



the network security company™

Palo Alto Networks In The Data Center: Eliminating Compromise

May 2011

Executive Summary

In principle, data center network security is easy – prevent threats, comply with regulations and enterprise policies, and do so without hindering business. In practice, however, the ever-increasing demands for application availability and performance, the constantly evolving threat landscape, and the need to understand what is happening with applications from a security perspective combine to make the seemingly easy demands for data center network security much more difficult to meet. Indeed, most organizations have been forced into significant compromises – trading security, function, and visibility for performance, simplicity, and efficiency. Compounding the issue, not all data centers are the same – internal enterprise data centers have very different missions and security requirements from Internet-facing data centers. Application and user characteristics, regulatory requirements, and additional, unique security concerns all vary between these two types of data center. For the enterprise data center, the ability to integrate network security into a variety of network architectures, the ability to segment networks based on business-relevant terms (e.g., application and user), and the ability to keep up with application developers are critical, specific demands of network security infrastructures. In contrast, managers of Internet-facing data centers, find themselves looking for increased flexibility, greater visibility, better, more easily integrated threat prevention, and yet more reliable operations.

Palo Alto Networks uniquely delivers on the network security demands found in both types of data center environments. With a foundation of innovative technologies (App-ID™, User-ID, and Content-ID) built into an architecture designed for data center performance and reliability, Palo Alto Networks next-generation firewalls offer organizations a data center network security solution that eliminates many of the unacceptable compromises previously endemic to data center network security.

Data Center Network Security = Prevent Threats, Comply, Perform

Famously, when asked why he robbed banks, American bank robber Willie Sutton famously replied, “because that’s where the money is.” While according to Sutton, the story is myth, data centers are attractive to criminals for the same reasons – because that’s where the data is (and data is either money or something equally valuable). Conceptually, data center network security is easy. In talking to enterprise customers, we find that there are three major demands in modern data center network security:

- Prevent threats
- Comply and compartmentalize
- Maintain application performance and availability

The first element, preventing threats, admittedly, has become more difficult in the last several years. Basic attacks on the infrastructure have given way to multi-vector, application-borne, sophisticated attacks that are stealthy, profit-driven, unwittingly aided by enterprise users, and in many cases, polymorphic. The level of organization associated with the development of these threats is also unprecedented. The second requirement, compliance, continues to be a major influence in data center architectures and network security. Whether it’s PCI, US health care regulations, or European privacy regulations, there are significant regulatory and compliance requirements that are pushing network segmentation deeper into organizations generally, and into data centers specifically. Finally, maintaining performance and availability is two requirements rolled into one – one of which usually translates to simplicity. Complexity usually means more integration issues, more chances for outage, and more latency. Keeping it simple is essential. The second requirement is speed – rapid processing that doesn’t introduce delays to the business. If security solutions can’t keep up, they don’t stay in the data center very long.

Data Center Network Security is Fraught With Unacceptable Compromises

Data center network security has traditionally lagged perimeter network security – for good reason: availability and performance of applications trump security. If an application hosted in a data center isn't available or responsive to users, an organization is often missing revenue opportunities – so network security controls, which all too often introduce delays and outages, are typically “streamlined.” When enterprises were building perimeter network security infrastructure using stateful inspection, data center network security used ACLs on routers. Later, when enterprise perimeter network security infrastructure embraced IPS, proxies, DLP, and other devices, some data centers just started adopting stateful inspection technology. This historical perspective exemplifies the significant trade-offs common in data center network security:

- Performance OR Security
- Simplicity OR Function
- Efficiency OR Visibility

These compromises are often “hardwired.” For example, an organization with an Internet-facing data center has had to choose between performance or security in their equipment choice: it could either choose a service provider-class firewall with ample performance capacity, but is light on security, or it could choose a perimeter-class firewall with plenty of security functionality, but is light on performance and some of the required reliability features. The problem is, once an organization made a choice, it was stuck – new designs and new products had to be implemented to alter the balance between security and performance.

