Identity-Powered Microsegmentation

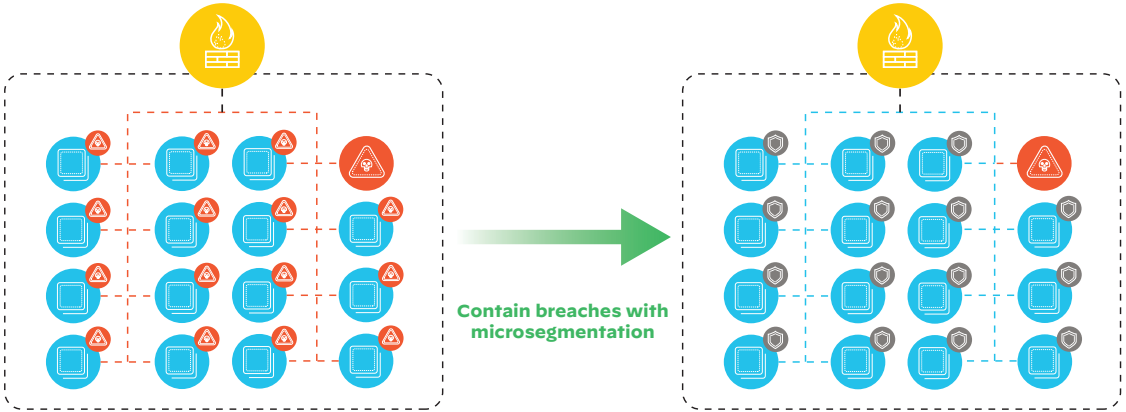
Going beyond network boundaries to protect cloud native applications

paloalto NETWORKS

Introduction

As organizations accelerate their cloud workload adoption—such as with on-premises virtual machines, public cloud virtual machines, and containers—they have greater responsibility to protect their digital assets from bad actors. Enterprises have shifted their cybersecurity methodology to ask when will a breach happen instead of if one will occur. When there is a breach, the best option is to contain the blast radius, and many enterprises turn to microsegmentation to prevent breaches from spreading laterally through their cloud environments.

At Palo Alto Networks, we have taken a unique approach to microsegmentation while allowing our customers to consume cloud compute, build cloud native applications, and increase application deployment velocity. De Facto workload/application segmentation technologies in the industry are tied to the IP infrastructure. An IP-centric approach to security introduces an overwhelming amount of complexity and a failure to establish a strong security posture. Our assertion is that segmentation must be abstracted from IP infrastructure to address application segmentation requirements and improve your application risk posture.



Network Segmentation Is a Quadratic Problem

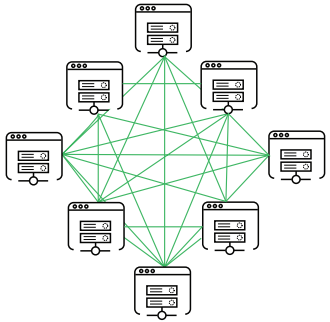
Network segmentation technologies today can shrink a large IP perimeter behind a firewall into smaller perimeters utilizing IP allow lists. Unfortunately, an IP-based perimeter still exists and is ineffective in reducing the attack surface for your high-value assets. Let's address some of the limitations of this approach to segmenting your cloud native applications.

Modern network security frameworks commonly use an IP allow-list policy model communications are blocked by default until someone explicitly writes an IP allow rule. While this model is simple, it boils down to a (N*(N-1))/2 or N2 problem where N is the number of workloads and the outcome is the number of rules. If you were to consider a set of eight interconnected workloads, you would need at least 28 IP allow rules. As the number of workloads increases the number of rules grow at an exponential rate. Some vendors hide the complexity of defining these IP rules with abstractions but hiding complexity does not address scale limitations—especially in highly dynamic environments running containers or microservices. The allow-list model is the right approach, but the complexity has to move toward an order of N problem—more on that approach later.

