# The Next Generation Firewall

## The Policy and Security Control Point

***By Jim Metzler***
***Jim@Kubernan.Com***

## Introduction

In the IT industry, the phrase *next generation* is used quite frequently. Vendors often use the phrase to describe an upgrade to one of their products. In most cases, the only difference between the upgraded product and its predecessor is that the upgraded product has somewhat increased performance or supports some new feature or two. A truly next generation product should be fundamentally different than anything that is currently available on the market. However, just being fundamentally different is not that compelling. Implicit in the definition of *next generation* is that the product does a significantly better job of solving a problem that IT organizations truly care about than is done by the existing generation of products.

Security has been an important issue for IT management for the last two decades. However, as will be described in this brief, over that time frame both the types and the sophistication of security threats have increased significantly. Unfortunately, the traditional firewall is built on some key assumptions that were valid twenty years ago, but no longer are valid. As a result, the current generation of firewalls cannot adequately protect organizations from the existing and emerging set of security threats.

The goal of this brief is to describe the existing and emerging set of security threats and will discuss the limitations of the current generation of firewalls. This brief will also describe what is needed in a next generation firewall to ensure that the product can do a fundamentally better job of protecting the organization from security threats than is possible with the current generation of firewalls.

As part of the creation of this brief, two IT professionals were interviewed. One is a network architect for a global semiconductor company and the other is the senior director of IT at a medical center. They will be referred to in this brief as The Global Architect and The Senior Director respectively.

## Current Generation Firewalls

As noted, security has been a top of mind issue for IT organizations for the last two decades. The first generation of firewalls was referred to as packet filters. These devices functioned by inspecting packets to see if the packet matched the packet filter's set of rules. Packet filters acted on each individual packet (i.e., 5-tuple consisting of the source and destination addresses, the protocol and the port numbers) and did not pay any attention to whether or not a packet was part of an existing stream or flow of traffic.

Today most firewalls are based on stateful inspection. According to Wikipedia [1], "A stateful firewall is able to hold in memory significant attributes of each connection, from start to finish. These attributes, which are collectively known as the state of the connection, may include such details as the IP addresses and ports involved in the connection and the sequence numbers of the packets traversing the connection. The most CPU intensive checking is performed at the time of setup of the connection. All packets after that (for that session) are processed rapidly because it is simple and fast to determine whether it belongs to an existing, pre-screened session. Once the session has ended, its entry in the state-table is discarded."

[1] http://en.wikipedia.org/wiki/Stateful_firewall

One reason that traditional firewalls focus on the packet header is that firewall platforms generally have limited processing capacity due to architectures that are based on software that runs on an industry standard CPU. A recent enhancement of the current generation firewall has been the addition of some limited forms of application level attack protection. For example, some current generation firewalls have been augmented with IPS/IDS functionality that uses deep packet inspection to screen suspicious-looking traffic for attack signatures or viruses. However, limitations in processing power of current generation firewalls prevents deep packet inspection from being applied to more than a small minority of the packets traversing the device.

## The Use of Well-Known Ports, Registered Ports, and Dynamic Ports

In IP networks, TCP and UDP ports are endpoints to logical connections and provide the multiplexing mechanism to allow multiple applications to share a single connection to the IP network. Port numbers range from 0 to 65535. As described in the IANA Port Number document  www.iana.org/assignments/port-numbers , the ports that are numbered from 0 to 1023 are reserved for privileged system-level services and are designated as *well-known ports*. A well-known port serves as a contact point for a client to access a particular service over the network. For Example, port 80 is the well-known port for HTTP data exchange and port 443 is the well-known port for secure HTTP exchanges via HTTPS. Ports numbers in the range 1024 to 49151 are reserved for Registered Ports that are statically assigned to user-level applications and processes.  For example, SIP uses ports 5059-5061. A number of applications do not use static port assignments, but select a port dynamically as part of the session initiation process. Port numbers between 49152 and 65535 are reserved for Dynamic Ports, which are sometimes referred to as Private Ports. One of the primary reasons that stateful inspection was added to traditional firewalls was to track the sessions of whitelist applications that use dynamic ports. The firewall observes the dynamically selected port number, opens the required port at the beginning of the session, and then closes the port at the end of the session.

