Secure the Enterprise (Strata)

The networking infrastructure of an enterprise can be extraordinarily complex. The Security Operating Platform secures enterprise networks' perimeters, data centers, and branches with a fully integrated and automated platform that simplifies security. Simplifying your security posture allows you to reduce operational costs and the supporting infrastructure while increasing your ability to prevent threats to your organization and quickly adjust to your dynamic environment. The key Security Operating Platform elements for securing the enterprise are:

- **Next-generation firewall.** The foundation of the Security Operating Platform available in physical, virtual, and cloud-delivered deployment options to provide consistent protection wherever your data and apps reside
- **Subscription services.** Add-on enhanced threat services and next-generation firewall capabilities, including DNS Security, URL Filtering, Threat Prevention, and WildFire malware prevention
- **Panorama.** Provides centralized network security management, simplifying administration while delivering comprehensive controls and deep visibility into network-wide traffic and security threats
- **Okyo Garde**. Transforms the employee home network into a trusted enterprise edge by leveraging Prisma Access to bring Secure Access Service Edge (SASE) to employee home networks with unified corporate security policy management to enable secure work-from-home (WFH) models

Next-generation firewall

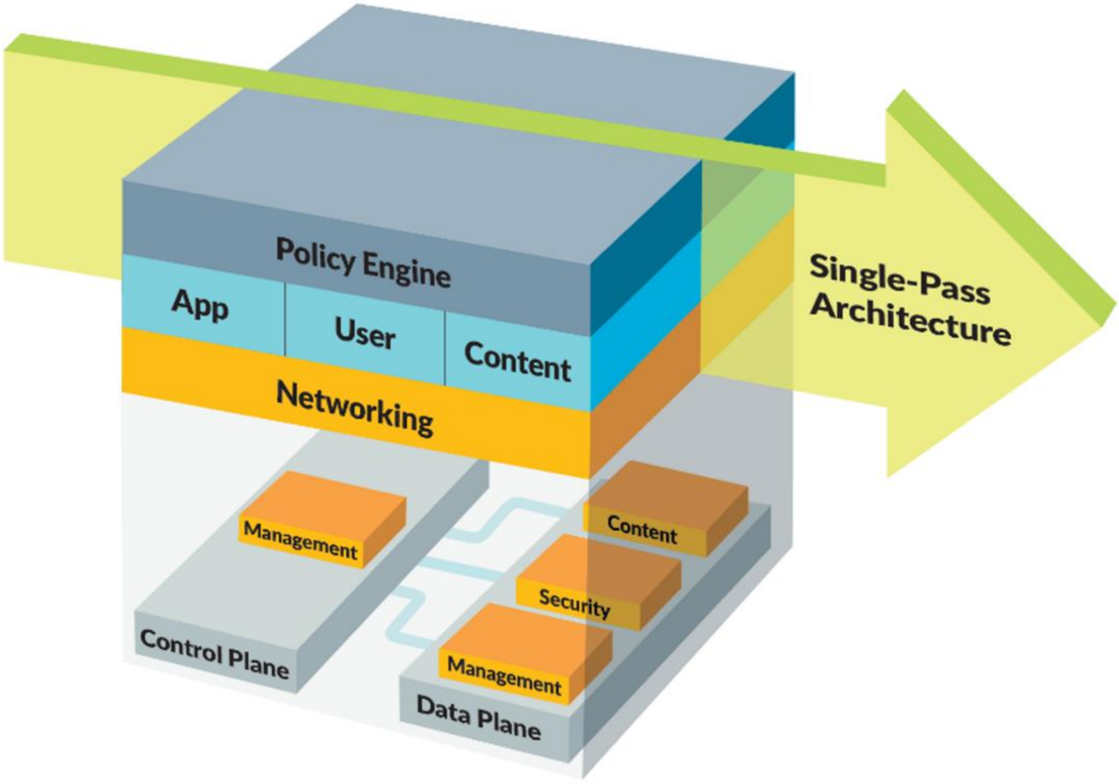
Fundamental shifts in application usage, user behavior, and complex network infrastructure have created a threat landscape that exposes weaknesses in traditional port-based network firewalls. End users want access to an ever-increasing number of applications, operating across a wide range of device types, often with little regard for the business or security risks. Meanwhile, data center expansion, network segmentation, virtualization, and mobility initiatives are forcing organizations to rethink how to enable access to applications and data, while protecting their networks from a new, more sophisticated class of advanced threats that evade traditional security mechanisms.

Palo Alto Networks next-generation firewalls are the core of the Security Operating Platform. The next-generation firewall inspects all traffic – including applications, threats, and content – and associates it with the user, regardless of location or device type. The application, content, and user become integral components of the enterprise security policy.

Palo Alto Networks next-generation firewalls are built on a single-pass architecture (see Figure 2-4), which is a unique integration of software and hardware that simplifies management, streamlines processing, and maximizes performance. The single-pass architecture integrates multiple threat prevention disciplines (IPS, anti-malware, URL filtering, etc.) into a single stream-based engine with a uniform signature format. This architecture allows traffic to be fully analyzed in a single pass without the performance degradation seen in multifunction gateways. The software is tied directly to a parallel processing hardware platform that uses function-specific processors for threat prevention, to maximize throughput and minimize latency.

