Zero Trust

OVERVIEW

MAY 2020

paloalto® NETWORKS

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:

- Minor updates to reflect changes in product naming and branding.

Purpose of This Guide

This guide describes how your organization can use the Palo Alto Networks Strata, Prisma™, and Cortex™ platforms in the design of a Zero Trust security policy in order to protect your sensitive and critical data, applications, endpoints, and systems.

This guide provides architectural guidance on Zero Trust for solution architects and engineers who are familiar with the platforms. Use this guide as a roadmap for architectural discussions between Palo Alto Networks and your organization.

AUDIENCE

This design guide is written for technical readers, including system architects and design engineers, who want to deploy the Palo Alto Networks platforms in support of a Zero Trust security model. It assumes the reader is familiar with the basic concepts of applications, networking, virtualization, and security, as well as a basic understanding of network, data center, and public cloud architectures.

RELATED DOCUMENTATION

The following documents support this guide:

- Securing Data in the Private Data Center and Public Cloud with Zero Trust—Details integrating the Palo Alto Networks platforms into the data center and public cloud, features, and capabilities, as well as example Zero Trust policies.

Introduction

Stories of breaches and data loss that expose private information are in the news almost every week. When these events happen, there can be a significant personal impact on those who have information exposed, as well as a loss of trust in the companies and applications that were breached. It doesn't matter if the loss occurred because of accidental exposure or malicious act; the impact to an organization that has a breach or data loss event is real.

These events have become so common that at a rapid pace, industry is developing new standards and governments are developing new regulations that are forcing organizations to evaluate their security posture and do everything they can to prevent these events from occurring. This process can be challenging for several reasons:

- **Location of security infrastructure**—Security infrastructure requires visibility and control of relevant activities in order to prevent data breaches. Many organizations deploy security as a function at the internet perimeter. Although security at the internet perimeter is important, sensitive data can be widely dispersed within an organization, including in the public cloud, which may be accessible to users without traversing a security device.
- **Capabilities of security infrastructure**—Even when relevant activities are visible to the security infrastructure, many organizations have a security infrastructure that cannot adequately characterize relevant traffic and activities. The infrastructure must have the ability to characterize traffic and activities based on business-relevant attributes such as application (not port and protocol) and user and group (not IP address). It also must be able to identify and stop threats as they happen.
- **Security component coordination**—When threat and policy information is siloed in multiple security devices, coordinating a comprehensive security posture to protect against breach and data loss requires manual coordination and operation. This reduces the effectiveness of the security infrastructure because the threats evolve faster than security operations can manually coordinate the security infrastructure.
- **Security framework**—The biggest challenge for many organizations is defining a security model that provides the required security across the organization holistically. Most security architectures focus on the protection at a specific location and use policy based on what is known to be a risk.

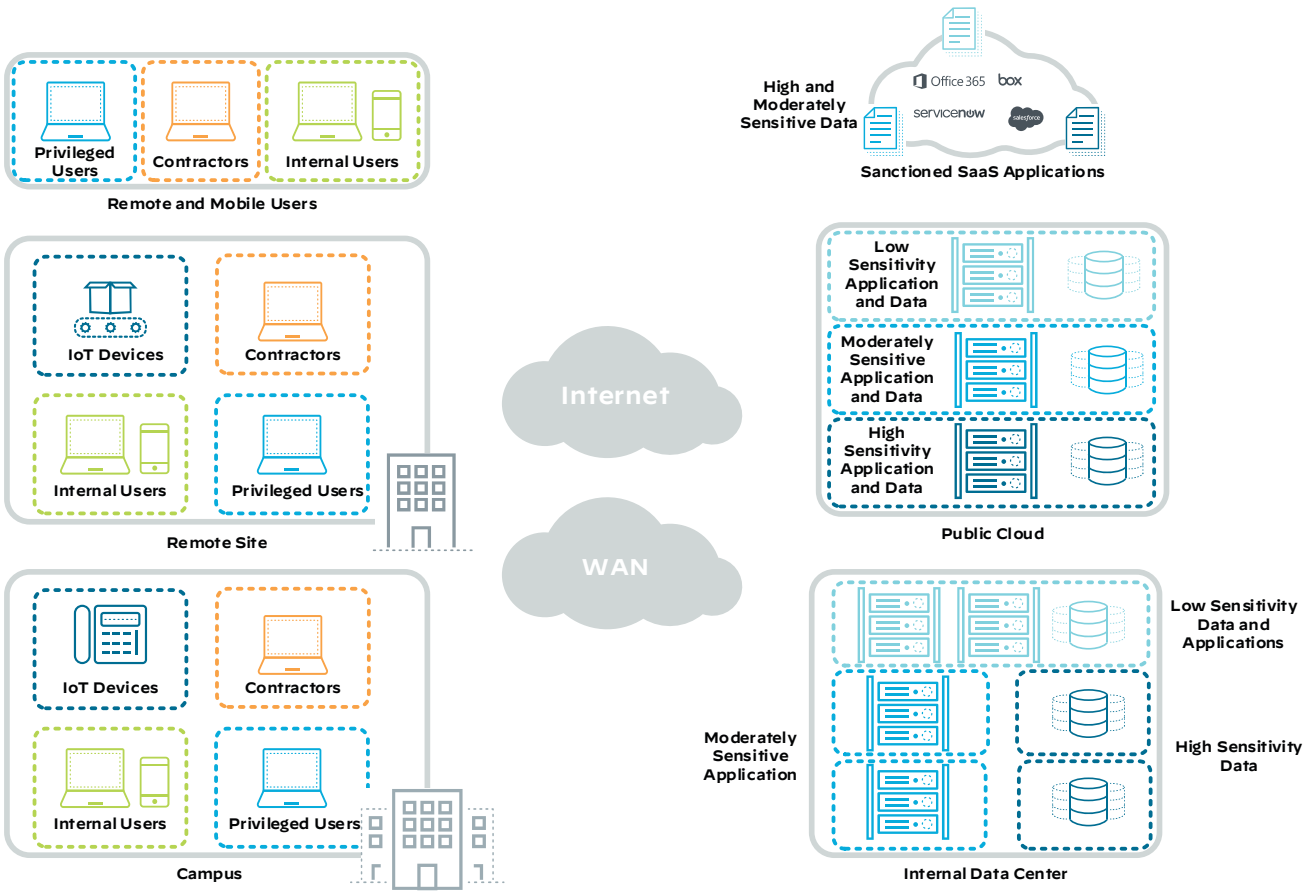
Zero Trust Concepts

Zero Trust is a security model developed specifically to address the security of sensitive data and critical applications in an enterprise organization. The primary goal of a Zero Trust security model is to prevent attackers and internal bad actors from successfully compromising the data, applications, and systems that are critical to an organization through exploits, malware, and credential- or user-based attacks.

Attacks on sensitive data rarely use a single exploit or compromised credential. Attackers use exploits, malware, compromised credentials, and other methods together to work their way from their beachhead in an organization to the data. Often attackers use one method after another repeatedly. Malware may provide a user's credentials, which in turn provides limited access to their organization's network. Network access allows the attacker to move around the private network and place additional malware on privileged devices that are closer to the data. Then the cycle repeats. The common thread in all these attack methods is that they take advantage of the trust inherent in the security posture of most enterprises. Trust has become a vulnerability as dangerous as any other.

Zero Trust remedies the deficiencies of the inherent trust in perimeter-centric security and the legacy devices and technologies used to implement them.

Figure 1 Zero Trust overview



Zero Trust policy leverages the capabilities of Palo Alto Networks automated platforms in order to prevent successful cyberattacks by moving away from inherent trust in the security policy to one based on least-privileged access. To accomplish Zero Trust the you must:

- **Segment sensitive data and critical data and applications**—To prevent successful cyber-attacks with Zero Trust, access to the critical data, applications, and systems must be visible to and controlled by the security platform.
- **Provide security on the endpoint**—To prevent attacks that aren't visible to the components of the security platform that protect a segment (through direct access to the endpoint or traffic that can't be decrypted), the security platform needs to have visibility and control in the endpoint operating system.
- **Detect sophisticated attacks through analysis**—To detect sophisticated attacks, the security platform must continuously analyze rich data collected from the components of the security platform that protect the segments and the endpoints. When analytics detect an attack, analytics must coordinate a response with the rest of the security platform.

