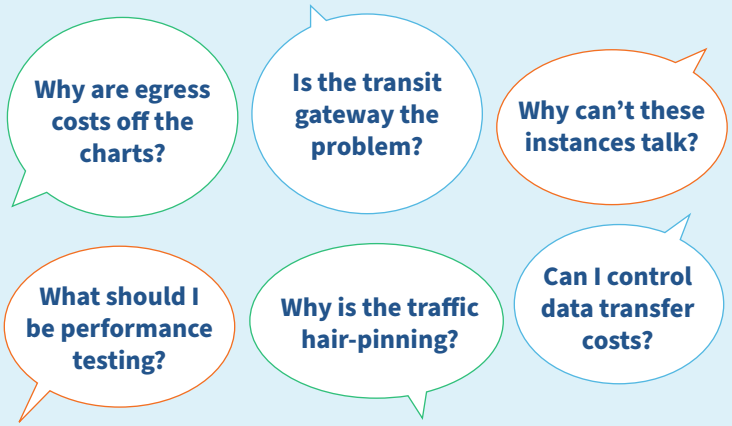
One thing needs to be clear: monitoring is not observability! You may hear the terms mixed up or used interchangeably, but they are entirely different concepts.

Monitoring tracks a predetermined dimension or expected event. It provides a view limited to what you already know to expect; if it happens, you will have answers.

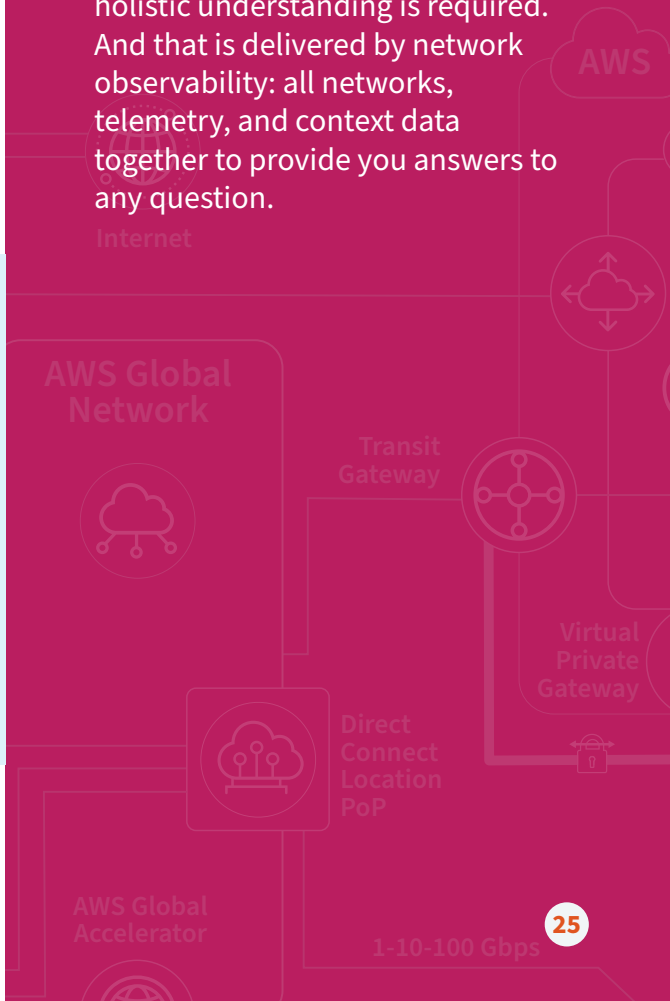
Monitoring is not effective in modern distributed cloud environments because teams mostly deal with what is called "unknown-unknowns." Without the combination of telemetry types and metadata to help achieve observability, unknown-unknowns cannot be addressed.

net•work ob•serv'a•bil'i•ty

(n) **1.** The ability to answer any question about your network.