Figure 2-4

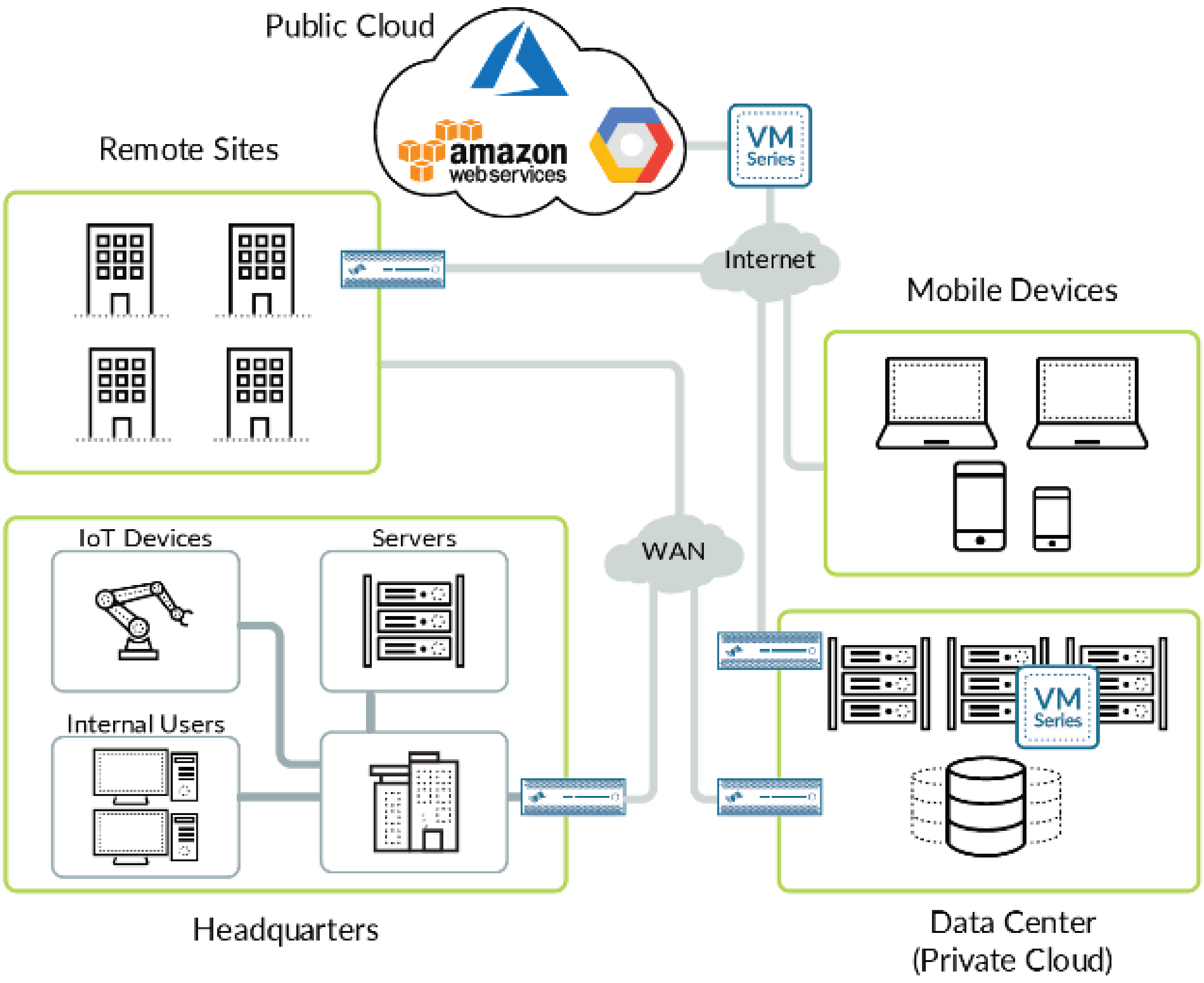
Palo Alto Networks next-generation firewalls use a single-pass architecture.



The use of one common engine means that two key benefits are realized. First, unlike file proxies that need to download the entire file before they can scan the traffic, a stream-based engine scans traffic in real time, only reassembling packets as needed and only in very small amounts. Second, unlike with traditional approaches, all traffic can be scanned with a single engine, instead of multiple scanning engines.

Organizations deploy next-generation firewalls at the network perimeter and inside the network at logical trust boundaries. All traffic crossing the next-generation firewall undergoes a full-stack, single-pass inspection, providing the complete context of the application, associated content, and user identity. With this level of context, you can align security with your key business initiatives (see Figure 2-5).

Figure 2-5 *Next-generation firewall locations in the enterprise network*



The next-generation firewall functions as a segmentation gateway in a Zero Trust architecture. By creating a micro-perimeter, the next-generation firewall ensures that only known, allowed traffic or legitimate applications have access the protect surface.

Next-generation firewalls include several key capabilities that enable complete visibility of the application traffic flows, associated content, and user identity and protect them from known, unknown, and advanced persistent threats. The essential functional capabilities in an effective next-generation firewall include:

- **Application identification.** Accurately identify applications regardless of port, protocol, evasive techniques, or encryption. Provide visibility of applications and granular policy-based control over applications, including individual application functions.
- **User identification.** Accurately identify users and subsequently use identity information as an attribute for policy control.
- **Content identification.** Content identification controls traffic based on complete analysis of all allowed traffic, using multiple threat prevention and data loss prevention techniques in a single-pass architecture that fully integrates all security functions.

Implementing Zero Trust with next-generation firewalls

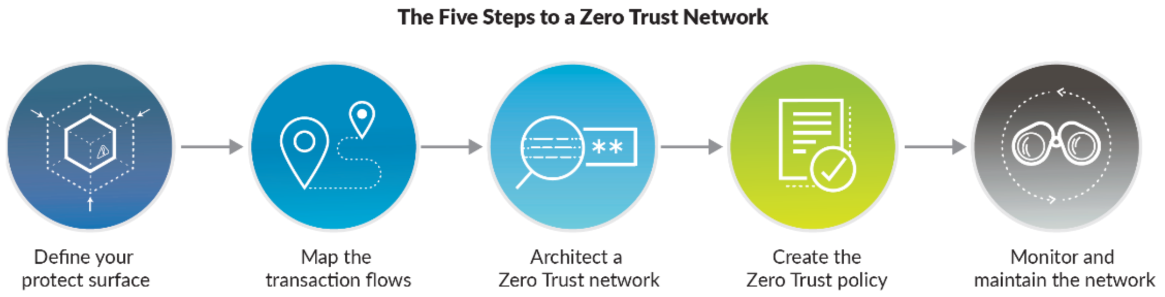
Companies are often reluctant to begin the Zero Trust journey because they believe it is difficult, costly, and disruptive. Twentieth-century design paradigms can create problems when designing a twenty-first-century Zero Trust network. However, building Zero Trust networks is actually much simpler than building legacy twentieth-century hierarchical networks. Because most of us learned to design networks from the outside in, based on classifying users as “trusted” and “untrusted” – an approach that has since proven unsecure – we struggle to adapt our design thinking to the Zero Trust methodology.