SEGMENTING SENSITIVE DATA AND CRITICAL APPLICATIONS

Wherever the sensitive data is located to achieve Zero Trust, the security platform must have visibility and control. This is primarily achieved through segmentation of the data behind the security platform. The Forrester Research Zero Trust documentation refers to the segmentation of sensitive data as a micro core and perimeter (MCAP). The requirements of the security platform to implement MCAPs differ based on the data's location.

Categorizing Data and Application Sensitivity

A Zero Trust security model protects the data, applications, services, and systems whose loss, or loss of availability, would cause harm to an organization or its customers. So, before segmenting the data, applications, services, and systems, you must understand their sensitivity.

There are a significant number of industry standards and government regulations such as GDPR, HIPAA, and PCI that help define how data and applications are categorized, but because every organization's data requirements are unique and standards differ based on industry and location, this guide uses the following example levels of data sensitivity:

- **Public**—Public information or service
- **Low sensitivity**—Information that would cause limited harm to the organization. Examples include:
  - Non-critical data and applications with a limited user base.
- **Moderate sensitivity**—Information that has a risk of causing serious harm to the organization or its customers. Examples include:
  - Business data and applications including email and voice and video communications.
  - Infrastructure, applications, and systems whose loss of integrity and availability would impact the organization.
- **High sensitivity**—Information that will cause severe harm to the organization or its customers. Examples include:
  - Any information falling under statutory requirements for notification in the case of a breach.
  - Personally identifiable information (financial, health and legal).
  - Critical intellectual property (code, designs, etc.).
  - Critical infrastructure and systems whose loss of integrity and availability would severely harm the organization.
  - Public key infrastructure, Active Directory.

Zero Trust in the Network

The concept of least-privileged access forms the basis of a Zero Trust security policy. Zero Trust is mainly implemented on the network security components of the security platform. Least-privileged access is a positive control model that allows organizations to control interactions with resources based on an extensive range of business-relevant attributes, including the specific application and application function being used, user and group identity, and the specific types or pieces of data being accessed (e.g., credit card or Social Security numbers). The result is granular control that safely enables the right applications for the right sets of users while automatically eliminating unwanted, unauthorized and potentially harmful interactions.

To be able to prevent attacks that use exploits and malware the network security components of the security platform must:

- Have visibility of and control over all the traffic that is destined to the organization's sensitive data. This includes all applications and encrypted traffic.
- Identify previously unknown and prevent known vulnerability exploits, malware, spyware, and command and control traffic that malware uses once in place.
- Prevent access to malicious websites that attempt to phish user credentials.

Additionally, to implement the Zero Trust least-privileged security model in the network, the security platform must:

- Have visibility of and control over the application and application functionality in the traffic. Traditional security infrastructure describes applications through port and protocol. Zero Trust's least-privileged access model requires precise control over application usage that a port and protocol definition cannot achieve.
- Be able to allow specific applications and block everything else. Allowing a specific set of applications through a whitelist and denying everything else significantly reduces the number of ways an organization can be attacked.
- Dynamically define access to sensitive data based on a user's group membership. Many traditional security policies define access based on the location of the endpoint in the network. Even if enterprise mobility didn't blur the traditional network boundaries, network location is a poor identifier for a user and their assigned privileges.
- Be able to validate a user's identity through authentication. For access to the most sensitive data, the security platform should validate user information obtained from the organization's authentication servers with another authentication method before allowing access. This ensures the traffic is coming from the expected user and not from someone impersonating them.
- Dynamically define the resources that are associated with the sensitive data or application. Many data centers and platform-as-a-service (PaaS) environments dynamically allocate resources to applications. To ensure the security posture matches the current resource allocation, the security platform needs to adjust along with the changing environment.
- Block data by file type and content. Blocking risky file types reduces the number of ways you can be attacked as well as reduces the number of ways attackers can exfiltrate data.
- Log detailed information on the traffic so security analytics can process it and identify abnormal behaviors.

Zero Trust for SaaS Applications and Cloud Storage Services

Sensitive data is not necessarily going to be sitting on a server in the private data center. In fact, public cloud storage services and software-as-a-service (SaaS) applications often host sensitive data. Because they are collaborative in nature, SaaS applications often become avenues for data leakage through inappropriate file sharing or more malicious actions. Additionally, their collaborative nature allows SaaS applications to be a new insertion point for malware into the organization. A key risk of many SaaS file storage applications is that they automatically synchronize files with users. Therefore, if a user downloads a malware file and stores it in the SaaS application, it may spread throughout the organization in minutes. The risk of malware being uploaded in the SaaS file share increases when users use personal devices, often without the same level of endpoint protection as their corporate devices. On top of that, when users outside your organization have access to SaaS file shares, you implicitly extend trust to those users.

Public cloud storage services don't have the collaborative nature of SaaS applications, but data breaches are still prevalent because of incorrect service configuration or a lack of governance on the content stored in the service. Users rarely have direct access to these services, but the data that is stored there such as source code, application configuration, large application data sets, and licensing information can be highly sensitive.

Although Zero Trust on the network security components of the security platform does enable the control over the SaaS applications and storage services that are available to internal users, when data is in a SaaS application or public cloud storage service, the network security components of the security platform do not have full visibility and control over the access to the data. Zero Trust must extend beyond the network. Fortunately, public cloud storage service and SaaS applications typically do still provide a way to secure data, through their APIs.

When sensitive data is stored in the public cloud or a SaaS application, Zero Trust requires the security platform to:

- Assess the appropriateness of the data stored in the SaaS application or storage service.
- Identify previously unknown malware and mitigate known malware and spyware stored in the SaaS application or storage service.
- Evaluate the accessibility of the data and mitigate unsafe sharing.
- Log detailed information on the access of the data so security analytics can process it and identify abnormal behaviors.

ZERO TRUST ON THE ENDPOINT

Although the network is the primary location of implementing Zero Trust, the exploits and malware attackers use to gain access to hosts and move closer to the organization's data aren't always visible to the network security components of the security platform. An internal user, maliciously or not, might bring in the malware through direct access to the endpoint, or the attack might be encrypted in a way the

network security components of the security platform can't decrypt. The Zero Trust model must extend to managed endpoints and servers to stop attackers from using exploits and malware to establishing a beachhead in your organization and accessing sensitive data.

At the OS level, exploits and malware are fundamentally different. *Exploits* are the results of techniques used against a system that are designed to gain access through vulnerabilities in the operating system or application code. *Malware* is a file or code that infects, explores, steals, or conducts virtually any behavior an attacker wants. To be able to prevent attacks that use exploits and malware the security platform must:

- Examine file characteristics before execution to determine whether a file is likely malicious or benign.
- In an evasion-resistant virtual environment, run unknown files to determine real-world effects and behavior.
- Run advanced threats that exhibit highly evasive characteristics in the virtual environment in a bare-metal analysis environment.
- Prevent the vulnerability-profiling techniques that exploit kits use before they launch attacks.
- Block the exploitation techniques attackers use to manipulate applications.
- Prevent exploits that leverage vulnerabilities in the operating system kernel to create processes with escalated (i.e., system-level) privileges.

Additionally, to implement the Zero Trust least-privileged security model on the endpoint, the security platform must be able to:

- Restrict the locations from which executable files can run.
- Restrict access to all network locations except for those that the application explicitly requires.
- Restrict the executable files that users can launch from external drives.
- Restrict the executable files that users can launch from optical disc drives.
- Log detailed information on the access of the data so security analytics can process it and identify abnormal behaviors.

ZERO TRUST SECURITY ANALYTICS

Although least-privileged access policies and protection on the endpoint stop the vast majority of attacks, sophisticated attacks may still be able to obtain a beachhead in your organization. Many organizations can't find these intrusions quickly. Often, this is because log messages generated by their infrastructure inundate security analysts. They try to find high-priority threats by correlating logs, but they rarely have the right data or tools to detect attacks accurately. This creates endless alerts to review, lots of false positives, and an unwieldy list of correlation rules to maintain.

Security analytics for Zero Trust must move beyond alerts and manual investigation to an automatic analysis based on rich data and machine learning. Machine learning requires a significant amount of detailed data to be able to distinguish between regular behavior and malicious attacks. This requires a shift in thinking around what must be logged. Most organizations don't log information that isn't suspicious, but machine learning needs to see the traffic that is harmless, as well as the malicious, to be able to identify the difference between them.

