Designing A Zero Trust Network With Next-Generation Firewalls

OVERVIEW

As enterprise boundaries blur due to an extended workforce of partners and contractors, and proliferation of mobile devices bring an increase in number and variety of devices connected to the network, the old paradigm of security controls at the perimeter or on user devices are no longer adequate. A new network security paradigm is needed – one that focuses on protecting the data at the heart of the enterprise network.

The Zero Trust Network Architecture is a security framework developed by John Kindervag of Forrester Research. In a series of reports[1], Kindervag introduces the concepts of Zero Trust, a new approach to network and device security that places security at the core of the network and makes it central to all network transactions. This security-centric approach advocates a number of principles to design a secure and flexible network that can protect against modern malware and threats.

This solution brief describes how Palo Alto Networks next-generation firewalls can be deployed as the key architecture component of the Zero Trust architecture – the network segmentation gateway. More importantly, the capabilities of the next-generation firewall technologies such as App-ID, User-ID, and Content-ID enable the identification, inspection and granular control of all applications in the enterprise, addressing key Zero Trust concepts.

FACTORS AFFECTING SECURITY TODAY

The enterprise has become distributed. Empowered users are accessing the network from a variety of devices, including laptops, tablets, and phones and from a variety of locations. The expectation of anytime anywhere “workspaces” for these users enable new gains in productivity, but also now leads to new security challenges in differentiating access based on user, application, device-type or access type (wired, wireless, VPN).

The extended enterprise now also includes an ecosystem of partners, contractors, supply chains, and customers requiring access to the network for collaboration on business projects. Security and compliance challenges abound in tracking data, and enforcing access control for these extended users.

Compounding security complexities, the threat landscape has shifted dramatically from notoriety-based attacks to cybercriminal- and nation state attacks with the objective of gaining data to exploit for financial or geopolitical purposes. Threats are delivered via a variety of different vectors, from application-enabled vectors and exploits to high-risk URLs, and malware. In many cases, enterprises are exposed to targeted and customized malware, which can easily pass undetected through traditional antivirus solutions. The impact of targeted attacks has been significant on organizations such as Sony, Nortel, Symantec, RSA, Comodo,

DigiNotar. The financial implications from these attacks have been in the millions and in the case of DigiNotar, have brought the company to bankruptcy and insolvency.

A security strategy focused just on protecting end-users, devices and the perimeter may not be adequate or effective due to the dynamic nature of these elements. A more effective approach is to protect and control access to the one entity that remains constant – the data, and ensure that the security travels with that data[2].

WHAT IS ZERO TRUST?

The Zero Trust model, proposed by John Kindervag of Forrester Research, is a conceptual model for IT security that tackles modern threats, and provides better security in an efficient way that eases compliance burdens and operational costs[3]. As the name implies, the Zero Trust model advocates[4] stripping away all previous assumptions about trust in the network, and not trusting users, packets, interfaces or the network. It also proposes that controls be consistently applied through the entire network, whether it is internal or external malicious users.

The following foundational concepts[5] are recommended to deliver effective security in the evolving borderless environments:

- Concept #1 – Ensure that all Resources are Accessed Securely Regardless of Location
  The recommendation here is access for all users, internal or external, should only be performed securely using encrypted tunnels.

- Concept #2 – Adopt a Least Privilege Strategy and Strictly Enforce Access Control
  While attacks can come from external hackers, insider threats are also prevalent. Role-based access control - access to specific applications based on the user role and his/her security privileges is required for all users. This not only protects against malicious attacks but also insider abuse of access rights.

- Concept #3 – Inspect and Log All Traffic
  In combination with the two concepts above, inspecting and logging of all traffic ensures validation of user activity, that users are indeed accessing allowed resources, and are not bringing vulnerabilities into the network.

ZERO TRUST ARCHITECTURE COMPONENTS

In order to fully implement the concepts of Zero Trust, Kindervag believes that security should be built into the network as a default instead of the existing pattern of retrofitting security into the network. He believes existing designs to create a secure network are unsuccessful because of the inability to rethink preconceived network designs.