It’s not necessary to rip and replace your existing network to deploy a Zero Trust network. Zero Trust augments your existing network, with each Zero Trust network designed for a specific protect surface. The Zero Trust network is interconnected with your existing network to take advantage of the technology you already have. Then, over time, you iteratively move your additional datasets, applications, assets, or services from your legacy network to your Zero Trust network. This phased approach helps make deploying Zero Trust networks manageable, cost-effective, and non-disruptive.

The following five-step methodology describes a Zero Trust deployment with next-generation firewalls and other tightly integrated Palo Alto Networks security solutions (see Figure 2-19).

Figure 2-19

The Palo Alto Networks Zero Trust methodology


The Five Steps to a Zero Trust Network
Define your protect surface
Map the transaction flows
Architect a Zero Trust network
Create the Zero Trust policy
Monitor and maintain the network

Step 1: Define your protect surface

When defining the protect surface, you need to consider all critical data, application, assets, or services (DAAS). Your protect surface could include:

- **Data.** Payment card information (PCI), protected health information (PHI), personally identifiable information (PII), and intellectual property
- **Applications.** Off-the-shelf or custom software
- **Assets.** Supervisory control and data acquisition (SCADA) controls, point-of-sale terminals, medical equipment, manufacturing assets, and IoT devices
- **Services.** Domain Name System (DNS), Dynamic Host Configuration Protocol (DHCP), and Active Directory

Palo Alto Networks next-generation firewalls, in physical or virtualized form, provide comprehensive Layer 7 visibility to help you determine your data, applications, assets, and service profile. Palo Alto Networks also has extensive partnerships with leading third-party companies to help with additional data and asset discovery. Cortex XDR detection and response utilizes network, cloud, and endpoints as sensors, feeding data into Cortex Data Lake to provide visibility into the activity of users, devices, applications, and services for greater insight into the individual protect surfaces across your enterprise environment.

Step 2: Map the transaction flows

To properly design a network, it's critical to understand how systems should work. The way traffic moves across the network (specific to the data in the protect surface) determines how it should be protected. This understanding comes from scanning and mapping the transaction flows inside your network in order to determine how various data, application, asset, and service components interact with other resources on your network.

It's common to approximate flows by documenting what you know about how specific resources interact. Even without a complete picture, this information still provides valuable data so that you don't arbitrarily implement controls with zero insight.

Zero Trust is a flow-based architecture. After you understand how your systems are designed to work, the flow maps will tell you where you need to insert controls.

Remember that Zero Trust is an iterative process. Start with what you know. As you move through the steps in this methodology, you'll gather more information that will enable more granularity in your design. You shouldn't delay your Zero Trust initiative just because you don't have perfect information.

Palo Alto Networks next-generation firewalls deliver deep, application-layer visibility with granular insight into traffic flows. Policy Optimizer (discussed in Section 2.6.1.2) gives deep visibility into applications to help you prioritize rule migration, identify rules that allow unused or overprovisioned applications, and analyze rule usage characteristics.

Additionally, Cortex Data Lake collects telemetry from the network via next-generation firewall appliances, the cloud via VM-Series virtualized next-generation firewalls (discussed in Section 2.6.1.6), and endpoints via Cortex XDR. With this data centralized, Cortex XDR taps into Cortex Data Lake to validate established interaction and provide details around that interaction to help refine the use of communication and understanding of the flow.

Step 3: Architect a Zero Trust network

Traditionally, the first step of any network design is to architect it. Individuals get "reference architectures" for the network and must work to make them usable for their business. In the Zero Trust journey, architecting the network is the third step. Further, Zero Trust networks are bespoke, not some universal design. After the protect surface is defined and the flows mapped, the Zero Trust architecture will become apparent.

The architectural elements begin with deploying a next-generation firewall as a segmentation gateway to enforce granular Layer 7 access as a micro-perimeter around the protect surface. With this architecture, each packet that accesses a resource inside the protect surface will pass through a next-generation firewall so that Layer 7 policy can be enforced, simultaneously controlling and inspecting access. There is a significant misunderstanding that Zero Trust is only about access control: Least-privileged access control is only one facet of Zero Trust. Another facet is the inspection and logging of every single packet, all the way through Layer 7, to determine if packets are clean. This determination is made by inspecting all network traffic for malicious content with multiple integrated security services, including IPS, sandboxing, DNS security, URL filtering, and data loss prevention (DLP) capabilities.

Palo Alto Networks next-generation firewalls take advantage of App-ID, User-ID, and Content-ID to define authoritative Layer 7 policy controls and prevent compromise of protect surfaces. Because these segmentation gateways are offered in both physical and virtual form factors, this architectural model can work everywhere you may have a protect surface, whether in on-premises or off-premises physical data centers, or in private, public, or hybrid cloud environments.

Endpoint security, such as Cortex XDR (discussed in Section 4.4.1), can prevent compromise of the protect surface by known and unknown threats, whether from malware, fileless attacks, or exploits. Secure access offerings, such as Prisma Access (discussed in Section 3.5.2), extend the policy of each micro-perimeter down to the endpoints attempting to access protect surface resources.

The Security Operating Platform delivers telemetry from all core Palo Alto Networks technologies to Cortex Data Lake (discussed in Section 4.4.3), enabling machine learning based policy optimization and automation via Cortex XDR for improvement in later stages of your deployment.

The architecture would still be incomplete without important third-party offerings. Palo Alto Networks integrates with multiple multi-factor authentication (MFA) providers to add fidelity to User-ID. To round out and simplify Zero Trust architectures, a powerful API provides deep integrations with more than 250 third-party partners, including anti-spam/anti-phishing technologies, DLP systems, software-defined wide-area networks (SD-WAN), and wireless offerings.