For security analytics to be able to detect and respond to sophisticated attacks, you must:

- Use the components from the cloud, and on the endpoint in order to obtain a complete view of the activities in the organization through rich and detailed logging information.
- Based on the logging data provided, you can model the behavior of users and devices.
- Detect anomalies that trigger a more detailed analysis of the historical data in order to distinguish between malware/exploit attacks and credential attacks.
- Recognize deviations from expected behavior.

Designing for Zero Trust with Palo Alto Networks Platforms

Because the most critical and sensitive data typically resides on resources within the private data center or public cloud environment, the design of Zero Trust begins at these locations and then migrates towards the user. However, Zero Trust isn't an all-or-nothing proposition. Although the security platform should protect every device in an organization, Zero Trust policies should focus on preventing attacks on the most sensitive data and applications. Specifically:

- In the data center and public cloud, Zero Trust policies are used to prevent successful attacks on the sensitive data and applications from within the data center itself. All traffic within the data center is inspected for malware, vulnerabilities, and command-and-control traffic by the next-generation firewall. Least-privileged access policies differ based on the sensitivity of the data.
- Low-sensitivity data is not protected with Zero Trust. Instead, base security policies on the next-generation-firewall protect against threats and block applications that are known to be malicious.
- Data with moderate sensitivity have Zero Trust policies that define broad sets of allowed application traffic within the data center.
- Highly sensitive data has the most restrictive policies that ensure traffic to the application. Even traffic that originates from other application components within the data center is tightly controlled.

Server operating systems, regardless of the sensitivity of the data they store, are protected by Cortex XDR Agent from exploits and malware. For data that is in the public cloud, Prisma SaaS and Prisma Cloud Storage Security monitors access to the data and remediate issues with access or the configuration of the cloud environment when they appear.

Zero Trust policies on the next-generation firewalls in the data center, public cloud, and internet perimeter also control user access to data and applications, whether they are on-premises, in a public cloud environment, or in a sanctioned SaaS application. Zero Trust policies control access as follows:

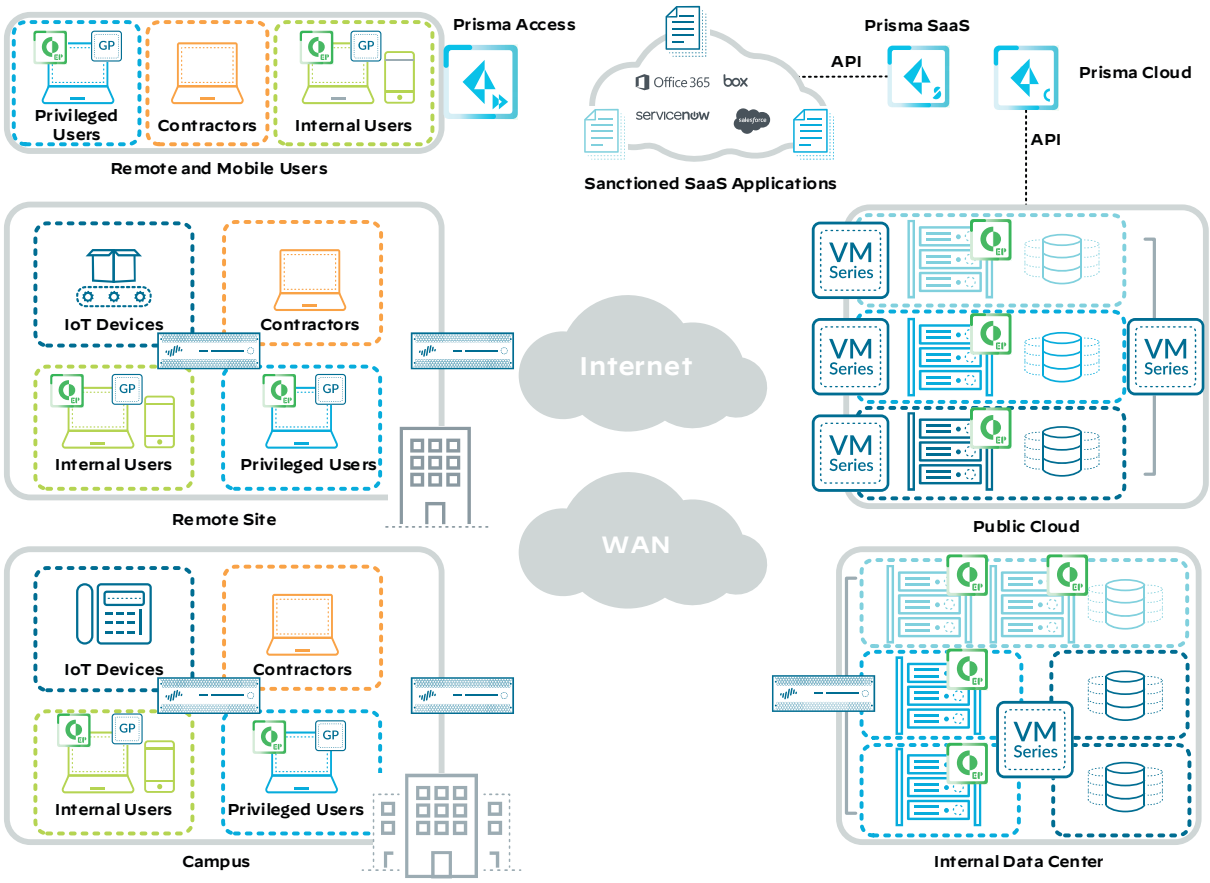
- Low sensitivity applications and data policies are inspected for malware and vulnerabilities.
- Data with moderate sensitivity have policies that define the appropriate user groups that should have access to the applications and broad sets of allowed applications into the segment.
- Highly sensitive data is controlled by the most restrictive policies. Multi-factor authentication (MFA) ensures that the user accessing the data isn't using stolen credentials; users must be in an expected location, and traffic can only use the expected applications.

Zero Trust policies on the next-generation firewalls in the campus, remote site, and mobile ensure privileged users are segmented from other users to stop lateral movement and the impersonation of a privileged user by an attacker. To protected endpoints from becoming a beachhead for attackers, all

endpoint operating systems are protected with Cortex XDR Agent against malware and exploits while the next-generation firewalls at the internet perimeter inspect for malware, vulnerabilities, and command-and-control traffic. The internet perimeter also blocks websites and applications that are known to be malicious.

The Cortex XDR analyzes the data logged from the next-generation firewalls, Cortex XDR Agent, and Prisma SaaS in order to identify sophisticated attacks that aren't detected directly by the platform components. Cortex XDR provides one single place for data center, cloud, user, and SaaS information.

Figure 2 Zero Trust with the Palo Alto Network platforms



ZERO TRUST WITHIN THE DATA CENTER AND PUBLIC CLOUD

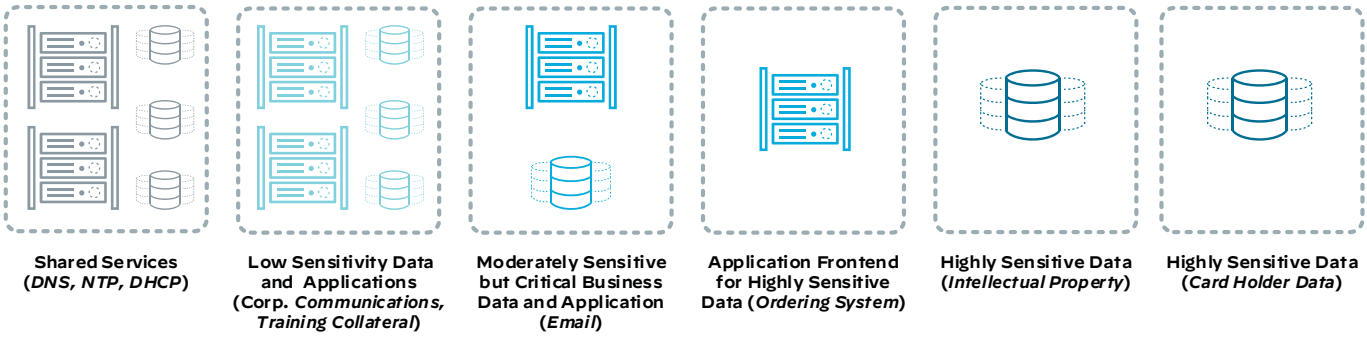
Starting with the most sensitive data, consider the servers that store the data and the applications that utilize it, as well as the services with which those applications communicate in order to function. Evaluate each of the resources and determine if they are dedicated to the data or application or if they are shared with other applications.