Most current generation firewalls make two fundamental assumptions, both of which are flawed.  The first assumption is that the information contained in the first packet in a connection is sufficient to identify the application and the functions being performed by the application.  In many cases, it takes a number of packets to make this identification because the application end points can negotiate a change in port number or perform a range of functions over a single connection.

The second assumption is that the TCP and UDP well-known and registered port numbers are always used as specified by IANA.  Unfortunately, while that may well have been the case twenty years ago that is often not the case today.  Some applications, for example, have been designed with the ability to hop between ports.  A good example of this is instant messaging (IM) software such as AOL's Instant Messenger (AIM).  AOL has been assigned well-known ports 5190 – 5193 for its Internet traffic and AIM is typically configured to use these ports. However, if these ports are blocked by a firewall, AIM endpoints will revert to TCP port 80.  Part of the security challenge associated with IM traffic using port 80 is that most IM services support file transfer and hence, potentially infected files are passing unnoticed through the firewall.  As will be discussed below, Skype is an example of an application that uses dynamic port assignments (rather than a well-known port) and occasionally selects port 80 to ensure its ability to traverse firewalls without interference.

## Port 80 and Port 443 Blind Spots

Many security experts have warned about other dangers associated with peer-to-peer networks.  For example, Antonio Nucci[2] wrote "In order to avoid detection, many peer-to-peer applications, including Skype, change the port that they use each time they start. Consequently, there is no standard (*i.e., well-known or registered*) 'Skype port' like there is a 'SIP port' or 'SMTP port'. In addition, Skype is particularly adept at port-hopping with the aim of traversing enterprise firewalls. Entering via UDP, TCP, or even TCP on port 80, Skype is usually very successful at bypassing typical firewalls. Once inside, it then intentionally connects to other Skype clients and remains connected, maintaining a 'virtual circuit'. If one of those clients happens to be infected, then the machines that connect to it can be infected with no protection from the firewall. Moreover, because Skype has the ability to port-

[2] Skype:  The Future of Traffic Detection and Classificationhttp://www.pipelinepub.com/0906/VC1.html

hop, it is much harder to detect anomalous behavior or configure network security devices to block the spread of the infection."

The key point here is that a growing number of applications, both sanctions and unsanctioned, are exploiting the firewall's blindness to the content of packets arriving on port 80. This makes it possible for network attacks to piggyback over applications like peer-to-peer file sharing, IM, or IP telephony to enter the network from the Internet and spread within the internal network.

Another blind spot of current generation firewalls is for HTTP traffic that is secured with SSL (HTTPS). HTTPS is normally assigned to well-known TCP port 443. Since the payload of these packets is encrypted with SSL, the traditional firewall cannot use deep packet inspection to determine if the traffic either poses a threat or violates enterprise policies for network usage. These two blind spots are growing in importance because they are being exploited with increasing frequency by application-based intrusions and policy violations.

## Enterprise Requirements

The Senior Director stated that his network encompasses several external sites including hospitals and clinics as well as various vendors and suppliers. He said that while they do not control those external sites they must provide access to them and that the approach they have adopted is that they "do not trust anything outside of our campus". The Global Director stated that they found themselves in the situation where they had a security policy and no ability to enforce it as they could not be sure of what applications were being used. In particular, they house their Web servers inside their DMZ so they can control the applications that run on those servers. As such, they were not concerned about managing the inbound connection to their Web servers. In contrast, they do not know what applications are running on their other servers and so they do not know what outbound traffic is being generated. Part of the concern of The Global Director is that if they were running programs such as BitTorrent they were vulnerable to being charged with breaking copyright laws. In addition, if they were supporting recreational applications such as Internet Radio, they were wasting a lot of WAN bandwidth. He summed up his feelings by saying, "You think that you are in a secure environment. However, at the end of the day a lot of applications that were declared as outlaws are still running on your network."