Step 4: Create the Zero Trust policy

After you've architected your Zero Trust network, you need to create the supporting Zero Trust policies, following the Kipling Method, to answer the who, what, when, where, why, and how of your network and policies. For one resource to talk to another, a specific rule must whitelist that traffic. The Kipling Method of creating policy enables Layer 7 policy for granular enforcement so that only known allowed traffic or legitimate application communication is allowed in your network. This process significantly reduces the attack surface while also reducing the number of port-based firewall rules enforced by traditional network firewalls. With the Kipling Method, you can easily write policies by answering:

- **Who** should be accessing a resource? This defines the "asserted identity."
- **What** application is the asserted identity of the packet using to access a resource inside the protect surface?
- **When** is the asserted identity trying to access the resource?
- **Where** is the packet destination? A packet's destination is often automatically pulled from other systems that manage assets in an environment, such as from a load-balanced server via a virtual IP address.

- **Why** is this packet trying to access this resource within the protect surface? This question relates to data classification, where metadata automatically ingested from data classification tools helps make your policy more granular.

- **How** is the asserted identity of a packet accessing the protect surface via a specific application?

To simplify the process, you should create policies primarily on your segmentation gateways' centralized management tool. Panorama provides this functionality, and Panorama is where the Kipling Method is applied.

Palo Alto Networks next-generation firewall technology and unique features enable you to write policies that are easy to understand and maintain while providing maximum security transparency to your end users. User-ID helps define the who, App-ID helps define the what, and Content-ID helps define the how, all of which is enforced throughout your deployment, including by the WildFire malware prevention service, as well as by the Threat Prevention, URL Filtering, and DNS Security services. PAN-OS delivers enhanced policy creation capability, notably through Policy Optimizer, which continuously helps you understand how to increase the fidelity of your Zero Trust policy. Additionally, you can create policies for Prisma SaaS based on how SaaS applications are accessed.

Step 5: Monitor and maintain the network

The last step in this iterative process is to monitor and maintain your network, which means continuously looking at all internal and external logs through Layer 7 and focusing on the operational aspects of Zero Trust. Inspecting and logging all traffic on your network is a pivotal facet of Zero Trust.

It's important to send the system as much telemetry as possible about your environment. This data will give you new insights into how to improve your Zero Trust network over time. The more your network is attacked, the stronger it will become, with greater insight into making policies more secure. Additional data provides insight into the protect surface – such as what you should include in it and the interdependencies of data within it – that can inform architectural tweaks to further enhance your security.

All telemetry generated by Palo Alto Networks endpoint, network, and cloud security technologies is sent to Cortex Data Lake, where the data is stitched together to enable machine learning based policy optimization and analytics.

Next-generation firewall and VM-Series data is consolidated into a singular view under Panorama, which raises an alert when a malicious or suspicious occurrence should be investigated.

AutoFocus contextual threat intelligence service enables this investigation with a combination of machine intelligence from WildFire and human intelligence provided by the Palo Alto Networks Unit 42 threat research team, resulting in policy improvement and a more refined protect surface. The MineMeld engine within AutoFocus can aggregate, enforce, and share threat intelligence from third-party sources, providing further context for improved Zero Trust policy. MineMeld can seamlessly integrate with your next-generation firewall inside or outside your Palo Alto Networks deployment.

Prisma Cloud provides public cloud security and compliance monitoring, scanning all audit and flow logs across multicloud environments for root user and overly permissive administrator activities. Prisma Cloud builds a deep contextual understanding of your cloud environment, allowing detection of user anomalies – based on activity and location – that could signal compromised credentials, brute-force attacks, and other suspicious activities. Prisma Cloud also correlates threat intelligence data to provide visibility into suspicious IP addresses and host vulnerabilities across your resources, which can quickly be isolated to avoid additional exposure. This data provides insight that enables you to fine-tune Zero Trust privileges.

Cortex XDR takes advantage of Cortex Data Lake to create profiles of users and devices, acting as a baseline of normal use. This baseline allows the behavioral analytics engine to detect threats based on anomalies targeting your protect surface. In evaluating current or additional protect surface policies, Cortex XDR allows you to search the telemetry within Cortex Data Lake for communication and interactions between entities. You can also analyze the telemetry to prove the condition or get valuable insight into how your policy should be modified. In rare instances, the search can identify an unknown threat vector not factored into the protect surface. Cortex XDR will then facilitate a deep investigation of the newfound threat so that you can uncover what occurred and react accordingly.

Next-generation firewall deployment options

The Palo Alto Networks family of next-generation firewalls includes physical appliances, virtualized firewalls, and 5G-ready firewalls.

Physical appliances (PA-Series)

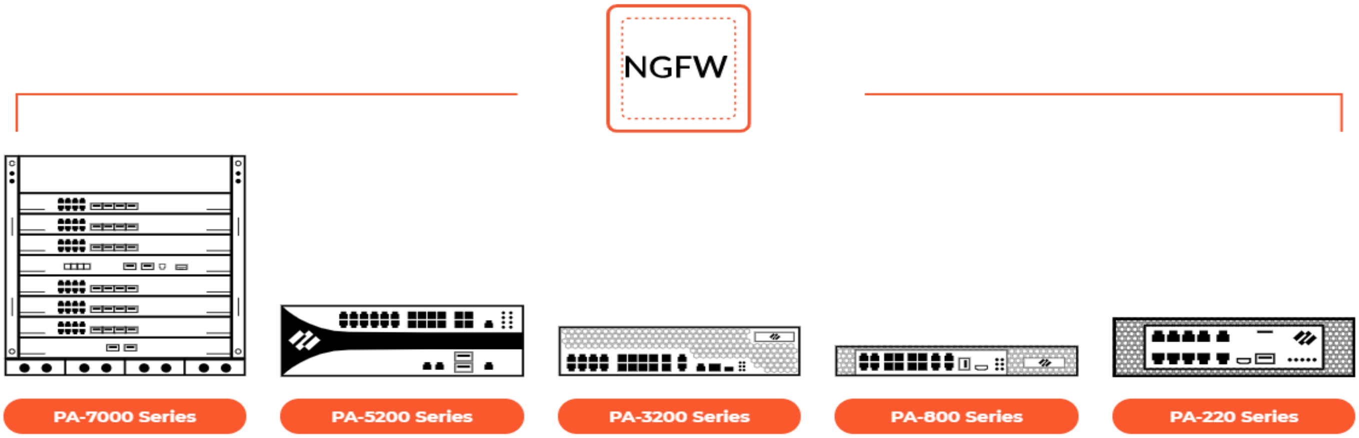
The full range of Palo Alto Networks physical next-generation firewalls is easy to deploy into your organization's network. They are purposefully designed for simplicity, automation, and integration. PA-Series firewalls support a variety of data center and remote branch deployment use cases. Available PA Series firewalls include (see Figure 2-19):

- **PA-7000 Series.** The PA-7000 Series next-generation firewalls enable enterprise-scale organizations and service providers to deploy security in high-performance environments, such as large data centers and high-bandwidth network perimeters. Designed to handle growing throughput needs for application-, user-, and device-generated data, these systems offer amazing performance, prevention capabilities to stop the most advanced cyberattacks, and high-throughput decryption to stop threats hiding under the veil of encryption. Built to maximize security-processing resource utilization and automatically scale as new computing power becomes available, the PA-7000 Series offers simplicity defined by a single-system approach to management and licensing.