Segmenting Sensitive Data within the Data Center or Public Cloud IaaS

To provide the security platform visibility of and control over all of the traffic that is destined to the organization's sensitive data, the sensitive data must be segmented behind a next-generation firewall. Each segment contains the resources that have a common level of data security and have common functionality allowing the resources in the segment to share a security policy. The Forrester Research Zero Trust documentation refers to the next-generation firewall a *network segmentation gateway*. Because all traffic entering and leaving the segment must traverse the firewall, the firewall can inspect the traffic and enforce control. Segment your data as follows:

- Highly sensitive data should reside on dedicated resources, and you should group those dedicated resources in a dedicated network segment. If you have multiple sets of highly sensitive data, each of them should have dedicated sets of resources and a dedicated network segment. For highly sensitive data, group the resources that interact with the data (frontend application, etc.) in a separate segment.
- Moderately sensitive data should reside in a segment that contains all the resources associated with it including frontend applications.
- Data that has low sensitivity can reside with the other low sensitivity data and resources.

Figure 3 Network segments

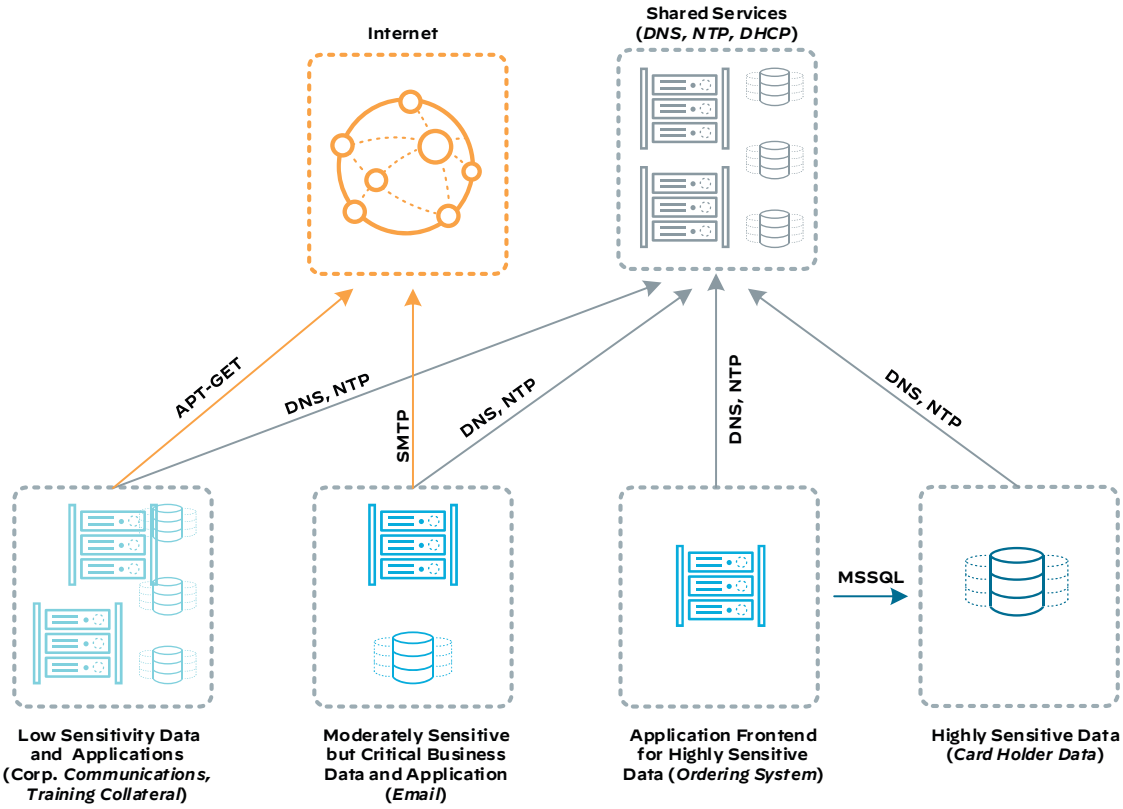


Zero Trust Policy for East-West Flows within the Data Center or Public Cloud IaaS

After you have the segments defined, you need to map out how those segments interact. Because the easiest traffic flows to describe are typically the flows between resources within an application, start with east-west traffic between segments associated with highest sensitivity data. Consider which segments need to communicate, which segments initiate the communication, and what applications communicate. Don't forget to consider the infrastructure shared services that almost every resource requires. For example, without access to DNS, NTP, and DHCP, many servers won't be able to communicate with other resources on the network. Account for configuration and patch management as well as other services that help operate and maintain the resources in the segments.

Finally, consider which segments need to initiate communication to the internet. The segments with the highest-sensitivity data should not be able to communicate to the internet, and moderate and low sensitivity segments may need to reach services that aren't hosted within the organization.

Figure 4 East-west and outbound traffic flow



Implement the Zero Trust policy on the next-generation firewall for traffic between segments (east-west flows) as a whitelist policy that explicitly defines what is allowed and denies everything else. The least-privileged policy for inter-segment traffic within the data reduces the ability for attackers to move laterally within the data center.

The next-generation firewall uses zones and dynamic address groups to define the source and destination networks for each traffic flow. 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 next-generation firewall to limit the communication between network segments to specific applications. App-ID does not map ports and protocols to applications. Instead, it uses application signatures, decoders, and heuristics to identify the application regardless of the port or protocol. 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.

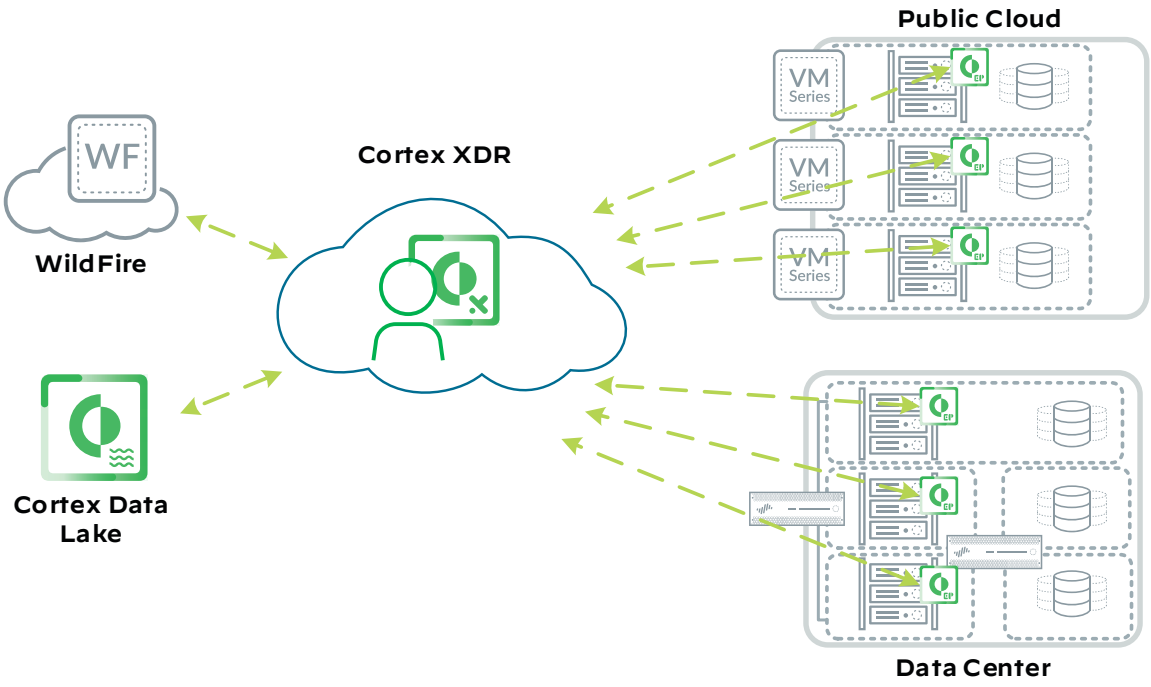
The Zero Trust model prevents attacks in traffic allowed by the least-privileged access policy through security inspection and the Threat Intelligence Cloud. Inspection identifies malware, vulnerabilities, command-and-control traffic, data exfiltration, and threats previously identified by the Threat Intelligence Cloud. The Threat Intelligence Cloud identifies previously unknown threats to the organization.

The next-generation firewalls send their logs to Cortex Data Lake . Cortex Data Lake acts as the centralized logging destination for all components of the Palo Alto Networks platforms. It not only allows for storing a significant amount of log data; that data is also available to applications in the hub to process and run security analytics.