The key architecture components[6] of Zero Trust, as shown in Figure 1 and Figure 2 include:

- Network segmentation gateway – This network segmentation gateway serves as the nucleus of the network, and integrates the features and functionality of individual standalone security devices, from firewalls, IPS and web-application firewalls to network access control, VPN gateways and encryption products. Kindervag's vision of this gateway is a "firewall on steroids" that supports high-speed multiple 10 Gb interfaces and delivers high performance capabilities to inspect all traffic

- Microcore and microperimeters– For better security and compliance, every interface connected to the "segmentation gateway" is its own switching zone. Each switching zone is a microcore and microperimeter (MCAP) with resources sharing similar functionality and global policy attributes, in other words, the same security trust levels. This essentially partitions the network into parallel, secure network segments that can scale to address specific compliance or regulatory needs.

- Centralized management – Because the network segmentation gateway is the nucleus of the system and the switch infrastructure is placed around this element, the "network backplane" is now the unified management of these microcores and microperimeters. The management backplane now supports centralized management of network and security devices.

- Data acquisition network– For complete network visibility, network data needs to be inspected and analyzed in real time for better network visibility. Traffic can be mirrored and forwarded to a data acquisition network MCAP that can capture, analyze and log all traffic traversing the network.

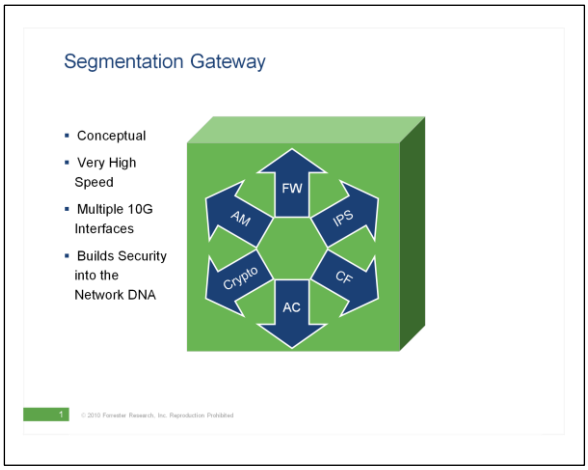

Figure 1: Forrester "Segmentation Gateway"

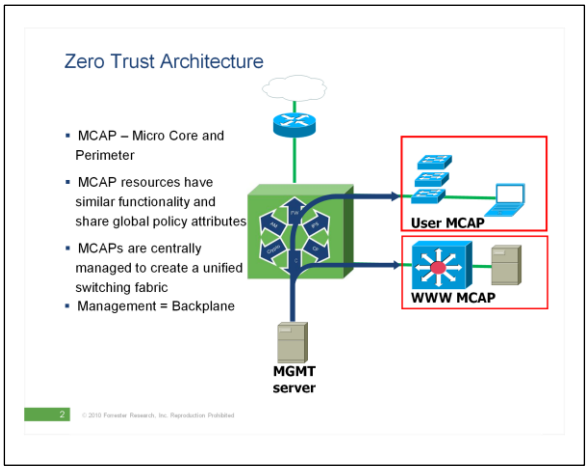

Figure 2: Forrester Zero Trust Architecture

PALO ALTO NETWORKS' NEXT-GENERATION FIREWALL IS A SEGMENTATION GATEWAY

The Palo Alto Networks next-generation firewall brings a unique combination of hardware and software functionality that makes it ideal as a Zero Trust network segmentation gateway. The integration of key security functions in the network segmentation gateway as described in Zero Trust makes logical sense and is what next-generation firewalls like Palo Alto Networks' provide.

The benefits of integration include:

- **Functional holes:** There are several basic pieces of information useful for setting security policy – irrespective of the function: source user, application, application function, URL category, and traffic destination. But each device acquires this information in unique ways, or in many cases, isn't capable of acquiring some of the pieces. These gaps and inconsistencies significantly impact the security effectiveness. Integration allows the information to be collected once and applied in a single security policy, and thus delivers better security.

- **Operational management:** Managing the complexity of loosely interconnected set of devices is not a simple task. Separate management systems, functional holes, unknown functional overlaps, and network complexity all contribute to costs and potentially ineffective network security. Integration simplifies security management through fewer consoles and fewer functional gaps.