Network segmentation does not reduce your risk posture

$$\text{Minimum rules} = \frac{N * (N-1)}{2}$$

$$\text{Maximum rules} = N^2$$

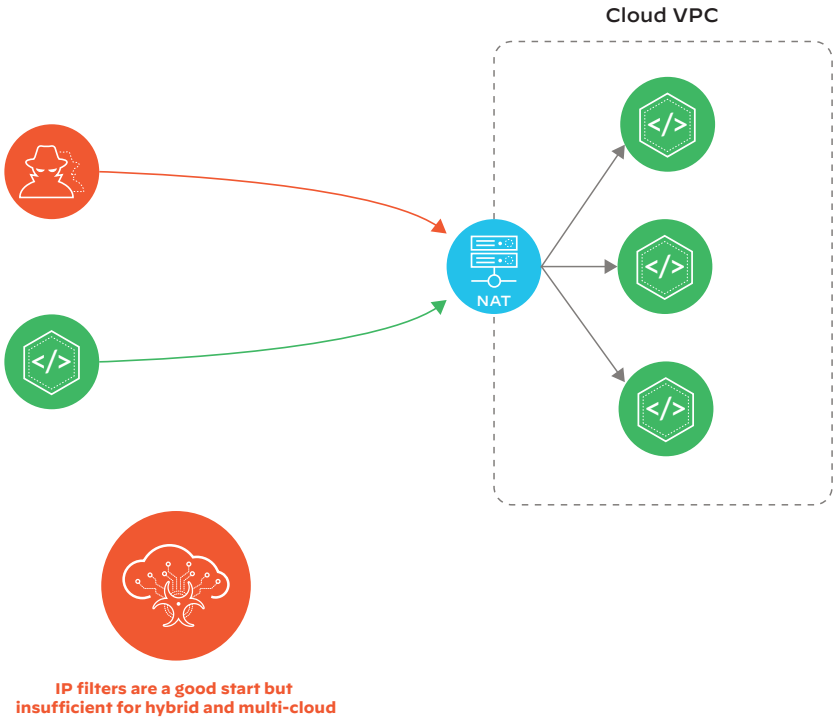


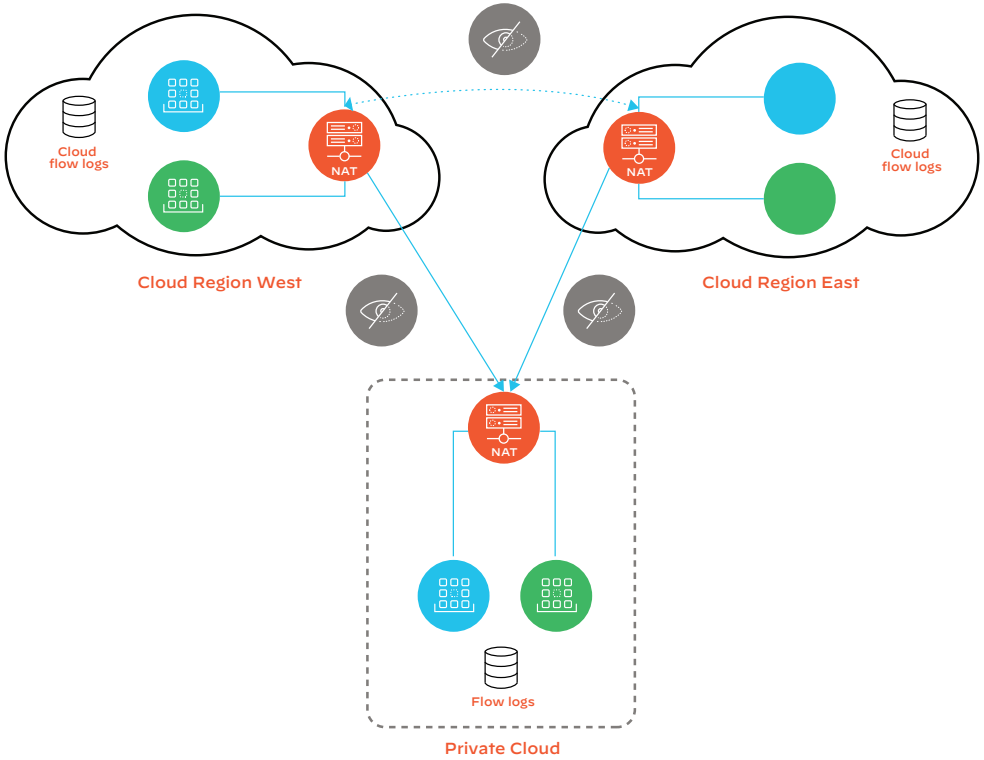
Number of workloads (N)	4	8
Range of rules	6-16	28-64

Multi-Clouds and Hybrids Span Multiple IP Domains

IP networks are not one flat layer today. Network Address Translation (NAT), proxies, and load balancers are common technologies that mask IP addresses. Most NAT technologies mask an entire IP domain using a single shared IP address. Traditional IP-based rules cannot span multiple IP domains. At best, network security teams can enforce coarse security policies that allow all traffic from a shared IP address or single IP domain.