Asked about the limitation of traditional firewalls, The Senior Director said that traditional firewalls do not provide any application layer filtering so if you are attacked above Layer 3 "you are toast". The Global Architect said that in theory an IT organization could mitigate the limitations of a traditional firewall by implementing a traditional firewall combined with other security related functionality such as an IPS. However, he stressed that this is only a theory because the IT organization would never have enough knowledge of the applications to make this work. This point was picked up on by The Senior Director who stated that his organization had been looking at adding other security functionality such as IDS, IPS and NAC. What he wanted, however, was to avoid the complexity of having a large number of security appliances. He preferred to have a "firewall on steroids" provide all this functionality.

## A Next Generation Firewall

The comments of The Global Architect and The Senior Director serve to underscore some of the unnatural networking that has occurred over the last decade. In particular, firewalls are typically placed at a point where all WAN access for a given site coalesces. This is the logical place for a policy and security control point for the WAN. Unfortunately due to the lack of a 'firewall on steroids' that could provide the necessary security functionality, IT organizations have resorted to implementing myriad firewall helpers [3].

It is understandable that IT organizations have deployed work-arounds to attempt to make up for the limitations of traditional firewalls. This approach, however, has serious limitations including the fact that the firewall helpers often do not see all of the traffic and the deployment of multiple security appliances significantly drives up the operational costs and complexity.

[3] Now Might Be a Good Time to Fire Your Firewall, http://ziffdavisitlink.leveragesoftware.com/blog_post_view.aspx?BlogPostID=603398f2b87548ef9d51d35744dcdda4

In order for the firewall to avoid these limitations and reestablish itself as the logical policy and security control point for the WAN, what is needed is a next generation firewall with the following attributes:

**Application Identification:** The firewall must be able use deep packet inspection to look beyond the IP header 5-tuple into the payload of the packet to find application identifiers. Since there is no standard way of identifying applications, there needs to be an extensive library of application signatures developed that includes identifiers for all commonly used enterprise applications, recreational applications, and Internet applications. The library needs to be easily extensible to include signatures of new applications and custom applications. Application identification will eliminate the port 80 blind spot and allow the tracking of port-hopping applications.

**Extended Stateful Inspection:** By tracking application sessions beyond the point where dynamic ports are selected, the firewall will have the ability to support the detection of application-level anomalies that signify intrusions or policy violations.

**SSL Decryption/Re-encryption:** The firewall will need the ability to decrypt SSL-encrypted payloads to look for application identifiers/signatures. Once this inspection is performed and policies applied, allowed traffic would be re-encrypted before being forwarded to its destination. SSL proxy functionality, together with application identification, will eliminate the port 443 blind spot.

**Control:** Traditional firewalls work on a simple deny/allow model. In this model, everyone can access an application that is deemed to be *good*. Analogously, nobody can access an application that is deemed to be *bad*. This model had more validity at a time when applications were monolithic in design and before the Internet made a wide variety of applications available. Today's reality is that an application that might be *bad* for one organization might well be *good* for another. On an even more granular level, an application that might be *bad* for one part of an organization might be *good* for other parts of the organization. Also, given today's complex applications, a component of an application might be *bad* for one part of an organization but that same component might well be *good* for other parts of the organization.

What is needed then is not a simple deny/allow model, but a model that allows IT organizations to set granular levels of control to allow the *good* aspects of an application to be accessed by the appropriate employees while blocking all access to the *bad* aspects of an application.

**Multi-gigabit Throughput:** In order to be deployed in-line as an internal firewall on the LAN or as an Internet firewall for high speed access lines, the next generation firewall will need to perform the above functions at multi-gigabit speeds. These high speeds will be needed to prevent early obsolescence as the LAN migrates to 10 GbE aggregation and core bandwidths, and as Internet access rates move to 1 Gbps and beyond via Metro Ethernet. Application Identification and SSL processing at these speeds requires a firewall architecture that is based on special-purpose programmable hardware rather on than industry standard general-purpose processors. Firewall programmability continues to grow in importance with the number of new vulnerabilities cataloged by CERT hovering in the vicinity of 8,000/year.