- **Network performance:** With every new device comes additional latency, throughput, chokepoints, routing issues, and more. Integration can reduce network latency and the number of chokepoints traffic must pass through.

- **Network Complexity Reduction:** Over the last several years, every new security need has resulted in a new security device to solve it. As the number of security requirements increased, the number of devices deployed at key network junction points has increased to an unmanageable point. Not only are significant network changes (VLANs, port changes, IP addresses, network space, power) needed, but there is added complexity to managing these disparate systems. Integration reduces network complexity. With simplicity also comes better security, as configuration errors are less likely to occur.

- **Total cost of ownership:** The cost of purchasing separate devices for each security functional requirement, maintaining the equipment, and operational costs all add significantly to the total cost of ownership. Integration can significantly reduce each of these costs.

More important than the ability to integrate multiple security functions in a single platform is Palo Alto Networks' next-generation approach for doing so--- with full visibility and control, processed in a single pass software architecture, and using a single unified security policy. Palo Alto Networks next-generation firewalls provide the ability to control the applications, users and content that can traverse the "microcore and microperimeter" (MCAP) segments, which are essentially security zones.

The ability to have visibility into all traffic regardless of what port, encryption or evasive technique is employed is one of the key benefits of using Palo Alto Networks next-generation firewalls as Zero Trust network segmentation gateways. By understanding what applications are being used with App-ID, appropriate differentiated access policies can be applied in combination with User-ID. Comprehensive threat protection leveraging Content-ID will further ensure there are no threats within approved

applications in a MCAP.

As the data center grows and more resources are deployed, additional MCAPs or security zones can be added. Each MCAP will be anchored at a firewall pair, but each firewall pair can connect to more than one MCAP. As additional MCAP capacity is needed, more firewall pairs are added.

To meet the high-performance requirements of the Zero Trust "segmentation gateway, the Palo Alto Networks next-generation firewalls offer a single-pass software architecture that processes functions in a single pass to reduce latency. Physical appliances combine the single-pass software architecture with parallel processing hardware architecture, with dedicated, specialized processing for networking, security, and content scanning so that the full suite of next-generation features can be enabled with high throughput and reliability.

The management of the network segmentation gateway can be centralized via several options. The first is using Panorama, a centralized security management system that provides global control over a network of Palo Alto Networks next-generation firewalls. For convenience, Panorama has the same look and feel as individual device management interface. Panorama also provides the ability to view logs and run reports across dynamic or locally queried data aggregated from managed devices. Distributed reporting can be done without a need to forward logs from firewalls to Panorama.

The other central management option is via a powerful XML management API that enables external software to centrally manage and configure Palo Alto Networks firewalls. The exhaustive and fully-documented REST-based API allows configuration parameters to be seen, set and modified as needed.

Here's how the Palo Alto Networks' next-generation firewall technologies address the concepts of Zero Trust:

- **Ensure that all resources are accessed securely regardless of location**

  GlobalProtect™ secures access to the enterprise network via VPN, effectively establishing a secure connection for all users, whether internal or partner, irrespective of location or devices used. GlobalProtect ensures that the same secure application enablement policies that protect users at the corporate site are enforced for all users when they access any resources within the enterprise network. In addition, site-to-site IPSec VPN secures access by external business partners.

- **Strictly enforce access control – App-ID™ and User-ID™**

  Using a combination of App-ID and User-ID, strict enforcement of access control to these MCAPs can be accomplished. App-ID is used to identify the applications in each MCAP, and only allow traffic relevant to that specific segment. If App-ID determines that encryption (SSL or SSH) is in use and a decryption policy is in place, the application is decrypted and application signatures are again applied to the decrypted flow.

  In many enterprise networks, custom or home-grown applications will be used in the MCAPS, in which case custom App-

IDs or application-override can be used to manage the traffic. Finally, any application that cannot be identified is characterized as “unknown” traffic, allowing IT administrators to isolate that traffic to understand its impact to the Zero Trust environment. User-ID, integrated with existing enterprise user repositories such as Active Directory or LDAP, ensures that only authorized users can access the MCAP data center resources.

