paloalto NETWORKS

Zero Trust Enterprise

REFERENCE ARCHITECTURE GUIDE

MARCH 2022

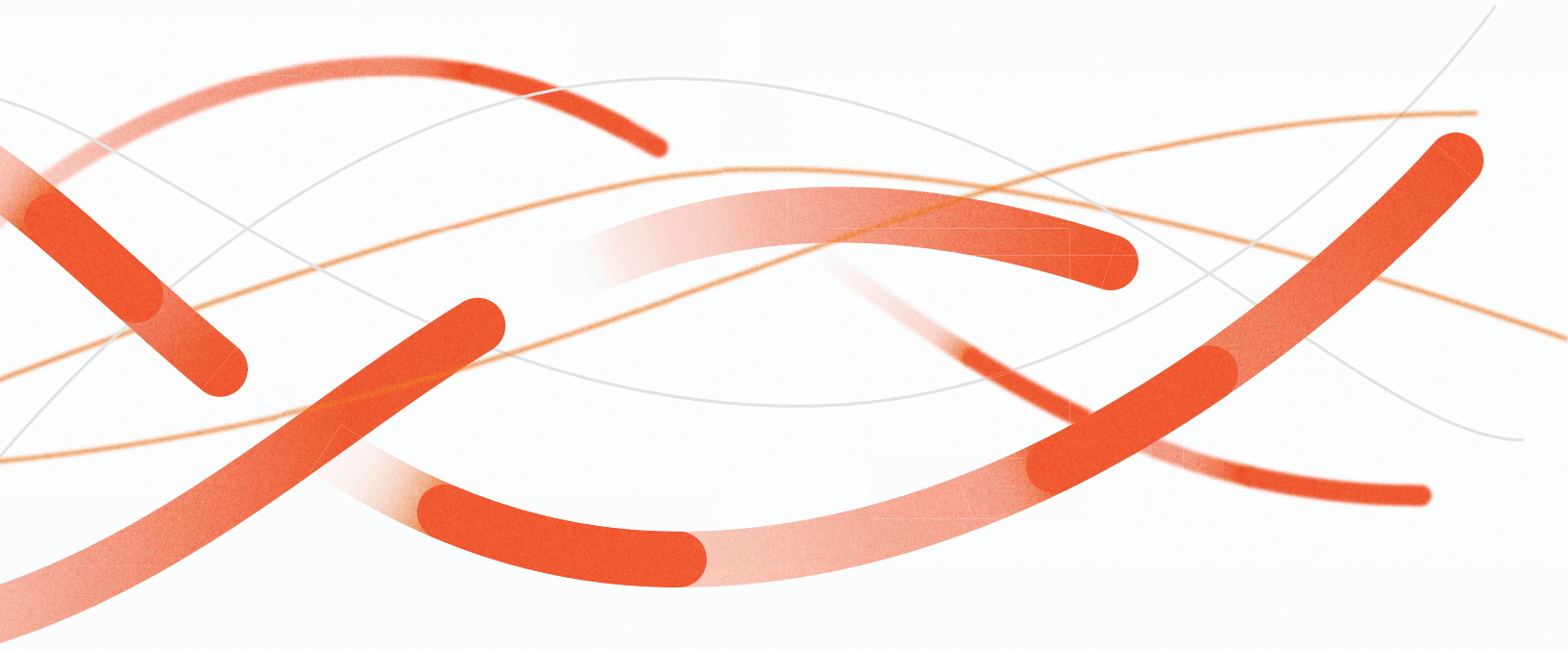


Table of Contents

Preface

GUIDE TYPES

Overview guides provide high-level introductions to technologies or concepts.

Reference architecture guides provide an architectural overview for using Palo Alto Networks® technologies to provide visibility, control, and protection to applications built in a specific environment. These guides are required reading prior to using their companion deployment guides.

Deployment guides provide decision criteria for deployment scenarios, as well as procedures for combining Palo Alto Networks technologies with third-party technologies in an integrated design.

DOCUMENT CONVENTIONS

Notes provide additional information.

Cautions warn about possible data loss, hardware damage, or compromise of security.

Blue text indicates a configuration variable for which you need to substitute the correct value for your environment.

In the **IP** box, enter `10.5.0.4/24`, and then click **OK**.

Bold text denotes:

- Command-line commands.

  ```
  # show device-group branch-offices
  ```

- User-interface elements.

  In the **Interface Type** list, choose **Layer 3**.

- Navigational paths.

  Navigate to **Network** > **Virtual Routers**.

- A value to be entered.

  Enter the password **admin**.

Italic text denotes the introduction of important terminology.

> An *external dynamic list* is a file hosted on an external web server so that the firewall can import objects.

Highlighted text denotes emphasis.

> Total valid entries: 755

ABOUT PROCEDURES

These guides sometimes describe other companies' products. Although steps and screen-shots were up-to-date at the time of publication, those companies might have since changed their user interface, processes, or requirements.

GETTING THE LATEST VERSION OF GUIDES

We continually update reference architecture and deployment guides. You can access the latest version of this and all guides at this location:

https://www.paloaltonetworks.com/referencearchitectures

WHAT'S NEW IN THIS RELEASE

Palo Alto Networks made the following changes since the last version of this guide:

- This is a new guide.

Purpose of This Guide

Zero Trust is a cybersecurity strategy that secures an organization by eliminating implicit trust and continuously validating every stage of a digital interaction. This guide describes the Palo Alto Networks Zero Trust Enterprise approach to securing users, applications and infrastructure. Use this guide as a roadmap for architectural discussions between Palo Alto Networks and your organization.

AUDIENCE

This guide is for technical readers, including system architects and design engineers, who want to implement Zero Trust principles and optimize security practices. It assumes the reader is familiar with the basic concepts of applications, networking, virtualization, and security and has a basic understanding of network, data center, and public-cloud architectures.

RELATED DOCUMENTATION

The following documents support this guide:

- **Network Security: Overview**—Describes how organizations can prevent cyberthreats by using the Palo Alto Networks next-generation firewall and Prisma™ Access to protect and secure their network.
- **Public Cloud Security: Overview**—Describes the key challenges in approaching public-cloud security and securing cloud-native applications. Details how organizations can leverage Palo Alto Networks platforms to discover resources, detect risks, mitigate network threats, highlight suspicious behaviors, prevent malware and data leakage, and identify host vulnerabilities.
- **Securing Containerized Applications in Kubernetes: Reference Architecture Guide**—Presents a detailed discussion of the design considerations and deployment options for using Prisma Cloud and the CN-Series next-generation firewall to secure applications in Kubernetes.
- **SASE for Users: Reference Architecture Guide**—Describes how to use Prisma Access to secure mobile users as they access applications hosted in the internet or on-premises, regardless of from where they connect.
- **SASE for Branch: Reference Architecture Guide**—Describes how to use Prisma Access and Prisma SD-WAN to bring visibility, control, and protection to branch office Internet traffic.
- **Securing IoT Environments**—Provides architectural guidance for the Palo Alto Networks internet of things (IoT) security solution, which allows discovery, identity, and inventory of your organization's deployed IoT devices. You can then use the solution to assess risks and enforce policies that mitigate them.

Introduction

Stories of security breaches that expose sensitive data are in the news every week. These events can result in significant personal impact on those who have information exposed, as well as a loss of trust and financial penalties for the compromised companies. New industry standards and government regulations are developing at a rapid pace, forcing organizations to constantly evaluate their security posture and increase the overall level of security.

Digital transformation is also driving changes that require a different approach to security. To accommodate the hybrid workforce, enterprises are transforming their cybersecurity infrastructure, migrating applications to cloud, while also looking to automate security operations. To access applications from anywhere, users require fast and convenient access to products and services from a wide variety of devices. This guide describes how you can implement Zero Trust principles and optimize your security practices.

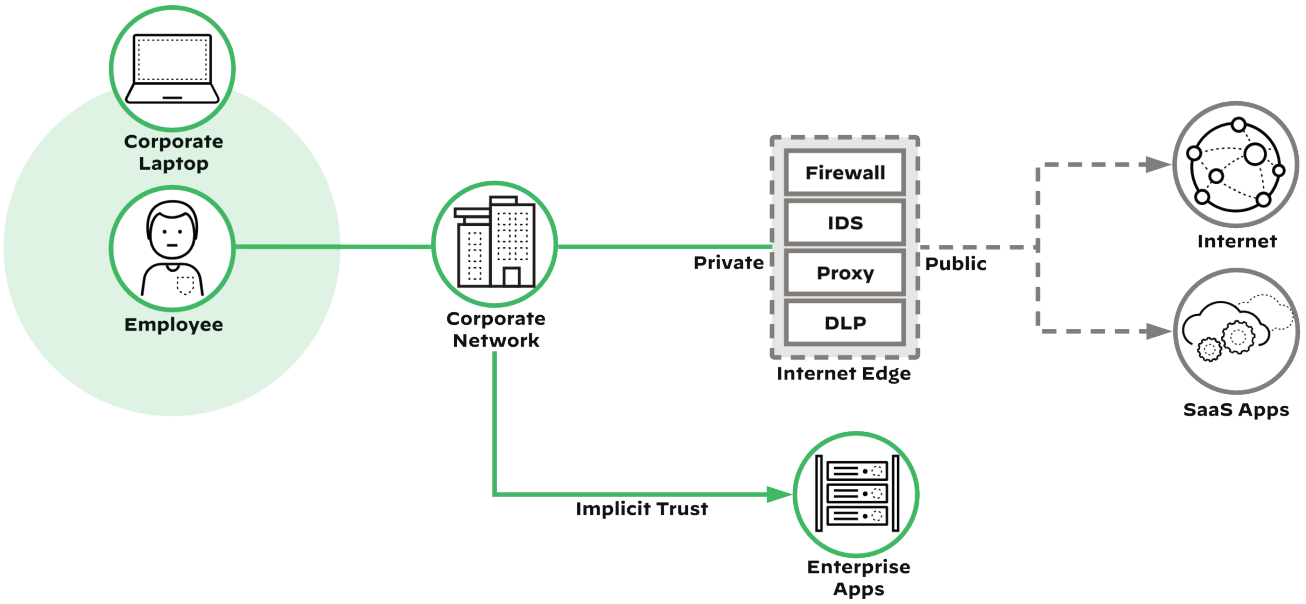
THE IMPLICIT-TRUST PROBLEM

Implicit trust is a term used to describe the elimination of security controls in specific contexts, the most common being user location. For example, you might allow a user located inside the office full access to internal applications with only a single verification of identity, but from a remote location, the same user would require additional security controls—like two-factor authentication, threat prevention, and data loss prevention (DLP)—to access the same applications. Used in traditional security models, implicit trust is a vulnerability as dangerous as any other.

Corporate Network Implicit Trust

Traditional perimeter-based security wrongly assumes that all users and devices inside the corporate network can be trusted and that a full security stack at the internet edge is sufficient for securing corporate data. In this approach, implicit trust is granted in the private zone of the perimeter firewall. Only transaction flows destined to the public zone for internet and SaaS applications are considered untrusted and inspected.

Figure 1 Traditional corporate-perimeter security



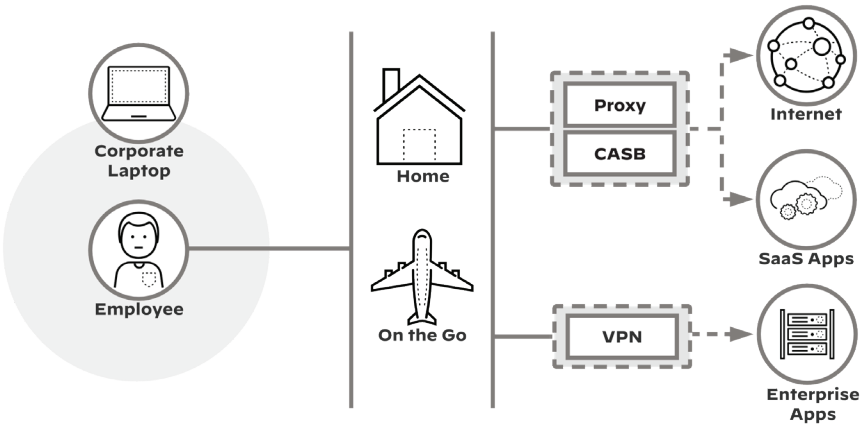
Traditional perimeter-based security is no longer adequate for protecting an organization's assets. Mobile devices moving on and off the corporate network, data and applications moving to the cloud, stealth malware, and attacks masquerading as legitimate applications or hiding in encrypted traffic have blurred the edges of the perimeter.

Attacks on sensitive data rarely use just a single exploit or compromised credential. Attackers use a composite of exploits, malware, compromised credentials, and other methods together to work their way from their beachhead in an organization to the target system. Often attackers use one method after another. Malware might supply a user's credentials, which in turn provide limited access to their organization's network. Once inside, the attacker moves around the private network and places other malware on privileged devices. The attacker eventually steals data, denies service, or encrypts the target system so that they can demand a ransom.

Remote Access Implicit Trust

Traditional virtual private networks (VPNs) for remote access also assume implicit trust, allowing user access to all corporate applications after the user authenticates into the corporate VPN. With implicit trust, you might apply one set of security controls when users access internet and SaaS applications via a remote access VPN, and you might apply a separate set of security controls when users access the same resources from the corporate network.

Figure 2 Traditional remote-access security



The Need for a Strategic Approach

Many traditional security policies are built around implicit trust, and these policies vary between locations, focusing on blocking what is considered a risk at each location. The use of disparate point solutions can result in threat and policy information that is siloed within the different enforcement points. Due to the manual correlation of the non-integrated solutions, coordinating a comprehensive security posture for protecting against breaches and data loss is slow and ineffective.

The biggest challenge for many organizations is defining a consistent security model that provides the required security controls holistically across the organization. Adopting a Zero Trust approach helps remedy the vulnerabilities associated with implicit trust in current security policies.

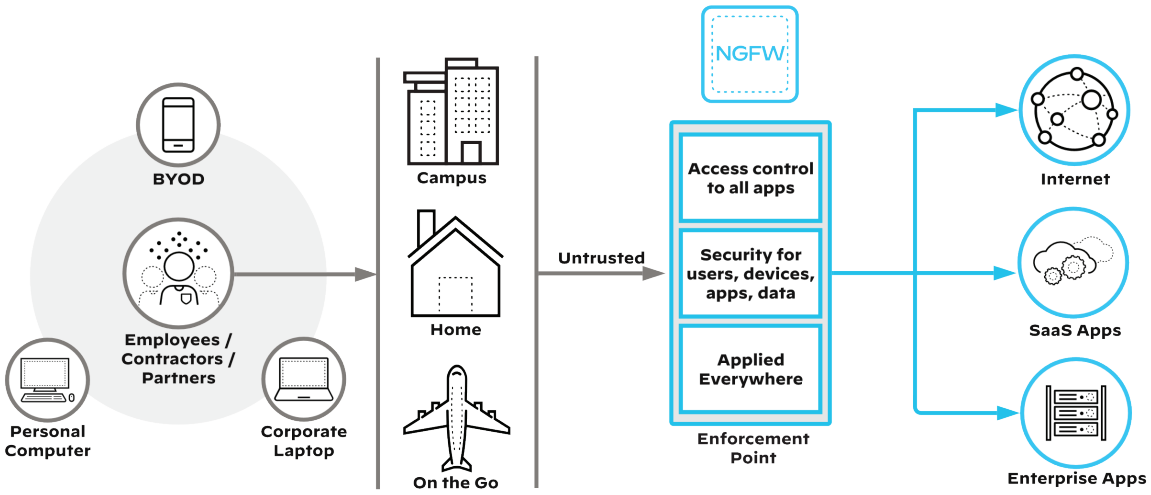
WHAT IS ZERO TRUST

The Zero Trust approach is based on the principle that no user, device, or transaction from inside or outside of the network can be trusted. The elimination of implicit trust promotes a consistent security policy regardless of the situation. The framework focuses on resource protection and the premise that trust is never granted implicitly but must be continually evaluated. In Zero Trust, authentication and authorization are critical, not just in the initial connection but at every stage of the digital interaction.

Traditional security models target the protection of the entire attack surface, which is difficult to identify and constantly evolving. In a Zero Trust framework, you define a *protect surface*, which is made up of the most critical and valuable data, assets, applications, and services (DAAS). Because it contains what is most critical to an organization's operations, the protect surface is orders of magnitude smaller than the *attack surface*, and it is always knowable.

In Zero Trust, only known, allowed traffic can access the protect surface. Users have access to the data and applications they need in order to perform their tasks but nothing more. This is known as *least-privileged access* and enforced using a segmentation gateway implemented with a next-generation firewall (NGFW).

Figure 3 Zero Trust approach



Note

You should consider Zero Trust a strategic initiative that you build over time. To become familiar with the process, tools and operations, start with a small, non-critical protect surface. After you become familiar with the process, you should prioritize mission-critical data and applications.

ZERO TRUST FRAMEWORKS

There are several standard frameworks that provide guidance on how to implement Zero Trust strategies. Example frameworks include **NIST 800-207**, **Google's BeyondCorp**, and **Microsoft's Zero Trust framework**. You can use the guidelines in these frameworks to evaluate your posture and formulate a strategy to secure your critical assets. These Zero Trust frameworks do not prescribe a specific product or technology, but helps you evaluate your specific protect surface which is key to identify the right security controls to put in place.

NIST 800-207 defines a framework in which you grant access to a resource through a policy decision point and corresponding policy enforcement point. This architecture framework defines the following basic tenets as targets for a Zero Trust deployment:

- All data sources and computing services are considered resources.
- All communication is secured regardless of network location.
- Access to individual enterprise resources is granted on a per-session basis.
- Dynamic policy determines access to resources.
- The enterprise monitors and measures the integrity and security posture of all owned and associated assets.
- All resource authentication and authorization are dynamic and strictly enforced before access is allowed.
- The enterprise collects as much information as possible about the current state of assets, network infrastructure, and communications and uses it to improve its security posture.

NIST 800-207 recommends that you also develop a Zero Trust network architecture with the following assumptions:

- The entire enterprise private network is not considered an implicit trust zone.
- The enterprise might not own or configure devices on the network might not be owned or configurable by the enterprise.
- No resource is inherently trusted.
- Not all enterprise resources are on enterprise-owned infrastructure.
- Remote enterprise subjects and assets cannot fully trust their local network connection.
- Assets and workflows moving between enterprise and non-enterprise infrastructure should have a consistent security policy and posture.

Google's BeyondCorp allows for single sign-on, access-control policies, access proxy, and user- and device-based authentication and authorization. The BeyondCorp principles are:

- Access to services must not be determined by the network from which you connect.
- Access to services is granted based on contextual factors from the user and their device.
- Access to services must be authenticated, authorized, and encrypted.

Microsoft's Zero Trust framework defines the following guiding principles for Zero Trust:

- **Verify explicitly**—Always authenticate and authorize based on all available data points, including user identity, location, device health, service or workload, data classification, and anomalies.
- **Use least-privileged access**—Limit user access with just-in-time and just-enough-access, risk-based adaptive policies, and data protection to help you secure both data and productivity.
- **Assume breach**—Minimize blast radius and segment access. Verify end-to-end encryption and use analytics to get visibility, drive threat detection, and improve defenses.

When comparing frameworks defined by standard bodies, analysts and security vendors, there is a common set of guidelines:

- Validate users and verify device integrity.
- Secure the access and enforce least-privileged user and device access to data and applications.
- Secure the transactions, prevent threats, and protect data.

Zero Trust Enterprise with the Palo Alto Networks Portfolio

The Palo Alto Networks Zero Trust Enterprise is a modern, strategic, platform-based approach to security. Zero Trust Enterprise is an end goal for security teams that want to implement zero trust principles, guide their security practices, and optimize their procurement across an entire enterprise.

PILLARS AND CAPABILITIES

At its core, Zero Trust Enterprise is about eliminating all trust and verifying all digital transactions. In Zero Trust, authentication and authorization are critical, not just in the initial connection but at every stage of the digital interaction.

The Zero Trust Enterprise approach is organized into the following three pillars and security capabilities:

- **Zero Trust for users**—Step one of any Zero Trust effort requires strong authentication for verifying user identity. Complementing strong authentication with the verification of user device integrity provides another layer of security by ensuring the user's device software has not been compromised. After the user identity and device security have been verified, least-privileged access policies ensure controlled network access to critical data and applications. Finally, to achieve Zero Trust, scanning of all transactions protects against malicious activity.
- **Zero Trust for applications**—To secure private data centers and cloud applications, organizations must remove all implicit trust and enforce cybersecurity checks across the entire application development lifecycle. To ensure least-privileged access to applications and infrastructure, the identity and entitlements granted to the developers, DevOps, and admins must be validated. Workloads accessing other workloads should mutually verify identity and apply least-privileged connectivity for the application. To prevent lateral movement of malware, you should enforce microsegmentation. You should continuously monitor workloads for misconfigurations, vulnerabilities, and indicators of compromise.
- **Zero Trust for infrastructure**—All critical infrastructure, IT/OT systems, points of sale, medical devices, supply chains, and more must be secured with a Zero Trust approach. The main difference between the Zero Trust approaches for users and infrastructure is the constraint that IoT devices are headless and most have limited authentication capabilities. To accurately verify the identity of the device in order to create least-privileged access policies, additional inspection is required.