When asked about the attributes that he expects in a next generation firewall, The Global Director said that the ability to learn about applications on the fly was a requirement as was the need to run a multi-gigabit speed. Critical to The Global Director is the ability to tie an event to a user. To exemplify that he said, "If somebody is communicating using BitTorrent, my ability to tie that application to a user is critical. I can do that with a traditional firewall, but it is a management nightmare."

The Senior Director agreed on the importance of application level visibility and high performance. He also stressed the importance of reporting and alerting when he said, "I need the ability to push security-related information from engineering to the help desk. The next generation firewall must not be so complicated that the average help desk analyst cannot input a rule set. It must also be simple enough for the people at the help desk to be able to use it to analyze what is going on."

## Summary

Twenty years is a very long time in IT. Twenty years ago CERT was just being established and now it catalogs roughly 8,000 new security vulnerabilities a year. Twenty years ago simple packet filtering was sufficient for the vast majority of enterprises. It has not been for a long time. More recently, firewalls based on stateful inspection were sufficient for the vast majority of enterprises. As pointed out in this brief, however, today the limitations of these firewalls are significant. And, as expressed by both The Global Director and The Senior Director, merely bolting functionality such as IDS/IPS onto a traditional firewall is helpful, but does not allow IT organizations to truly cope with current security vulnerabilities.

For most IT organizations, the firewall is a primary component of their security strategy. One of the main reasons for that is due to the typical placement of a firewall at a point were all WAN access for a given site coalesces. Given this placement in the network, the firewall is positioned to be a primary policy and security control point – but only if it can provide the necessary functionality.

To provide the necessary functionality requires a next generation firewall with an architecture that allows it to overcome the limitations of the traditional firewall. Some of the primary attributes of that firewall include application identification at wire speed, the detection of anomalies, and the ability to decrypt and re-encrypt SSL traffic. However, as stressed by The Senior Director, a next generation firewall has to provide value to more than a small set of sophisticated engineers. A next generation firewall must be usable by a wide array of people with ranging needs and technical abilities.

## A Word from the Sponsor – Palo Alto Networks

Palo Alto Networks™ enables visibility and policy control of applications running on enterprise networks. Based on innovative App-ID(tm) application classification technology, the Palo Alto Networks PA-4000 Series is a next-generation firewall that accurately identifies applications - regardless of port, protocol, evasive tactic or even SSL encryption - at 10Gbps with no performance degradation. Enterprises can now set and enforce application usage policies to meet compliance requirements, improve threat mitigation and lower operational costs. The Palo Alto Networks team includes security and networking industry veterans from Check Point, NetScreen, McAfee, Cisco, Juniper and Blue Coat. It is backed by investors Globespan Capital Partners, Greylock Partners and Sequoia Capital. For more information, visit www.paloaltonetworks.com.

## About Kubernan™

Kubernan™, a joint venture of industry veterans Steven Taylor and Jim Metzler, is devoted to performing in-depth analysis and research in focused areas such as Metro Ethernet and MPLS, as well as in areas that cross the traditional functional boundaries of IT, such as Unified Communications and Application Delivery. Kubernan's focus is on providing actionable insight through custom research with a forward looking viewpoint. Through reports that examine industry dynamics from both a demand and a supply perspective, the firm educates the marketplace both on emerging trends and the role that IT products, services and processes play in responding to those trends.

Kubernan is the Greek root word for *helmsman* as well as the phrases to guide and to steer. As such, the name Kubernan reflects our mission of guiding the innovative development and usage of IT products and services.

**Kubernan Briefs**
Vol.1, Number 6

**Published by Kubernan**
www.Kubernan.com

**Cofounders:**
Jim Metzler
*jim@kubernan.com*
Steven Taylor
*taylor@kubernan.com*

**Professional Opinions Disclaimer**
All information presented and opinions expressed in this publication represent the current opinions of the author(s) based on professional judgment and best available information at the time of the presentation. Consequently, the information is subject to change, and no liability for advice presented is assumed. Ultimate responsibility for choice of appropriate solutions remains with the reader.

For editorial and sponsorship information, contact Jim Metzler or Steven Taylor. Kubernan is an analyst and consulting joint venture of Steven Taylor and Jim Metzler.