



Understanding and Selecting an Enterprise Firewall

Version 1.2

Released: October 18, 2010

Author's Note

The content in this report was developed independently of any sponsors. It is based on material originally posted on the Securosis blog <<http://securosis.com>>, but has been enhanced, reviewed, and professionally edited.

Special thanks to Chris Pepper for editing and content support.

Licensed by Palo Alto Networks

Palo Alto Networks™ is the network security company. Its next-generation firewalls enable unprecedented visibility and granular policy control of applications and content – by user, not just IP address – at up to 10Gbps with no performance degradation. Based on patent-pending App-ID™ technology, Palo Alto Networks firewalls accurately identify and control applications – regardless of port, protocol, evasive tactic or SSL encryption – and scan content to stop threats and prevent data leakage. Enterprises can for the first time embrace Web 2.0 and maintain complete visibility and control, while significantly reducing total cost of ownership through device consolidation. For more information, visit www.paloaltonetworks.com.



Copyright

This report is licensed under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0.



<http://creativecommons.org/licenses/by-nc-nd/3.0/us/>

Table of Contents

Introduction	5
Application Awareness	7
Blind Boxes and Postmen	7
Use Case: Visibility	8
Use Case: Blocking	8
Perimeter Disruption	9
Complexity is not your friend	10
Technical Architectures	11
Packet Processing Evolves	11
Application Profiles	12
Identity Integration	12
Management Evolution	13
Scalability	13
Product Line Consistency	14
Embedded Firewalls	15
Trust, but Verify	15
Deployment Considerations	16
Bandwidth Matters	16
High Availability Clusters	17
Internal Deployment	17
Migration	17
Firewall Management	19

What to Manage?	19
Checking the Policy	20
Reporting	20
Other Management Considerations	21
Advanced Features	22
Application Visibility	22
Application Blocking	23
Overlap with Existing Web Security	23
Bot Detection	24
Content Inspection	24
Vulnerability Integration	25
To UTM or not to UTM?	26
Selection Process	29
Define Needs	29
Formalize Requirements	30
Evaluate Products	30
Selection and Deployment	31
Conclusion	32
About the Analyst	33
About Securosis	34

Introduction

What? A research report on enterprise firewalls. Really? Most folks figure firewalls have evolved about as much over the last 5 years as ant traps. They're wrong, of course, but most people think of firewalls as old, static, and generally uninteresting. In fact, most security folks begin their indentured servitude looking after the firewalls, where they gain seasoning before anyone lets them touch important gear like the IPS.

But this perception is unfounded. Firewalls continue to evolve and these new capabilities can and should impact your perimeter architecture and firewall selection process. That doesn't mean we will be advocating yet another rip and replace job at the perimeter (sorry vendors), but there are definitely new capabilities that warrant consideration, especially as the maintenance renewals on your existing gear come due.

To state the obvious, the firewall tends to be the anchor of the enterprise perimeter, protecting your network from most of the badness out there on the Intertubes. We also see some use of internal firewalls, driven mostly by network segmentation requirements. Pesky regulations like PCI mandate that private data is at a minimum logically segmented from non-private data, so some organizations use firewalls to keep their *in scope* systems separate from the rest, although most organizations use network-level technologies to implement their segmentation.

Firewalls continue to evolve and these new capabilities can and should impact your perimeter architecture and firewall selection process.

In the security market, firewalls reside in the *must have* category along with anti-virus (AV). There must be organizations that don't use firewalls to protect their Internet connections, but we have yet to come across one. Those are the same companies which give a blank, vacant stare when asked what they do to protect critical data. The ubiquity of the technology means we see a huge range of price points and capabilities across firewalls.

Consumer products aside, firewalls range in price from about \$750 to over \$250,000. Yes, you can spend a quarter of a million dollars on a firewall. It's not easy, but you can do it. Obviously there is a huge difference between the low end boxes protecting branch and remote offices and the gear protecting the innards of a service provider network, but ultimately they both do the same thing. A firewall protects one network from another based on a defined set of rules. In this report we focus on the *enterprise firewall*, which is designed for use in larger organizations (2,500+ employees). That doesn't mean our research isn't relevant to smaller companies, but enterprise is the focus.