The tools and techniques for enforcing Zero Trust Enterprise capabilities might vary for each protect surface. For example, although you can use an inline NGFW in order to secure all user transactions to an application in a private data center, for visibility into in-cloud transactions from the internet, a SaaS application might require an API-based approach.

The following table summarizes the security capabilities required for each of the three Zero Trust Enterprise pillars.

Table 1 Key Zero Trust Enterprise capabilities

	Identity validated	Device/workload	Access	Transaction
Zero Trust for users	Users, with strong authentication	Verifies user's device integrity	Enforces least-privileged user access to data and applications	Scans all content for malicious activity and data theft
Zero Trust for applications	Developers, DevOps, and admins, with strong authentication	Verifies workload integrity	Enforces least-privileged access for workloads accessing other workloads	Scans all content for malicious activity and data theft
Zero Trust for infrastructure	All users who have access to infrastructure	Identifies all devices, including IoT devices	Enforces least-privileged access segmentation for native and third-party infrastructure	Scans all content within the infrastructure for malicious activity and data theft

PLATFORMS FOR ZERO TRUST

Using the network security platform, Prisma Cloud, and Cortex™ XDR solutions from Palo Alto Networks, you can implement Zero Trust strategies. These solutions are more secure, more consistent, and provide a simpler architecture by using cloud-delivered services. For Enterprise Identity and Access Management (IAM), Palo Alto Networks Cloud Identity Engine (CIE) integrates with many of the IAM solutions in the market. CIE is a cloud-based identity synchronization service that you can use to consistently authenticate and authorize your users, regardless of location and where the user identity stores live.

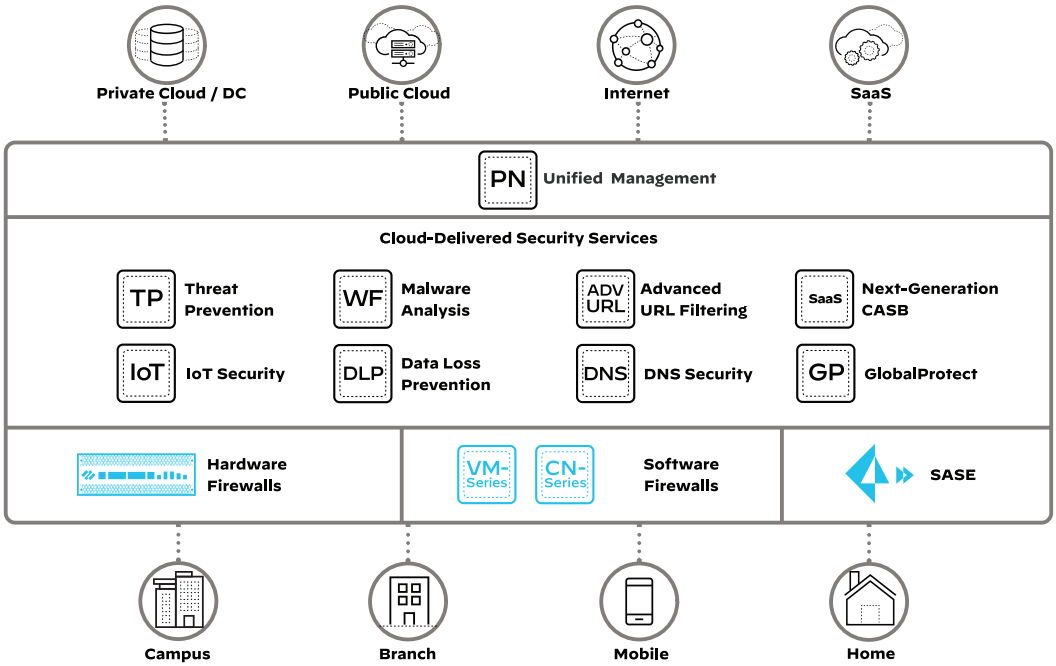
Table 2 Palo Alto Networks platforms for Zero Trust Enterprise

	Identity	Device/workload	Access	Transaction
Zero Trust for users	Enterprise IAM	Cortex XDR	Network security platform	
Zero Trust for applications	Enterprise IAM/ Prisma Cloud	Cortex XDR/ Prisma Cloud	Prisma Cloud & software NGFWs	
Zero Trust for infrastructure	Enterprise IAM	Network security platform		

Network Security Platform

The network security platform provides consistent protection and experience wherever users, devices and applications reside. The network security platform is offered as a cloud-delivered Secure Access Services Edge (SASE) solution or with on-premises hardware and software. The network security platform consists of NGFW, Prisma Access, and a set of cloud-delivered security services.

Figure 4 Network security platform



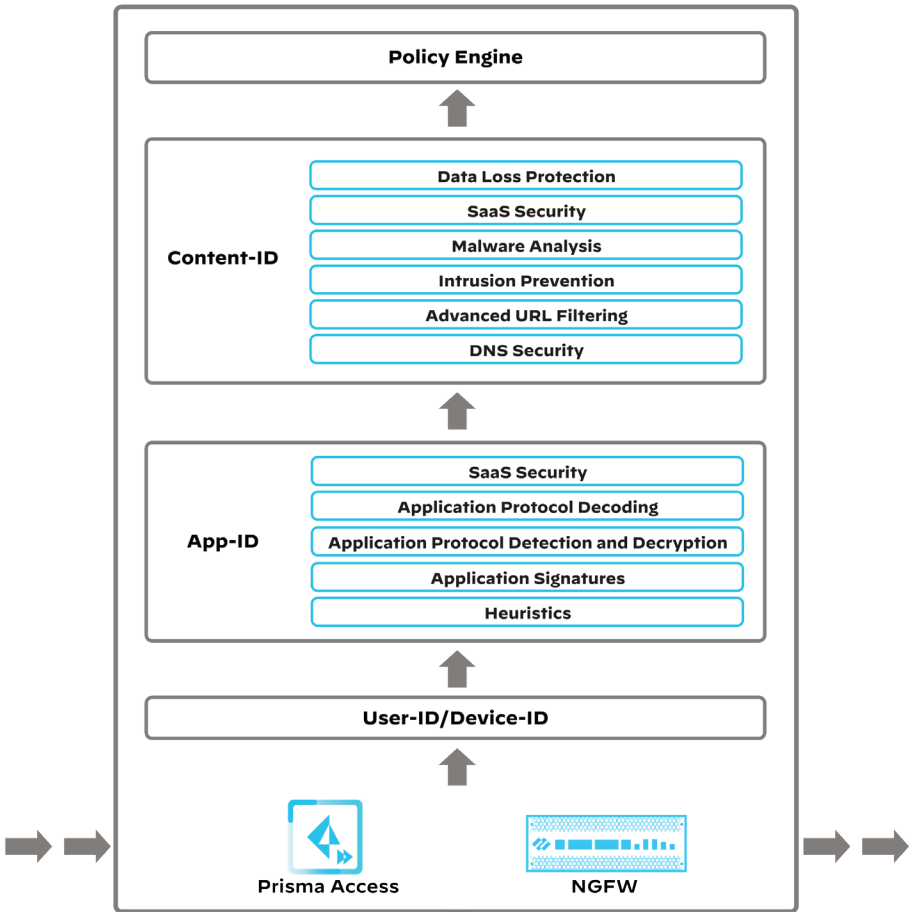
NGFW

Powered by the Palo Alto Networks operating system, PAN-OS®, the Palo Alto Networks NGFW gives you complete visibility, threat protection, and control of applications in use by all users, in all of your locations, all the time. The NGFWs have three flexible deployment options:

- **PA-Series**—The PA-Series are physical appliance NGFWs. All PA-Series NGFWs have feature parity across the range. You can deploy them at the internet perimeter, the data-center perimeter, colocation facilities, and other locations (within the campus and branch) in which high-speed and/or high-density Ethernet connectivity is required.
- **VM-Series**—The VM-Series are virtualized form-factor NGFWs. You can deploy them as virtual machines (VMs) in a wide range of compute platforms, such as public and private cloud. These virtual NGFWs help you secure cloud environments with the same levels of protection offered by the physical NGFWs.
- **CN-Series**—The CN-Series are containerized NGFWs. Securing containers presents new and unique challenges, due to their ephemeral behavior and how they share compute platforms. For advanced protection and compliance for your applications, the CN-Series NGFWs give you the ability to deploy Layer 7 network security and threat protection inside your Kubernetes clusters.

Each of the NGFWs performs multiple security functions with a single-pass architecture. This parallel processing system applies all elements of threat protection with a single packet scan, which increases the performance and flexibility of the NGFW. Eliminating multiple lookups benefits performance by reducing latency, and multiple processor components then work in parallel in order to provide individual security functions. This differs from legacy firewalls, which typically follow a sequence of separate functions in packet processing; as features get enabled, this deteriorates performance.

Figure 5 Key processes of single-pass architecture

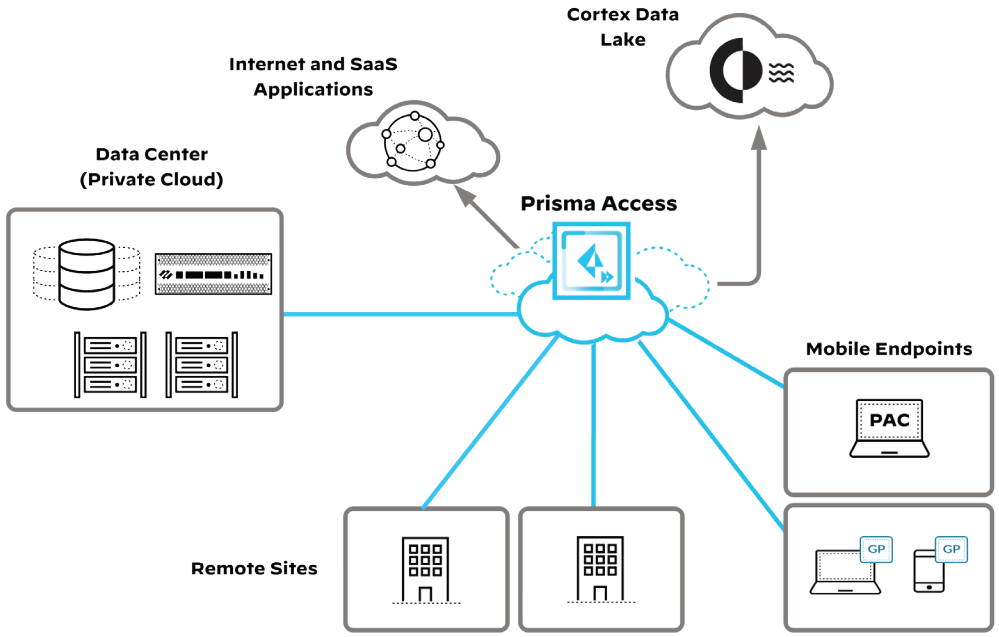


Prisma Access

Also powered by PAN-OS, Prisma Access is a cloud-hosted NGFW service that provides secure access to internet and business applications hosted in SaaS, as well as access to corporate HQ, data center, and public cloud. For secure access to SaaS and private applications, mobile users with managed or unmanaged devices can connect to Prisma Access. To protect user access to internet applications, remote sites can connect to Prisma Access with a VPN tunnel. The firewall-as-a-service capabilities of Prisma Access inspect all traffic, not just HTTP and HTTPS, in order to identify applications, threats, and content.

Prisma Access is a cloud service that allows you to avoid the challenges of sizing firewalls and compute resources and minimizes coverage gaps or inconsistencies associated with a distributed organization. As demand shifts and traffic patterns change, the elasticity of the cloud scales. The cloud service operationalizes security deployment to remote networks and mobile users by leveraging a cloud-based security infrastructure delivered by Palo Alto Networks.

Figure 6 Prisma Access



There are two options available as part of Prisma Access: Prisma Access for networks and Prisma Access for users. Prisma Access offers the industry's most comprehensive SASE solution, enabling your organization to connect and secure any user, device, or application.

Cloud-Delivered Security Services

Organizations often find themselves with many-point solutions, each one designed to address specific security threats. This approach usually results in many physical or virtual appliances, one for each solution, and also prevents them from fully realizing both the value and capabilities of these solutions. Palo Alto Networks offers multiple security services that are specifically designed to complement and enhance each other and ensure that you can confidently secure all traffic that traverses any networks or clouds. These security capabilities aid and support the Zero Trust Enterprise approach.

Seamlessly integrated with NGFWs and Prisma Access, cloud-delivered security subscriptions coordinate intelligence and provide protections across all attack vectors, providing best-in-class functionality while eliminating the coverage gaps disparate network security tools create. The NGFW/Prisma Access becomes a multi-function platform that can deliver multiple security services in a single unit. The security subscriptions include Enterprise DLP, next-generation Cloud Access Security Broker (CASB), Threat Prevention, Advanced URL Filtering, WildFire®, DNS Security, IoT Security, and GlobalProtect™.

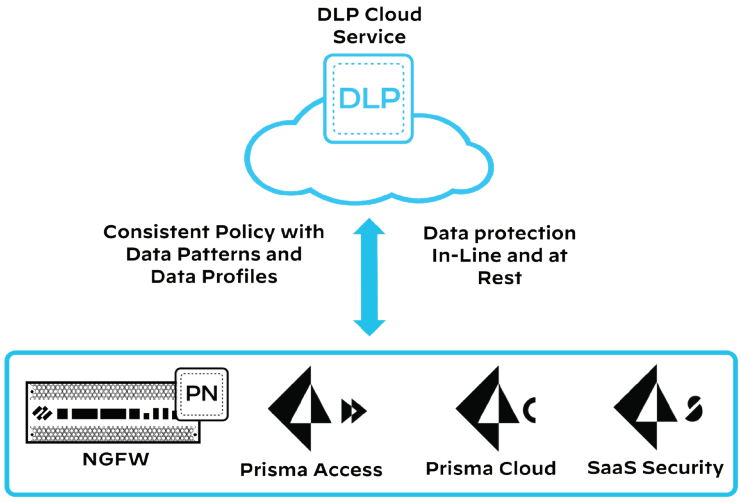
Enterprise Data Loss Prevention

The DLP cloud service is a subscription service delivered across multiple products, including next-generation CASB, Prisma Access, Prisma Cloud, and all of the NGFW's deployment options. Because many Palo Alto Networks products deliver the service, you do not need separate probes or physical servers deployed in a few selected parts of the network. Separate probes would limit your view and control of data leakage. The advantage is a single DLP engine across all products. You create the configuration once, and then you automatically synchronize it everywhere you use the DLP cloud service. The service provides advanced analytics, machine-learning and advanced remediation of workflows, providing operational simplicity. The service discovers, monitors, and protects your sensitive data, avoiding data loss and data theft.

The DLP cloud service provides detection and response through data policies. Detection rules find and classify sensitive information based on data patterns. Response rules are actions that are used to mitigate the risk of data loss, such as blocking an action for example. The service uses pre-defined patterns, as well as DLP profiles, to provide a much more granular data-matching option than just using search patterns. Today, over 380 patterns and seventeen data profiles are available, including profiles for GDPR, CCPA, PII, and many others. You also can create your own custom and file property data patterns. You create custom data patterns with regular expressions and keywords. File property data patterns are created to match a name-value pair, metadata, and the attributes of a file.

Because the DLP cloud service is embedded across multiple Palo Alto Networks control points, you can secure both data-at-rest and data-in-motion. As an organization, you receive consistent and comprehensive protection for all sensitive data, regardless of location. This reduces the time needed to implement the protection and makes it much easier to manage.

Figure 7 DLP cloud service



Next-Generation CASB

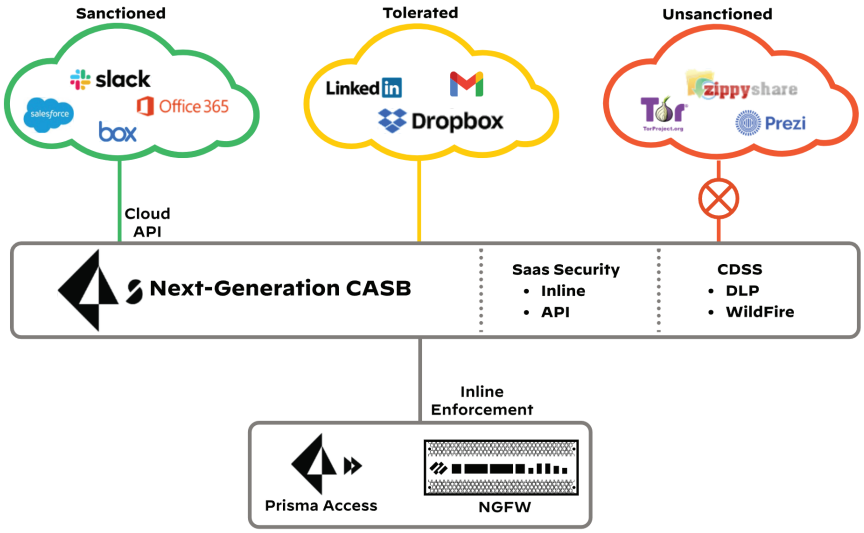
With the emergence of the hybrid workforce, organizations have rapidly increased their SaaS application adoption to maintain and accelerate employee productivity anywhere they work. As SaaS adoption continues to increase, so does more uploading, storing, and sharing of business information with these applications. Collaboration applications require a different security approach due to the increased use of shorter messages, group chats and screenshots. Organizations need an easier way to discover, control user access and protect SaaS applications from all threats, both sanctioned and unsanctioned applications.

Natively integrated with the Palo Alto Networks NGFW platform (cloud-based, virtual, and hardware form factors), next-generation CASB delivers granular visibility and control of SaaS applications, their use within your organization, and their risks. A full and complete view into shadow IT risks enables security teams to intelligently keep up with their growth and prevent unsanctioned apps from becoming another conduit of data loss. This solution integrates with other cloud-delivered security services, such as DLP for scanning files for sensitive information and WildFire for preventing known and unknown threats from spreading through sanction SaaS applications.

Two of the key components of the next-generation CASB platform are:

- **SaaS Security Inline**—SaaS Security Inline provides visibility and control of all SaaS application use from your corporate network and managed endpoints. It can tell you what SaaS applications are being accessed, who is accessing them, and the risks associated with those applications. Furthermore, it enables you to create and distribute granular policy in order to control or block access to those applications.
- **SaaS Security API**—SaaS Security API secures sanctioned SaaS applications. Without any configuration on endpoints, it provides complete visibility across all users, folders, and activity within a sanctioned SaaS application, and it enables detailed analysis and analytics of application use in order to prevent data risk and compliance violations. More importantly, SaaS Security API allows granular, context-aware policy control within these SaaS applications in order to drive enforcement and quarantine users and data as soon as a violation occurs.

Figure 8 Next-generation CASB



Threat Prevention

Organizations face a multitude of attacks from threat actors driven by various motives, including profit, ideology/hacktivism, or even organizational discontent. Today's attackers are well-funded and well-equipped. They use evasive tactics to gain footholds in target networks and launch advanced attacks at high volume. Their methods are highly targeted, leveraging sophisticated playbooks to breach an organization, move laterally, and extract valuable data, all while remaining invisible to traditional independent defenses.

The Threat Prevention service protects your network by providing multiple layers of prevention, confronting threats at each phase of an attack. The service provides comprehensive protection from all threats irrespective of port, protocol, and encryption. To block threats, the service inspects network traffic for vulnerability exploits, malware, spyware, command-and-control (C2) communication, and even unknown threats. To apply content scanning policy rules to security policies, the NGFWs and Prisma Access use security profiles. Default profiles are available, or you can create your own custom profiles. The following profiles are used for threat prevention:

- **Antivirus profiles**—Protect against downloads of common malware types, such as viruses, worms, trojans, and spyware. The profile scans for a wide variety of malware found in executables and other common file types.
- **Anti-spyware profiles**—Block spyware on compromised hosts reaching out to external C2 servers. You can enable DNS sinkholing within the anti-spyware profile in order to enable the NGFW to respond to DNS queries for known malicious domains and identify the infected hosts.
- **Vulnerability protection profiles**—Detect and block exploit attempts and evasive techniques, including port scans, buffer overflows, remote code-execution, and protocol fragmentation. They stop attempts based on threats that have patterns related to exploits' attacks on system vulnerabilities. For example, the profiles protect against buffer overflows and illegal code execution.