This Illustrates the consequences of this approach. A group of cloud workloads exist in the same cloud VPC and provide an application service. Hosts outside of the VPC depend on this service and must pass traffic through a load balancer—doing NAT—before reaching the service. The hosts inside the VPC must trust the load balancer IP in order to allow outside hosts to communicate inbound, but what happens if an untrusted host tries to communicate with the load balancer? All communications from outside will be masked with the load balancer IP address. The hosts inside the VPC will not know the difference between trusted and untrusted hosts.





Understanding Applications Requires Visibility

Understanding applications and how they communicate is critical to operations. Organizations use network and security tools to gain visibility into application dependencies; however, traditional methods such as taps, SPAN, and cloud flow logs report IP addresses as the end-point identifier. Oftentimes, the logs neither identify who initiated the connection nor provide application context.

Network and security teams are burdened with combing through IP logs and stitching the IPs against a source of truth—such as an IP address manager (IPAM)—to determine which applications communicate with each other. Applications spanning multiple cloud environments communicate across multiple IP domains or move across NAT technologies. This makes IP flow stitching across clouds more challenging—sometimes impossible.

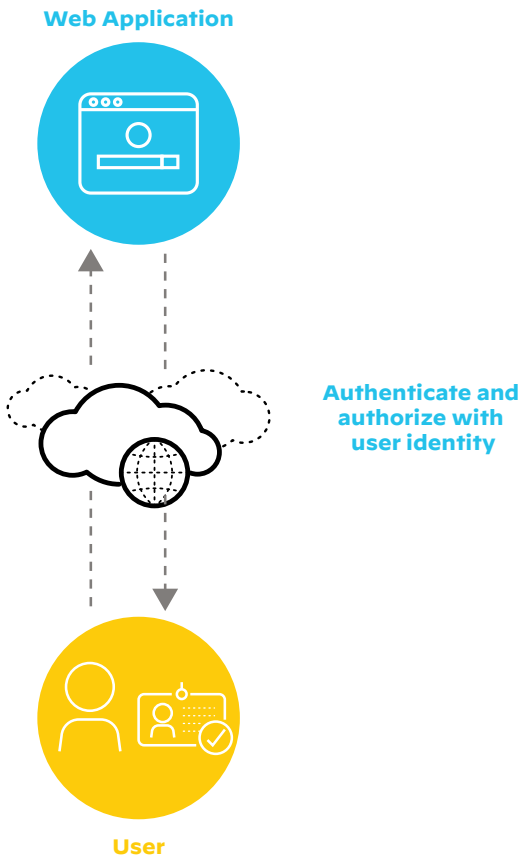
Network visibility tools do not visualize application dependencies

Identity: Separating Security and the IP

We, as users, access web applications for banking, social media, and email from the internet every day. Imagine if your home's public IP address—assigned by your internet service provider—was your only form of identification on the internet, and web applications used that IP to authorize access to personal data. You couldn't check your email inbox when traveling because your requests would come from a different IP. When friends and family enter your home, you wouldn't allow them on your guest Wi-Fi because their phones and laptops would use your public IP and have access to your banking information. This means of identification would minimize user experience quality and dramatically increase cyber risk.

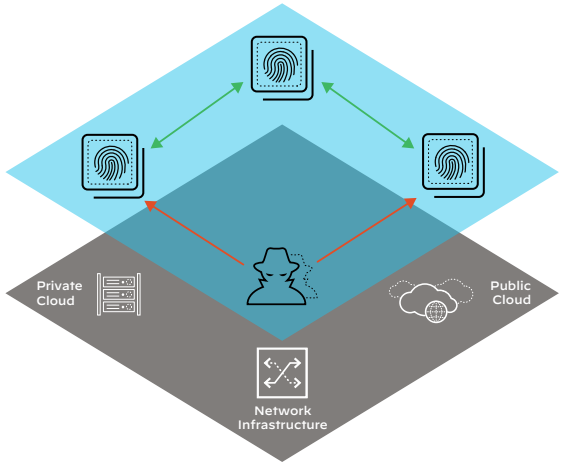
Now step back into reality: users remotely access web applications from any masked network or location. Web applications authenticate user identity before authorizing access to digital data.

Apply this concept to cloud native workloads. Applications depict the same characteristics as users; workloads can now exist on any cloud and any masked network. Application workloads need a persistent identity that is independent of an IP address and can be used for security policy enforcement.