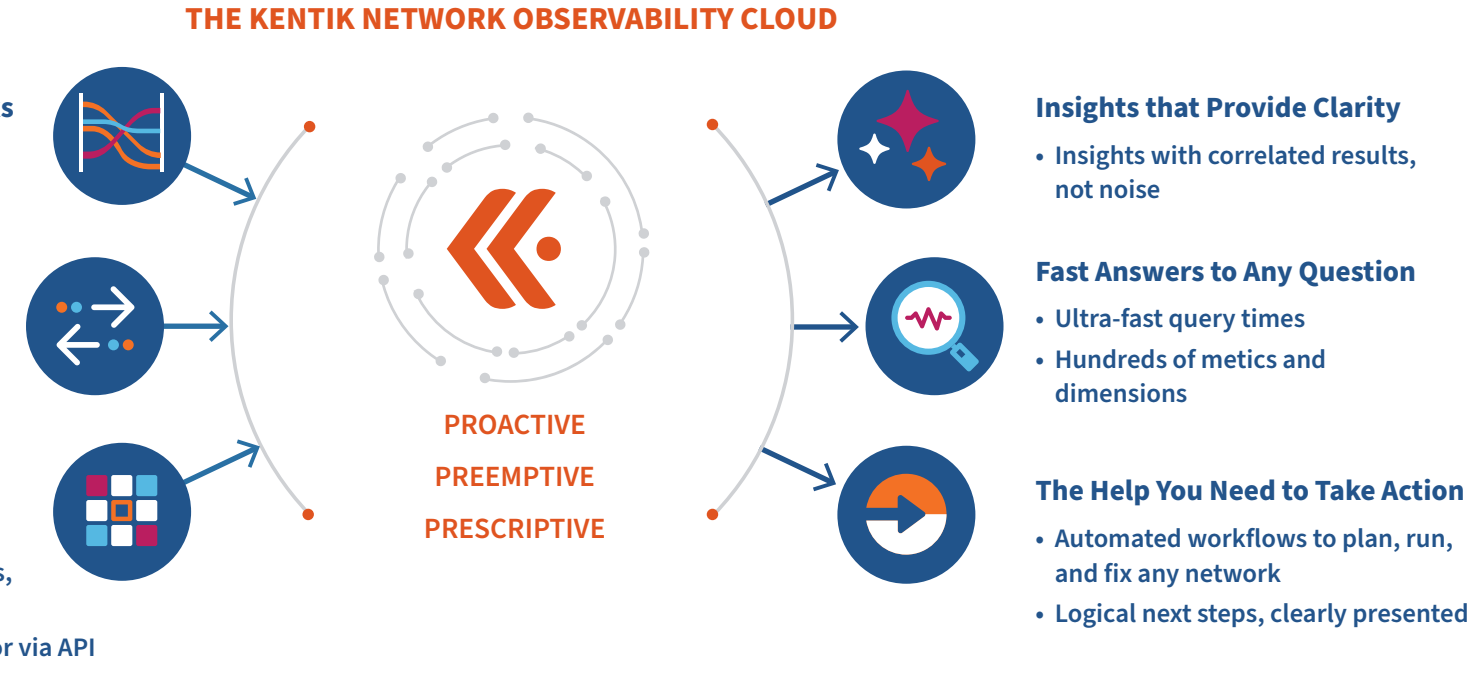


States in complex public and hybrid cloud networking environments are not understood from anomalies in dashboards. A holistic understanding is required. And that is delivered by network observability: all networks, telemetry, and context data together to provide you answers to any question.



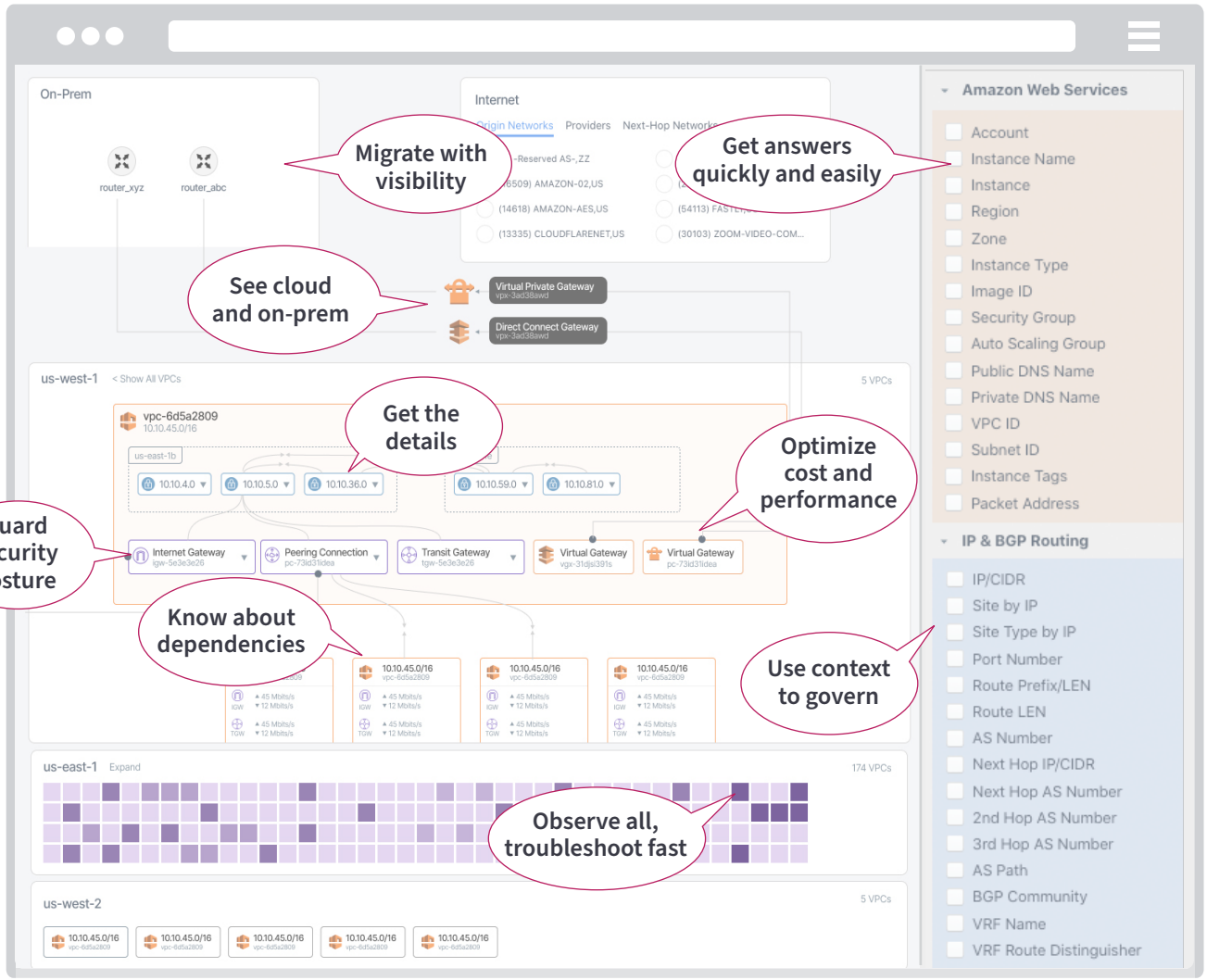
Network observability is proactive, preemptive, and prescriptive