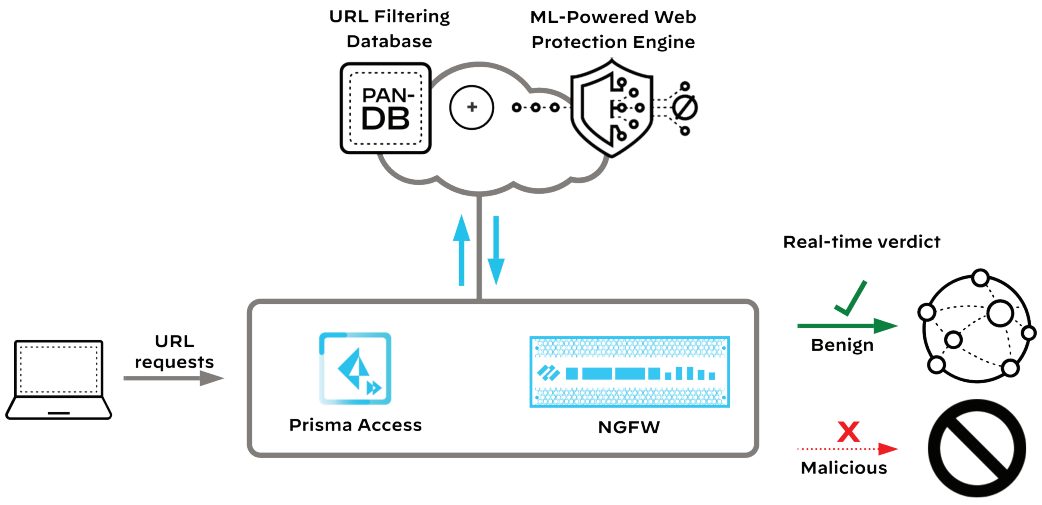
Advanced URL Filtering

A large majority of edge traffic is web traffic. This is due to the increase of business and personal applications moving to the internet. Traditional web-security controls rely strictly on URL databases that are populated using web-based crawlers that analyze and classify web pages' content. These techniques work well for acceptable-use policy enforcement but are insufficient for preventing targeted web attacks like phishing and data exfiltration. Evasion techniques used by modern adversaries include single-use links, very short-lived links, hiding malicious content behind CAPTCHA challenges, malicious URLs hidden in compromised legitimate websites, and more. This problem is made worse because these techniques are available in phishing-as-a-service and web attack kits, putting advanced evasion techniques into the hands of less-skilled attackers.

Advanced URL filtering detects and prevents new and unknown malicious web-based attacks in real-time, stopping initial infections and protecting users from phishing, malware, grayware, and C2 attacks. Detection runs inline, in real time web traffic, rather than relying on databases and after-the-fact crawling. This approach eliminates patient-zero attacks that can happen with URL databases that block only previously known malicious URLs.

As the user's web traffic ingresses, the NGFW/Prisma Access sends the request to PAN-DB cloud to identify and categorize the URL. If the URL is unknown, or a high-risk category, the URL is sent to the Web Protection Engine, where it is further analyzed using machine learning (ML) models. Based on the results of the analysis, the traffic is blocked if it is malicious. This real-time protection is delivered without affecting the user experience. Because this is a cloud-native service, there is no resource limitation; the service can expand and support additional ML models and data over time.

Figure 9 Advanced URL filtering



WildFire

Today's adversaries have easy access to cloud scale, legitimate infrastructure, and ML and can quickly distribute evasive malicious files to end users. To mitigate risks associated with evasive attacks, organizations turn to network sandboxing solutions for malware analysis. Traditional sandboxing approaches affect user productivity, because they are slow to deliver verdicts. They can protect against new threats only after the first victim has been compromised.

Palo Alto Networks offers a security service that provides cloud scale ML analytics capabilities to prevent known and unknown threats. WildFire is a threat intelligence cloud and virtual sandbox the NGFWs use in order to send unknown files and URLs for analysis. WildFire determines if the sample is benign, grayware, malware, or a phishing threat and generates the necessary verdicts and signatures.

When the NGFW receives a file, it checks whether WildFire has seen the file before and if there is a verdict. If WildFire has not seen the file, the NGFW sends the file to WildFire, which examines the file and provides a verdict immediately. WildFire provides content signatures for prevention and a single signature protects against the polymorphic variants of an individual malware.

To uncover the true nature of malicious files and URLs, WildFire identifies hundreds of potentially malicious behaviors, including:

- **Changes made to host**—WildFire monitors all processes for modifications to the host, including file and registry activity, code injection, memory heap spraying (exploits), mutexes, Windows service activity, the addition of auto-run programs, and other potentially suspicious activities.
- **Suspicious network traffic**—WildFire performs analysis of all network activity produced by the suspicious file, including creating backdoors, downloading of next-stage malware, visiting low-reputation domains, performing network reconnaissance, and more.
- **Anti-analysis detection**—WildFire monitors techniques used by advanced malware that is designed to avoid VM-based analysis, such as debugger detection, hypervisor detection, code injection into trusted processes, disabling of host-based security features, and more.

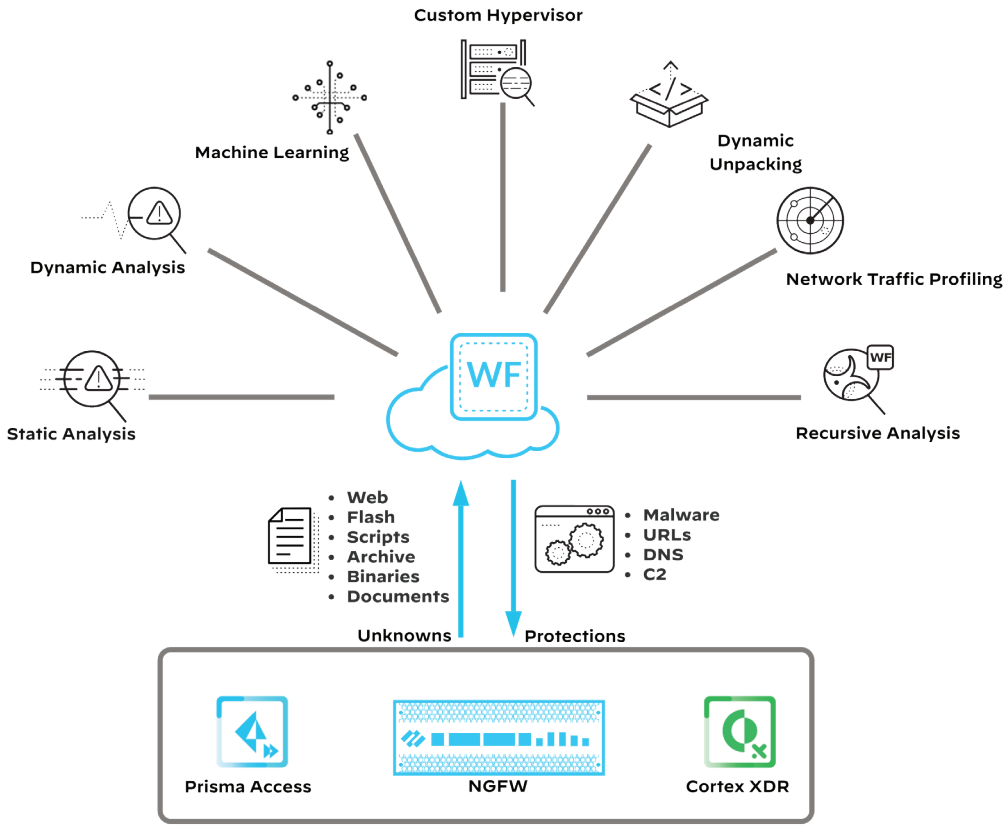
In addition to protecting from malicious and exploitive files and links, WildFire looks deeply into malicious outbound communication, disrupting C2 activity with anti-C2 signatures and DNS-based callback signatures. WildFire also feeds this information into URL filtering with PAN-DB, which automatically blocks newly discovered malicious URLs. This correlation of threat data and automated protections are key to identifying and blocking ongoing intrusion attempts and future attacks on your organization, without requiring policy updates and configuration commits. Unlike other solutions, WildFire can follow several stages of an attack, unifying analysis across both web and file vectors in order to prevent multi-stage threats.

To uncover and prevent new threats, WildFire provides multiple techniques, including:

- **Dynamic analysis**—Observes files as they execute in a purpose-built, evasion-resistant virtual environment, enabling detection of previously unknown malware by using hundreds of behavioral characteristics.
- **Machine learning**—Extracts thousands of unique features from each file, training a predictive ML model to identify new malware, which is not possible with static or dynamic analysis alone.
- **Static analysis**—Complements dynamic analysis with effective detection of malware, providing instant identification of malware variants. Static analysis further leverages dynamic unpacking in order to analyze threats attempting to evade detection through the use of packing tool sets.
- **Custom-built hypervisor**—Prevents evasion techniques with a robust, proprietary hypervisor that does not depend on open-source projects or proprietary software to which attackers have access.

Leveraging innovations in ML, artificial intelligence, and big-data analytics is the only way to stay ahead of a fast-moving adversary. However, all analytic solutions depend on massive amounts of data from many sources, to identify new threats and exploit techniques and to generate and share threat intelligence. This sharing model enables rapid response across the entire base of WildFire customers in order to prevent successful cyberattacks.

Figure 10 WildFire analysis



DNS Security

DNS is required for domain-name-to-IP-address translation when users are accessing resources such as websites, IP-based services, and applications. At the same time, DNS is a massive and frequently overlooked attack surface that adversaries use to their advantage. Malicious actors can compromise DNS in order to steal data or establish connections with C2 servers.

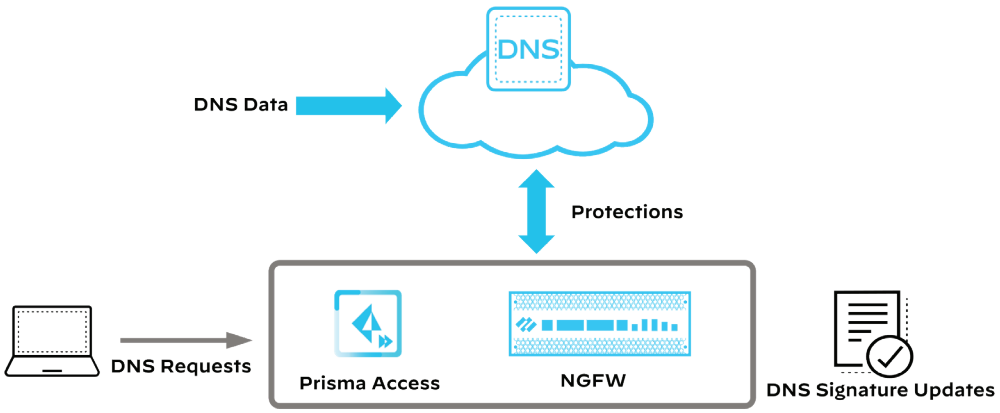
Adversaries use DNS tunneling in order to encode data of non-DNS based programs with DNS queries and responses, which often is used as a channel for slowly exfiltrating data and bypassing detection. DNS tunneling detection can detect a tunneling-based attack and block it with security polices, avoiding data theft.

The DNS Security service uses techniques to defend against domain-generation algorithms (DGAs). DGA domains are used for C2 communication and are always changing, and simple block-lists cannot keep up with the pace or scale. Because malicious domains are often autogenerated by machines, a DGA analysis can determine whether a domain was likely generated by a person or machine. By reverse-engineering and analyzing other frequently used techniques, the DNS Security service can identify and block previously unknown DGA-based threats in real time.

The DNS Security service allows the NGFW to sinkhole internal DNS requests, which allows it to generate a response to a DNS query for a known malicious domain/URL and causes the malicious domain name to resolve to a fake IP address that is given to the client. If the client attempts to access the fake IP address, a security rule blocks traffic to the IP address and logs the information, allowing further analysis of the end device.

Static lists and manual responses do not scale. The DNS Security service from Palo Alto Networks is a subscription-based service that is designed to protect and defend your network from advanced threats that are using DNS. The DNS Security service leverages ML and predictive analytics to provide real-time DNS request analysis. The analysis enables production and distribution of DNS signatures that are specifically designed to defend against malware that uses DNS for C2 and data exfiltration.

Figure 11 DNS Security service



IoT Security

Organizations are deploying large volumes of IoT devices for increasing productivity, helping with digital transformation, and providing operational efficiency. With the massive increase of connected devices comes a huge security risk, and securing IoT devices can be difficult.

IT departments do not always know what devices are deployed and what vulnerabilities these devices introduce. Because most IoT devices are connected to the internet, they present multiple security risks from unpatched and outdated software. The most frequent attacks are exploits (using long-known vulnerabilities) and password attacks (using default device passwords).

Palo Alto Networks offers an IoT Security solution, delivered as a cloud service, that takes a life cycle approach to securing your IoT environment. Your existing NGFWs perform discovery, visibility, and enforcement tasks, versus other solutions that require you to buy separate probes in the network. The lifecycle approach consists of the following steps:

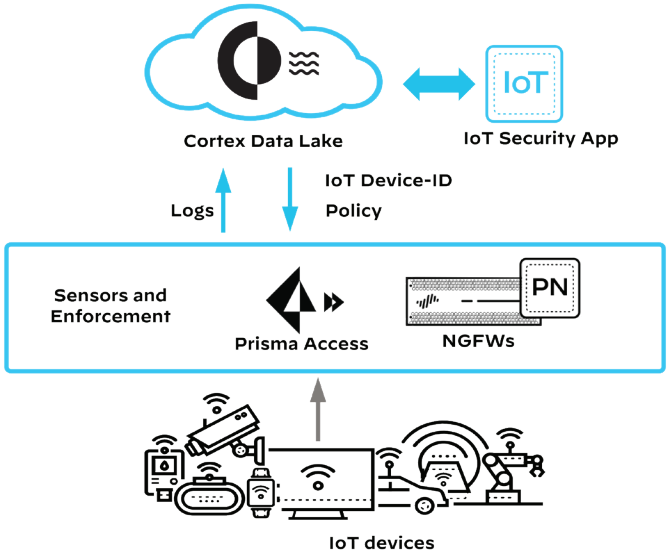
- **Understand IoT assets**—Provides full device discovery, including manufacturer, device type, software version, serial number, and multiple other attributes, as well as using with ML and device profiling for continuous ongoing detection and classification of all IoT and OT devices.
- **Assess IoT risks**—Evaluates the risks and vulnerabilities, using vendor information for specific patching or software updates that are required. In addition, IoT Security provides continuous ongoing and 100% passive risk assessment and automated risk-based policy recommendations.
- **Apply risk reduction policies**—Provides recommendations for policy enforcement based on risk and behaviors, using NGFW features (including App-ID™, User-ID™ and Device-ID) to reduce the attack surface.
- **Prevent known threats**—Provides full detail of device context for alerts, protection from exploits, C2, spyware, malware, and other known threats through other available subscriptions on the NGFW, such as Threat Protection, WildFire, URL Filtering, and DNS Security.
- **Detect and respond to unknown threats**—Leverages ML with threat-modeling in order to detect threats, zero-day detection, incident response, isolation, and device quarantine.

The IoT Security solution is cloud delivered and provides complete visibility, in-depth risk analysis, and automated enforcement with the NGFW/Prisma Access. The components of the IoT solution consist of an IoT Security app residing on the Palo Alto Networks hub, data storage, and log retention, as well as an IoT security subscription.

The NGFWs and Prisma Access nodes behave as sensors and generate enhanced application logs, which they send to Cortex Data Lake, where the IoT Security app leverages this data. The IoT Security app analyses the data, provides IP-address-to-device mappings, and creates recommendations for policy rules to implement. From the IoT Security app, you can create security policy rules, which you can then import to the NGFWs or Panorama™ for enforcement.

The IoT subscription add-on security offering from Palo Alto Networks is easy to deploy. Because the add-on leverages your existing NGFW, you do not need any extra infrastructure. To reduce risk and secure your environment, the add-on provides complete IoT security, quickly discovers all your devices, and understands the full device context.

Figure 12 IoT security solution



GlobalProtect

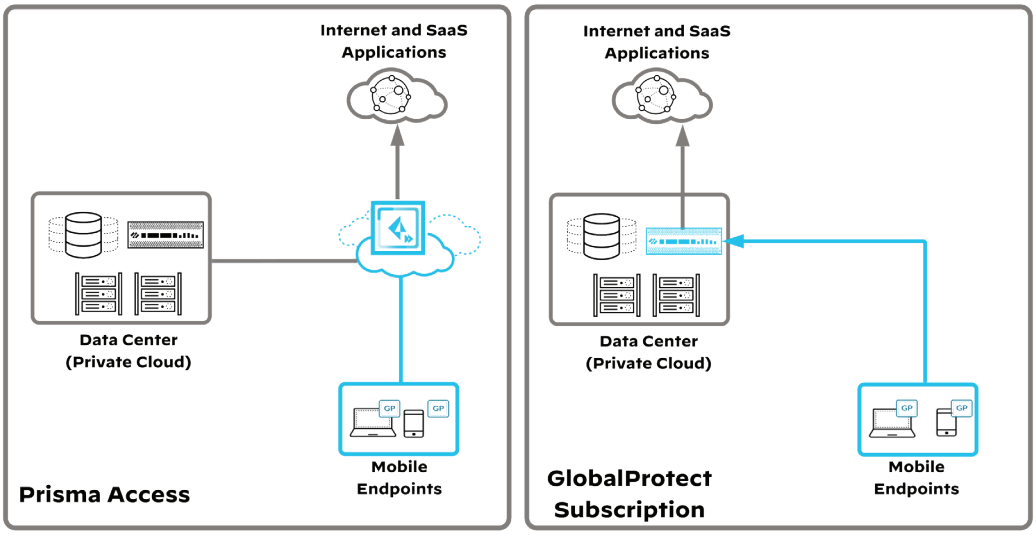
Organizations need to allow mobile users access to corporate resources, regardless of where the users are located, while ensuring that users are secure and that the organizations have visibility and control of the traffic. Traditional technologies used to protect mobile endpoints, such as host endpoint antivirus software and remote access VPNs, cannot stop the advanced techniques employed by today's sophisticated attackers. GlobalProtect network security for endpoints enables you to protect your mobile workforce by extending the Palo Alto Networks NGFWs (hardware, software, and Prisma Access) to all users, regardless of location. GlobalProtect enables security teams to build Zero Trust policies that are consistently enforced whether the user is internal or remote.

GlobalProtect extends Zero Trust Enterprise capabilities to the mobile workforce by inspecting all traffic using your NGFWs deployed as VPN gateways, whether at the perimeter, in the demilitarized zone (DMZ), or in the cloud (Prisma Access). With the GlobalProtect app, laptops, smartphones, and tablets automatically establish an IPsec/ SSL VPN connection to the NGFW using the best gateway, thus providing full visibility of all network traffic, applications, ports, and protocols. By eliminating the blind spots in mobile-workforce traffic, your organization can maintain a consistent view into applications.

To get an inventory of how the endpoint is configured, GlobalProtect checks it and builds a host-information profile (HIP) that is shared with the NGFW. The NGFW uses the HIP to enforce application policies that only permit access when the endpoint is properly configured and secured.

GlobalProtect supports clientless SSL VPN for secure access to applications in the data center and the cloud from unmanaged devices. This approach allows you to enable secure access for third-party users and employees connecting from bring-your-own-device (BYOD) devices by providing access to specific applications through a web interface, both without requiring users to install a client and without setting up a VPN tunnel.

Figure 13 Deployment options for GlobalProtect



Prisma Cloud

Public-cloud adoption is increasing dramatically as organizations transform their business models and shift their focus toward speed-to-market. Public cloud is becoming the dominant way that organizations develop and deploy applications and store data. Organizations moving to public-cloud services gain several benefits, including scalability, lower cost, reduced maintenance, reliability, and data recovery. However, the security and privacy of data remain a major concern. As your organization's cloud presence expands, its attack surface is increasing at the same rate. Poor visibility, fragmented tooling, data security, and limited threat-prevention increase your organization's potential to experience a critical security incident.

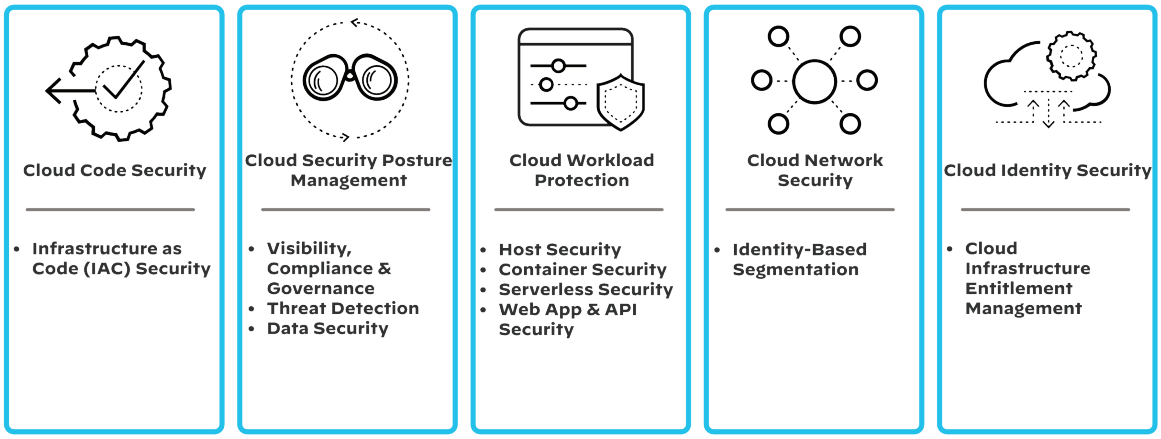
Software development tools and processes are also evolving in order to take advantage of public cloud capabilities. The DevOps model is frequently used to reduce time to market by uniting development and operations teams throughout the entire software delivery process. One problem in DevOps is that security often ends up unintentionally neglected or ignored. Developers move quickly, their workflows are automated, and they deploy applications without going through the proper security channels, inevitably making risky security mistakes.