Not All Data Centers Are Created Equal

Looking at a data center that hosts an insurance company's internal claims processing application and a data center that host a retail website, it's easy to see significant differences. These differences include the nature and number of applications, the quality and quantity of users, and the priority and possibility of certain security controls. Below, this paper discusses the different network security attributes and needs of the two main different types of data centers:

- Enterprise/Internal Data Centers
- Internet-Facing Data Centers

While Palo Alto Networks next-generation firewalls enable organizations to achieve the main elements of data center network security in both types of data centers, because of the significant differences, the innovative technologies from Palo Alto Networks (see Appendix) apply differently.

Enterprise/Internal Data Centers

Relatively speaking, enterprise data centers host more applications, but have fewer users. Applications come from a variety of origins – applications might be packaged, home-grown, or customized. Applications might be browser-based, but are just as likely to be client/server. Applications may be terminal-based, or virtualized. Regarding users, they are typically known – and are generally employees, contractors, or partners. See Figure 1.

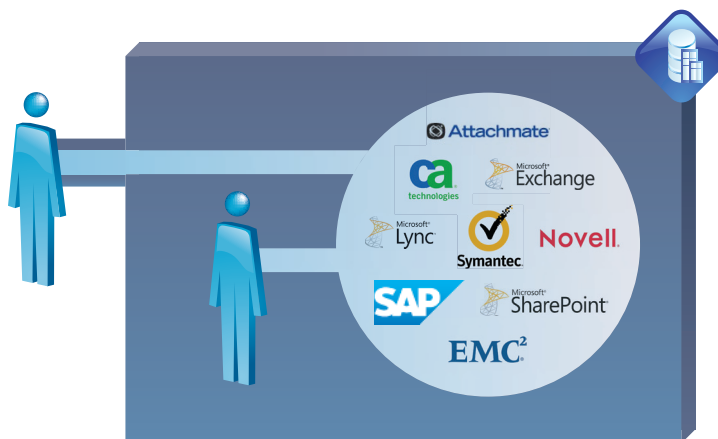


Figure 1: The Enterprise Data Center

The unique network security issues facing enterprise data centers include the need for network segmentation (typically driven by compliance), the need to keep pace with application developers, and the requirement that network security “fit” into a variety of data center designs. Another issue that enterprises are increasingly highlighting in enterprise data centers is surge in “rogue” applications. Whether it’s a rogue SharePoint installation, or an administrator using SSH on non-standard ports, these applications need to be identified and controlled on data center networks.

Palo Alto Networks in the Enterprise Data Center

As noted above, network security in the enterprise data center is often about network segmentation. While any firewall can do network segmentation, port- and IP address-based segmentation is as meaningless in the data center as it is in the perimeter – which is to say, practically worthless in the face of an application and threat mix that can, for the most part, use any open port. Furthermore, controlling access by IP address or IP address pool is an equally poor approximation for users. What enterprises are looking for, whether it’s to comply with regulations or other external requirements (e.g., PCI Appendix F), is network segmentation by user and application. So, for example, an organization can segment off the servers containing cardholder data, and only permit access to that segment to finance users employing the payments application – thus containing and limiting access, and maintaining for individual accountability. Having that level of control, and perhaps most importantly, auditability, has proven to be indispensable for many large enterprises.

Another key attribute of enterprise data centers is diversity of architectures. Part of this is due to the fact that in some organizations, internal “data centers” aren’t necessarily a single place. Which means that the usual stack of routers, core switches, access switches, and other network resources can look a little different in the face of extensive use of VLANs and distributed application components. This is another strong suit of Palo Alto Networks next-generation firewalls – the ability of these firewalls to integrate at L1 (virtual wire), L2, and L3, even operating in mixed mode across a port-dense appliance. Furthermore, the ability to trunk VLANs, aggregate ports, and perform role-based administration across security zones and virtual firewalls enables organizations to integrate next-generation firewalls into any architecture and operational model. See Figure 2 for a picture of a “firewall on a stick” design incorporating VLANs and L2 integration – it’s still inline, but flexible integration enables the enterprise customer to use its preferred network architecture.