- **Inspect and log all traffic**

  All traffic into and out of MCAPs is inspected with Content-ID. Internal threat protection is important in internal MCAPs that may be accessed by infected internal users. External protection is also critical for Internet-facing MCAPs that may face automated kiddie script attacks or denial-of-service attacks.

  With Palo Alto Networks next-generation threat prevention strategy, application-specific threat prevention begins first by reducing the threat footprint through an explicit deny policy for unwanted applications such as external proxies, circumventors, and P2P file-sharing. Then, as you enable specific applications and associated features, you enable virus, vulnerability exploit, spyware and modern malware protection features to extend the application-specific context into threat prevention. For example, in a finance server MCAP, you can allow Oracle on its standard port only for finance and operations and protect against SQL injection attacks and Oracle specific vulnerability exploits. The threat prevention features that are part of Content-ID use a single, uniform signature format to scan for (and block, by policy) all manner of threats, in a single pass.

  To address security attacks that utilize targeted and customized malware, WildFire™ provides automated sandbox analysis of suspicious files to reveal unknown malware and the Behavioral Botnet Report can identify the unique patterns of botnet infections in the network.

  In addition, logging information provided with User-ID visibility will create accurate records of who accessed MCAP resources and when. This will also enable network analysis and visibility. Specific visibility into what applications are used and not used within an MCAP allows periodic refinements of security policy.

DEPLOYING A ZERO TRUST ARCHITECTURE WITH PALO ALTO NETWORKS FIREWALLS

This section provides an example of how to deploy a zero-trust architecture with Palo Alto Network’s next-generation firewalls. We use an example of an enterprise, Acme Enterprise, that is interested in using a pair of Palo Alto Networks’ firewalls to secure an enterprise data center environment that supports both Internet-facing applications as well as internal data center applications.

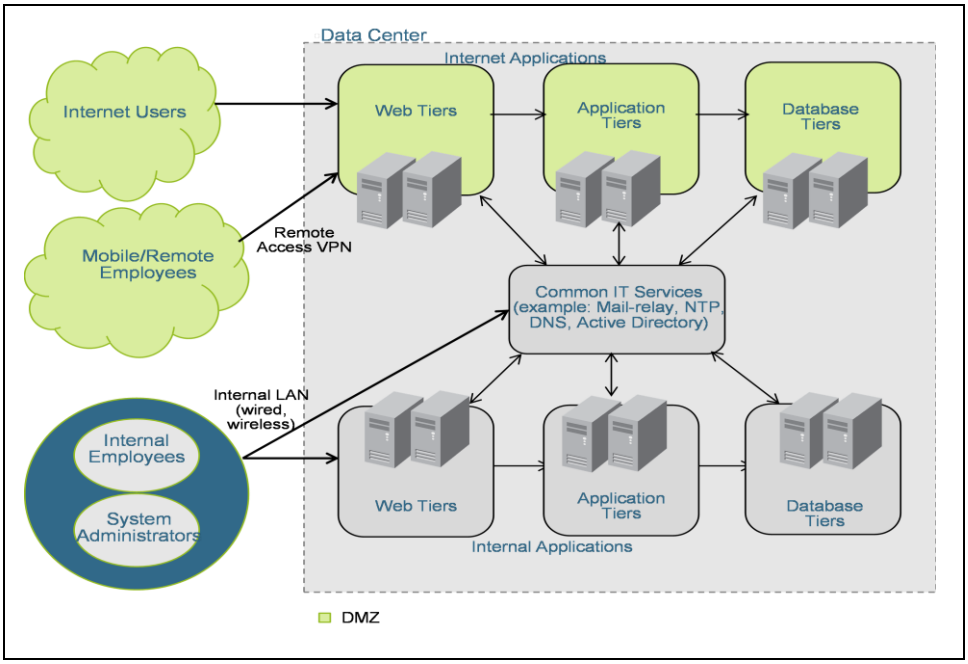

Figure 3: Data Center Traffic Flow

In Figure 3, we show a high-level, simplified traffic flow of the data center. Two sets of data center servers are deployed. One set is to be accessed by internet users such as remote/mobile employees and general internet users. The other set of servers are internal servers that will be accessed by employees on premise as well as mobile or remote employees.