Prisma Cloud is the industry's most comprehensive cloud-native security platform, with the broadest security and compliance coverage for applications, data, and the entire cloud-native security stack. Prisma Cloud integrates and centralizes otherwise disparate security functions into a single UI. This

approach provides visibility across silos and ensures that security, cloud infrastructure, and DevOps teams can deliver full-stack security. To identify and fix flaws early in the application lifecycle, a single platform can protect applications at runtime while also integrating security into development workflows.

At application runtime, Prisma Cloud offers a defense-in-depth approach by securing the cloud infrastructure environment, applications within it, and related cloud services and entitlements. To improve overall security outcomes, Prisma Cloud also integrates security into development and DevOps workflows, identifying and preventing vulnerabilities and misconfigurations for applications and infrastructure-as-code (IAC) templates. If a misconfigured IaC template is used dozens or hundreds of times, it could easily add hundreds or thousands of misconfigurations and alerts for security teams to address later in the process.

Figure 14 Prisma Cloud platform capabilities



Cloud Code Security

To quickly provision and update cloud applications and infrastructure, development and DevOps teams rely on containers and IaC templates. At each stage of application development—from code, to build, to deploy, to run—Prisma Cloud directly provides the developer feedback about misconfigurations and vulnerabilities. Integrated security into the developer's toolchain ensures security testing is automated as code is created and moves through the development pipeline. This approach reduces the friction of adoption by developers and runtime alerts in production.

The platform offers full-stack security from code to cloud, covering:

- **IaC security**—You can identify and prevent vulnerabilities, misconfigurations, and compliance violations in popular IAC frameworks, such as Terraform, Amazon Web Services (AWS) CloudFormation, Kubernetes manifests, and more. To check against hundreds of policies, Prisma Cloud scans IaC files via the command line interface (CLI) or integrated development environment. These policies are built on industry benchmarks such as CIS, NIST, PCI, and HIPPA, as well as community use cases. To catch misconfigurations during code development, Prisma Cloud integrates these controls into developer tools. This means there are fewer misconfigurations when code is checked back into a version control system and fewer misconfigurations for the cloud infrastructure when the code is applied.
- **Secrets scanning**—Prevents exposing passwords and API keys so the secrets never make it into public repositories. Prisma Cloud finds hardcoded secrets in IAC templates, container images, registries, and CI/CD scans. To locate common and uncommon secrets such as AWS access keys and database passwords, you can use regular expressions, keywords, or entropy-based identifiers.
- **Container image scanning**—Container images are a key component of cloud-native applications. However, they typically include many resources outside the control of developers, such as operating systems and configurations. Prisma Cloud allows security teams to provide actionable feedback and guardrails for vulnerabilities and compliance violations in container images to keep these components secure. Prisma Cloud identifies vulnerabilities in container images, provides remediation guidance, and alerts about and/or blocks images with vulnerabilities.
- **Repository scanning**—A majority of modern application code is composed of open-source dependencies. Prisma Cloud locates dependencies in repositories and builds a software bill-of-materials of the packages in use for vetting. Lack of awareness and breaking changes prevent developers from using the latest packages that minimize vulnerabilities. Prisma Cloud scans git and non-git-based repositories for package vulnerabilities and compares them against public databases. To prioritize updating libraries, the output of the findings includes the fix status, the minimum version to remediate, and the time since the fix was released.

Cloud Security Posture Management

Effective cloud security requires visibility into every deployed resource, along with confidence in their configuration and compliance status. Prisma Cloud takes a unique approach to cloud security posture management (CSPM), going beyond compliance and configuration management. Prisma Cloud Intelligence Stream provides vulnerability intelligence from more than thirty sources so that security teams can prioritize risks and quickly respond to issues.

Prisma Cloud offers the following CSPM capabilities:

- **Visibility, compliance, and governance**—Prisma Cloud delivers asset inventory management that provides visibility and control over the security posture of every deployed resource. Although some solutions simply aggregate asset data, Prisma Cloud analyzes and normalizes disparate data sources in order to provide risk clarity. Prisma Cloud continuously monitors cloud compliance posture and supports one-click reporting from a single console. More than fifteen compliance frameworks are included out of the box, and you can build additional custom frameworks.
- **Threat detection**—Prisma Cloud provides user and entity behavior analytics (UEBA). It analyzes millions of audit events and then uses machine learning to detect anomalous activities that could signal account compromises, data exfiltration, insider threats, stolen access keys, and other potentially malicious user activities. Network anomaly detection monitors cloud environments by ingesting flow logs in order to detect unusual network activities, classify and view suspicious IP addresses, detect port scan, and port sweep activities that probe a server or host for open ports. Detailed visualization tools help security operations quickly determine the scope of threats.
- **Data security**—Prisma Cloud extends Enterprise DLP and WildFire security services in order to provide complete visibility and classification into all AWS Simple Storage Service (S3) buckets and objects, including contents by region, owner, and exposure level. To identify and monitor sensitive content, you can fine-tune data identifiers such as driver's license number, Social Security number, credit card number, or other patterns. Prisma Cloud includes specific data policies in order to quickly determine your risk profile based on data classification and exposure/file types. Prisma Cloud helps users identify and protect against known and unknown file-based threats that have infiltrated S3 buckets, leveraging the WildFire malware prevention service to flag any objects that contain malware. Prisma Cloud automatically generates alerts for each object based on data classification, data exposure, and file types. Analysts can take action on alerts in order to quickly remediate exposure, tag individual DevOps teams for violations, and delete any objects that contain malware.

Cloud Workload Protection

The Prisma Cloud Compute (PCC) module is the cloud workload protection platform that provides a comprehensive view into every host, container, and serverless function. The PCC module provides vulnerability management, compliance checks, runtime security, network visibility and access control for cloud workloads. Prisma Cloud Web Application and API Security auto-discovers unprotected web apps and APIs, with full coverage across the OWASP Top 10 threats, in any public or private cloud.

You easily integrate PCC into your source-code development, container-build process, and runtime deployment, providing security and compliance capabilities at each stage. Security teams can set policies that allow only compliant and fully remediated images to progress down the DevOps pipeline. Upon deployment, PCC immediately begins working to secure your workloads.

Prisma Cloud Compute is composed of two components: a console and an agent. You access the PCC console from within the Prisma Cloud Enterprise Edition console. The PCC console provides a centralized dashboard where you define policy and monitor your environment. To help prioritize risks in real time, the PCC console shows key metrics, including vulnerability status, remediation guidance, and real-time alerts.

Defender is the PCC agent component that runs on each host. Defender enforces the policies defined in the console and sends event data to the console for correlation. Defender is completely containerized and uses a least-privileged security design.

Cloud Network Security

Network protection must be adapted for cloud native environments while still enforcing consistent policies across hybrid environments. Cloud-native deployments can present the following additional challenges to traditional enterprise network security policies:

- Cloud workloads can quickly move across hosts, subnets and/or locations. Network enforcement points might not be able to follow the workload and apply consistent policy as it moves.
- IP addresses are no longer as persistent and in many cases, become obscured to external devices due to extensive use of NAT, proxies, and load balancers. Network security devices might lack visibility for granular policy-enforcement.
- It is difficult to determine which applications communicate with each other. Network and security teams must comb through IP logs and stitch the IPs together, and those can change over time. Creating segmentation policies without the correct dependency mapping can break applications.

Identity-Based Microsegmentation in Prisma Cloud gives your organization the ability to base security policies on strong, machine-generated persistent identity for individual workloads instead of broad IP addresses. This reduces the attack surface of cloud networks to the individual workload level and makes it possible to track a workload as it moves through your environment, even if IP addresses and other traditional identifiers change. East-west traffic segmentation between workloads traversing multiple IP domains is no longer an issue because IP reachability no longer assumes application access.

Container and virtual-machine workloads are appointed an identity that can be dynamically ascribed from cloud-native sources, including system information attributes such as operating system or hostname; cloud provider information, including IAM role; or Kubernetes objects such as namespace, app labels, service accounts, and more. After workload identities are assigned, Prisma Cloud can automatically discover and learn application communication behaviors inside and across clouds. You can then monitor the application flows in real time by using a visualized map of the network presented on the dashboard. Security teams can then set policies to authenticate and authorize connection requests across a generated set of identities and automatically deny unknown requests. Compliance teams, for example, can create policies to isolate systems that might be subject to specific regulations.

Cloud Identity Security

Controls for who is authenticated and who is authorized to use cloud resources and make changes to configurations are implemented inconsistently across cloud service-providers. There are many different (and sometimes overlapping) policies and levels of permissions that can be attached to each user, which makes the seemingly simple task actually rather complicated. Traditional manual methods for determining least-privileged access make it difficult for security teams to keep up with the growing number of entitlements across cloud services.

Common threads across security breaches include:

- Overly permissive roles.
- Reuse of the same IAM role for many users and machines.
- Resources exposed to the public internet.

Prisma Cloud continuously detects and automatically remediates identity and access risks across infrastructure-as-a-service (IaaS) and platform-as-a-service (PaaS) offerings. It discovers all human and machine identities across cloud environments and then analyzes entitlements, roles, and policies. Prisma Cloud provides permissions visibility, IAM governance, automated response and UEBA.

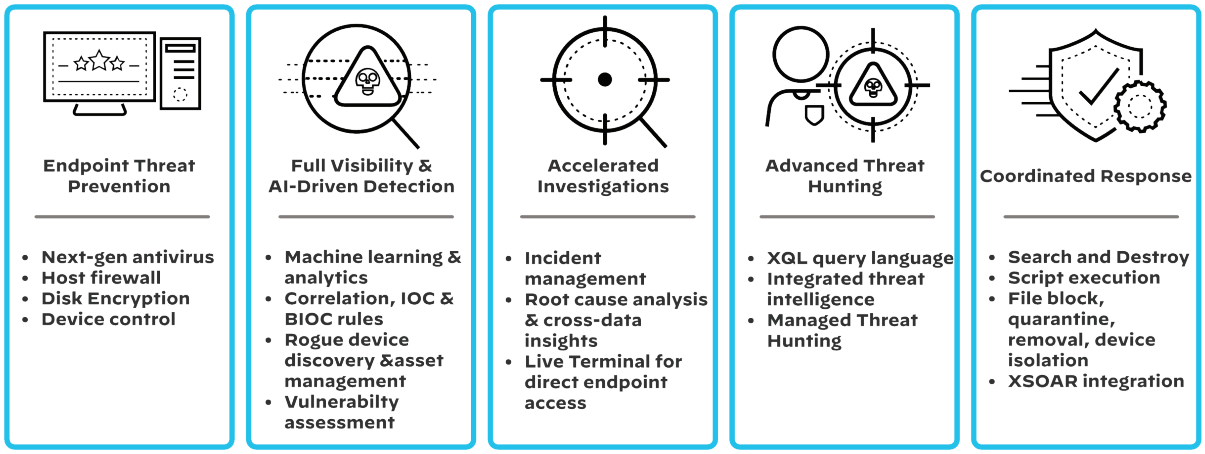
Cortex XDR

Security teams face an endless array of threats, from ransomware and cyberespionage to fileless attacks and damaging data breaches. However, the biggest challenge for many security analysts is the repetitive tasks they must perform every day as they triage incidents and attempt to reduce an endless backlog of alerts.

Cortex XDR is the industry's first extended detection and response platform that integrates network, endpoint, cloud, and third-party data to stop sophisticated attacks. Cortex XDR was designed from the ground up to help organizations secure their digital assets and users while simplifying operations. Using behavioral analytics, it identifies unknown and highly evasive threats. ML and AI models uncover threats from any source, including managed and unmanaged devices.

Cortex XDR helps you accelerate investigations by providing a complete picture of each alert. It stitches diverse types of data together and reveals the root cause and timeline of alerts, allowing analysts of all experience levels to perform triage. Tight integration with enforcement points lets you respond to threats anywhere in your organization or restore hosts to a clean state easily. With Cortex XDR, you can use your existing network, endpoint, and cloud security as sensors and enforcement points, eliminating the need to deploy new software or hardware.

Figure 15 Cortex XDR capabilities



Endpoint Threat Prevention

The Cortex XDR agent safeguards endpoints from malware, exploits, and fileless attacks with industry-best, AI-driven local analysis and behavior-based protection. Integration with WildFire boosts security accuracy and coverage. The lightweight agent is cloud-delivered and secures your endpoints without slowing them down or impacting network performance. The agent starts protecting your endpoints immediately without requiring a reboot. The following endpoint threat prevention capabilities are available with the Cortex XDR agent:

- **Next-generation antivirus**—Antivirus signatures cannot keep up with fast-moving threats. To correctly find and block known and unknown malware, AI-driven local analysis analyzes thousands of attributes of a file.
- **Host firewall**—You can reduce the attack surface of your endpoints by controlling network access.
- **Disk encryption**—You can secure endpoint data by managing BitLocker and FileVault encryption policies on your endpoints.
- **Device control**—You can securely manage USB devices and protect your endpoints from malware and data loss.

Visibility and Threat Detection

To stop sophisticated attacks, Cortex XDR gathers data from network, endpoint, cloud, and third-party data. All data from across your organization is stitched together so you gain complete visibility, eliminate blind spots, and root out advanced threats. Visibility and threat detection capabilities include the following:

- **ML-driven detection**—Using machine learning, Cortex XDR continuously profiles user and endpoint behavior in order to detect anomalous activity indicative of attacks.
- **Correlation, IOC & BIOC rules**—Correlation enables you to broaden the scope of threat detection across your entire environment. Pre-defined and custom indicator of compromise (IoC) rules make it easy for analysts to find malicious or suspicious artifacts. Behavioral-based indicators of compromise (BIOCs) address a more complex problem, which is to identify threats by evaluating pre-defined and custom endpoint behavior rules.
- **Asset management**—Cortex XDR Asset Management can locate and manage your assets more effectively and reduce the amount of investigation required to identify unmanaged devices on your network.
- **Vulnerability assessment**—Cortex XDR identifies and prioritizes the security vulnerabilities on your endpoints. You can view the vulnerabilities detected by CVE or by host.

Accelerated Investigations

Cortex XDR allows you to manage and investigate threats quickly by providing a complete picture of each attack, including alerts, artifacts, and MITRE tactics. Threat investigation capabilities include the following:

- **Incident management**—Intelligent alert-grouping and incident-scoring reduces the number of individual alerts to review and alleviates alert fatigue while letting you focus on the threats that matter.
- **Root cause analysis**—Cortex XDR automatically reveals the root cause, reputation, and sequence of events associated with each alert, lowering the experience level needed to verify an attack.
- **Live Terminal**—Provides flexible response actions to run Python, PowerShell, or system commands or scripts on endpoints. Live Terminal also lets you review and manage active processes and view, delete, move, or download files.

Advanced Threat Hunting

Cortex XDR allows your security team to search, schedule, and save queries to identify hard-to-find threats. By integrating threat intelligence with an extensive set of security data, your team can catch malware, external threats, and malicious insiders. The following advanced threat-hunting capabilities are available with Cortex XDR:

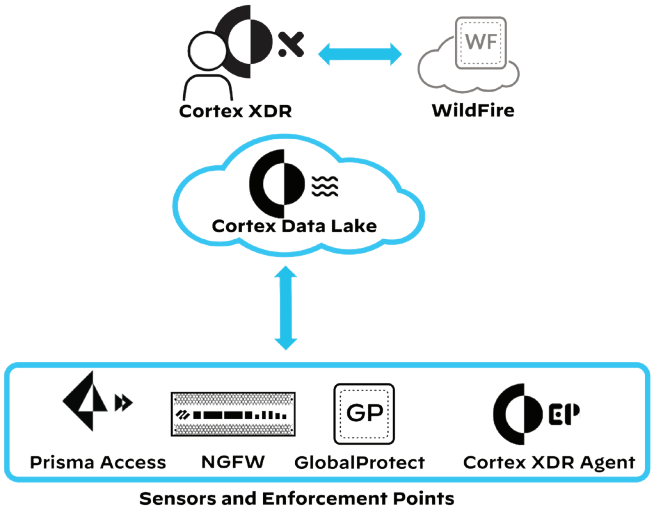
- **XQL**—For rigorous endpoint and network event analysis investigations, XDR Query Language (XQL) enables you to perform powerful queries on data ingested into Cortex XDR.
- **Integrated threat intelligence**—Cortex XDR displays the WildFire-issued verdict for each key artifact in an incident. To provide additional verification sources, you can integrate an external threat intelligence service such as AutoFocus™ and VirusTotal.
- **Managed Threat Hunting**—To augment your team and discover attacks anywhere in your environment, Cortex XDR Managed Threat Hunting offers round-the-clock monitoring from world-class threat hunters.

Coordinated Response

Cortex XDR lets your security team instantly contain endpoint, network, and cloud threats from one console. Your analysts can quickly stop the spread of malware by blocking files, stopping processes, restoring hosts to a clean state, and restricting network activity to and from devices. Tight integration with enforcement points allows you to update prevention lists, such as bad domains. The following threat-response capabilities are available with Cortex XDR:

- **Search and destroy**—Find, retrieve, and delete malicious files across all endpoints.
- **Script execution**—For enhanced remediation, run and track execution of Python 3.7 scripts on your endpoints directly from Cortex XDR.
- **File block**—Block additional executions of a given file by adding it to the block list in the policy.
- **File quarantine**—Quarantine malicious files and remove them from their working directories.
- **Endpoint isolation**—Disable all network access on compromised endpoints except for traffic to the Cortex XDR management console.
- **Cortex XSOAR integration**—Integrate with Cortex XSOAR and use automated playbooks for security orchestration, automation, and response.

Figure 16 Cortex XDR



Cortex XDR provides a seamless platform experience by combining endpoint policy management, detection, investigation, and response in one web-based management console.

ZERO TRUST READY INFRASTRUCTURE

Not all Zero Trust networks are created equal. For securing access to distributed applications and data, Palo Alto Networks provides flexibility with the following two reference architectures:

- Cloud-delivered network security (SASE solution)
- On-premises network security (NGFWs at the edge)

Both options provide industry-leading security subscriptions, such as DLP, WildFire threat prevention, and SaaS Security. They also provide high-performing connectivity services for networks and mobile users. You can choose to deploy one or a combination of both solutions.

SASE Reference Architecture

SASE is the convergence of security services and software-defined WAN (SD-WAN) services in a cloud-based solution. A SASE solution integrates these services seamlessly and provides secure access to applications and data no matter where they reside or from where they are being accessed. The solution provides that access from anywhere and provides security services along the path from the user to the application or data without unnecessary redirects, such as through a centralized data center. This reduces latency and improves the user experience. Security is improved because users no longer "turn off the VPN" to get the performance they want from their applications.