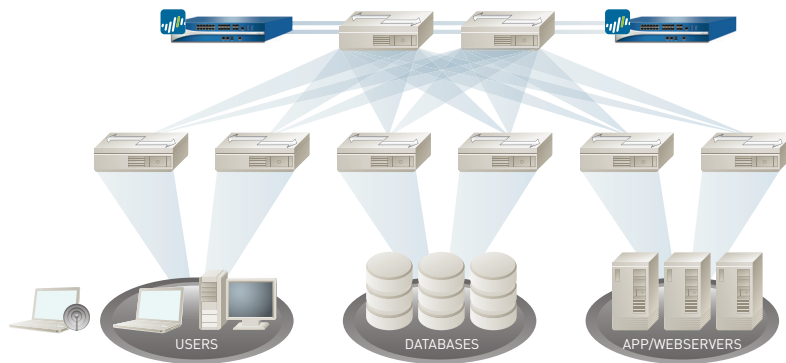


Figure 2: Enterprise Data Center Design in L2 Using VLANs with Palo Alto Networks

Finally, one of the other key uses of Palo Alto Networks next-generation firewalls in the enterprise data center is control of rogue applications. Rogue, misconfigured SharePoint deployments, unauthorized use of SSH on non-standard ports, and even P2P filesharing have been discovered and controlled in customer deployments. Another example is more operationally focused – application developers are known to implement databases and other application components on any port that is convenient. Rather than attempt to control application developers, control the applications – meaning if MySQL is an approved application between security zones, it’s allowed, regardless of which port it’s on. This greatly simplifies keeping up with developers, and safely enables key applications without increasing the attack surface.

Internet-Facing Data Centers

Conversely, in the Internet-facing data center, there are typically relatively few applications, and they’re usually web (i.e., browser-based) applications. Often, these applications will use one of the common web infrastructure “stacks” (e.g., IBM, LAMP, Microsoft, Oracle). Users are many, and often unknown/untrusted. See Figure 3.



Figure 3: The Internet-Facing Data Center

The unique network security issues Internet-facing data centers must contend with include design (e.g., the firewall is inline in a fault-tolerant, high-throughput deployment, but where do you put the IPS, given the common issues with IPS performance?), and those hardwired compromises mentioned previously, and visibility (e.g., what is being used, what is being attacked?).

Palo Alto Networks in the Internet Data Center

In the Internet-facing data center, Palo Alto Networks next-generation firewalls offer an important benefit – flexibility. Enterprises now have the option of choosing network security infrastructure that doesn’t force an either/or choice for performance and security. Changing the security posture is a policy setting – from full content scanning, to application- and user-specific firewall policies, to basic firewall policies.

This design flexibility comes with an important side benefit – simplification. See Figure 4. No longer do data center network architects have to figure out how to incorporate IPS (which is an historic choke-point). Palo Alto Networks next-generation firewalls have been tested by NSS for IPS – and not only did they achieve a top block rate (93.4%, with 100% resistance to evasion), but Palo Alto Networks next-generation firewalls outperformed datasheet claims (delivering 115% of rated throughput). So data center network architects can easily clean up their designs.