- **PA-5200 Series.** The PA-5200 Series next-generation firewalls – comprising the PA-5280, PA-5260, PA-5250, and PA-5220 firewalls – are ideal for high-speed data center, internet gateway, and service provider deployments. The PA-5200 Series delivers up to 64Gbps of throughput, using dedicated processing and memory, for the key functional areas of networking, security, threat prevention, and management.

- **PA-3200 Series.** The PA-3200 Series next-generation firewalls – comprising the PA-3260, PA-3250, and PA-3220 – are targeted at high-speed internet gateway deployments. PA-3200 Series appliances secure all traffic (including encrypted traffic), using dedicated processing and memory for networking, security, threat prevention, and management.
- **PA-800 Series.** The PA-800 Series next-generation firewalls – comprising the PA-850 and PA-820 firewalls – are designed to provide secure connectivity for organizations' branch offices as well as midsize businesses.
- **PA-220.** The PA-220 firewall brings next-generation firewall capabilities to distributed enterprise branch offices, retail locations, and midsize businesses in a small form factor.
- **PA-220R.** The PA-220R firewall is a ruggedized next-generation firewall that secures industrial and defense networks in a range of harsh environments, such as utility substations, power plants, manufacturing plants, oil and gas facilities, building management systems, and healthcare networks.

Figure 2-19 *Strata Next-Generation Firewalls*



Virtualized firewalls (VM-Series)

VM-Series virtual firewalls provide all the capabilities of Palo Alto Networks next-generation physical hardware firewalls (PA-Series) in a virtual machine form factor. VM-Series form factors support a variety of deployment use cases, including:

- **Micro-segmentation.** VM-Series virtual firewalls reduce your environment's attack surface by enabling granular segmentation and micro-segmentation. Threat prevention capabilities ensure that when threats do enter the environment, they are quickly identified and stopped before they can exfiltrate data, deliver malware or ransomware payloads, or cause other damage.

- **Multicloud and hybrid cloud.** VM-Series virtual firewalls eliminate the need for multiple security tool sets by providing comprehensive visibility and control across multicloud and hybrid cloud environments – including Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, and Oracle Cloud – and just as effortlessly in software-defined networks and virtualized environments, all managed from a single console.

- **DevOps and CI/CD pipelines.** VM-Series virtual firewalls provide on-demand, elastic scalability to ensure security when and where you need it most. With automated network security, security provisioning can now be integrated directly into DevOps workflows and CI/CD pipelines without slowing the pace of business.

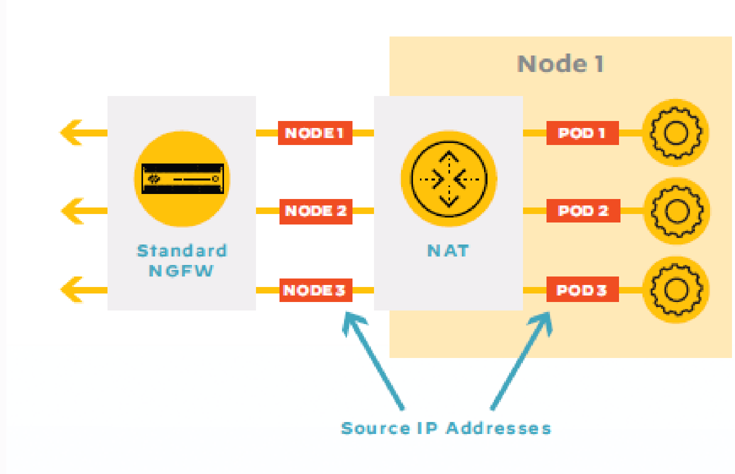
Cloud-native firewalls (CN-Series)

Standard next-generation firewalls play an indispensable role in securing on-premises deployments — few data centers can do without them. However, cloud-native environments pose unique challenges that next-generation firewalls were not designed to handle, especially when it comes to looking inside a Kubernetes environment.

In Kubernetes, pods (collections of containers) run on nodes, either physical or virtual machines. Developers rarely have to deal with nodes explicitly, but nodes impact how firewalls operate. Next-generation firewalls cannot determine which pod is the source of outbound traffic because all source IP addresses are translated to the node IP address. To a traditional firewall, all outbound traffic from the node looks the same (see Figure 2-21).

Figure 2-21

Due to the use of network address translation (NAT) in Kubernetes, all outbound traffic carries the node source IP address.



While Kubernetes creates challenges for traditional security tools, it also presents opportunities to enhance security by taking advantage of native constructs—most notably, namespaces.