Acme Enterprise would like to utilize the concepts of Zero Trust:

- Secure access to data center resources:
  - Access to data center resources by remote/mobile employees must be secured by remote access VPN.
  - Access to data center resources by external business partners must be secured via site-to-site VPN
- Segmentation is critical to security:
  - All data center servers must be segmented appropriately based on attributes such as similar risk factors and security classification. In this example, we are segmenting them according to their functions such as web, application and database.
  - The Internet servers in the DMZ must be segmented from the internal servers to prevent internet users from being able to access internal resources.
  - Common infrastructure such as Active Directory, NTP in “Common IT Services” are sometimes the most vulnerable and critical, because they can typically communicate with all other services. These common services must be segmented from other server tiers.
- Access control is on a “need-to-know” basis and is strictly enforced
  - Access control to data center resources must be enforced based on the employee role and responsibilities
- All traffic should be inspected and logged

The logical firewall placement to achieve these objectives is shown in Figure 4:

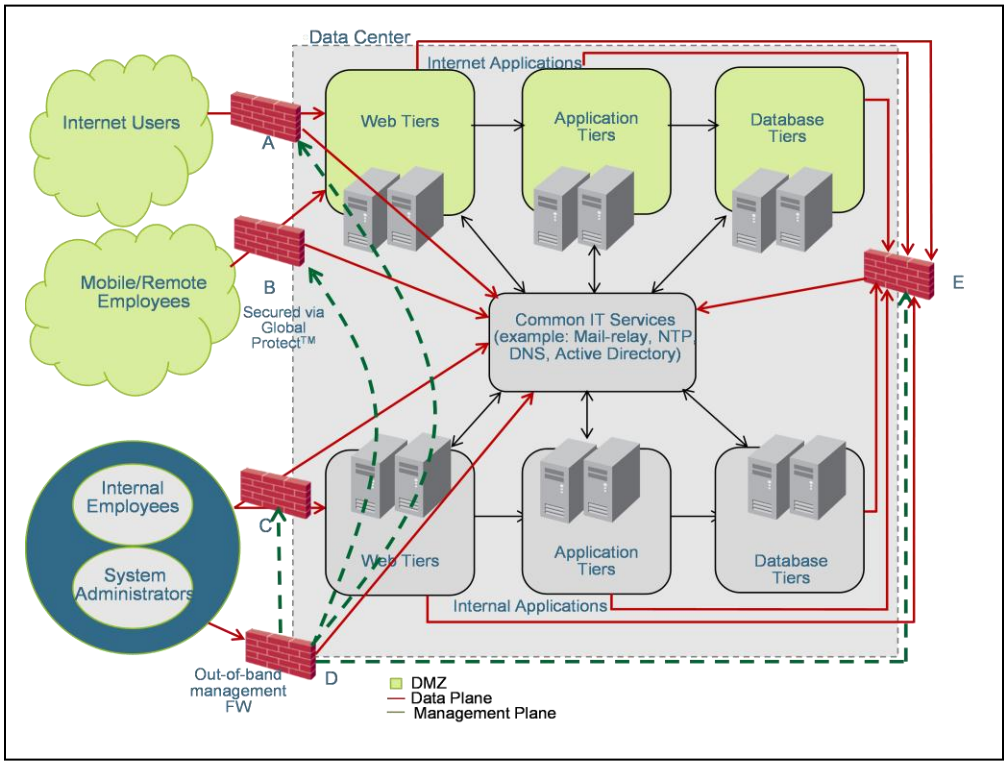

Figure 4: Data Center with Logical Firewall Placement

Each of the logical firewalls address specific objectives:

Logical Firewall	Role
A	○ Access control and threat prevention for Internet users accessing Internet servers ○ Site-to-site VPN termination for partners/contractors
B	○ Access control and threat prevention for the DMZ ○ Remote access termination for mobile/remote employees.
C	Access control and threat prevention for internal employees
D	Access control for out-of-band management network for system administrators
E	○ Segment between DMZ servers (outside) and Internal servers (inside) ○ Segment between each individual server tier (web, application, database) ○ Segment common IT services from other server tiers

Table 1: Logical Firewall Functions

Note that with the Palo Alto Networks next-generation firewall, all of these logical firewalling functions can be enabled on a single platform, i.e. a single network segmentation gateway, as shown in Figure 5. This provides not only cost savings by reducing the number of hardware appliances needed, but also provides operational savings through more efficient management and the ability to leverage the same hardware to address multiple security objectives.