Identity-Based Microsegmentation with Prisma Cloud

1. Separate security and the underlying network infrastructure. Assign cryptographic identity to cloud workloads.
2. Discover end-to-end application dependencies.
3. Manage security policies and monitor policy behavior without impact to applications.
4. Segment every app. Authenticate all requests before authorizing network access.



Palo Alto Networks brings a concept of identity to applications. Our approach is to assign every workload a cryptographically signed identity. Identity can be dynamically ascribed from cloud native sources.

System Information

Use operating systems, services, hostnames, and other system attributes as identity

Cloud Provider

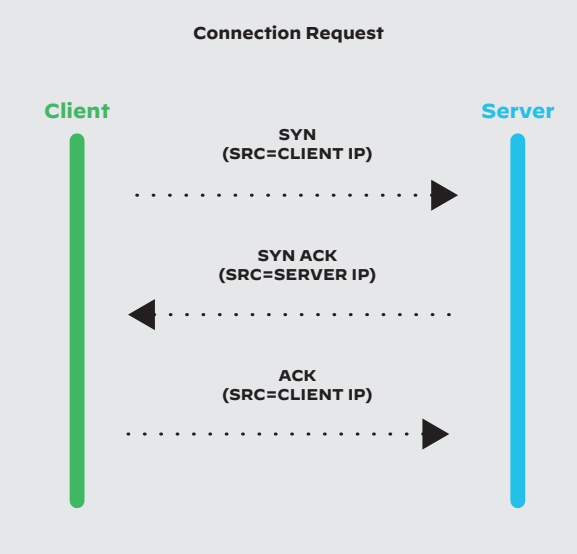
Derive identity from IAM roles and other cloud metadata sourced from Amazon, Azure, and Google clouds

Container Orchestrator

Extract Kubernetes® service accounts, app labels, namespaces, docker images, and more

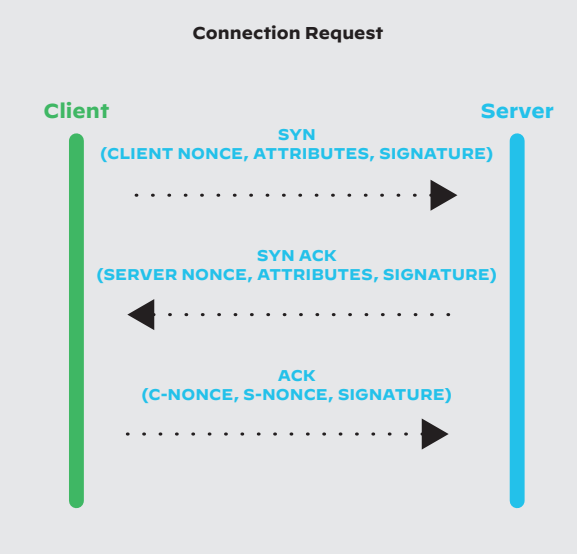
How It Works

Traditional IP Approach



IP is the application identifier. Reachability assumes network authorization to data

Identity-Based Microsegmentation



Cryptographic identity is used to first authenticate both workloads and then authorize network access

Identity-Based Microsegmentation Solves Problems

Applying segmentation across private and public clouds does not have to be impossible. Here is how Prisma Cloud resolves all of the issues highlighted earlier.

Learn More About Prisma Cloud

Prisma™ Cloud is a comprehensive cloud native security platform with the industry's broadest security and compliance coverage—for applications, data, and the entire cloud native technology stack—throughout the development lifecycle and across multi-cloud and hybrid cloud environments. The integrated Prisma Cloud approach enables SecOps and DevOps teams to stay agile, collaborate effectively, and accelerate cloud native application development and deployment securely.

To learn more about Prisma Cloud, visit paloaltonetworks.com/prisma/cloud.

No More Quadratic Rules

Prisma Cloud utilizes an allow list approach combined with identity. The use of an identity reduces policy enforcement from N2 rules to N rules. As applications scale up or scale down, other workloads do not need policy updates.

Purpose-Built Segmentation for Multi-Cloud and Hybrid Environments

East-west traffic segmentation between workloads in heterogeneous environments: traversing multiple IP domains is no longer an issue since IP reachability no longer assumes application access.

End-to-End Visibility into Application Dependencies

Getting visibility for your applications across any cloud now becomes possible since you have a common workload identifier that is abstracted from infrastructure.

Dynamic cloud native applications now have a trusted identity that is used for security policy enforcement

paloalto NETWORKS

3000 Tannery Way
Santa Clara, CA 95054

Main: +1.408.753.4000
Sales: +1.866.320.4788
Support: +1.866.898.9087

www.paloaltonetworks.com