Kubernetes namespaces help to simplify cluster management by making it easier to apply certain policies to some parts of the cluster without affecting others. However, they are also a valuable security tool. Security teams use namespaces to isolate workloads, which reduces the risk of attacks spreading within a cluster, and establish resource quotas to mitigate the damage that can be caused by a successful cluster breach.[1]

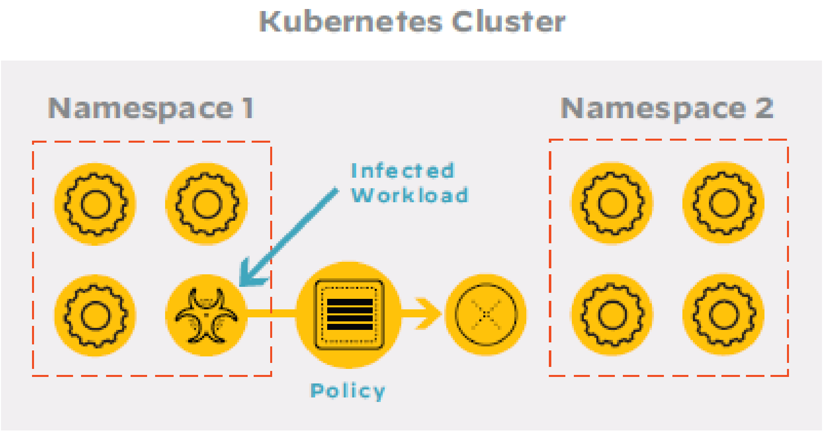
A secure cloud-native architecture requires the ability to secure traffic that crosses namespace boundaries or travels outbound to legacy workloads such as bare metal servers. However, doing so requires knowing the internal state of objects such as namespaces, pods, and containers. Because

[1] "Kubernetes Security Best Practices," Twistlock, June 6, 2019.

that information is not available outside the environment, the only effective solution is to take the security solution inside the Kubernetes walls (see Figure 2-22).

Figure 2-22

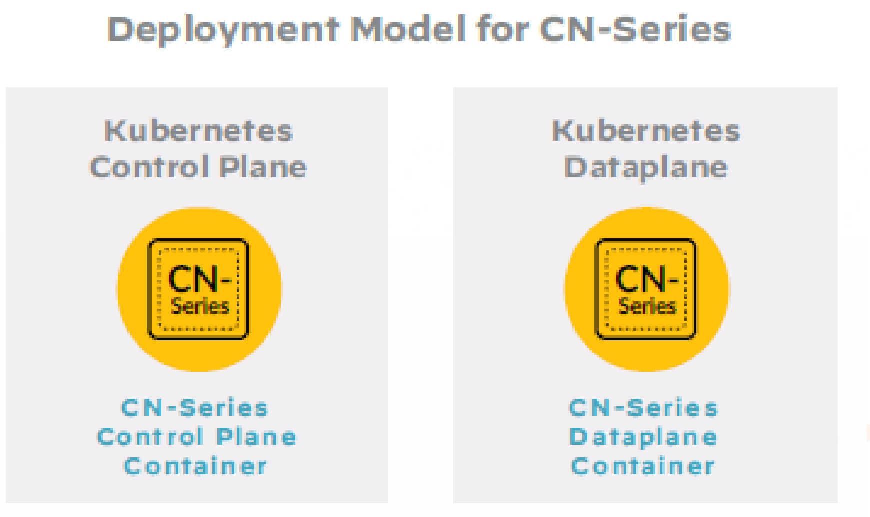
Security policies based on namespaces prevent spread of exploits within a physical cluster.



Palo Alto Networks CN-Series next-generation firewalls deploy as two sets of pods: one for the management plane (CN-MGMT), and another for the firewall dataplane (CN-NGFW), as shown in Figure 2-23. The management pod always runs as a Kubernetes service. The dataplane pods can be deployed in two modes: distributed or clustered. In distributed mode, the firewall dataplane runs as a daemon set on each node. Administrators can deploy next-generation firewalls on all cluster nodes with a single command, placing security controls as close to the workloads as possible. In clustered deployment mode, the firewall dataplane runs as a Kubernetes service in a dedicated security node. When deployed in clustered mode, CN-Series next-generation firewalls take advantage of the native auto scaling capabilities of Kubernetes to ensure security in even the most dynamic Kubernetes environments. Clustered deployments are best suited for large Kubernetes environments where a distributed deployment would be resource-intensive and cost-prohibitive.

Figure 2-23

The CN-Series deploys natively as control and dataplane pods within the Kubernetes environment.


Deployment Model for CN-Series
Kubernetes Control Plane
CN-Series
CN-Series Control Plane Container
Kubernetes Dataplane
CN-Series
CN-Series Dataplane Container

Native integration with Kubernetes enables CN-Series next-generation firewalls to leverage contextual information about the containers in the environment in the formulation of security policies. For instance, container namespaces can be used to define a traffic source in a firewall policy.

5G-ready firewalls (K2-Series)

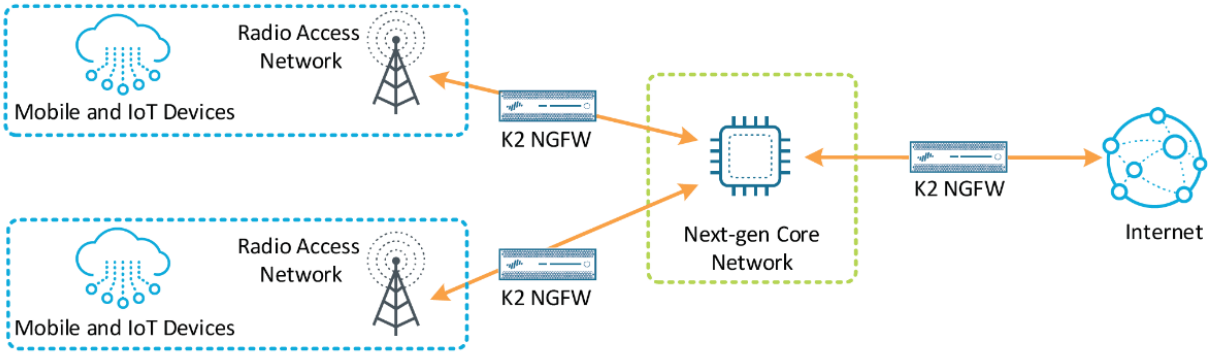
5G creates disruptive business opportunities for mobile network operators because it can move beyond delivering connectivity and use security as a business enabler and competitive advantage. The evolution to 5G opens the door to exciting new services, but it also increases the number of potential intrusion points, amplifying the security impact. To tap into the 5G business opportunities with minimal risk of being exploited by bad actors, you need complete visibility and automated security across all network locations.

Palo Alto Networks has developed, as part of the next-generation firewall platform, a 5G-ready platform, called the K2-Series, to prevent successful cyberattacks from targeting mobile network services. The K2-Series firewalls are designed to handle growing throughput needs due to the increase of application-, user-, and device-generated data. The K2-Series offers amazing performance and threat prevention capabilities to stop advanced cyberattacks and secure mobile network infrastructure, subscribers, and services.