From an innovation standpoint, not much happened on firewalls for a long time. But then three major trends hit, which are forcing a general re-architecting of firewalls:

- **Performance/Scale:** Networks aren't getting slower and that means the perimeter must keep pace. Where Internet connections used to be sold in multiples of T1 speed (1.5megabits/sec), now we see speeds in the hundreds of megabits/sec or gigabits/sec, and supporting internal network segmentation and carrier uses requires the ability to scale up to and past 10gbps. This is driving new technical architectures which better utilize advanced packet processing and custom chips.
- **Integration:** Most network perimeters have evolved along with the threats. That means the firewall/VPN is there, along with an IPS, but also an anti-spam gateway, web filter, web application firewall, and probably 3-4 other types of devices. Yeah, this perimeter sprawl creates a management nightmare, forcing customers to push for integration of some of these capabilities (if not all) into a single device. Most likely it's firewall and IDS/IPS, but there is clearly growing interest in broader integration (UTM: unified threat management) even at the high end of the market.
- **Application Awareness:** It seems everything nowadays gets encapsulated into port 80 or 443. That means your firewall has no visibility into what's really happening, making the device like making the device as blind as the proverbial mouse to a large portion of your traffic. This is clearly problematic, and causes much of the perimeter sprawl described above, as you deploy a whole bunch of different boxes to protect different types of traffic. But through the magic of Moore's law and some savvy integration of IPS-like capabilities, the firewall can enforce rules on specific applications. This climbing of the stack by the firewall will have a dramatic impact on not just firewalls, but also IDS/IPS, web filters, WAFs, and network-layer DLP before it's over. This is one of our key research positions, so we'll be spending a lot of time on application awareness.

Application Awareness

We see three main forces driving firewall evolution. The first two are pretty straightforward and don't require a lot of explanation or debate. First, networks are getting faster and so the perimeter gateways need to get faster. That's not brain surgery.

Second, most end users have been dealing with significant perimeter security sprawl, meaning where they once had a firewall they now have 4-5 separate devices, so they are looking for integrated capabilities. Depending on performance requirements, organizational separation of duties, and good old fashioned politics, some enterprises are more receptive than others to integrated gateway devices (yes, UTM-like things) for a simple reason. Less devices = less complexity = less management angst = happier administrators. WOOT! Again, not brain surgery.

But these drivers really just fall into the category of bigger and faster — not really *different*. The one aspect of perimeter protection we see truly changing is the need for these devices to become *application aware*. That means you keep policies and rules based on not just port, protocol, source, destination, and time — but also on application, and perhaps even specific activities within an application.

This one concept will drive a total overhaul of the enterprise perimeter. Not today or tomorrow — regardless of vendor propaganda — but certainly over the next 5 years. The old saying, “*We overestimate progress over a 1-2 year period, but usually significantly underestimate progress over a 10 year period.*” applies here. We believe that is true for application awareness within network security devices.

Blind Boxes and Postmen

Encapsulated traffic makes your firewall blind to most of the traffic coming through it.

A good analogy from the email security space helped to describe the need for an anti-spam appliance a few years ago. Think about the security guards in a typical large enterprise. They are sitting in the lobby, looking for things that don't belong. That's your firewall. But think about the postman, who shows up every day with a stack of mail. That's port 25 traffic (SMTP). The firewall says, “Hey, Mr. Postman, come right in,” regardless of what is in the mail bin. Most of the time that's fine, but sometimes a package is ticking and the security guard will miss it.

So the firewall is blind to what happens within port 25. Now replace port 25 with port 80 (or 443), which represents web traffic, and you are in the same boat. Your security guard (firewall) expects that traffic, so it goes right on through, regardless of the payload. And application developers know that, so it's much easier to just encapsulate application-specific data and/or protocols within port 80 or 443 so they can easily traverse most firewalls. On the other hand, that makes *your* firewall blind to most of the traffic coming through it. As a bat.