Network observability is the ability to answer any question about your network. Network observability is critical to reliable, high-performance public cloud networks because it empowers IT teams to get the answers they need to plan, run, and fix the network. Network observability also assists cross-functional teams with the agility to collaborate and respond.



With network observability you can...

- Visualize and streamline network performance troubleshooting across public clouds and hybrid infrastructures.
- Optimize cloud services delivery by identifying best workload regional distribution and route to services dependencies.
- Migrate applications to cloud with a knowledge of dependencies, a traffic profile baseline, and an estimate of data egress costs.
- Manage cloud traffic costs, detecting load shifts in egress routes and automating cost impact notifications.
- Secure cloud deployments observing and auditing traffic through security groups and network ACLs. Alert on anomalies and violations. Identify and neutralize attacks and attackers.

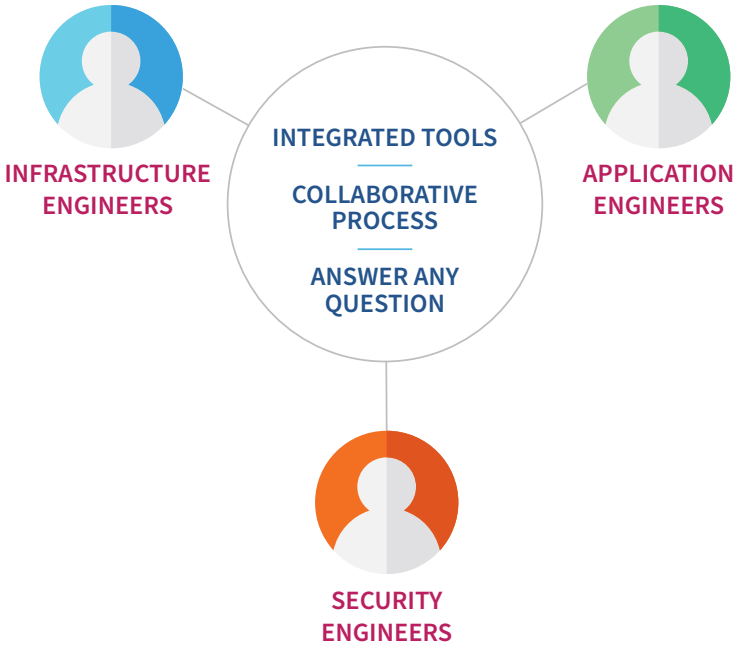


Get it right from the start

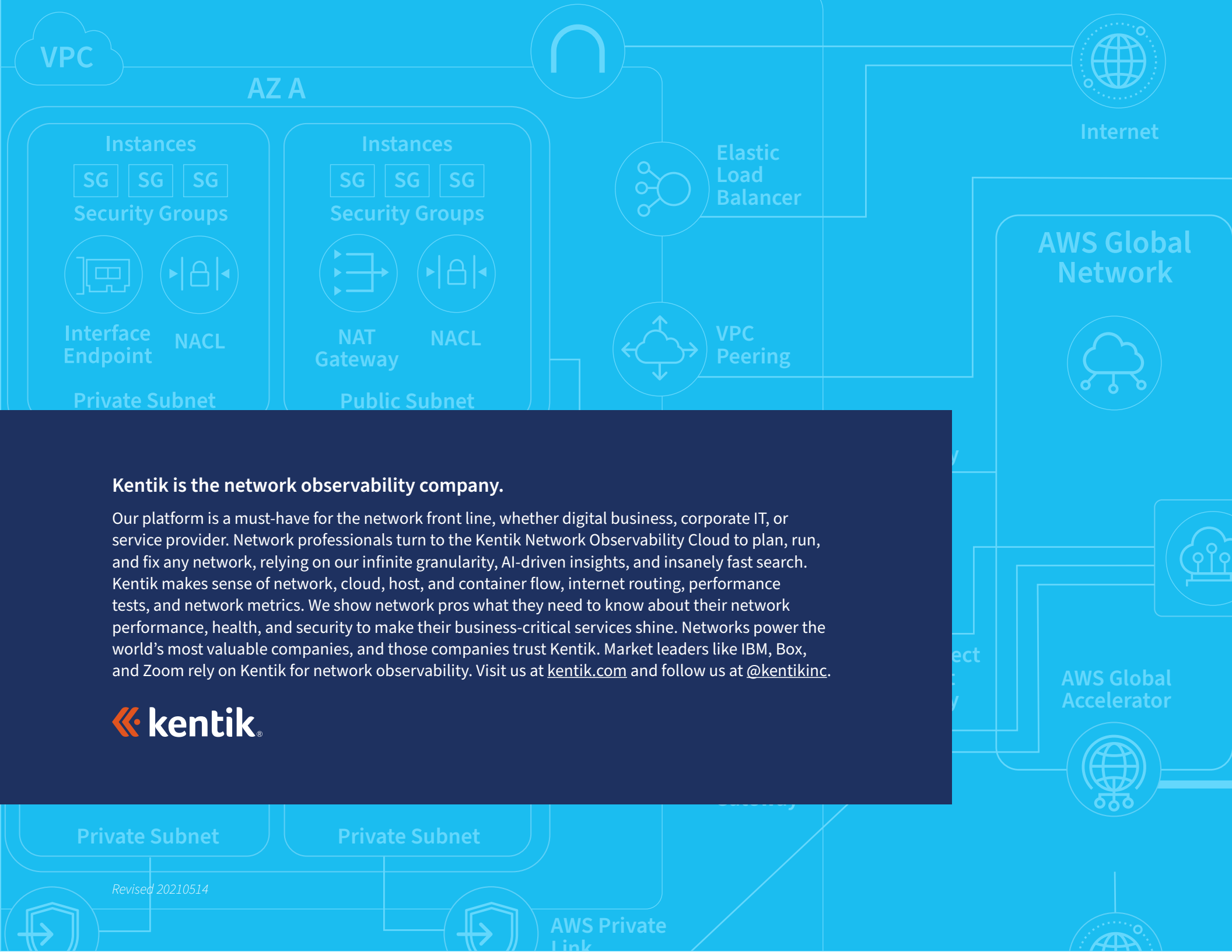
Network observability closes gaps in visibility, control, and operations in cloud networking.

Kentik's network observability platform provides unified visibility and insights into public cloud and hybrid networks. It collects and ingests enriched telemetry (synthetic tests, metrics, events, and metadata) from all logical and physical networks, from source to destination, and all paths in between.

Data is collected in real time, kept in an infinitely granular data store, presented in maps and made promptly available for fast querying. For the network pro, Kentik provides AI-driven insights and automated workflows that help answer any question about your cloud network. With network observability from Kentik you get to the root of cloud performance and connectivity issues before they become a problem for your users.



Success in cloud networking means establishing the correct governance and organizing for teamwork. To be successful, your teams need the right processes and tools to plan, run, and fix your cloud network.



Kentik is the network observability company.

Our platform is a must-have for the network front line, whether digital business, corporate IT, or service provider. Network professionals turn to the Kentik Network Observability Cloud to plan, run, and fix any network, relying on our infinite granularity, AI-driven insights, and insanely fast search. Kentik makes sense of network, cloud, host, and container flow, internet routing, performance tests, and network metrics. We show network pros what they need to know about their network performance, health, and security to make their business-critical services shine. Networks power the world's most valuable companies, and those companies trust Kentik. Market leaders like IBM, Box, and Zoom rely on Kentik for network observability. Visit us at kentik.com and follow us at @kentikinc.

kentik®

Revised 20210514