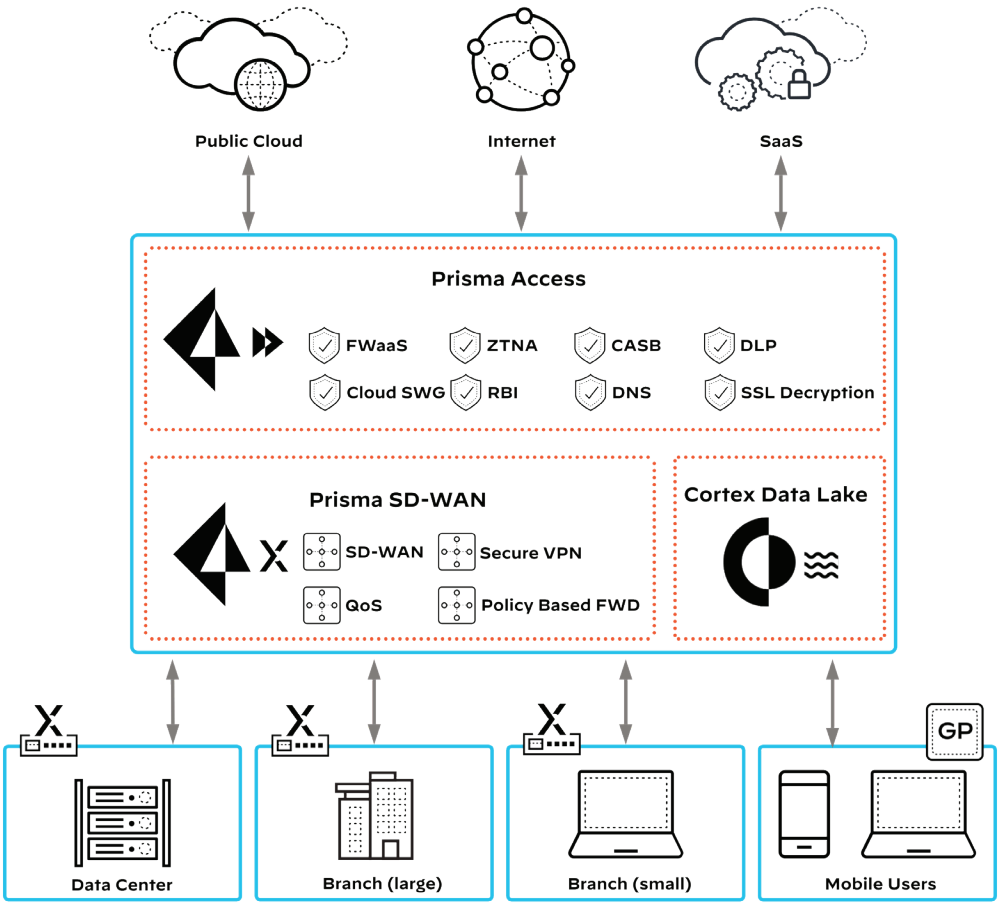
Prisma SASE includes the following main components:

- **Prisma Access**—A NGFW delivered as a cloud-native service. Use this for securing mobile-user and remote-site access to internet, SaaS, and public-cloud services.
- **Prisma SD-WAN**—Next-generation, software-defined, application-aware wide-area networking combined with cloud-orchestration. Use this for secure data transport between sites and private cloud services.

The capabilities of the Palo Alto Networks SASE solution include:

- **Cloud-native, cloud-based delivery**—Over 100 points of presence worldwide reduces latency, enhancing the user experience, and includes support of in-country or in-region resources and regulatory requirements.
- **Scalability**—Quickly and easily onboard users without overloading your existing infrastructure or needing to acquire and deploy additional resources. Bring new sites online quickly with the Prisma SD-WAN ION device.
- **Line-rate security**—Prisma Access's single-pass architecture provides a suite of security services without adding the latency that you would have when daisy-chaining security products.
- **Single vendor management**—The components of Palo Alto Networks SASE solution work together seamlessly. There is no need to figure out how to piece together various products from multiple vendors using multiple APIs and orchestration applications.

Figure 17 Cloud-delivered, Zero Trust–ready infrastructure



On-Premises Reference Architecture

Some customers might prefer an on-premises solution for the following reasons:

- An already existing investment in NGFWs in their branch and central locations
- Requirements for local segmentation inside the network
- Regulatory restrictions that prevent the use of cloud services for specific locations or verticals

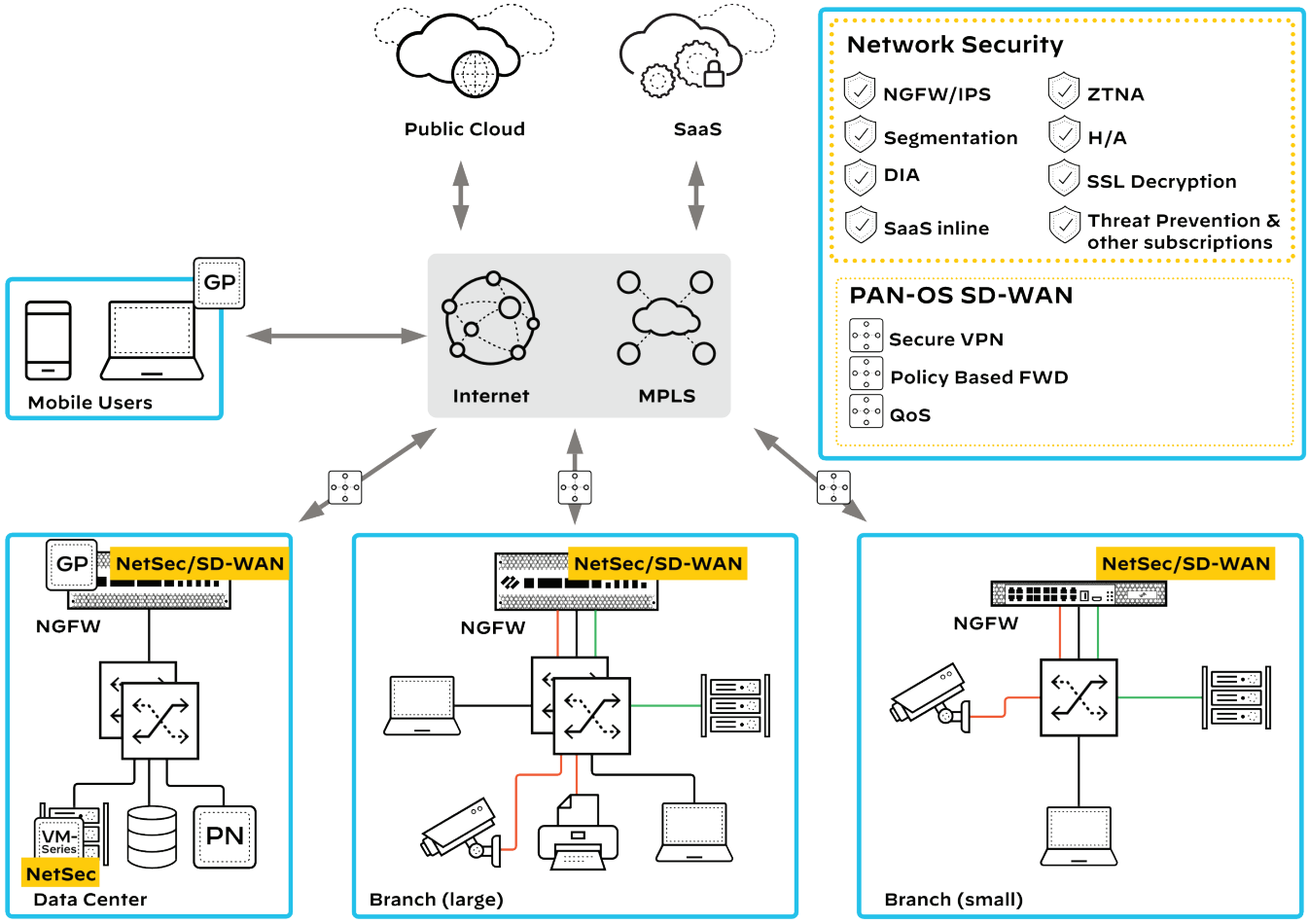
The on-premises network security solution offers self-managed and self-deployed capabilities. Application-aware policy determines how the local LAN traffic flows and whether outbound traffic is sent to the corporate WAN tunnels or directly to the internet. The local NGFW performs the Layer 7 traffic inspection, access control, threat prevention, and security services when accessing internet-based applications and data. Access for mobile users is provided by VPN tunnels (from the GlobalProtect client on the endpoint), which are terminated on a GlobalProtect gateway at the central site or a regional access site.

The capabilities of the Palo Alto Networks on-premises solution include:

- **SD-WAN integration in NGFW**—This model consolidates SD-WAN and security functions on a single device at remote sites and data center.
- **Line-rate security**—The NGFW single-pass architecture provides a suite of security services without adding the latency that you would have when daisy-chaining security products.
- **Panorama management**—Use a single management platform for all PAN-OS solutions.

The following figure shows the on-premises infrastructure for a Zero Trust network.

Figure 18 On-premises, Zero Trust–ready infrastructure



Implementing Zero Trust Enterprise

The Palo Alto Networks portfolio provides the tools, technologies, and products you need to turn your Zero Trust strategy into a practical implementation. Achieving Zero Trust is often perceived as costly and complex, but it does not have to be when using the right infrastructure. With Palo Alto Networks, Zero Trust capabilities are built in your existing architecture and do not require you to rip and replace existing technology. As you begin planning your Zero Trust implementation, you should understand the five-step methodology and apply the concepts.

FIVE-STEP METHODOLOGY

You implement and maintain Zero Trust by using a simple, iterative, five-step methodology. This guided process helps identify where you are and where to go next. Palo Alto Networks professional services and certified partners can assist you with personalized consulting resources for implementing Zero Trust Enterprise in your environment.

Identify the Protect Surface

The Zero Trust Enterprise security model protects your most sensitive assets by controlling access and verifying all transactions. Because the most critical and sensitive data typically resides on resources within the private data center or public cloud, you typically start the design of Zero Trust at these locations and then migrate towards the user.

There are a considerable number of standards and regulations (such as GDPR, HIPAA, and PCI) that can help you define how data and applications are categorized. When identifying the protect surface, you should assess the business impact to your organization if sensitive data gets exposed.

When defining the protect surface, you need to consider all critical DAAS in your environment:

- **Data**—Payment card information, protected health information, personally identifiable information, and intellectual property
- **Applications**—Off-the-shelf or custom software
- **Assets**—Networking equipment, point-of-sale terminals, medical equipment, manufacturing assets, and IoT devices
- **Services**—DNS, DHCP, and identity stores

Map the Transaction Flows

To apply Zero Trust most effectively, you need to understand the application flows within your organization. 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 DAAS components interact with other resources on your network. It is a common practice to approximate flows by documenting what you know about specific resource interactions, even without having a complete picture. This information still provides valuable data so that you do not arbitrarily implement controls with insufficient insight. You can also deploy NGFWs in monitor mode in order to gather precise data about traffic flows and applications without disrupting the network.

After you understand how your systems work, the flow maps tell you where you need to insert controls. To become familiar with the process, tools and operations, start with a small, non-critical protect surface. You should then prioritize mission-critical DAAS. As you move through the steps in this methodology, you gather more information about what works for your situation, which allows you to enable more granularity in your design as you move your security controls closer to your most important assets.

Pick a Zero Trust Infrastructure

Zero Trust is not a product; however, there are products that work well in Zero Trust environments, along with many that do not. Selecting the right infrastructure, one that provides the required security controls for your DAAS, is key when enabling Zero Trust policies across your locations.

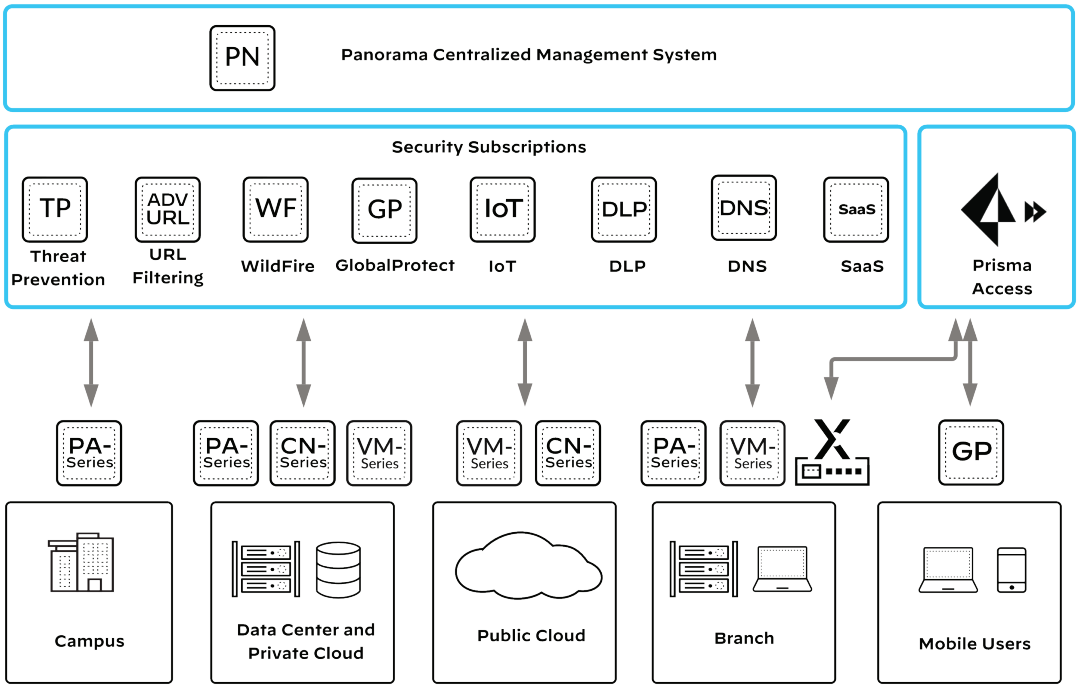
The Palo Alto Networks NGFWs offer flexibility in covering your deployment location needs. This is due to the variety of consumption options such as the PA-Series, VM-Series, CN-Series, and Prisma Access. The NGFWs are typically deployed in the following types of locations:

- **Campus**—In the campus, IT departments face many challenges with infrastructure management, visibility, and security. This is due to the increased number of IoT devices, wireless networks, BYOD devices, and multi-floor and multi-building connected networks, as well as increased demand for internet, data center, public cloud, and SaaS access. Protecting data, applications, devices and users is paramount. Palo Alto Networks NGFWs can secure the campus environment, protecting your key assets and providing visibility and threat protections that secure all east-west and north-south traffic flows.
- **Private data center**—You must deploy applications rapidly and with consistent protection, no matter where data and applications reside in the data center. Using a prevention-focused architecture, Palo Alto Networks secures data centers with PA-Series and VM-Series NGFWs. Supported integrations include Cisco ACI, VMware NSX-T, VMware ESXi, and many others. You can deploy the VM-Series within your private cloud deployments in order to provide inline security and threat protection across multiple environments, including KVM, OpenStack, VMware, and Nutanix.

- **Public cloud**—To reduce the attack surface and secure both north-south and east-west traffic flows, you can deploy VM-Series NGFWs in multiple public-cloud providers. The VM-Series provides comprehensive visibility and control across multiple cloud providers, including AWS, Azure, Google, and Oracle Cloud. As your organization's cloud presence expands, its attack surface is increasing at the same rate. Prisma Cloud offers a defense-in-depth approach by securing the cloud infrastructure environment, applications within it, and related cloud services and entitlements. To improve overall security outcomes, Prisma Cloud also integrates security into development and DevOps workflows in order to identify and prevent vulnerabilities and misconfigurations during the full development process.
- **Container clusters**—Container clusters are used both in private and public cloud. Containers often cause a security challenge due to the lack of granular visibility and control. The CN-Series from Palo Alto Networks provides container-level security for inbound, outbound, and east-west traffic flows. The CN-series are natively integrated into Kubernetes in order to provide full Layer 7 visibility, threat protection, and the ability to isolate and protect container workloads in both private and public clouds. Prisma Cloud provides vulnerability management, compliance checks, container runtime security, network visibility, and identity-based microsegmentation for container workloads. You can use one or both solutions, depending on the controls you want to put in place.
- **Branch**—Branch locations need access to applications that might be in the data center, the cloud, or through SaaS providers. You can achieve secure branch internet connectivity with Prisma SASE (Prisma SD-WAN and Prisma Access) or with an on-premises NGFW that provides SD-WAN and security services. The secure branch internet deployment models are:
  - **Cloud-delivered NGFW as a service**—In this model, Prisma Access and Prisma SD-WAN deliver internet security services. Lightweight edge-security services complement the full suite of security services in the cloud. This model also provides scalable, secure WAN transport between your organization's locations.
  - **On-premises NGFW**—In this model, a local Palo Alto Networks NGFW provides local network segmentation with zones to secure localized east-west traffic as well as internet-bound traffic. Deploying a local NGFW provides threat protection, visibility, and control of all traffic and provides IoT security in the local branch. To optimize WAN transport to other NGFW SD-WAN locations and to internet-based applications, you can leverage the SD-WAN services provided by PAN-OS on the NGFW.
- **Remote locations**—Prisma Access for users provides security services, including App-ID and threat prevention for mobile users, as a cloud-delivered service. You can also leverage GlobalProtect running on NGFWs as an alternative to or in combination with Prisma Access.

The following figure shows Zero Trust deployment options from Palo Alto Networks.

Figure 19 Zero Trust deployment options



Create the Zero Trust Policy

The concept of least-privileged access forms the basis of a Zero Trust security policy. For one resource to talk to another, a specific policy must allow 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. With the Kipling Method, you can easily write policies by answering:

- **Who**—What user or group of individuals should access the protected resource? This defines the asserted identity.
- **What**—What data and services from the protected resource should the asserted identity be allowed to access?
- **When**—When should the asserted identity have access to the protected resource?
- **Where**—From where should the asserted identity have access to the protected resource?
- **Why**—Why does the asserted identity need access to the protected resource?
- **How**—What application and/or device should the asserted identity use to access the protected resource?

The NGFWs and Prisma Access contain tightly integrated technologies for uniquely identifying users, applications, and devices across different network segments or security zones. To create your Zero Trust policy, use these tools as building blocks.

App-ID

App-ID uses multiple identification techniques such as application signatures, application protocol decoding, and heuristics to determine the exact identity of applications on your network, including those that try to evade detection by disguising themselves as legitimate traffic, by hopping ports, or by using encryption. App-ID allows you to use application information in your security policies rather than relying on port numbers and protocols.

User-ID

User-ID enables you to uniquely identify users on your network, even through changes in IP address. It leverages user information stored in a wide range of repositories such as LDAP and user authentication via SAML. When using a Zero Trust policy, high-fidelity sources of User-ID are essential because the IP-to-user mappings need to be in place before any inbound traffic from the user reaches the NGFW. High-fidelity resources for users on the private network include infrastructure authentication services (AAA) and GlobalProtect internal gateways. For mobile users, preferred sources include GlobalProtect external gateways and Prisma Access. You can create policies based on individual users or groups of users.

Device-ID

Device-ID enables you to use device information in your security policies rather than using an IP address. Device-ID uses metadata from logs, network protocols, and sessions to identify devices by their attributes, such as a device type (for example, a printer), model, software version or vendor. Because Device-ID identifies devices by their attributes, you can use Device-ID in security policies in order to control access to and from devices. This is especially important in IoT-heavy environments. You might need to create policies to allow specific types of devices access to sensitive data, such as an x-ray machine sending imaging data to a patient record. Grouping and segmenting IoT devices by type reduces their risk of being compromised and used as a beachhead for exploiting higher-sensitivity data.

To simplify the process, you should create policies by primarily using your segmentation gateways' centralized management tool. Panorama and the Prisma Access app provide this functionality and are where you apply the Kipling Method. Palo Alto Networks NGFW technology and unique features let you write policies that are easy to understand and maintain while providing maximum security transparently to your end users.

Monitor and Maintain the Network

The last step in this iterative process is to monitor and maintain your network. This means analyzing internal and external logs through Layer 7 and focusing on verifying the operation of Zero Trust policies. Inspecting and logging all traffic on your network is a pivotal facet of Zero Trust. It is important to send the system as much telemetry data as possible about your environment. This data gives you new insights

into how to improve your Zero Trust network over time. The more your network is attacked, the stronger it becomes, with greater insight into making policies more secure. Additional data gives more insight into the protect surface, such as what is included and the interdependencies of data within it. This detail leads to architectural tweaks that 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 ML and analytics. NGFW data from all sources is also consolidated into a singular view under Panorama.

Cortex XDR takes advantage of Cortex Data Lake to create profiles of users and devices, acting as a baseline of normal use. This 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 then facilitates a deep investigation of the newfound threat so you can uncover what occurred and react accordingly.

Prisma Cloud provides public cloud security and compliance monitoring, scanning all audit and flow logs across multi-cloud 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 IPs and host vulnerabilities across your resources, which can quickly be isolated to avoid additional exposure. This data provides insight that allows you to fine-tune Zero Trust privileges.

ZERO TRUST APPROACHES

Identifying the security controls that work best for enforcing your Zero Trust policy allows you to turn your policy into a practical implementation. This section describes the Palo Alto Networks portfolio's capabilities for enforcing Zero Trust policies for users, applications, and infrastructure. This is not an exhaustive list but covers the most common security controls used in Zero Trust environments.

Zero Trust for Users

The Zero Trust for users approach allows you to secure user access to critical data and applications by removing all implicit trust and verifying all digital transactions. As shown in Table 1, the Zero Trust for users security controls are:

- **Identity**—Use strong authentication to validate users
- **Device/workload**—Verify user device integrity.
- **Access**—Enforce least-privileged user access to data and applications.
- **Transaction**—Scan all content for malicious activity and data theft.

User Identity Validation

Implementing Zero Trust for users must start by using strong authentication to verify user identity. A Zero Trust Enterprise needs to be able to allow or block access to data and applications based on verified user identity. To protect against stolen credentials, you should use multifactor authentication (MFA) against all critical assets.

User-ID

User-ID enables security teams to define policy rules on NGFWs in order to safely enable applications and control access based on users or groups of users. Because IP addresses can change, using IP address assignments to define static policies is inefficient, difficult to manage, and can pose a security risk. User-ID enables you to uniquely identify users on your network, even through changes in IP address. High-fidelity sources of User-ID are essential when using a Zero Trust policy because the IP-to-user mappings need to be in place before any inbound traffic from the user reaches the NGFW.