Securing the Server Operating System

Although the next-generation firewall configured with a Zero Trust security policy prevents the majority of attacks from ever reaching your servers, new never-before-seen attacks can slip through before being identified by the Threat Intelligence Cloud. Cortex XDR Agent prevents malware and exploits from executing when installed on the servers in your private data center and public cloud infrastructure-as-a-service (IaaS) environments.

Figure 5 Cortex XDR Agent protecting servers



Cortex XDR Agent advanced endpoint protection stops threats on the server and minimizes server infections by blocking exploits, malware, and ransomware. Like with the next-generation firewall, you can use a Zero Trust policy in Cortex XDR Agent in order to extend the least-privileged access model to protecting the server operating system.

Some of the Cortex XDR Agent capabilities critical to the least-privileged policy include:

- **Exploit-prevention profile**—Because attackers most often target application vulnerabilities when attempting to compromise servers, the Cortex XDR Agent exploit-prevention profile is key to extending the Zero Trust security model to servers and blocking the core techniques used by Zero Day exploits.
- **Restriction profiles**—Similar to traffic flows, applications that run on the servers in your private data center and public cloud environment should be well defined. Restriction profiles reduce the avenues an attacker can use to compromise your servers.
- **Malware protection**—Ransomware, script-based attacks, and malicious executables must be prevented from executing on the server.

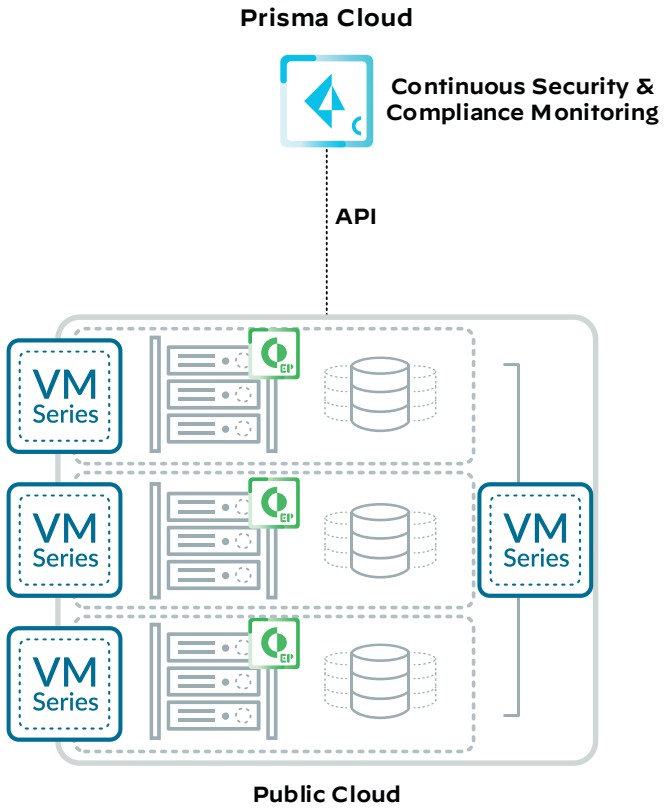
Securing the Public Cloud Environment

The policies and capabilities used to secure data and applications in the public cloud IaaS environments are effectively the same as those used in the private data center. The one major area of difference is that you do not have complete control over the infrastructure. This is troublesome because a single mistake in the configuration of the cloud environment can leave it open to attack and loss of sensitive data. When high and moderate sensitivity data and applications live in the public cloud, beyond segmenting the data behind a next-generation firewall, you must secure the cloud infrastructure as part of Zero Trust.

Prisma Cloud helps secure your public cloud IaaS environments and automatically validates best security practices against your public cloud resources. Prisma Cloud continuously monitors your distributed multi-cloud environments, proactively alerting you of any misconfigurations.

Prisma Cloud also can monitor the cloud environment for compliance with industry and regulatory standards. Configure Prisma Cloud to monitor your cloud environment for compliance with the standards that are relevant to your organization. Supported compliance benchmarks include including CIS, NIST, PCI, HIPAA, GDPR, ISO and SOC 2.

Figure 6 Prisma Cloud



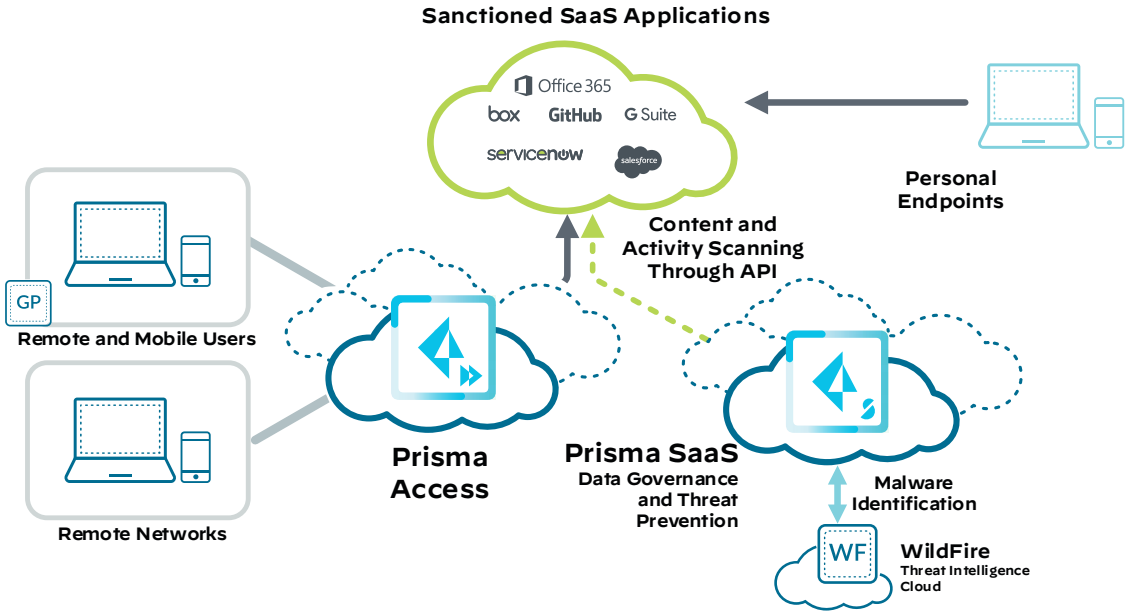
Protecting Data Stored in PaaS Storage Services and SaaS Applications

After your data leaves your network and is stored in a PaaS storage service or SaaS application, in-line network security devices can't see access and changes to the data. The Zero Trust security model requires visibility and control over sensitive data regardless of its location. Building on the Zero Trust protections at the internet perimeter:

- Prisma SaaS secures the data stored in sanctioned SaaS applications.
- Prisma Cloud Storage Security secures the data stored in PaaS storage.

Prisma SaaS and Prisma Cloud Storage Security are cloud services that connect directly to sanctioned SaaS applications and PaaS storage services using APIs. This connection provides visibility and control over the data, allowing for Zero Trust policy to extend into SaaS and PaaS. Visibility and control even extends to data and activities that originate on personal devices and collaborators who aren't part of your organization.

Figure 7 Prisma SaaS API integration with SaaS applications



When you first connect a sanctioned SaaS application to Prisma SaaS, the application's API allows Prisma SaaS to discover and retroactively inspect all files and data (called *assets* in Prisma SaaS) managed by the application. Prisma SaaS 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.

As Prisma SaaS identifies incidents, you can assess them and define automated actions to 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, Prisma SaaS continuously monitors the SaaS application and applies the policy to new or modified assets.

Zero Trust Policies That Assess Risk in SaaS and PaaS

Prisma SaaS allows you to define policies that automatically assess risk. There are two types of policies: content-based and activity-based.

- **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 the exposure-level categorization of 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, may indicate abnormal movement of data out of the SaaS application or other compliance violations.

You can remediate any 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 to educate them about SaaS application acceptable-use policies.

Automatic remediation can be used to address incidents across large numbers of assets that Prisma SaaS finds. 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, Prisma SaaS replaces the original asset with a placeholder (tombstone) file. The placeholder is a customizable plain-text file that contains a simple description to explain that Prisma SaaS quarantined the asset. Also, when Prisma SaaS automatically quarantines an assert, 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 Prisma SaaS automatically changes an asset's sharing settings, you can send the asset owner a Remediation Digest Email that describes the changes Prisma SaaS 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 can have a significant impact on a user's experience and may affect productivity. For example, if you change the sharing settings on a parent folder 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 only use them on new assets stored in SaaS application.