That's why most folks aren't so interested in firewall technology any more. It's basically a traffic cop, more of a networking technology than a security device. It tells you where you can go, but doesn't really protect much of anything. This has driven web application firewalls, web filters, email gateways, and even IDS/IPS devices to sit behind the firewall to actually protect things. Of course, that isn't the most efficient way to do things.

This is also problematic for one of the key fundamentals of network security — [Default Deny](#): rejecting all traffic that is not explicitly allowed. Obviously you can't just block port 80, which is why so many applications (both good and bad) use it to get that free ride around default deny policies.

So that's the background for why application awareness is important. Now let's get into some tangible use cases to further illuminate its importance.

Use Case: Visibility

Do you know what's running on your networks? Yeah, we know that's a loaded question, but most network/security folks don't. They may lie about it, and some actually do a decent job of monitoring, but most don't. They have no idea the CFO is watching stuff he shouldn't be. They have no idea the junior developer is running a social network from the high-powered workstation under his desk. They also don't know the head of engineering is sending critical intellectual property to an FTP server outside the country.

Well, they don't know until it's too late. So one of the key drivers for application awareness is visibility. We've seen this before, haven't we? Remember how web filters were first positioned? Right, as employee productivity tools — *not security devices*. It was about making sure employees weren't violating Internet acceptable use policies. Only afterwards did folks realize how much bad stuff is out there on the web that should be blocked.

In terms of visibility, you want to know not just how much of your outbound traffic is Facebook, or how much of your inbound traffic is from China, or from a business partner. You want to know what Mike Rothman is doing at any given time. And how many folks (and from where) are hitting your key Intranet site through the VPN. The questions are endless once you can actually peek into the application traffic and really understand what is happening. And alert on it. Cool, right?

The possibility for serious eye candy is also compelling. We all know senior management likes pie charts. This kind of visibility enables some pretty cool pie charts. You can pinpoint exactly what folks are doing on both ingress and egress connections, and isolate issues that cause performance and security problems. Did we mention that senior management likes pie charts?

Use Case: Blocking

As described above, the old-style firewall doesn't really block sophisticated attacks nowadays because it's blind to the protocols comprising the bulk of inbound and outbound traffic. OK, that's a bit harsh, but it certainly doesn't block what we need it to block. We rely on other devices (WAF, web filter, email security gateway, IPS) to do the blocking. Mostly via a *negative* security model, meaning you are looking for specific examples of bad behavior — this is how IPS, web filters, and email gateways work. Obviously that means you need to profile every bad thing that can possibly happen, learn to recognize it, and then look for it in every packet or message that comes in or goes out. Given the infinite possibilities for badness that's a tall order — okay, it's completely ridiculous and impossible.

But if we have the ability to look into the traffic and profile applications, we can build policies and rules to govern how they can be used. We can block traffic unless the rules are followed, which represents a *positive* security model. Now that would be cool. In fact, this kind of capability really enhances another of the Network Security Fundamentals, [egress filtering](#). Being able to both profile and block traffic going out, based on application characteristics, provides a lot of power to disrupt exfiltration before you have to disclose a breach to all your pissed-off customers.

The other main use case for application awareness is to block certain traffic (both ingress and egress) that violates policy. Obviously this opens up a world of possibilities in terms of integration with identity stores. For example, the marketing group can use Facebook during business hours, but the engineering team cannot. You could also enforce specific application activity, so perhaps Finance can enter payroll into the payroll SaaS system, but factory workers can only view pay stubs. You can even enforce privileged user monitoring via this type of capability, monitoring DBA attempts to access the back-end database from remote locations and (possibly) allowing them, but blocking anyone else. The possibilities are endless. But this kind of functionality will have a significant effect on your perimeter architecture.

Perimeter Disruption

As cool as application awareness on the firewall is, it's not without challenges and overlap with existing technology. Whether you want to call it *disruptive* or *innovative* or something else, introducing new capabilities into your technology stack tends to have a ripple effect on everything else. This is no exception.

Let's run through the other security devices usually present on your perimeter and get a feel for whether these newfangled firewalls can replace and supplant, or just supplement, these other devices. Clearly you want to simplify the perimeter where you can, and part of that is reducing the device footprint.