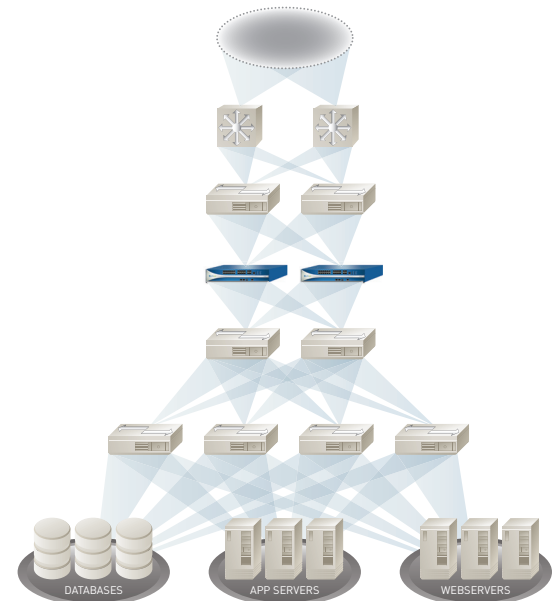


Figure 4: Internet Data Center with Palo Alto Networks

Finally, the visibility available in one spot has significant benefits. Usually, “visibility” means reviewing multiple log files, looking for the needle in a haystack. But Palo Alto Networks data center customers have found that the application visibility, the traffic visibility, coupled with the inbound URL and threat logs – all available in one user interface – eliminate the either/or choice between visibility and efficiency.

Next-Generation Firewalls Reinvent Data Center Network Security

With a handful of innovative technologies (App-ID, User-ID, Content-ID and a single-pass, parallel processing architecture – see Appendix for detail), Palo Alto Networks next-generation firewalls eliminate many of the traditional data center network security compromises enterprises have traditionally struggled with. For the first time, organizations can:

- Prevent threats
- Comply and compartmentalize
- Maintain application performance and availability

All while enabling performance, simplicity, and efficiency with security, function, and visibility.

Appendix

Palo Alto Networks Unique Technologies

Palo Alto Networks has four unique technologies to help customers achieve the network security requirements of both types of data center, while eliminating the need to compromise security, function, and visibility for performance, simplicity, and efficiency.

- App-ID
- User-ID
- Content-ID
- Single-pass, parallel processing architecture

App-ID is the Primary Classification Mechanism

Accurate traffic classification is the heart of any firewall, and the result is the basis of the security policy. Traditional firewalls classify traffic by port and protocol, which, at one point, was a satisfactory mechanism for securing the data center. Today, applications and threats can easily bypass a port-based firewall; hopping ports, using SSL and SSH, sneaking across port 80, or using non-standard ports. App-ID, a patent-pending traffic classification mechanism that is unique to Palo Alto Networks, addresses the traffic classification limitations that plague traditional firewalls by applying multiple classification mechanisms to the traffic stream, as soon as the device sees it, to determine the exact identity of applications traversing the network – whether that be a standard application, a packaged application, or a custom application.

User-ID Integrates Directory User/Group Into Network Security Policy

A standard feature on every Palo Alto Networks firewall platform, User-ID technology links IP addresses to specific user identities, enabling visibility and control of network activity on per-user basis. Tightly integrated with Microsoft Active Directory (AD) and other LDAP directories, the Palo Alto Networks User Identification Agent supports this objective in two ways. First, it regularly verifies and maintains the user-to-IP address relationship using a combination of login monitoring, end-station polling, and captive portal techniques. Next, it communicates with the AD domain controller to harvest relevant user information, such as role and group assignments. These details are then available to:

- Gain visibility (and auditability) into who specifically is responsible for all application, content, and threat traffic on the network;
- Enable the use of user identity as a variable within access control policies; and,
- Facilitate troubleshooting/incident response and be used in reports.

With User-ID, IT departments get another powerful mechanism to help control the use of applications in an intelligent manner. For example, an application holding regulated data can be enabled for individuals or groups that have a legitimate need to use it, but scanned to mitigate risks, and access logged for audit and retention requirements.

Content-ID Scans Allowed Traffic for Threats

Like its counterpart technologies, Content-ID infuses the Palo Alto Networks next-generation firewall with capabilities previously unheard of in an enterprise firewall. In this case, it's real-time prevention of threats within permitted application traffic, granular visibility of web application use, and file and data filtering.