Preventing Malware in SaaS Applications

Prisma SaaS and Prisma Cloud Storage Security use the Threat Intelligence Cloud to detect both known and unknown malware. Prisma SaaS submits a hash of the asset to WildFire® to determine whether WildFire has seen it before. If WildFire has seen the asset previously, WildFire uses the existing verdict. If it has not seen the asset before, Prisma SaaS forwards the asset to WildFire for analysis.

Unlike the next-generation firewall, Prisma SaaS uses this verdict and not additional signatures to identify assets with malware. WildFire still generates signatures for the malware and makes them globally available to the next-generation firewall and Cortex XDR Agent, so if a user in your organization stores the malware elsewhere and another user attempts to access it, the firewall and Cortex XDR Agent can prevent it.

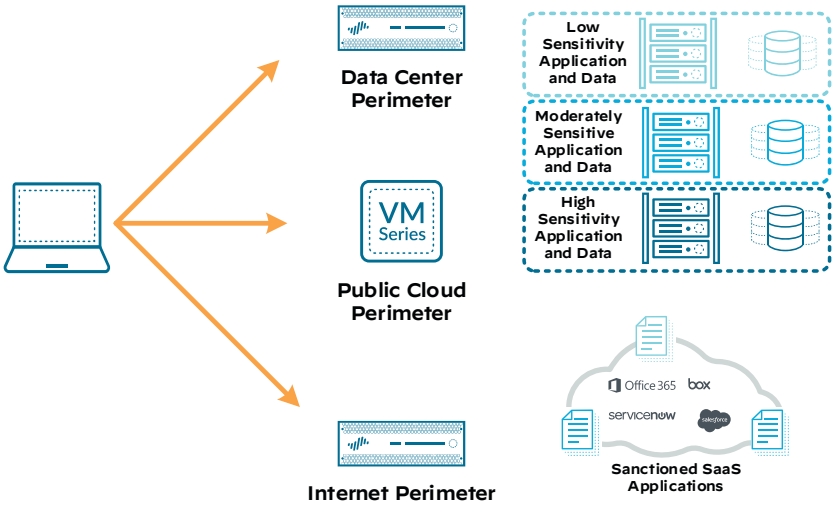
If WildFire returns a Malware verdict, the asset is in violation of the content policy rules, and Prisma SaaS applies the configured action settings of Quarantine, Change Sharing, Notify, or Log. Detected malware shows up as a risk in the Dashboard, Risks pane, and the SaaS Risk Assessment Report.

CONTROLLING ACCESS TO DATA AND APPLICATIONS WITH A ZERO TRUST POLICY

Whether or not the data and applications live in the private data center or as applications on the internet, designing a Zero Trust policy that allows users to access data and applications through the next-generation firewall relies on the following detail for each application:

- Who should be able to access the resources within it
- Where they should be able to access from within the network
- What are the required resources in the segment
- How users access the resources

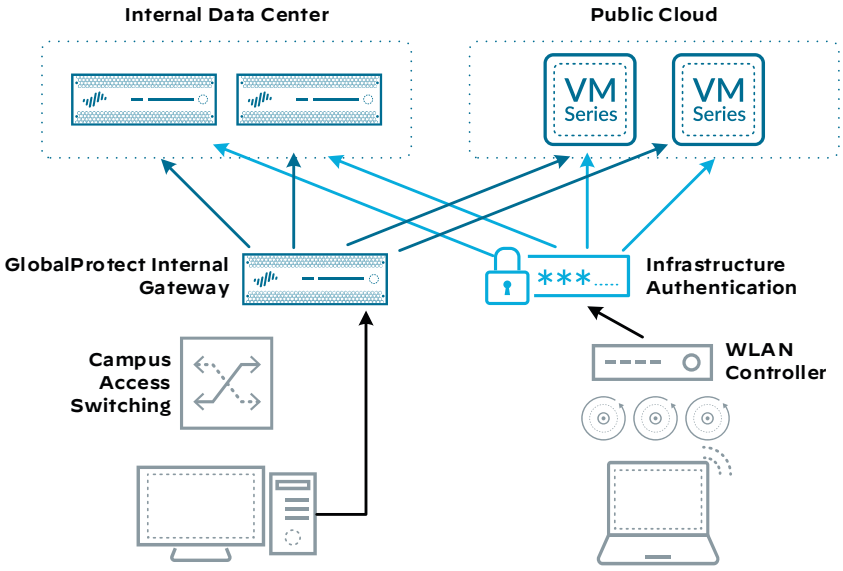
Figure 8 Zero Trust Policy for on-premises and internet applications



User-ID

You do not need to place users into separate network segments to identify one user group from another in policy. User-ID™ is the primary source of mapping users and groups to a Zero Trust policy. User-ID not only gives fine-grained control of user identification in policy, but it is also simpler than configuring the access layer infrastructure to segment out all the user groups from each other. User-ID maps IP addresses to users by integrating with a variety of user repositories. 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 firewall. 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.

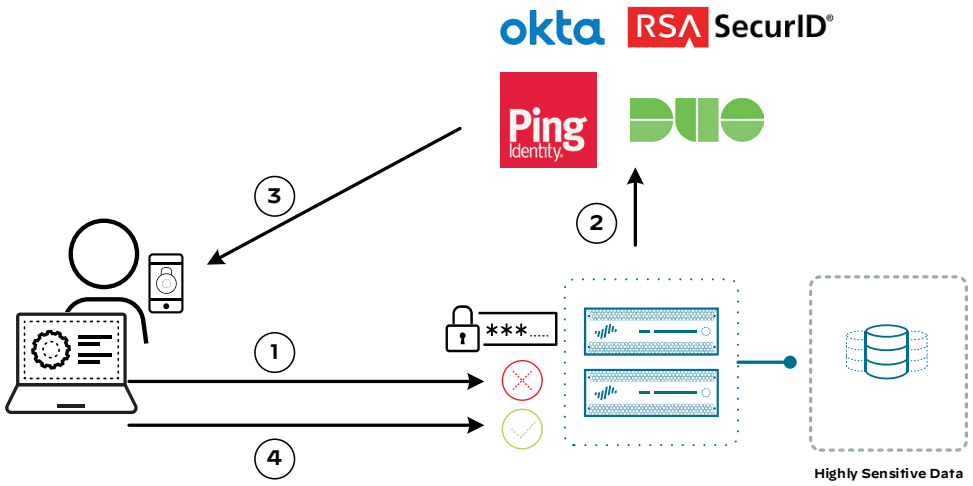
Figure 9 User-ID sources



Authentication Policies

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 a multi-factor authentication of the user before allowing access and ensure that the originator isn't using stolen credentials. Because the next-generation firewall 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 doesn't natively support or is difficult to configure for MFA.

Figure 10 MFA authentication for high-sensitivity segments



App-ID

Like the inter-segment flows within the data center, App-ID is used to control the traffic between the users and the data by explicitly defining the applications that are your critical business applications and the SaaS applications where your users store their data. Use application filters to dynamically define application groups based on the risk attribute and characteristic. Application filters for low-risk applications allow you to safely enable access to those applications without having to define each one in policy. As new applications emerge, and new App-IDs get created, these new applications automatically match the filter. Any application that matches the filter will be safely enabled without additional changes to the policy.

Inspection

The Zero Trust model prevents threats in traffic allowed by the least-privileged access policy through inspection and the Threat Intelligence Cloud. Inspection is used to identify malware, vulnerabilities, data exfiltration, and threats previously identified by the Threat Intelligence Cloud. The Threat Intelligence Cloud is used to identify previously unknown threats to the organization. Inspection in the next-generation firewall includes:

- **Antivirus and WildFire**— In a Zero Trust security policy, the firewall inspects traffic for known antivirus signatures.
- **Anti-spyware**—Anti-spyware profiles prevent infected endpoints from sending malicious traffic to command and control systems.
- **URL filtering**—In a Zero Trust security policy, URL filtering blocks command-and-control traffic and access to known-malicious websites.
- **File blocking**—Blocks files that are known to carry threats.

PROTECTING ENDPOINTS IN THE CAMPUS, REMOTE SITE, AND MOBILE

After you secure the most sensitive data in your data center and public cloud environments by using the Zero Trust model, consider the security of your endpoints. Because exploitation of endpoints or end users is often the starting point of an attack on sensitive data, protecting the endpoints is critical to an overall Zero Trust security design.