NGFW supports several methods for determining user-to-IP-address mapping:

- **GlobalProtect**—Because it provides both user authentication and device identification, this is the recommended solution for Zero Trust deployments with managed endpoints. This method integrates directly with the NGFW and is considered a high-fidelity source.
- **Extensible markup language API**—This NGFW API captures login events from third-party VPNs and network access control (NAC) servers. Use this method for environments in which users and devices perform network access authentication. This method integrates directly with the NGFW and is considered a high-fidelity source.
- **Captive portal**—When accessing applications using a captive portal, inline NGFW can authenticate users. Use this method in open-access networks or as an extra layer of authentication for sensitive applications. This method integrates directly with the NGFW and is considered a high-fidelity source.
- **Server monitoring**—The user IP address is provided by either a Windows-based agent (running on a domain server in your network) or the integrated PAN-OS User-ID agent running on the NGFW that monitors security event logs.
- **Port mapping**—In multi-user systems (Microsoft Terminal Server, Citrix), many users share the same IP address. The Terminal Services agent provides user-to-IP-address mapping.
- **X-Forwarded-For (XFF) headers**—When NGFW is between the users and a proxy server, the source IP address is the proxy IP address. To map the actual address of the clients behind the proxy, you can use XFF header.
- **Syslog**—If your environment has other systems that authenticate users (802.1X, NAC, proxies), you can configure these systems to send logging information for the NGFW to parse in order to obtain the IP information.
- **Client probing**—Client probing is supported for Windows client but is not recommended, because it can expose security risks and generate a large amount of traffic if misconfigured.

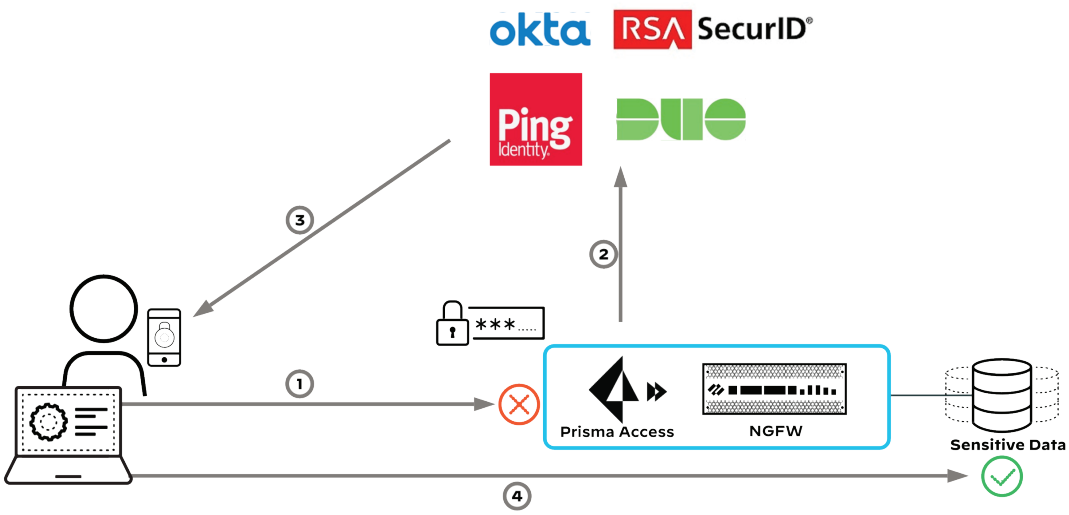
Authentication Policy

Authentication policy enables you to authenticate users before evaluating the security policy. Authentication policies are useful tools for validating User-ID before allowing access to sensitive applications. Authentication policies can force an MFA of the user before allowing access and ensure that the originator isn't using stolen credentials. Because the NGFW is proxying the authentication, the application does not require any configuration to support MFA. In fact, MFA is transparent to the application and is especially useful for administrative access that does not natively support or is difficult to configure for MFA.

The following figure describes the use of MFA to protect sensitive data and applications:

1. The user attempts to access the sensitive data and is redirected to an authentication form.
2. NGFW sends the authentication request to the MFA vendor. The methods used might vary between providers.
3. The user receives notification on his mobile device and acknowledges the authentication request.
4. The user has access to the sensitive data or application.

Figure 20 MFA authentication for sensitive data

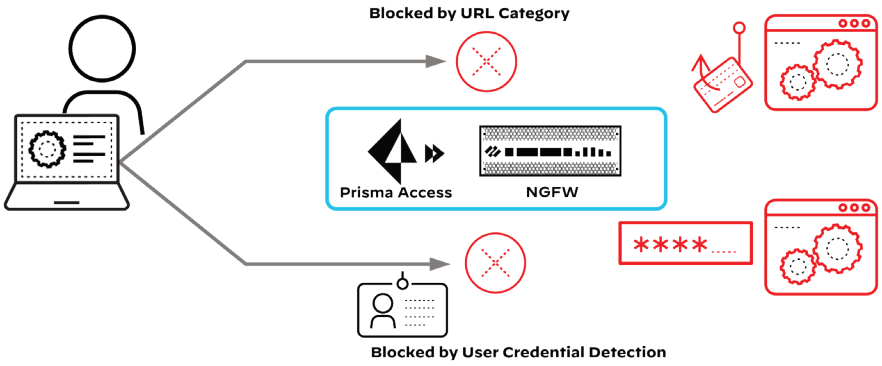


Credential-Based Attack Protection

To prevent phishing attacks from stealing your user's credentials and providing an avenue of access to your sensitive data and applications, you configure credential-phishing protection on all security policy rules that allow user access to the internet. The first part of preventing phishing attacks is configuring URL filtering on internet-bound rules that block known phishing sites.

Second, to stop phishing attempts from sites that are not part of the current URL database, configure the NGFW to use its IP-address-to-user mapping table in order to detect if a user is submitting their corporate username when they submit website forms. Enable credential phishing protections on all URL categories except for the categories that contain your sanctioned and tolerated SaaS applications.

Figure 21 Phishing prevention

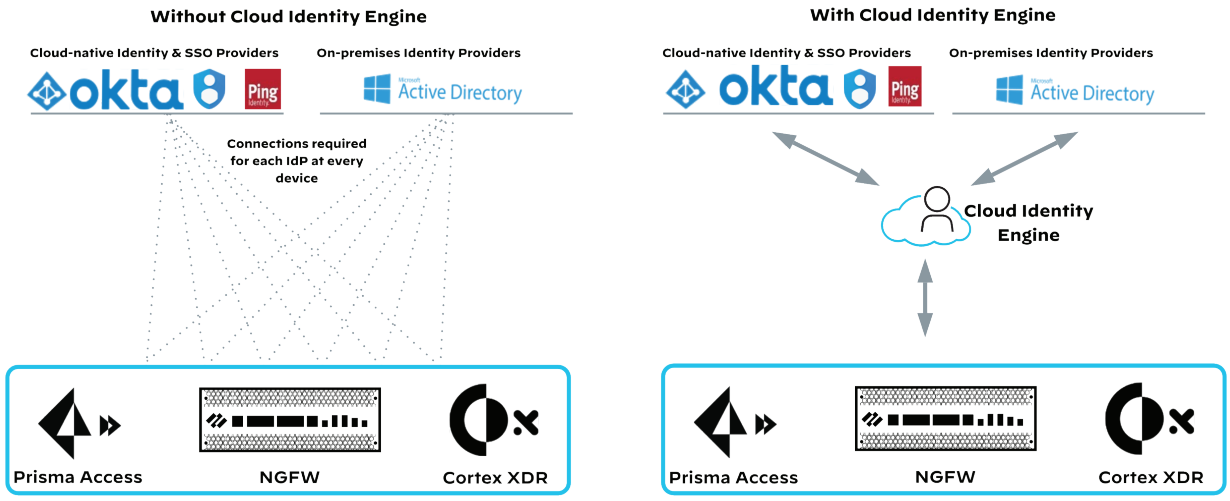


Identity Management

The storing of identity and group information for user authentication, although traditionally implemented in on-premises identity stores (examples: Active Directory, LDAP), is now also implemented in cloud-based identity providers (examples: Okta, Azure Active Directory, Ping Identity). Most organizations have multiple solutions for identity management. Distributing verified identity from the Enterprise IAMs to the security enforcement points is a key capability for Zero Trust Enterprise.

User-group information facilitates the creation of Zero Trust policies. In most organizations, user-group information is present in multiple identity providers (IdPs). The Palo Alto Networks Cloud Identity Engine simplifies the process of keeping identity information up-to-date and synchronized by pulling and correlating the information from multiple IdPs and providing unified identity information to the NGFWs, Prisma Access, GlobalProtect, and Cortex XDR. The Cloud Identity Engine also provides user authentication by using Samuel 2.0 IdPs. This simplifies NGFW configuration because they can point to the Cloud Identity Engine as a single identity store. Adding a new IdP is a very simple process at the Cloud Identity Engine, without having to involve the security enforcement points.

Figure 22 User-ID with Cloud Identity Engine



User-Device Integrity

While the Zero Trust protect surface typically resides on resources within the private data center or public cloud and not the endpoints; the use of advanced endpoint security is an important baseline protection that reduces risk and ensures businesses continuity.

In Zero Trust for users, complementing strong authentication with the verification of user-device integrity provides another layer of security by ensuring that the user's device software has not been compromised. The Cortex XDR agent safeguards endpoints from malware, exploits, and fileless attacks with industry-best, AI-driven local analysis and behavior-based protection.

GlobalProtect extends Zero Trust Enterprise capabilities to the mobile users by inspecting all traffic using your NGFWs deployed as VPN gateways. To provide visibility into the integrity of the user endpoint, the GlobalProtect agent builds a HIP. Both the Cortex XDR and GlobalProtect agents can coexist on the same device and, to enforce policy in case the integrity of the device is compromised, are integrated with NGFW and Prisma Access.

Cortex XDR Agent Protection

The Cortex XDR agent safeguards endpoints from malware, exploits, and fileless attacks with industry-best, AI-driven local analysis and behavior-based protection. Integration with WildFire boosts security accuracy and coverage. The lightweight agent is cloud-delivered and secures your endpoints without slowing them down or impacting network performance. The agent starts protecting your endpoints immediately, without requiring a reboot.

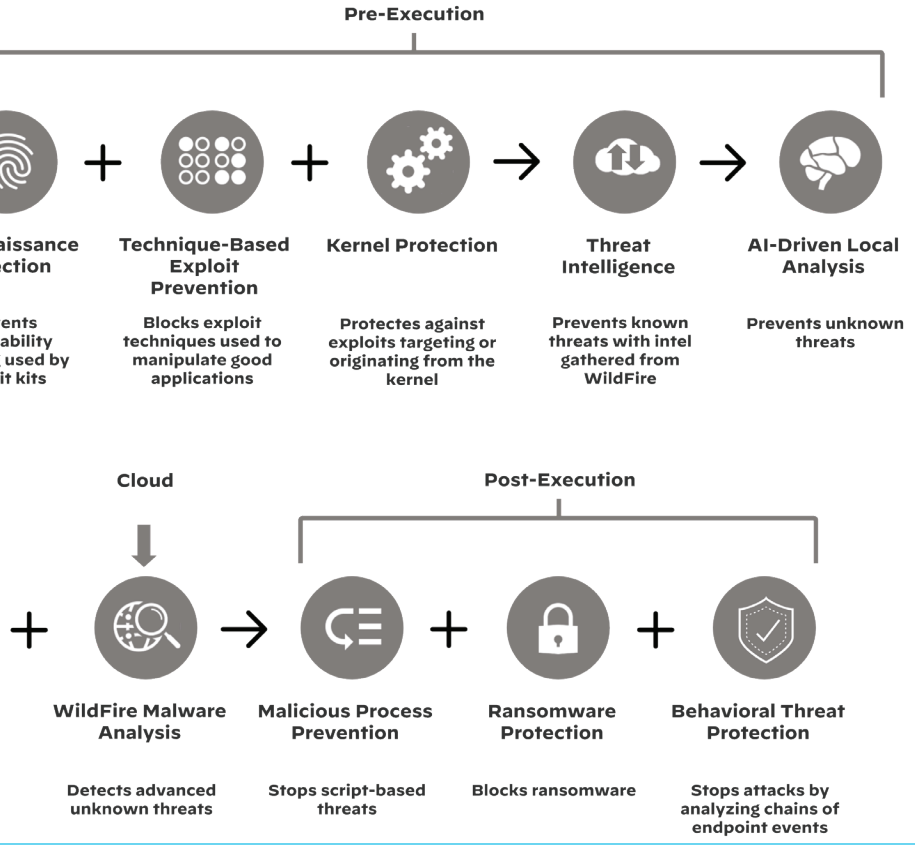
The following endpoint threat prevention capabilities are available with the Cortex XDR agent:

- **Next-generation antivirus**—Antivirus signatures cannot keep up with fast-moving threats. To correctly find and block known and unknown malware, AI-driven local analysis analyzes thousands of file attributes.
- **Host firewall**—You can reduce the attack surface of your endpoints by controlling network access.
- **Disk encryption**—Secure endpoint data by managing BitLocker and FileVault encryption policies on your endpoints.
- **Device control**—Securely manage USB devices and protect your endpoints from malware and data loss.

To compromise organizations, attackers often blend two primary attack methods, targeting application vulnerabilities through exploits and deploying malicious files. They can use these methods individually or in various combinations, but they are fundamentally different. Due to the differences between malware and exploits, effective prevention requires an approach that protects against both.

Cortex XDR detects and stops each step of an endpoint attack, from the initial reconnaissance and exploit to runtime analysis with our unique Behavioral Threat Protection engine. A deep network inspection engine blocks the spread of network threats, such as worms, while a ransomware protection module blocks ransomware attacks as they occur. To prevent successful cyber-attacks, the Cortex XDR agent coordinates enforcement with cloud and network security.

Figure 23 Cortex XDR Agent protection



Cortex XDR provides granular control over service-protection settings and the security of the Cortex XDR agent running on the endpoints. This default protection prevents attempts to disable or make changes to Cortex XDR agent processes, services, registry key values, and files.

GlobalProtect Host Information Profile

GlobalProtect extends Zero Trust Enterprise capabilities to the mobile workforce by inspecting all traffic using your NGFWs deployed as VPN gateways. In addition to VPN and strong authentication capabilities, GlobalProtect builds a HIP to provide visibility into the integrity of the endpoint. HIP attributes used in Zero Trust policies include:

- Managed/unmanaged device identification.
- Machine certificates present on device.
- Device information received from mobile device manager.
- Operating system and application patch level.
- Host anti-malware version and state.
- Host firewall version and state.
- Disk encryption configuration.
- Data backup product configuration.
- Customized host conditions (e.g., registry entries, running software)

Least-Privileged User Access

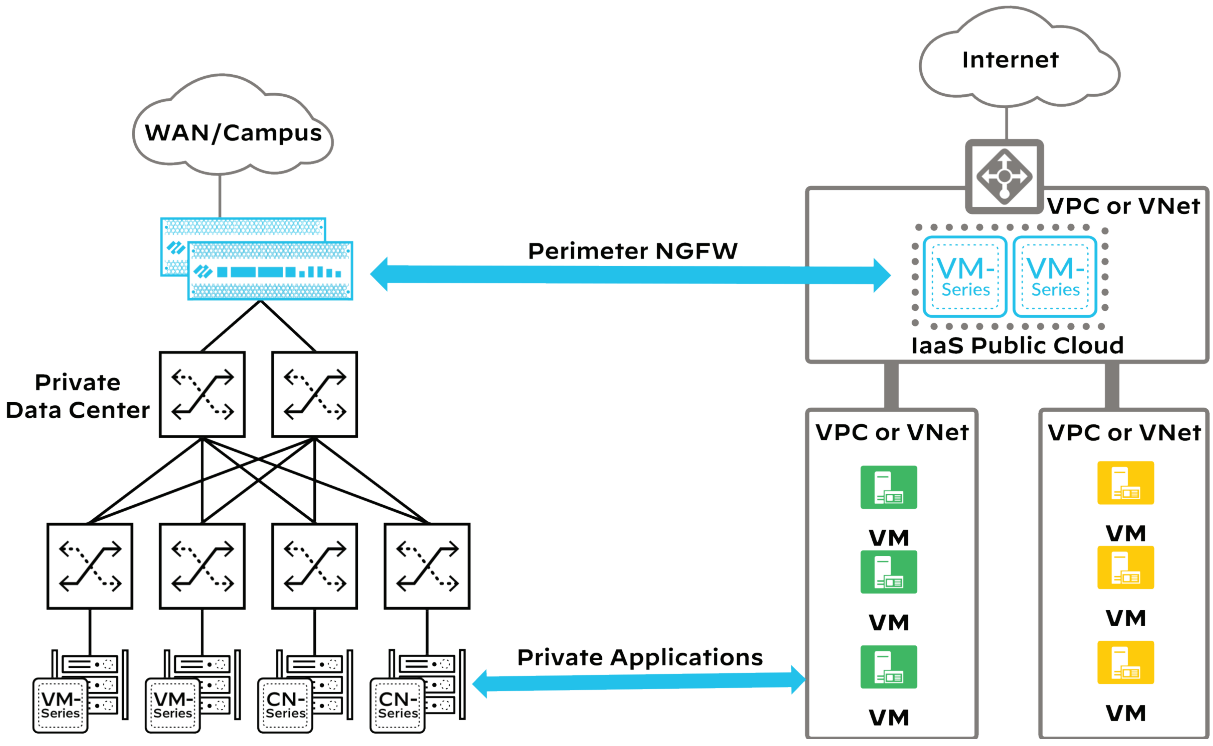
Once the user identity and device integrity have been verified, an application of a least-privileged access policy ensures access control to critical data and applications. Users should have access to the data and applications they need in order to perform their tasks but nothing more. Zero Trust policies should be enforced as close as possible to the protect surface.

Private Applications

In a private data center, because the users are located outside of the data center, user-to-application traffic is considered north-south traffic. A perimeter firewall at the edge of the data center is the optimal location to enable least-privileged access policies. A single point of enforcement at the perimeter of the data center has inline visibility into all users trying to access resources inside the data center.

IaaS public clouds—such as AWS, Google Cloud Platform (GCP), Microsoft Azure, Oracle Cloud Infrastructure, and Alibaba Cloud—provide servers, storage, and networking capabilities to host private applications. In IaaS, it is a best practice to implement a perimeter VM based NGFW in a virtual network such as an Amazon or Google VPC or AWS VNet. To restrict user access to data and applications in IaaS public clouds, the optimal location for enabling least-privileged policies is a virtual network perimeter that connects other application-hosting VPCs or VNETs.

Figure 24 Private data center and IaaS perimeter



Visibility into Encrypted Traffic

Most applications encrypt the traffic between client and server. The NGFW must decrypt encrypted traffic in order to have granular visibility and control for the application and to detect threats. The NGFW and Prisma Access can decrypt and inspect both inbound and outbound SSL/TLS connections traversing the NGFW. To ensure privacy during transport, the inspected traffic is re-encrypted on egress.

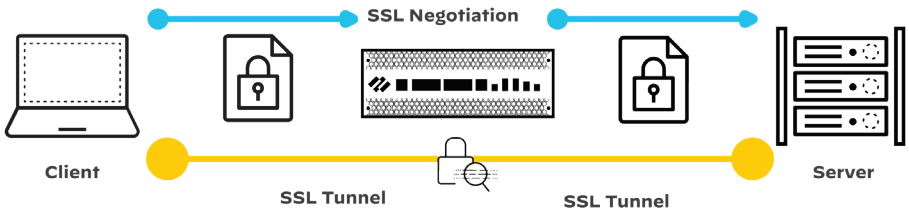
SSL Inbound Inspection decrypts and inspects traffic from internal endpoints to private applications in the data center and public cloud. SSL Inbound Inspection requires you to import the server certificate and private key for the servers that you are protecting into the NGFW. Because the NGFW has the server certificate and private key, it can transparently decrypt the SSL session between the server and the client. After traffic is decrypted, the NGFW can classify applications and apply security policies to the decrypted traffic in order to detect malicious content and control the applications running over the secure channel.

With SSL Forward Proxy, the NGFW decrypts outbound traffic (from internal users and servers in the data center and public cloud), which is encrypted using SSL/TLS. To secure the connection, SSL uses certificates to establish trust between the client and server. Certificates that are signed by a trusted certificate authority are installed on the NGFW to establish it as a trusted third party to your internal clients during the connection setup. Your public key infrastructure must trust the NGFW SSL certificate, or you must deploy the SSL certificate on each client participating in sessions decrypted by the NGFW.

The NGFW preserves the integrity of the SSL/TLS session by using the cryptographic settings of the original SSL/TLS negotiation as mandated by the client and the server. It does not change the

cryptographic parameters after the session has been negotiated. Further, to reduce risks associated with older versions of the protocols, PAN-OS allows you to specify the supported SSL/TLS protocol versions and cipher suites. Certificate Revocation List/Online Certificate Status protocol checks ensure that certificates presented during SSL decryption are valid.