Threat Prevention. This component of Content-ID leverages several innovative features to prevent spyware, viruses, and application vulnerabilities from penetrating the network, regardless of the type of application traffic – legacy or new-age – with which they hitch a ride.

- **Application decoder.** Content-ID leverages this App-ID component, using it to pre-process data streams that it then inspects for specific threat identifiers.
- **Uniform threat signature format.** Performance is further enhanced by avoiding the need to use separate scanning engines for each type of threat. Viruses, spyware, and vulnerability exploits can all be detected in a single pass.
- **Vulnerability attack protection (IPS).** Robust routines for traffic normalization and de-fragmentation are joined by protocol-anomaly, behavior-anomaly, and heuristic detection mechanisms to provide comprehensive protection from the widest range of both known and unknown threats.
- **Integrated URL logging.** Understand which elements of the web application in the data center are being used, or attacked.

File and Data Filtering. Taking advantage of the in-depth application inspection performed by App-ID, this set of features enables enforcement of policies that reduce the risk associated with unauthorized file and data transfer. Specific capabilities include the ability to block files by their actual type (i.e., not based on just their extension), and the ability to control the transfer of sensitive data patterns such as credit card and social security numbers.

The bottom line is that with Content-ID, IT departments gain the ability to stop known and unknown threats, increase visibility, and insure appropriate use – all without having to compromise performance, simplicity, or efficiency.

Single Pass, Parallel Processing Architecture Forms A High-Performance Foundation

First and foremost, data center network security infrastructure must perform. As stated previously, anything that doesn't perform doesn't get installed in the data center. In order to implement a true next-generation firewall, Palo Alto Networks had to develop a new architecture that could perform computationally intensive functions (e.g., application identification) at wire speed.

Palo Alto Networks next-generation firewalls use a single-pass parallel processing (SP3) architecture to protect datacenter environments at speeds of up to 20 Gbps.

The two key elements that make up the SP3 architecture are the single pass software architecture and the custom-built hardware platform. Palo Alto Networks SP3 architecture is a unique approach to hardware and software integration that simplifies management, streamlines processing and maximizes performance.

Single Pass Software

Palo Alto Networks single pass software is designed to accomplish two key functions within the Palo Alto Networks next-generation firewall. First, the single pass software performs operations once per packet. As a packet is processed, networking functions, policy lookup, application identification and decoding, and signature matching for any and all threats and content are all performed just once. This significantly reduces the amount of processing overhead required to perform multiple functions in one security device.

Second, the content scanning step in Palo Alto Networks' single pass software is stream-based, and uses uniform signature matching to detect and block threats. Instead of using separate engines and signature sets (requiring multi-pass scanning) and instead of using file proxies (requiring file download prior to scanning), the single pass software in our next-generation firewalls scans content once and in a stream-based fashion to avoid latency introduction.

This single pass traffic processing enables very high throughput and low latency – with all security functions active. It also offers the additional benefit of a single, fully integrated policy, enabling simplified management of enterprise network security.

Parallel Processing Hardware

The other critical piece of Palo Alto Networks SP3 architecture is hardware. Palo Alto Networks next-generation firewalls use multiple banks of function specific processing operating in parallel to ensure that the single pass software operates as efficiently as possible.

- **Networking:** routing, flow lookup, stats counting, NAT, and similar functions are performed on network-specific processor.
- **Security:** User-ID, App-ID, and policy lookup are all performed on a multi-core security-specific processing engine with acceleration for encryption, decryption, and decompression.
- **Threat prevention:** Content-ID uses a dedicated content scanning processor to analyze content for all manner of malware.
- **Management:** A dedicated management processor drives the configuration management, logging, and reporting without touching data processing hardware.

The final element of the architecture revolves around built-in resiliency which is delivered by the physical separation of data and control planes. This separation means that heavy utilization of one won't negatively impact the other – for example, an administrator could be running a very processor- intensive report, and yet the ability to process packets would be completely unhindered, due to the separation of data and control planes.