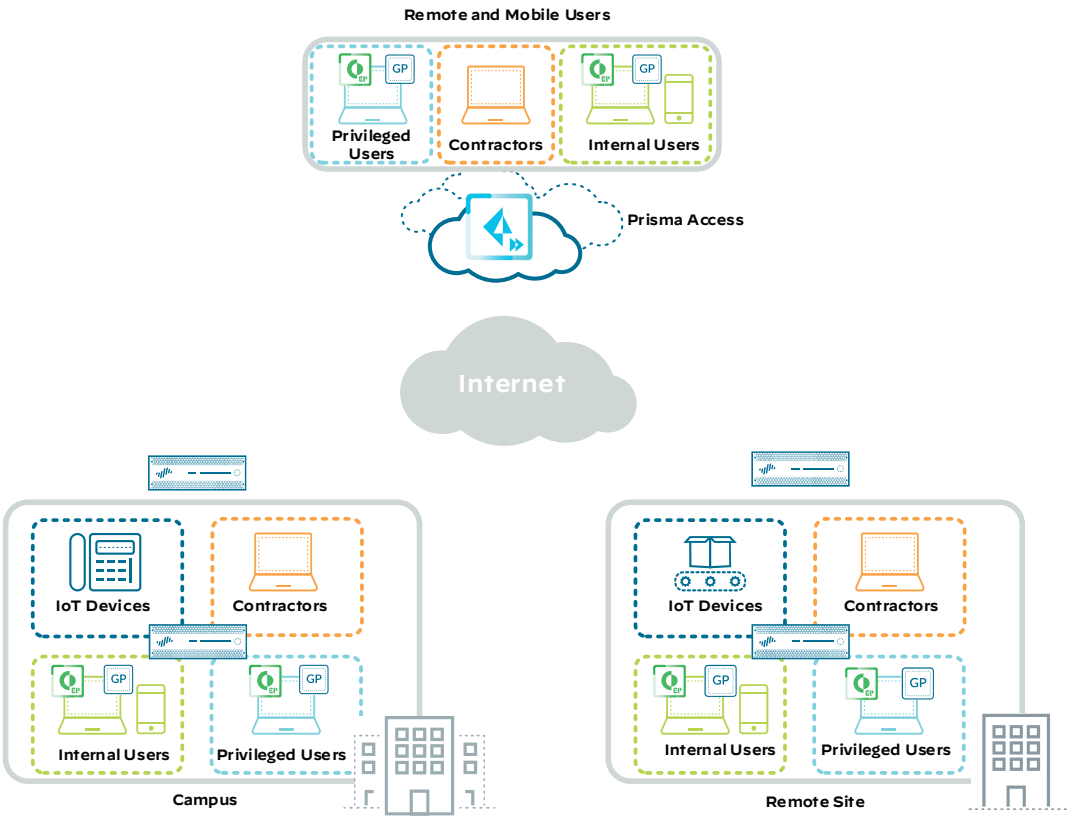
Protecting Endpoints of Privileged Users

Although endpoints should rarely store high sensitivity data, when a user has access to the most sensitive data, the endpoints they use must also be considered when designing a Zero Trust policy to protect the data.

The next-generation firewall does not require endpoint segmentation in the infrastructure in order to protect data and applications in the data center. User-ID and authentication policies allow the firewall to control access to sensitive data and properly segment users in policy without relying on infrastructure segmentation of the source endpoint. However, segmenting the endpoints that privileged users use from other endpoints and devices in your network is still recommended. Segmentation of the privileged user's devices behind a next-generation firewall reduces the chances that an attacker can successfully gather sensitive data that may live on the device temporarily or compromise the endpoint to gain privileges necessary to reach the sensitive data in the data center or public cloud.

When privileged endpoints leave the internal network, configure policies in Prisma Access to segment them like the campus and remote site. Segmenting mobile users ensures endpoints are protected from other endpoints regardless of their location.

Figure 11 User segmentation across the organization



Tightly control the inbound policies for the privileged user segments. Because the endpoints do not host applications or provide services the policy should be straightforward. In fact, in many instances, the inbound policy should not allow any connections initiated from the rest of your organization's internal network.

The outbound policy should be a combination and mirror of the internet perimeter firewall and the inbound policies on the next-generation firewall that protects the data and applications that the administrators should have access to.

Protecting Endpoints at the Internet Perimeter

To minimize the chance of a successful cyber-attack, every endpoint requires protection. Although the data might not live on the endpoint, every endpoint can become the launching point for an attack on the data from within your private network. Protect every managed device in your organization with a next-generation firewall when accessing the internet, even when the devices are mobile. Endpoints in the campus, at remote sites, and mobile should traverse a next-generation firewall for access to the internet.

For mobile users, the GlobalProtect client establishes an always-on, secure SSL/IPsec VPN connection to a GlobalProtect gateway running on the next-generation firewall or Prisma Access.

Like in the data center and public cloud, Zero Trust security policy on the internet perimeter firewalls uses App-ID, User-ID, and Content-ID™ to inspect the traffic and detect and block known threats and send unknown files to WildFire analysis to identify new threats and generate signatures to block them.

In addition to whitelisted SaaS applications, the internet perimeter can limit the functionality of any applications through functional App-IDs. Limiting application functionality is useful when you want to tolerate an application. Tolerated applications are important to your organization but are either not fully visible to the security platform, or the platform cannot secure them because a third-party controls them or a partner uses them to share data with your users.

Application filters dynamically define application groups based on the risk attribute and characteristic. Application filters for high-risk applications allow you to deny access to those applications without having to define each one in policy. As new applications emerge, and new App-IDs get created, these new applications automatically match the filter. The firewall denies any application that matches the filter without additional changes to the policy.

Because a Zero Trust security model uses least-privileged access, when only a select group of your users require access to an application use User-ID and group membership to limit who can access them. internet perimeter firewalls can share the same User-ID sources as the data center and public cloud firewalls. High-fidelity sources are just as important at the internet perimeter as in the data center and public cloud. In the event the firewall does not have an IP address mapping to a user, use an authentication policy to send that traffic to the captive portal. The captive portal allows the user to authenticate so that User-ID can map the IP address to a user before evaluating policy.

Inspection

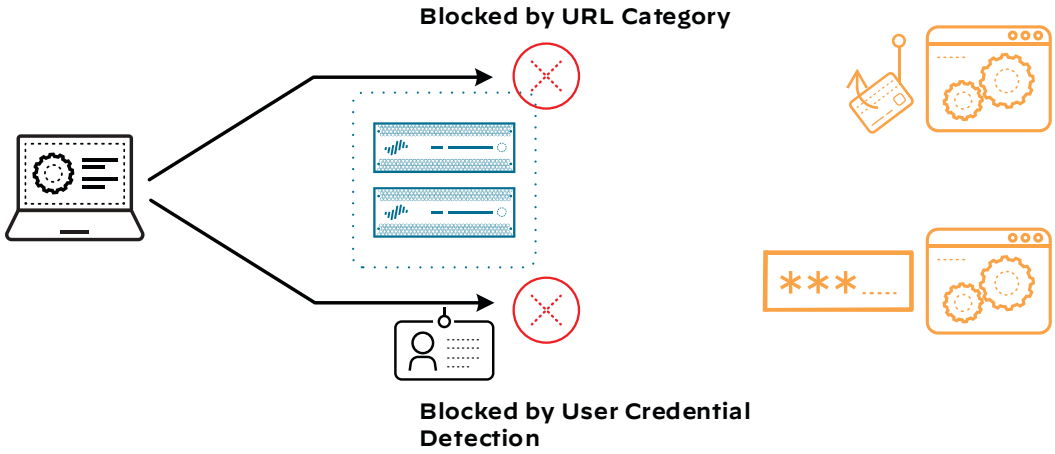
Like the next-generation firewalls in the data center and public cloud, internet perimeter firewalls use inspection to identify malware, vulnerabilities, data exfiltration, and threats previously identified by the Threat Intelligence Cloud. The Threat Intelligence Cloud is used to identify previously unknown threats to the organization. Inspection in the next-generation firewall includes antivirus and WildFire, anti-spyware, URL filtering, and file blocking.