Figure 25 SSL Forward Proxy



A decryption policy rule allows you to define traffic that you want the NGFW to decrypt and to define traffic that is excluded from decryption. Excluding traffic from decryption is typically done for applications that contain sensitive data and/or data that is restricted by local regulations.

SaaS Applications

With the emergence of the hybrid workforce, organizations have rapidly increased their SaaS application adoption in order to maintain and accelerate employee productivity anywhere they work. Organizations do not have access to the infrastructure or the application; you cannot place a perimeter NGFW inside the SaaS-provider cloud. Users can access the applications from managed and unmanaged devices and from both within and outside of the corporate network, bypassing standard inline protections and controls that are available at the corporate network.

Natively integrated with the Palo Alto Networks NGFW platform (cloud-based, virtual, and hardware form factors), next-generation CASB delivers granular visibility and control of SaaS applications, their use within your organization, and their risks. A full and complete view into shadow IT risks enables security teams to intelligently keep up with their growth and prevent unsanctioned apps from becoming another conduit of data loss. This solution integrates with other cloud-delivered security services, such as DLP for scanning files for sensitive information and WildFire for preventing known and unknown threats from spreading through sanctioned SaaS applications.

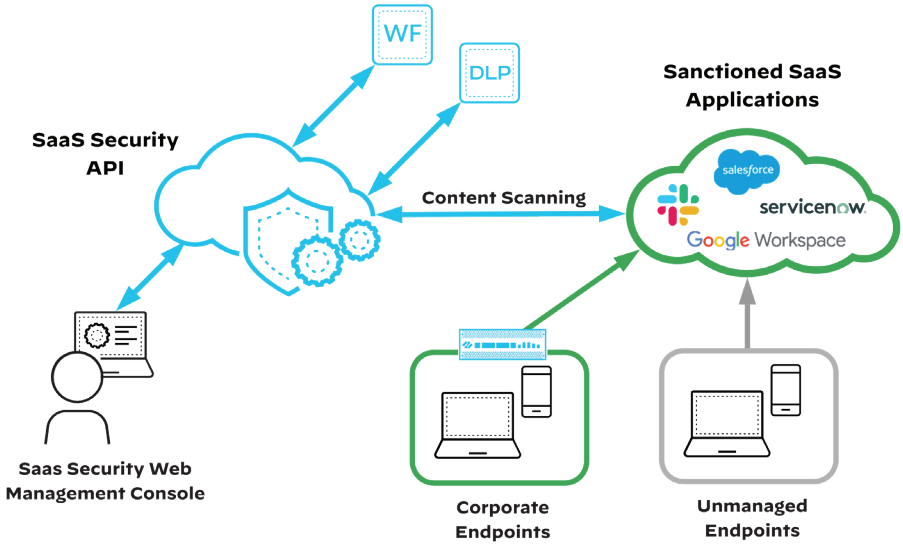
Two components of the solution help you establish least-privileged user access to SaaS applications:

- **SaaS Security Inline**—SaaS Security Inline provides visibility and control of all SaaS application use from your corporate network and managed endpoints. It can tell you what SaaS applications are being accessed, who is accessing them, and the risks associated with those applications. Furthermore, it enables you to create and distribute granular policy to control or block access to those applications.
- **SaaS Security API**—SaaS Security API secures sanctioned SaaS applications. Without any configuration on endpoints, it provides complete visibility across all users, folders, and activity within a sanctioned SaaS application, and it enables detailed analysis and analytics of application use to prevent data risk and compliance violations. More importantly, SaaS Security API allows granular, context-aware policy control within these SaaS applications in order to drive enforcement and quarantine users and data as soon as a violation occurs.

API-Based Data Visibility and Control

The SaaS Security API provides security for data-at-rest in your sanctioned applications. When you first connect a sanctioned SaaS application to SaaS Security, the application's API allows SaaS Security to discover and retroactively inspect all files and data (called *assets* in SaaS Security) managed by the application. SaaS Security inspects and analyzes all assets and identifies exposures, external collaborators, risky user behavior, and sensitive documents, as well as identifying the potential risks associated with each asset. The service also performs deep content-inspection and protects both historical assets and new assets from malware, data exposure, and data exfiltration in near real-time. SaaS Security leverages DLP (to categorize sensitive and regulated data) and the WildFire malware analysis engine (to identify and protect against all file-based threats).

Figure 26 SaaS Security API integration with SaaS applications



As SaaS Security identifies incidents, you can assess them and define automated actions that remediate the incidents or alert users and administrators to the risks. For ongoing incident assessment and protection, in addition to the initial inspection of historical assets, SaaS Security continuously monitors the SaaS application and applies the policy to new or modified assets.

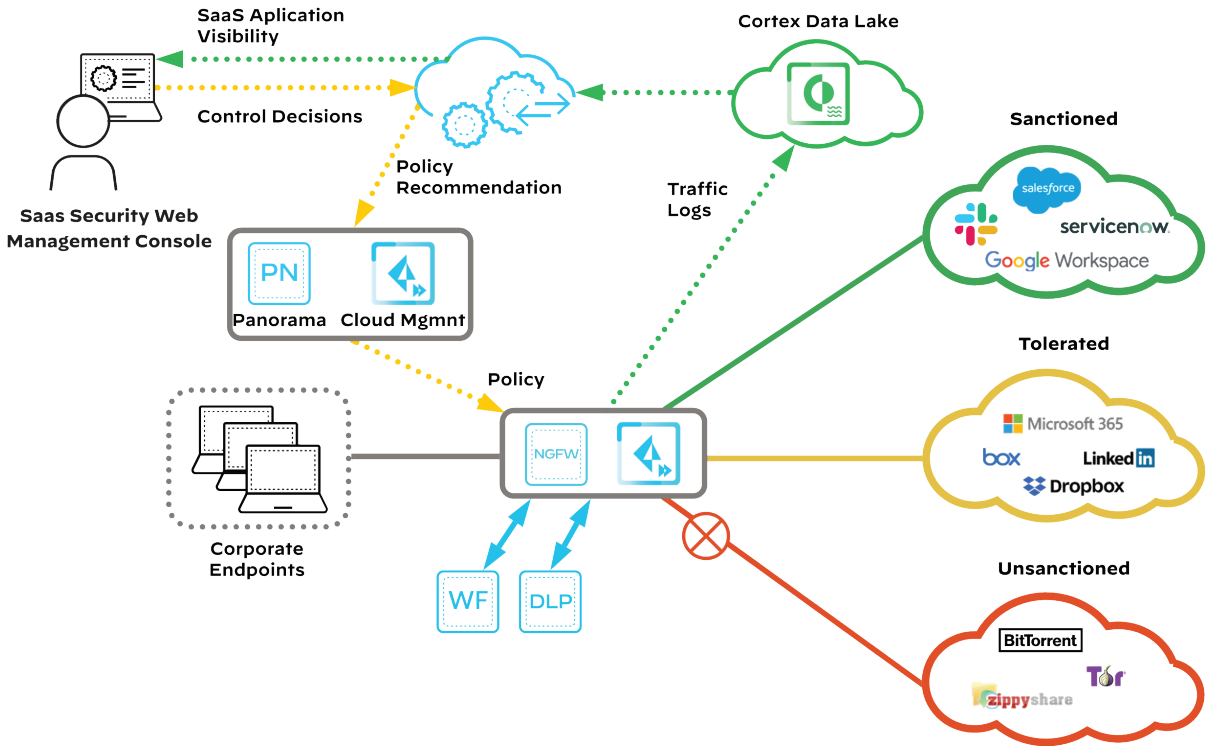
Inline Application Visibility and Control

SaaS Security Inline provides visibility into the use of SaaS applications on the network and the ability to control users' access to those applications. Prisma Access and the next-generation firewalls send network traffic logs to Cortex Data Lake. SaaS Security Inline examines those logs for SaaS application use.

Key to both visibility and control is App-ID functionality. By inspecting the session and payload information of the traffic traversing the NGFW, App-ID identifies applications and granular application functionality. To retrieve SaaS application information, SaaS Security Inline uses the App-ID Cloud Engine (ACE). ACE contains over 15,000 SaaS application IDs and is adding to the list daily. ACE uses machine learning and crowdsourcing in order to identify new SaaS applications as they become available.

SaaS Security Inline also provides risk information for each application in the ACE. When defining policy, you can use this information to decide which applications to allow without restrictions and which applications to block without exception. You can also conditionally allow access to specific application functions, such as file upload, download, or sharing, depending upon the application. You can also limit who has access to an application or application-function based on user and group information. SaaS Security Inline facilitates creating granular policies based on App-ID, User-ID, and DLP content profiles and deploying those policies to your next-generation firewalls.

Figure 27 SaaS Security Inline



Secure User Transactions

In Zero Trust, authentication and authorization are critical, not just in the initial connection but also at every stage of the digital interaction. To achieve Zero Trust and protect against malicious activity, scanning of all transactions is required.

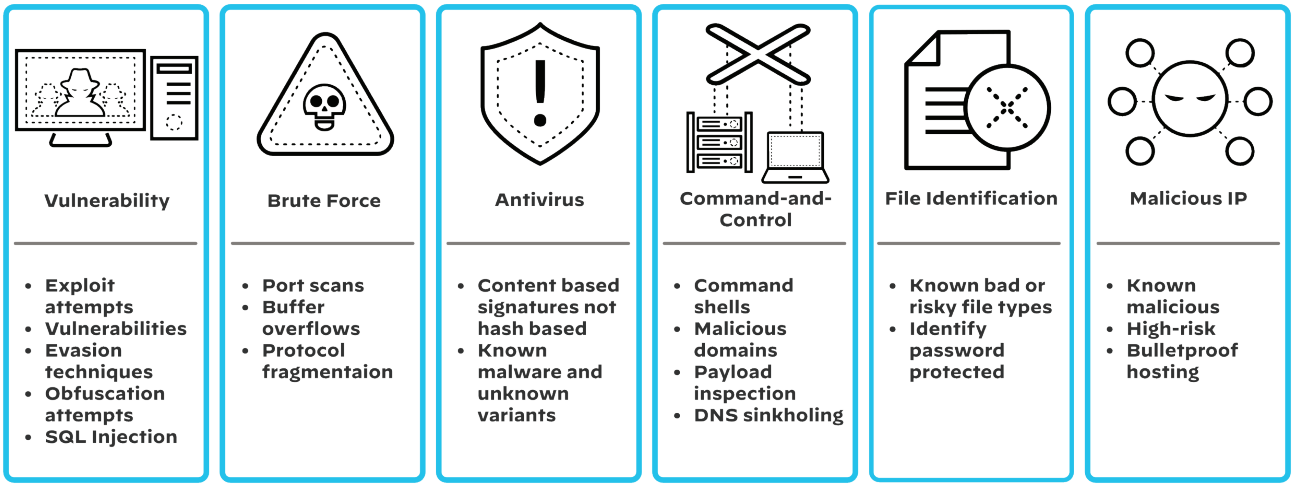
As described in the "Cloud-Delivered Security Services" section, NGFW and Prisma Access support a set of subscriptions that coordinate intelligence and provide protections across all attack vectors, providing best-in-class functionality while eliminating the coverage gaps that disparate network security tools create. These security subscriptions allow you to prevent threats, detect malware, and protect data.

NGFW Threat Prevention

Using cloud-delivered security services, the NGFW and Prisma Access prevent threats in traffic allowed by Zero Trust least-privileged access policy. Traffic inspection identifies malware, vulnerabilities, data exfiltration, and previously identified threats by using the following subscriptions:

- **Antivirus and WildFire**—In a security policy, the NGFW inspects traffic for known malware types.
- **Anti-spyware**—Anti-spyware profiles prevent infected endpoints from sending malicious traffic to C2 systems.
- **URL Filtering**—In a Zero Trust security policy, URL filtering blocks C2 traffic and access to malicious websites.
- **DNS Security**—Secures all DNS traffic, including unexpected DNS resolvers, malicious and suspicious domains, and DNS rebinding and tunneling attacks.
- **Data Loss Prevention**—The Enterprise DLP engine inspects files for common patterns of sensitive data, such as social security numbers, credit card numbers, and so on.
- **File blocking**—Blocks files that are known to carry threats.

Figure 28 NGFW threat protections



Transaction Security for SaaS

SaaS Security API allows you to define policies that automatically assess risk. There are two types of policies:

- **Content policies**—Visibility into the assets' content allows you to ensure that information stored in the application is appropriate. It also allows you to secure content that is critical to the organization, sensitive, or subject to compliance based on its exposure-level categorization: internal, company, external, or public.
- **Activity policies**—Assessing the risk of asset-related activities helps identify abnormal user behavior. Activity policies identify where excessive activity, such as downloading or exporting data, might indicate abnormal movement of data out of the SaaS application or other compliance violations.

You can remediate incidents yourself or send notifications to owners of the identified files and folders, requesting they fix the problems. When incidents require remediation from the owner, you can email the user and educate them about SaaS application acceptable-use policies.

You can also use automatic remediation in order to address incidents across large numbers of assets that SaaS Security identifies. Auto-remediation provides four automatic remediation actions:

- **Quarantine**—If an asset poses an immediate threat to intellectual property or proprietary data, you can automatically move the compromised asset to a quarantine folder. Depending on the SaaS application, that quarantine folder can either be in the asset owner's root directory or a special admin quarantine folder that only admin users can access. When you quarantine an asset, SaaS Security replaces the original asset with a placeholder (tombstone) file. The placeholder is a customizable, plain-text file that contains a simple description explaining that SaaS Security quarantined the asset. Also, when SaaS Security automatically quarantines an asset, you can send the asset owner a Remediation Digest email that describes the changes made.
- **Change sharing**—You can automatically change sharing to remove public links from an asset. You have the option to remove either the direct link on the asset only or the links from parent folders that expose the asset due to inheritance. When SaaS Security automatically changes an asset's sharing settings, you can send the asset owner a Remediation Digest email that describes the changes SaaS Security made.
- **Notification**—Instead of automatically fixing the issue, send the asset owner a Remediation Digest email that describes what actions the owner can take to remediate the risk (recommended actions).
- **Log**—Identify potential risks but take no further action. After you uncover specific issues that are high-compliance risks on the network, you can modify the rule or add a new rule to remediate the risk automatically or notify the owner.

Auto-remediation can be a valuable tool in resolving data-governance risks, but you should use it carefully. Quarantining assets and changing sharing attributes can have a significant impact on a user's experience and might affect their productivity. For example, if you change the sharing settings on a parent folder in order to remove a file from being accessible publicly, it affects all files in that folder, not just those with content in violation of policy. Use these two remediation methods very deliberately, and in many cases, you should use them only on new assets stored in SaaS applications.

Zero Trust for Applications

The Zero Trust for applications approach allows you to secure applications by removing all implicit trust and securing all transactions between workloads. As shown in Table 1, the Zero Trust for applications security controls are:

- **Identity**—Validate developers, DevOps, and admins with strong authentication.
- **Device/Workload**—Verify workload integrity.
- **Access**—Enforce least-privileged access for workloads accessing other workloads.
- **Transaction**—Scan all content for malicious activity and data theft.

User Identity Validation

Developers, DevOps, and admin users who require access to application-development environments and private- and public-cloud infrastructure require strong authentication. To protect against stolen credentials, you should use MFA against all critical assets. As discussed in the "User Identity Validation" section, perimeter NGFWs safely enable applications and control access based on users or groups of users.

You should also validate the entitlements granted to the developers, DevOps, and admins who seek access to cloud infrastructure. Prisma Cloud continuously detects and automatically remediates identity and access risks across infrastructure-as-a-service and platform-as-a-service offerings. It discovers all human and machine identities across cloud environments and then analyzes entitlements, roles, and policies. Prisma Cloud scans all audit and flow logs across multi-cloud environments for root user and overly permissive administrator activities.

Workload Integrity

Verification of workload integrity requires securing all hosts, containers, and serverless functions across private and public clouds. Securing workloads effectively requires enforcing cybersecurity checks across the entire application development cycle.

The Prisma Cloud Compute (PCC) module is the cloud workload protection platform that provides a comprehensive view into every host, container and serverless function. PCC module provides vulnerability management, compliance checks, runtime security, network visibility and access control for cloud workloads. Prisma Cloud Web Application and API Security auto-discovers unprotected web apps and APIs with full coverage across OWASP Top 10 threats in any public or private cloud.

You easily integrate PCC into your source-code development, container build process, and runtime deployment, providing security and compliance capabilities at each stage. Security teams can set policies that allow only compliant and fully remediated images to progress down the DevOps pipeline. Upon deployment, PCC immediately begins working to secure your workloads.

Prisma Cloud Compute is composed of two components: a console and an agent. You access the PCC console from within the Prisma Cloud Enterprise Edition console. The PCC console provides a centralized dashboard where you define policy and monitor your environment. To help prioritize risks in real time, the PCC console shows key metrics, including vulnerability status, remediation guidance, and real-time alerts.

Defender is the PCC agent component that runs on each host. Defender enforces the policies defined in the console and sends event data to the console for correlation. Defender is completely containerized and uses a least-privileged security design.

To support the full variety of workloads in cloud-native environments, there are three types of Defenders. Depending on the assets in your environment that require protection, you might deploy one or more of the following:

- **Container Defender**—You deploy this type of Defender as a container on every asset running containers in your infrastructure.
- **Host Defender**—You deploy this type of Defender on VMs that do not run containers.
- **Serverless Defender**—You deploy this type of Defender as part of your serverless functions, and it provides runtime-application self-protection capabilities.

In addition to agent-based scanning, Prisma Cloud also offers agentless scanning capabilities. Agentless scanning helps users inspect the risks and vulnerabilities of a virtual machine without having to install an agent or affecting the execution of the instance. Ensuring that you have complete coverage requires a combined approach. For example, some workloads in your environment could be web-server applications that require web-application and API-protection capabilities in order to alert and prevent attacks like DoS requests, SQL injection, or cross-site scripting from affecting your service. To continuously monitor requests and block such attacks from affecting your application workloads, you need agents. Whereas if you have managed clusters or environments where you don't directly access host kernels or need blocking capabilities, agentless scanning provides deep insights, leveraging cloud service provider API calls and providing visibility into unpatched vulnerabilities or exposed risks.

Least-Privileged Workload Access

In Zero Trust for applications, workloads accessing other workloads should mutually verify identity and apply least-privileged connectivity for the application. To prevent lateral movement of malware and ensure least-privileged connectivity, you should enforce microsegmentation.

Network segmentation is a technique used to isolate application environments for security and connectivity reasons. Network segments are implemented using different virtual networking technologies like VLANs, VNETs, VPCs, VXLAN, and virtual routing functions (VRFs). Examples of typical applications for network

segments include critical applications, storage networks, development environments, DMZs, extranets, IoT applications, multi-tenancy, regulatory compliance, and more. Although network segmentation provides coarse-grained separation between application groups, very granular segmentation can be challenging because of the dependency to the network infrastructure design. Adding many VLANs, IP subnets, and VRFs creates scalability problems for the network infrastructure and operations.

Microsegmentation is a fine-grained application segmentation method that is decoupled from the network infrastructure design. This allows for a much higher degree of isolation and is ideal for ensuring least-privileged workload access. Prisma Cloud Identity-Based Microsegmentation and the CN-Series NGFWs support capabilities for enabling microsegmentation at the container level. The combination of both network segmentation and microsegmentation provides coarse-grained isolation of similar applications across your entire environment and fine-grained, identity-based microsegmentation that prevents lateral attacks for hosts and containers.

Network Segmentation

Securing applications and services depends upon the NGFW's ability to have visibility and control of the inbound traffic to the application, outbound traffic from the application, and traffic between applications' components. To provide the required visibility and control, you should segment data and applications in the private data center and public-cloud provider behind a next-generation firewall.

The data-sensitivity level of the application informs you on how to group applications and services with common security and traffic-flow requirements. When the data sensitivity increases, additional policies and protections are necessary, including a stricter definition of what is permitted to access the application. You should not group an application or service that is at the highest level of sensitivity with any other application. You should even separate high-sensitivity services from other components of their application if those other components have a reduced security requirement.

The sensitivity levels are:

- **Low**—Applications and information whose loss of availability would have limited impact on the organization or its customers.
- **Moderate**—Infrastructure, applications, and systems whose loss of integrity and availability would impact the organization or its customers.
- **High**—Any information falling under statutory requirements for notification in the case of a breach.