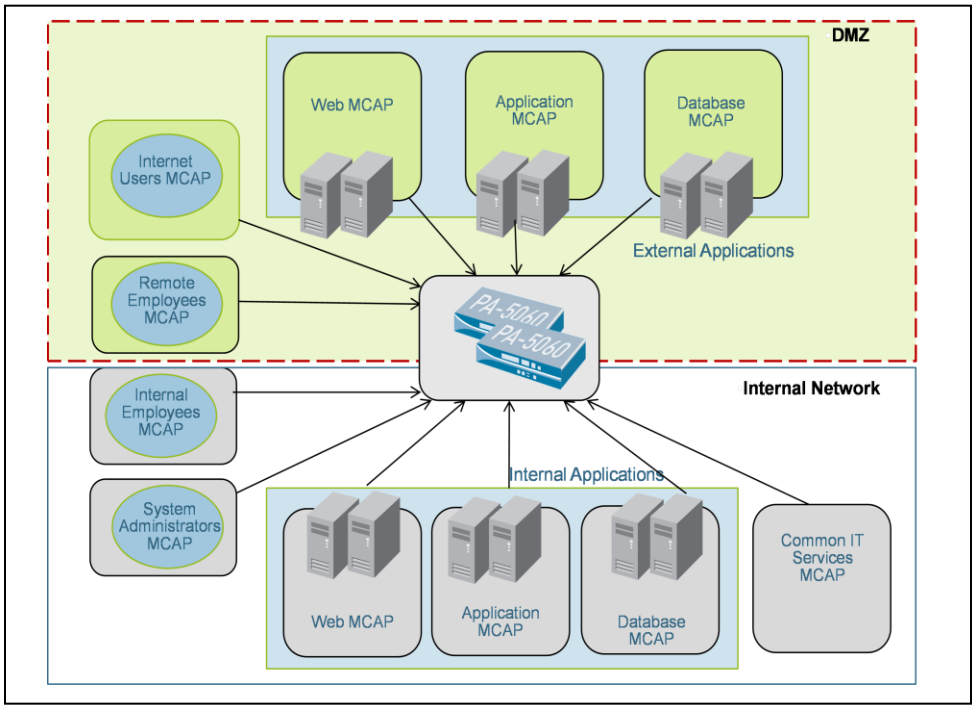

Figure 5: Palo Alto Networks next-generation Firewall as Zero Trust Network Segmentation Gateway

The servers and the users described earlier are placed in security zones. These security zones are the Zero Trust user MCAPs and server MCAPs. As a "segmentation gateway", the next-generation firewall inspects all traffic between MCAPS to enforce access control and to ensure that traffic meets security policies. Traffic between security zones traffic is selectively permitted in line with security policy and access control requirements. Note that in Figure 5, we show a pair of firewalls being deployed. Either active/passive or active/active high availability are supported, complete with session and configuration synchronization.

The advantage of the Palo Alto Network next-generation firewalls is the flexibility to support different firewall deployment options for the MCAPs. Palo Alto Networks next-generation firewalls allow you to mix and match multiple interface types– virtual wire (layer 1), layer 2 and layer 3 modes-- on a single chassis. In this example Acme network, you can deploy Palo Alto Networks as a Layer 3 firewall for the Internet user and remote employee MCAPs (logical firewalls "A" and "B" in the DMZ) while segmenting between the Internet server MCAPs in layer 2 mode (logical firewall "E").

In addition, the selection of the types of next-generation firewall services can be enabled depending on the policies. For example, it would be prudent to enable full Content-ID threat prevention capabilities for the Internet users and remote employees MCAPs.