You can deploy K2-Series firewalls on all 5G network interfaces in order to achieve scalable, complete protection with consistent management and full application visibility (see Figure 2-20). The fundamental shift in 5G network architectures further intensifies the impact on the security landscape, with growth in the number of intrusion points, including attacks inside mobile tunnels and threats within apps traversing cellular traffic. Mobile operators need consistent security enforcement across all network locations and all signaling traffic. This larger attack surface increases the need for application-aware Layer 7 security to detect known and unknown threats.

Figure 2-20

Securing 4G and 5G New Radio (NR) networks



K2-Series offers two modes: secure mode and express mode. Secure mode comes with all of the next-generation firewall features enabled, including threat prevention with the following enabled: App-ID, IPS, antivirus, antispyware, advanced malware analysis, and logging. Express mode is optimized for the highest throughput configuration; it is upgradable to secure mode.

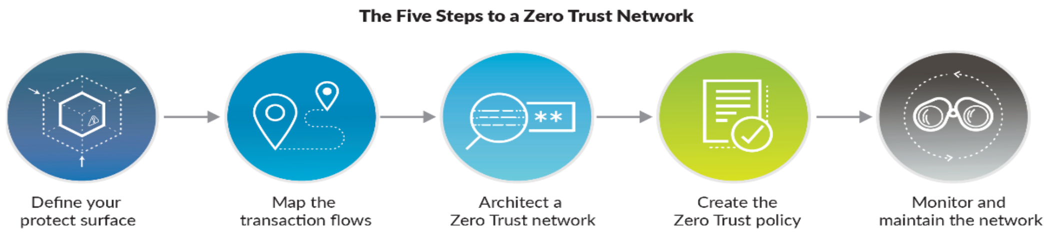
Implementing Zero Trust with next-generation firewalls

Companies are often reluctant to begin the Zero Trust journey because they believe it is difficult, costly, and disruptive. Twentieth-century design paradigms can create problems when designing a twenty-first-century Zero Trust network. However, building Zero Trust networks is actually much simpler than building legacy twentieth-century hierarchical networks. Because most of us learned to design networks from the outside in, based on classifying users as “trusted” and “untrusted” – an approach that has since proven unsecure – we struggle to adapt our design thinking to the Zero Trust methodology.

It’s not necessary to rip and replace your existing network to deploy a Zero Trust network. Zero Trust augments your existing network, with each Zero Trust network designed for a specific protect surface. The Zero Trust network is interconnected with your existing network to take advantage of the technology you already have. Then, over time, you iteratively move your additional datasets, applications, assets, or services from your legacy network to your Zero Trust network. This phased approach helps make deploying Zero Trust networks manageable, cost-effective, and non-disruptive.

The following five-step methodology describes a Zero Trust deployment with next-generation firewalls and other tightly integrated Palo Alto Networks security solutions (see Figure 2-21).

Figure 2-21 *The Palo Alto Networks Zero Trust methodology*


The Five Steps to a Zero Trust Network
Define your protect surface
Map the transaction flows
Architect a Zero Trust network
Create the Zero Trust policy
Monitor and maintain the network

Step 1: Define your protect surface

When defining the protect surface, you need to consider all critical data, application, assets, or services (DAAS). Your protect surface could include:

- **Data.** Payment card information (PCI), protected health information (PHI), personally identifiable information (PII), and intellectual property
- **Applications.** Off-the-shelf or custom software
- **Assets.** Supervisory control and data acquisition (SCADA) controls, point-of-sale terminals, medical equipment, manufacturing assets, and IoT devices
- **Services.** Domain Name System (DNS), Dynamic Host Configuration Protocol (DHCP), and Active Directory

Palo Alto Networks next-generation firewalls, in physical or virtualized form, provide comprehensive Layer 7 visibility to help you determine your data, applications, assets, and service profile. Palo Alto Networks also has extensive partnerships with leading third-party companies to help with additional data and asset discovery. Cortex XDR detection and response utilizes network, cloud, and endpoints as sensors, feeding data into Cortex Data Lake to provide visibility into the activity of users, devices, applications, and services for greater insight into the individual protect surfaces across your enterprise environment.

Step 2: Map the transaction flows

To properly design a network, it's critical to understand how systems should work. The way traffic moves across the network (specific to the data in the protect surface) determines how it should be protected. This understanding comes from scanning and mapping the transaction flows inside your network in order to determine how various data, application, asset, and service components interact with other resources on your network.

It's common to approximate flows by documenting what you know about how specific resources interact. Even without a complete picture, this information still provides valuable data so that you don't arbitrarily implement controls with zero insight.

Zero Trust is a flow-based architecture. After you understand how your systems are designed to work, the flow maps will tell you where you need to insert controls.

Remember that Zero Trust is an iterative process. Start with what you know. As you move through the steps in this methodology, you'll gather more information that will enable more granularity in your design. You shouldn't delay your Zero Trust initiative just because you don't have perfect information.

Palo Alto Networks next-generation firewalls deliver deep, application-layer visibility with granular insight into traffic flows. Policy Optimizer gives deep visibility into applications to help you prioritize rule migration, identify rules that allow unused or overprovisioned applications, and analyze rule usage characteristics.

Additionally, Cortex Data Lake collects telemetry from the network via next-generation firewall appliances, the cloud via VM-Series virtualized next-generation firewalls, and endpoints via Cortex XDR. With this data centralized, Cortex XDR taps into Cortex Data Lake to validate established interaction and provide details around that interaction to help refine the use of communication and understanding of the flow.

Step 3: Architect a Zero Trust network

Traditionally, the first step of any network design is to architect it. Individuals get “reference architectures” for the network and must work to make them usable for their business. In the Zero Trust journey, architecting the network is the third step. Further, Zero Trust networks are bespoke, not some universal design. After the protect surface is defined and the flows mapped, the Zero Trust architecture will become apparent.