How you create the network segments for an application depends upon the infrastructure on which it is built. The Palo Alto Networks portfolio allows segmentation in a variety of locations within the data center:

- **Data center**—The PA-Series and VM-Series are ML powered NGFWs. The PA-Series are physical appliances that you typically deploy at the data center perimeter. The VM-Series are virtualized form-factor, ML-powered next-generation firewalls that you typically deploy within the data center, providing a more granular layer of segmentation.
- **Public cloud**—The VM-Series are virtualized form-factor, ML-powered NGFWs. You deploy these in a variety of public, private, and hybrid cloud environments. The VM-Series images are often available from the public-cloud service-provider stores.
- **Containers**—Palo Alto Networks provides two methods for segmenting workloads within Kubernetes clusters: the CN-Series NGFW and Prisma Cloud Identity-Based Microsegmentation. The CN-Series are containerized form-factor NGFWs. They provide advanced Layer 7 network security and threat protection. In Kubernetes clusters, Prisma Cloud Identity-Based Microsegmentation gives you the ability to provide segmentation based on the individual workload identity instead of on IP addresses.

To define the source and destination networks for securing traffic flows, the NGFW uses zones and dynamic address groups. Zones are used in static environments, and dynamic address groups allow the security policy to stay in-sync with dynamic virtual environments both in the data center and public cloud.

App-ID identifies the applications in the traffic between network segments and enables the NGFW to limit the communication between network segments to specific applications. Because the Zero Trust security policy in the data center denies all traffic between segments, use App-ID to explicitly define the inter-segment traffic that is required for the applications to function and administrators to manage the applications.

Prisma Cloud Identity-Based Microsegmentation

Identity-Based Microsegmentation in Prisma Cloud gives your organization the ability to base security policies on strong, machine-generated persistent identity for individual workloads instead of broad IP addresses. This reduces the attack surface of private- and public-cloud networks to the individual workload level and makes it possible to track a workload as it moves through your environment, even if IP addresses and other traditional identifiers change.

Prisma Cloud identity-based microsegmentation assigns a cryptographically signed workload identity to every protected host and container across your cloud environments. This makes it possible to track a workload as it moves through your environment, even if IP addresses and other traditional identifiers change. Each identity consists of contextual attributes, including metadata from cloud-native services across AWS, Microsoft Azure, GCP, Kubernetes, and more. You can use these attributes to create and manage microsegmentation policies for your cloud applications.

Prisma Cloud provides the ability to enforce security policies on applications deployed on Linux hosts, Windows hosts, or Kubernetes infrastructure and to gain end-to-end visibility of ingress, egress, and pod-to-pod communications. Identity-Based Microsegmentation is an add-on license for Prisma Cloud that enables you to:

- Decouple security from the network by assigning every workload a cryptographic identity. The identity (which is derived using metadata from AWS, GCP, Azure, Kubernetes, and other application contexts) becomes the perimeter instead of an IP address.
- Discover applications and learn the communication patterns both inside and across clouds. Prisma Cloud then maps this information in real-time with workload identity context, not the IP address and port.
- Enable centralized policy management for endpoints that are distributed. Policies can be auto generated for you, or you can choose a more declarative approach to defining and testing segmentation policies without impacting runtime.
- Use workload identity to authenticate and authorize each connection request. The ability to control communications between workloads enables you to segment applications and implement Zero Trust.

Identity-Based Microsegmentation contains two components, a console and agent. The Microsegmentation console manages all microsegmentation resources, and you can access it from within the Prisma Cloud Enterprise Edition console, through a CLI interface or REST API. An Enforcer is the agent component that runs on each host. Enforcers run as a service on a virtual machine or as a DaemonSet on node servers in a Kubernetes cluster. To retrieve network rulesets and to send flow and DNS resolution logs, Enforcers connect to the Microsegmentation console API.

CN-Series NGFW for Kubernetes Clusters

You can deploy PA-Series and VM-Series NGFWs only at the edge of a Kubernetes environment and therefore cannot determine the specific pod where traffic originates. To overcome this challenge, you deploy CN-Series containerized next-generation firewalls on node servers within a Kubernetes cluster, giving the firewalls precise visibility into pod traffic. The CN-Series NGFW allows network security teams to manage network security and threat prevention policies for their Kubernetes environments in the same way as they manage their physical and cloud environments.

The CN-Series firewall uses Palo Alto Networks Panorama, a distributed PAN-OS architecture, and native Kubernetes constructs in the deployment. The core building blocks of the distributed PAN-OS architecture are the CN-MGMT pods (management plane), the CN-NGFW pods (data plane), and the PAN-CNI network plugin. To enable better runtime protection for applications and to support a smaller resource footprint, the CN-MGMT and CN-NGFW components of the containerized firewall are implemented separately.

Secure Data and Transactions

In the Zero Trust approach, authentication and authorization are critical, not just in the initial connection but at every stage of the digital interaction. To achieve Zero Trust, scanning of all transactions is required to protect against malicious activity.

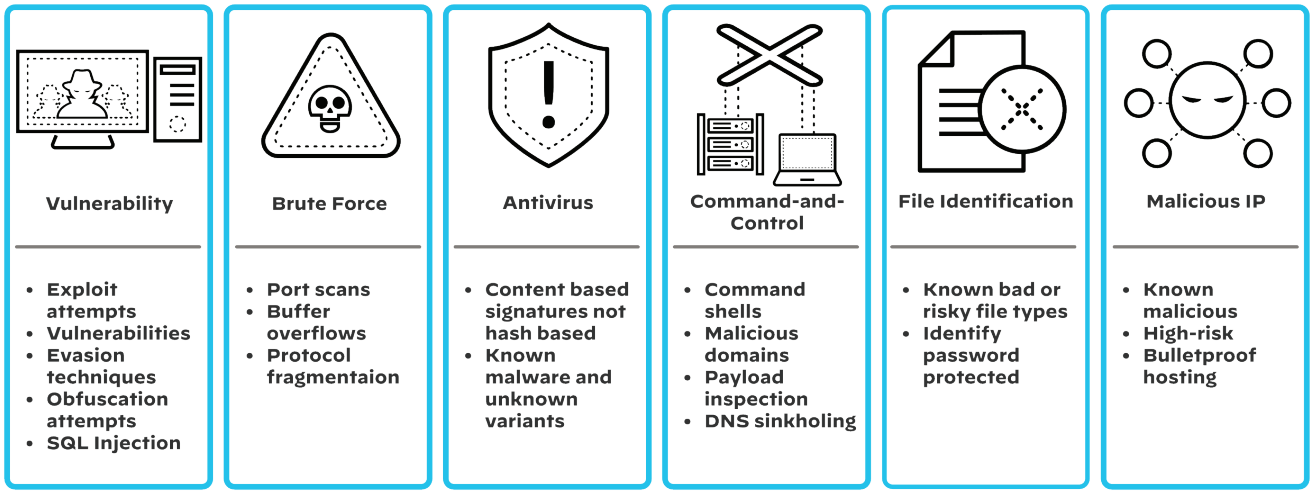
As described earlier, NGFW supports a set of cloud-delivered security subscriptions, which coordinate intelligence and provide protections across all attack vectors, providing best-in-class functionality while eliminating the coverage gaps that disparate network-security tools create. These security subscriptions allow you to prevent threats, detect malware, and protect data.

NGFW Threat Prevention

Using cloud-delivered security services, the NGFW and Prisma Access prevent threats in traffic allowed by Zero Trust least-privileged access policy. Traffic inspection identifies malware, vulnerabilities, data exfiltration, and previously identified threats by using the following subscriptions:

- **Antivirus and WildFire**—In a security policy, the NGFW inspects traffic for known malware types.
- **Anti-spyware**—Anti-spyware profiles prevent infected endpoints from sending malicious traffic to C2 systems.
- **URL Filtering**—In a Zero Trust security policy, URL filtering blocks C2 traffic and access to malicious websites.
- **DNS Security**—Secures all DNS traffic, including unexpected DNS resolvers, malicious and suspicious domains, and DNS rebinding and tunneling attacks.
- **Data Loss Prevention**—The Enterprise DLP engine inspects files for common patterns of sensitive data, such as social security numbers, credit card numbers, and so on.
- **File blocking**—Blocks files that are known to carry threats.

Figure 29 NGFW threat protections



Prisma Cloud Defender

Although the NGFW configured with a Zero Trust security policy prevents most attacks from ever reaching your servers, new never-before-seen attacks can slip through before being identified by the cloud-delivered security services. Prisma Cloud Compute prevents malware and exploits from executing when installed on the servers in your private data center and public cloud IaaS environments.

Prisma Cloud Compute Defender stops threats on the server and within container images, and it minimizes server infections by blocking exploits, malware, and ransomware. Like with the NGFW, you can use a Zero Trust policy to extend the least-privileged access model to protecting the server operating system. Some of the capabilities critical to the least-privileged policy include:

- **Vulnerability Explorer**—Because attackers most often target application vulnerabilities when attempting to compromise servers, Prisma Cloud Vulnerability Explorer is key to extending the Zero Trust security model to servers and blocking the core techniques used by Zero Day exploits.
- **Compliance Explorer**—Prisma Cloud helps enterprises monitor and enforce compliance for hosts, containers, and serverless environments. To understand and assess standard configurations and security best practices in your environment, use the compliance-management system.
- **Runtime defense**—Predictive protection for containers and threat-based active protection for running containers, hosts, and serverless functions provide additional security for the workloads running on your servers. Predictive protection includes capabilities like determining when a container runs a process not included in the original image or when the container creates an unexpected network socket. Threat-based protection includes capabilities like detecting when malware is added to a workload or when a workload connects to a botnet.
- **Continuous integration**—You need to find and fix problems before they enter production. Prisma Cloud's continuous integration plugins surface vulnerability and compliance issues directly in the build tool every time developers build their container images and serverless functions. Security teams can set policies that allow only compliant and fully remediated images to progress down the pipeline.
- **Malware scanning**—Ransomware, script-based attacks, and malicious executables must be prevented from executing on the server.

Zero Trust for Infrastructure

The Zero Trust for infrastructure approach allows you to secure critical infrastructure by removing all implicit trust and verifying all digital transactions. As described in the Five-Step Methodology section, the first step for implementing Zero Trust is to identify the protect surface. When identifying the protect surface, you should assess the business impact to your organization if your infrastructure is compromised. Critical infrastructure can be anywhere in your network. In Zero Trust for infrastructure, the placement of security controls is critical to ensuring the security of your infrastructure and business continuity.

As shown in Table 1, the Zero Trust for infrastructure security controls are:

- **Identity**—Validate all users with access to infrastructure.
- **Device/Workload**—Identify all devices, including IoT devices.
- **Access**—Enforce least-privileged access segmentation for native and third-party infrastructure.
- **Transaction**—Scan all content within the infrastructure for malicious activity and data theft.

User Identity Validation

Only IT admins usually require access to network infrastructure like routers, switches, and wireless equipment. In some cases, third party vendor access is required in order to support physical security systems and building management controls. You should implement role-based access control and MFA in order to ensure that only the right groups of users have access to critical infrastructure assets and to protect against stolen credentials. As discussed in the "User Identity Validation" section, perimeter NGFWs safely enable applications and control access based on users, or group of users.

Identify Devices

IoT devices are on the increase because many organizations are deploying large volumes into their environments for increasing productivity, helping with digital transformation, and providing operational efficiency. IT departments do not always know what devices are deployed and what vulnerabilities they introduce. Palo Alto Networks offers an IoT Security solution, delivered as a cloud service, that takes a life cycle approach to securing your IoT environment. Your existing NGFWs perform discovery, visibility, and enforcement tasks, versus other solutions that require you to buy separate probes in the network. The lifecycle approach consists of the following steps:

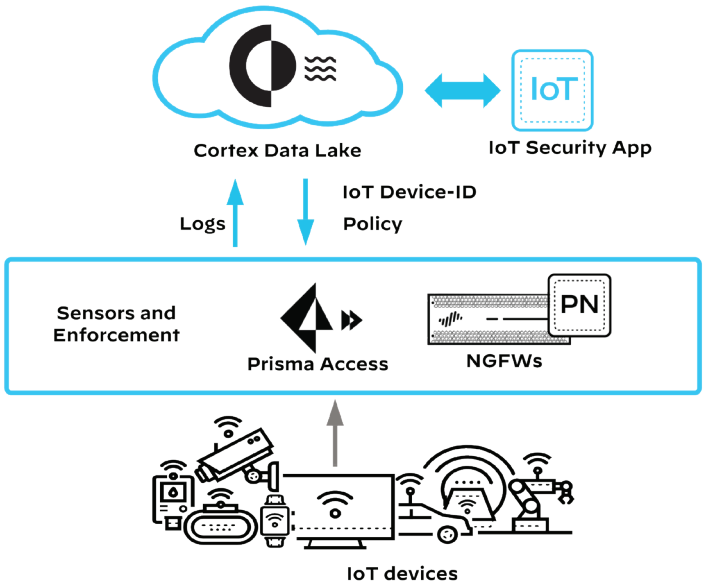
- **Understand IoT assets**—Provides full device discovery, including manufacturer, device type, software version, serial number, and multiple other attributes, as well as using with ML and device profiling for continuous ongoing detection and classification of all IoT and OT devices.
- **Assess IoT risks**—Evaluates the risks and vulnerabilities, using vendor information for specific patching or software updates that are required. In addition, IoT Security provides continuous ongoing and 100% passive risk assessment and automated risk-based policy recommendations.
- **Apply risk reduction policies**—Provides recommendations for policy enforcement based on risk and behaviors, using NGFW features (including App-ID, User-ID and Device-ID) to reduce the attack surface.
- **Prevent known threats**—Provides full detail of device context for alerts, protection from exploits, C2, spyware, malware, and other known threats through other available subscriptions on the NGFW, such as Threat Protection, WildFire, URL Filtering, and DNS Security.
- **Detect and respond to unknown threats**—Leverages ML with threat-modeling in order to detect threats, zero-day detection, incident response, isolation, and device quarantine.

The IoT Security solution is cloud delivered and provides complete visibility, in-depth risk analysis, and automated enforcement with the NGFW/Prisma Access. The components of the IoT solution consist of an IoT Security app residing on the Palo Alto Networks hub, data storage, and log retention, as well as an IoT security subscription.

The NGFWs and Prisma Access nodes behave as sensors and generate enhanced application logs, which they send to Cortex Data Lake, where the IoT Security app leverages this data. The IoT Security app analyses the data, provides IP-address-to-device mappings, and creates recommendations for policy rules to implement. From the IoT Security app, you can create security policy rules, which you can then import to the NGFWs or Panorama for enforcement.

The IoT subscription add-on security offering from Palo Alto Networks is easy to deploy. Because the add-on leverages your existing NGFW, you do not need any extra infrastructure. To reduce risk and secure your environment, the add-on provides complete IoT security, quickly discovers all your devices, and understands the full device context.

Figure 30 IoT security solution



Secure Infrastructure Access

Device-ID is derived from the classification engine in the IoT Security subscription and provides the capability to define policy rules that are based on a device, regardless of changes to its IP address or location. Some examples include limiting authorized users to accessing data from approved devices, limiting access to industrial control mechanisms to specific user groups, and controlling traffic from specific types of devices, such as security cameras, to prevent lateral attacks or data exfiltration from vulnerable IoT endpoints.

IoT devices and users are often deployed on the same network segments. You should segment IoT devices that pose significant security risk to the organization into their own network. This reduces the risk that they can be compromised or compromise devices in other segments.

Secure Transactions

IoT devices are usually headless. Most do not provide the capability to authenticate or run any endpoint-protection software. IT departments do not always know what devices are deployed and what vulnerabilities these devices introduce. Because most IoT devices are connected to the internet, they present multiple security risks from unpatched and outdated software. The most frequent attacks are exploits (using long-known vulnerabilities) and password attacks (using default device passwords).

Unlike data center and public cloud, critical infrastructure can be anywhere in your network. Enforcing Zero Trust policies requires the NGFW be as close as possible to the protect surface in order to build a security perimeter around it. Critical infrastructure perimeter could be located inside your network, protecting intermediate distribution frames and building distribution frames. Operational technology systems like manufacturing, mining, utilities, and warehouses also hold critical infrastructure. Branch locations might also have to protect devices that that handle business-sensitive data, such as medical devices and point-of-sale systems. A compromise on critical infrastructure is as impacting as a security breach of business applications.

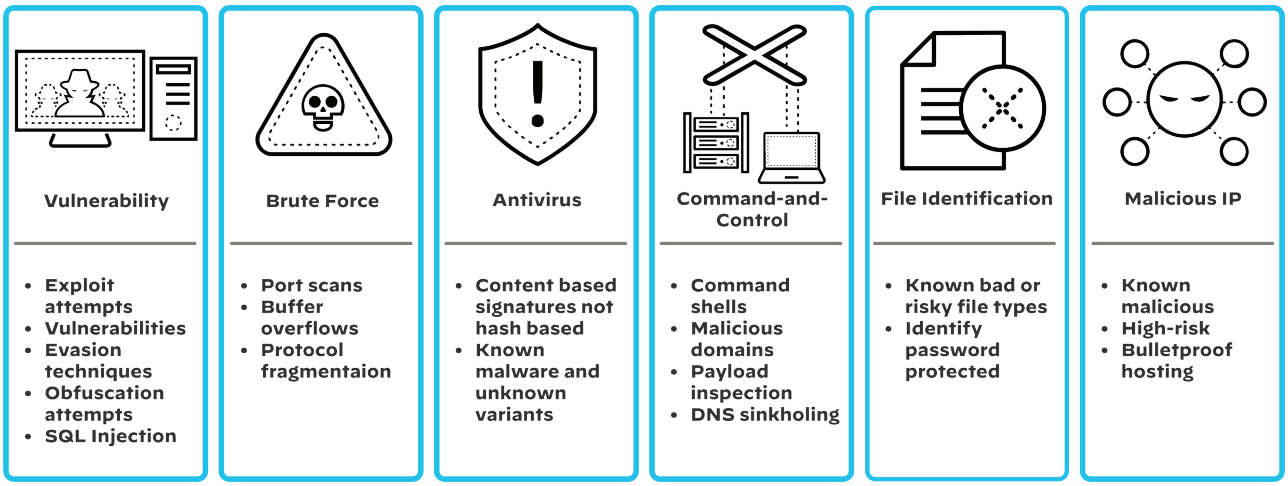
NGFW Threat Prevention

IoT Security enhances existing security subscriptions by providing device context (including content signatures for blocking known IoT malware), providing safe web access for IoT devices, and blocking DNS. For complete IoT security, you need to enable the security subscriptions in order to block these threats.

Using cloud-delivered security services, the NGFW and Prisma Access prevent threats in traffic allowed by Zero Trust least-privileged access policy. Traffic inspection identifies malware, vulnerabilities, data exfiltration, and previously identified threats by using the following subscriptions:

- **Antivirus and WildFire**—In a security policy, the NGFW inspects traffic for known malware types.
- **Anti-spyware**—Anti-spyware profiles prevent infected endpoints from sending malicious traffic to C2 systems.
- **URL Filtering**—In a Zero Trust security policy, URL filtering blocks C2 traffic and access to malicious websites.
- **DNS Security**—Secures all DNS traffic, including unexpected DNS resolvers, malicious and suspicious domains, and DNS rebinding and tunneling attacks.
- **Data Loss Prevention**—The Enterprise DLP engine inspects files for common patterns of sensitive data, such as social security numbers, credit card numbers, and so on.
- **File blocking**—Blocks files that are known to carry threats.

Figure 31 NGFW threat protections



Summary

Breaches and data loss have serious consequences for organizations and their customers. Zero Trust is based on the principle that no user, device, or transaction from inside or outside of the network can be trusted. The implicit trust in traditional security models is a vulnerability as dangerous as any other. The elimination of implicit trust promotes a consistent security policy across all situations. The Zero Trust framework focuses on resource protection and the premise that trust is never granted implicitly but must be continually evaluated.

The Palo Alto Networks Zero Trust Enterprise approach is a modern, platform-based security strategy—a strategic framework guiding security practices and procurement across an entire Enterprise. With the following Palo Alto Networks portfolio capabilities and functionality, you can implement an end-to-end Zero Trust model in your environment:

- NGFWs and cloud-native security products act as segmentation gateways. These products are available in a variety of form factors so that you can defend your protect surface wherever it is located, on-premises or in the cloud.
- App-ID, User-ID, and Device-ID provide reliable identification of your users, applications, and devices beyond traditional IP address or protocol/port identification. You can use these context-based security capabilities in order to create granular access-control policies that follow your users and devices as they move across your network.
- Prisma Cloud enables cloud security posture management, data security, and cloud workload protection, allowing comprehensive visibility and threat detection across your organization's hybrid, multi-cloud infrastructure.
- SaaS Security and Prisma Cloud Data Security inspect asset accessibility and risk through API integrations into the public cloud storage services and SaaS applications.
- Cortex XDR Agent advanced endpoint protection stops threats on the endpoint and coordinates enforcement with cloud and network security in order to prevent successful cyber-attacks.
- Cortex Data Lake serves as the central cloud-based repository for all security platform data and logs.

HEADQUARTERS

Palo Alto Networks
3000 Tannery Way
Santa Clara, CA 95054, USA
http://www.paloaltonetworks.com

Phone: +1 (408) 753-4000
Sales: +1 (866) 320-4788
Fax: +1 (408) 753-4001
info@paloaltonetworks.com

You can use the **feedback form** to send comments about this guide.

B-000300P-1-22a