In particular, denial-of-service protection against Internet attacks will be important for the Internet-facing servers. However, for the internal users MCAPs, the DOS protection profile can be disabled.

Internet and Remote/Mobile Employee MCAP

Access control for Internet and remote/mobile employee MCAPs will utilize Layer 3 mode firewalling due to the need for VPN routing capabilities. The same next-generation firewall is used to terminate GlobalProtect VPN for employees and external business partners. IPSec site-to-site VPN is another secure access method for external business partners. However, each group of MCAP (remote/mobile employees or business partners) can be routed to a different IP address on the same next-generation firewall for termination. This achieves further segmentation within the DMZ.

Access control with next-generation firewall technologies App-ID and User-ID is enabled on all employee and business partner MCAPs. Access control ensures that only remote/mobile employees can transition from the DMZ to the internal network. Threat prevention is enabled with Content-ID. Additional DOS protection profiles can be enabled to prevent any denial-of-service attacks.

Internal employee MCAPs

Access control for internal employee MCAPs can also utilize Layer 3 mode firewalling. Access control with next-generation firewall technologies App-ID and User-ID is enabled for access to the data center applications. It is critical within this internal data center environment to have complete visibility into all traffic to ensure only approved applications are enabled. Known data center applications should be allowed for authorized employees, management applications should be enabled only for a select group of IT users, and rogue or misconfigured applications should be remediated or dropped. A Custom-ID can be defined for custom or home-grown applications in the MCAPs.

Threat prevention via Content-ID continues to be critical to ensure that there are no threats in approved applications. Traffic categorized as “unknown” should be investigated further to understand the threat impact. In a properly designed data center environment, assuming all applications have been identified, there should be no unknown traffic. The unknown traffic category in the Application Command Center, unknown application reports, or detailed traffic and threat logs can provide more ways to drill into specific communications.

WildFire should be enabled to route unknown files to be executed in a virtual sandbox. In addition, by understanding the normal traffic patterns, anomalous traffic can be investigated further via Botnet Traffic reports.

Out-of-network management access is restricted only to system administrators. The next-generation firewall technologies play an important role here in controlling remote management applications such as RDP, Telnet and SSH that are often leveraged by attackers to access the data center. With AppID™, a group of management applications can be created and assigned to IT and supported through user-ID™, tying groups to policy. Examples include enabling SSH, SCP, SFTP for native usage, but denying

port-forwarding or tunneled SSH (local, remote, x11).

If there is a need for further management segmentation, virtual systems can be used. Each virtual system is an independent firewall within the device that is managed separately and cannot be accessed or viewed by any other system administrator. This enables Acme Enterprise to deploy a multi-tenant virtual system to support “IT as a service” for various departments in the organization.

Server MCAP segmentation

Server MCAP segmentation can utilize either Layer 2 or Layer 3 mode firewalling depending on the network configuration of the servers. Two types of segmentation are critical in the Acme Enterprise. The first is isolation between the Internet DMZ network and the internal network. The second is server to server segmentation. For example, segmentation between servers based on functions such as web, database and application servers, or based on security risk levels is important to reduce security risks and/or to meet compliance requirements. In addition, the vulnerable common IT services tier needs to be segmented from the servers.

SUMMARY

The Zero Trust model provides an innovative data-centric approach to security that protects against sophisticated and targeted attacks. The Palo Alto Networks next-generation firewalls, with a high-performance flexible platform that is able to identify, control, and safely enable applications while inspecting all content, are the right “Zero Trust Segmentation Gateways” to help organizations deploy a zero-trust network. They address the key Zero Trust concepts – securing access to data center resources, enabling granular access control and complete inspection of all traffic – without compromising the performance needs of the data center.

Citations:

[1] [2] [3] [4] [5] [6] – The following Forrester papers were referenced for this whitepaper:

- John Kindervag. “Building Security into Your Network’s DNA: The Zero Trust Network Architecture”, November 5, 2010, updated November 11, 2010
- John Kindervag. “No More Chewy Centers: Introducing the Zero Trust Model of Information Security”, September 14, 2010
- John Kindervag, “Applying Zero Trust to the Extended Enterprise”, August 5, 2010