The architectural elements begin with deploying a next-generation firewall as a segmentation gateway to enforce granular Layer 7 access as a micro-perimeter around the protect surface. With this architecture, each packet that accesses a resource inside the protect surface will pass through a next-generation firewall so that Layer 7 policy can be enforced, simultaneously controlling and inspecting access. There is a significant misunderstanding that Zero Trust is only about access control: Least-privileged access control is only one facet of Zero Trust. Another facet is the inspection and logging of every single packet, all the way through Layer 7, to determine if packets are clean. This determination is made by inspecting all network traffic for malicious content with multiple integrated security services, including IPS, sandboxing, DNS security, URL filtering, and data loss prevention (DLP) capabilities.

Palo Alto Networks next-generation firewalls take advantage of App-ID, User-ID, and Content-ID to define authoritative Layer 7 policy controls and prevent compromise of protect surfaces. Because these segmentation gateways are offered in both physical and virtual form factors, this architectural model can work everywhere you may have a protect surface, whether in on-premises or off-premises physical data centers, or in private, public, or hybrid cloud environments.

Endpoint security, such as Cortex XDR, can prevent compromise of the protect surface by known and unknown threats, whether from malware, file-less attacks, or exploits. Secure access offerings, such as Prisma Access, extend the policy of each micro-perimeter down to the endpoints attempting to access protect surface resources.

The Security Operating Platform delivers telemetry from all core Palo Alto Networks technologies to Cortex Data Lake, enabling machine learning based policy optimization and automation via Cortex XDR for improvement in later stages of your deployment.

The architecture would still be incomplete without important third-party offerings. Palo Alto Networks integrates with multiple multi-factor authentication (MFA) providers to add fidelity to User-ID. To round out and simplify Zero Trust architectures, a powerful API provides deep integrations with more than 250 third-party partners, including anti-spam/anti-phishing technologies, DLP systems, software-defined wide-area networks (SD-WAN), and wireless offerings.

Step 4: Create the Zero Trust policy

After you've architected your Zero Trust network, you need to create the supporting Zero Trust policies, following the Kipling Method, to answer the who, what, when, where, why, and how of your network and policies. For one resource to talk to another, a specific rule must whitelist that traffic. The Kipling Method of creating policy enables Layer 7 policy for granular enforcement so that only known allowed traffic or legitimate application communication is allowed in your network. This process significantly reduces the attack surface while also reducing the number of port-based firewall rules enforced by traditional network firewalls. With the Kipling Method, you can easily write policies by answering:

- **Who** should be accessing a resource? This defines the "asserted identity."
- **What** application is the asserted identity of the packet using to access a resource inside the protect surface?
- **When** is the asserted identity trying to access the resource?
- **Where** is the packet destination? A packet's destination is often automatically pulled from other systems that manage assets in an environment, such as from a load-balanced server via a virtual IP address.
- **Why** is this packet trying to access this resource within the protect surface? This question relates to data classification, where metadata automatically ingested from data classification tools helps make your policy more granular.
- **How** is the asserted identity of a packet accessing the protect surface via a specific application?

To simplify the process, you should create policies primarily on your segmentation gateways' centralized management tool. Panorama provides this functionality, and Panorama is where the Kipling Method is applied.

Palo Alto Networks next-generation firewall technology and unique features enable you to write policies that are easy to understand and maintain while providing maximum security transparent to your end users. User-ID helps define the who, App-ID helps define the what, and Content-ID helps define the how, all of which is enforced throughout your deployment, including by the WildFire malware prevention service, as well as by the Threat Prevention, URL Filtering, and DNS Security services. PAN-OS delivers enhanced policy creation capability, notably through Policy Optimizer, which continuously helps you understand how to increase the fidelity of your Zero Trust policy. Additionally, you can create policies for Prisma SaaS based on how SaaS applications are accessed.

Step 5: Monitor and maintain the network

The last step in this iterative process is to monitor and maintain your network, which means continuously looking at all internal and external logs through Layer 7 and focusing on the operational aspects of Zero Trust. Inspecting and logging all traffic on your network is a pivotal facet of Zero Trust.

It’s important to send the system as much telemetry as possible about your environment. This data will give you new insights into how to improve your Zero Trust network over time. The more your network is attacked, the stronger it will become, with greater insight into making policies more secure. Additional data provides insight into the protect surface – such as what you should include in it and the interdependencies of data within it – that can inform architectural tweaks to further enhance your security.

All telemetry generated by Palo Alto Networks endpoint, network, and cloud security technologies is sent to Cortex Data Lake, where the data is stitched together to enable machine learning based policy optimization and analytics.

Next-generation firewall and VM-Series data is consolidated into a singular view under Panorama, which raises an alert when a malicious or suspicious occurrence should be investigated.

AutoFocus contextual threat intelligence service, enables this investigation with a combination of machine intelligence from WildFire and human intelligence provided by the Palo Alto Networks Unit 42 threat research team, resulting in policy improvement and a more refined protect surface. The MineMeld engine within AutoFocus can aggregate, enforce, and share threat intelligence from third-party sources, providing further context for improved Zero Trust policy. MineMeld can seamlessly integrate with your next-generation firewall inside or outside your Palo Alto Networks deployment.

Prisma Cloud provides public cloud security and compliance monitoring, scanning all audit and flow logs across multicloud environments for root user and overly permissive administrator activities. Prisma Cloud builds deep contextual understanding of your cloud environment, allowing detection of user anomalies – based on activity and location – that could signal compromised credentials, brute-force attacks, and other suspicious activities. Prisma Cloud also correlates threat intelligence data to provide visibility into suspicious IP addresses and host vulnerabilities across your resources, which can quickly be isolated to avoid additional exposure. This data provides insight that enables you to fine-tune Zero Trust privileges.

Cortex XDR takes advantage of Cortex Data Lake to create profiles of users and devices, acting as a baseline of normal use. This baseline allows the behavioral analytics engine to detect threats based on anomalies targeting your protect surface. In evaluating current or additional protect surface policies, Cortex XDR allows you to search the telemetry within Cortex Data Lake for communication and interactions between entities. You can also analyze the telemetry to prove the condition or get valuable insight into how your policy should be modified. In rare instances, the search can identify an unknown threat vector not factored into the protect surface. Cortex XDR will then facilitate a deep investigation of the newfound threat so that you can uncover what occurred and react accordingly.