Phishing

To prevent phishing attacks from stealing your user's credentials and providing an avenue of access to your sensitive data and applications, configure credential phishing protection on all the 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 aren't part of the current URL database, configure the firewall to uses its IP-address-to-user mapping table 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 12 Phishing prevention

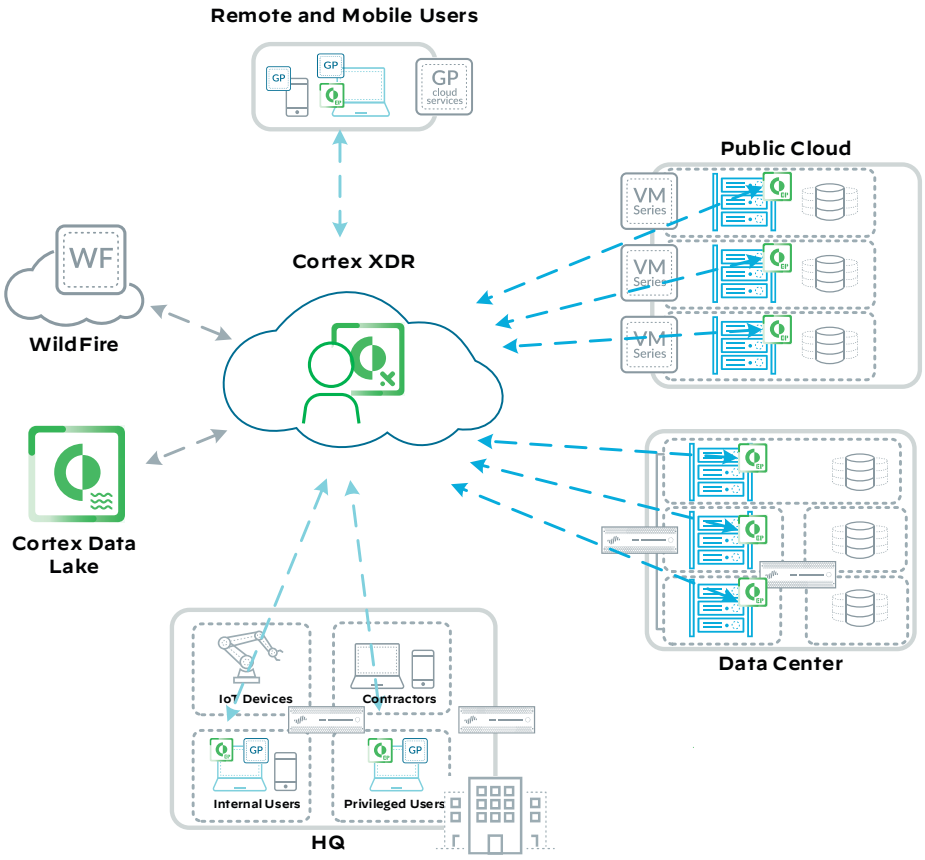


Protecting Endpoint Operating Systems

Cortex XDR Agent advanced endpoint protection stops threats on the endpoint and coordinates enforcement with cloud and network security to prevent successful cyber-attacks. Cortex XDR Agent minimizes endpoint infections by blocking malware, exploits, and ransomware.

Managed endpoints and servers connect to the same Cortex XDR but can be configured separately. Having a separate group allows you to define unique Zero Trust security profiles for managed endpoints and servers.

Figure 13 Cortex XDR with managed endpoints and servers



Attackers often blend two primary attack methods to compromise organizations: targeting application vulnerabilities through exploits and deploying malicious files. These methods can be used individually or in various combinations, but they are fundamentally different.

Due to the fundamental differences between malware and exploits, effective prevention requires an approach that protects against both. Cortex XDR Agent combines multiple methods of prevention at critical phases of the attack lifecycle to halt the execution of malicious programs and stop the exploitation of legitimate applications on Windows and macOS endpoints.

The Cortex XDR provides granular control over service protection settings and 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 keys and values, and files.

Cortex XDR Agent sends all its logs to Cortex Data Lake for storage and to be used as data for security analytics in the hub. Cortex XDR Agent logs provide detail on the processes and activities that generate and receive traffic giving additional information to traffic that the next-generation firewall logs.

SECURITY ANALYTICS

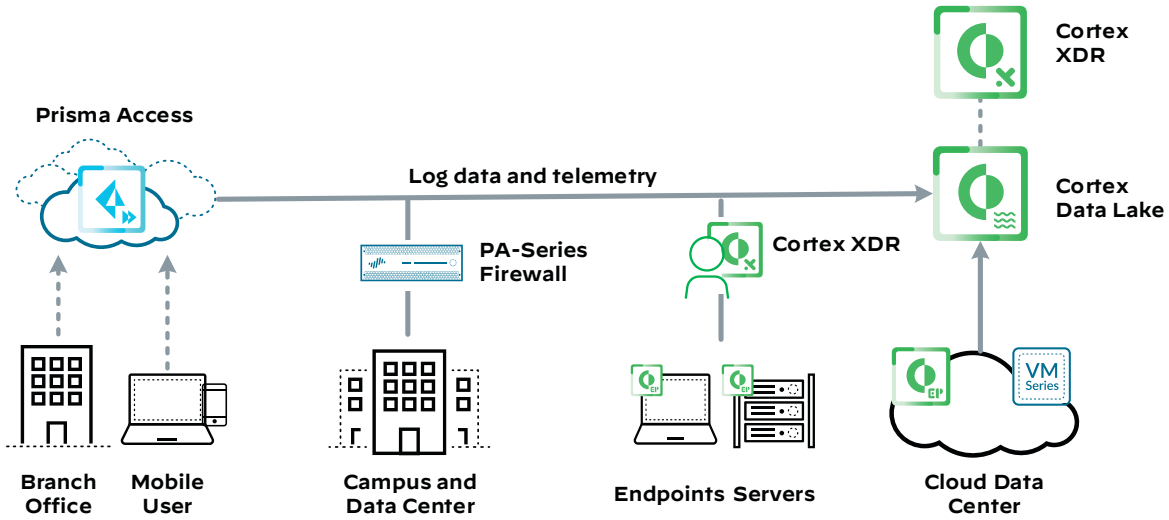
Cortex XDR empowers organizations to quickly detect and respond to the stealthiest network threats affecting the data center, cloud, SaaS applications, and users. By analyzing rich network, endpoint, and cloud data with machine learning, in one location Cortex XDR accurately identifies targeted attacks, malicious insiders, and compromised endpoints based on behavioral changes that attackers cannot conceal.

A complete view of the activities in the network, on the endpoint, and in the cloud is required for machine learning to profile behavior and automatically detect attacks accurately. Luckily, to prevent attacks Zero Trust requires a security platform that covers all these locations so that the security platform can provide the rich data without requiring additional sensors.

Because of the large amount of data required, Zero Trust security analytics require big data storage that is localized to your environment. Cortex Data Lake serves as the central cloud-based repository for all security platform data and logs. Cloud-based log storage not only allows organizations to collect the ever-expanding amounts of data required by Cortex XDR to identify threats. It can also serve as a data repository for analysis by other applications as they emerge in the hub.

Cortex XDR performs analysis based on a combination of unsupervised and supervised machine learning techniques. Cortex XDR uses unsupervised machine learning to model user and device behavior, perform peer group analysis, and cluster devices into relevant groups of behavior. Based on these profiles, Cortex XDR detects anomalies compared to past behavior and peer behavior. Cortex XDR also monitors multiple characteristics of network traffic to classify each device by type, such as a desktop computer, mobile device, or mail server. Cortex XDR also learns which users are IT administrators and which are regular users. With supervised machine learning, Cortex XDR recognizes deviations from expected behavior based on the type of user or device, reducing false positives.

Figure 14 Cortex Data Lake and XDR



Summary

Breaches and data loss have serious consequences for organizations and their customers. Zero Trust is a security model developed specifically to address the security of sensitive data and critical applications in an enterprise organization. Zero Trust policy leverages the Palo Alto Networks platforms' capabilities and functionality:

- PA-Series and VM-Series firewalls provide the inline protection required to segment data from other applications and endpoints in the network.
- Cortex XDR Agent advanced endpoint protection inspects system process execution and filesystem behavior through local and dynamic analysis.
- Prisma SaaS and Prisma Cloud for Storage Security inspect asset accessibility and risk through API integrations into the public cloud storage services and SaaS applications.
- Cortex Data Lake serves as the central cloud-based repository for all security platform data and logs.
- Cortex XDR performs analysis based on a combination of unsupervised and supervised machine learning techniques on the data in the Cortex Data Lake to detect attacks using behavioral analytics.

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

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

B-000300P-1-20a