- **IDS/IPS:** Are application aware firewalls a threat to IDS/IPS? In a nutshell, yes. In fact, as we'll see when we examine technical architectures, a lot of the application aware firewalls actually use an IPS engine under the covers to provide application support. In the short term, the granularity and maturity of IPS rules mean you probably aren't turning your IPS off, yet. But over time, the ability to profile applications and enforce a positive security model definitely will impinge on what a traditional IDS/IPS brings to the table.
- **Web application firewall (WAF):** Clearly being able to detect malformed web requests and other simple attacks is possible on an application aware firewall. But providing complete granular web application defenses, such as automated profiling of web application traffic and specific application calls (as a WAF does) are not as easily duplicated via the vendor-delivered application libraries/profiles, so we still see a role for the WAF to protect inbound traffic directed at critical web apps. Over time, though, it looks pretty clear that these granular capabilities will show up in application aware firewalls.
- **Secure Email Gateway:** Most email security architectures today involve a two-stage process of getting rid of the unsolicited email by first using reputation and connection blocking, then doing in-depth filtering and analysis of the

*Whether you want to call it **disruptive** or **innovative** or something else, introducing new capabilities into your technology stack tends to have a ripple effect on everything else.*

remaining message content. We clearly see a role for application aware firewalls to provide reputation and connection blocking for inbound email traffic, but believe it will be hard to duplicate the kind of analysis present on email security gateways. That said, end users increasingly turn to service providers for anti-spam capabilities, so over time this feature is decreasing in importance for the perimeter gateway.

- **Web Filters:** In terms of capabilities, there is a tremendous amount of overlap between the application aware firewall and web filtering gateways. Obviously web filters have gone well beyond simple URL filtering, which is already implemented on pretty much all firewalls. But some of the advanced heuristics and visibility aspects of the web security gateways are not particularly novel, so we expect significant consolidation of these devices into the application aware firewall over the next 18 months or so. We believe this will be the first category to fall in favor of the application aware firewall.

Ultimately the role of the firewall in the short and intermediate term is going to remain the coarse filter sitting in front of specialized devices. Over time, as customers get more comfortable with the overlap (and realize they may not need all the capabilities on the specialized boxes), we'll start to see significant cannibalization on the perimeter. That said, most of the vendors moving towards application aware firewalls already have many of these devices in their product portfolios. So it's likely to be about neutral to the vendor whether IPS capabilities are implemented on the perimeter gateway or a device sitting behind it.

Complexity is not your friend

Yes, these new devices add a lot of flexibility and capability in terms of how you protect your perimeter. But with that flexibility comes potentially significant complexity. With your current rule base probably numbering in the thousands of rules, think about how many more you'd need to set up to control specific applications. And then to control how specific groups use specific applications. Right, it's mind numbing. And you'll also have to revisit these policies far more frequently, because apps are always changing and so identifying and enforcing acceptable behavior may also need to change.

Don't forget the issues around keeping application support up to date, either. It's a monumental task for the vendor to constantly profile important applications, understand how they work, and be able to detect the traffic as it passes through the gateway. This kind of endeavor never ends because the applications are always changing. There are new applications being implemented and existing apps change under the covers, which impacts protocols and interactions. So one of the key considerations in choosing an application aware firewall is comfort with the vendor's ability to stay on top of the latest application trends.

The last thing you want is to either lose visibility or not be able to enforce policies because Twitter changed their authentication process (which recently happened). It pretty much defeats the purpose of having an application aware firewall in the first place.

All this potential complexity means application blocking technology still isn't simple enough for widespread deployment. But it doesn't mean you shouldn't be playing with these devices or thinking about how application visibility and blocking can bolster existing defenses for well known applications. It's really more about figuring out how to gracefully introduce the technology without totally screwing up the existing security posture. We'll talk a lot more about that when we get to deployment considerations.

Technical Architectures

Now let's dig into the technical goodies that enable firewall evolution. We won't rehash the history of the firewall — that's what Wikipedia is for. Suffice it to say the [firewall](#) started with [application proxies](#), which led to [stateful inspection](#), which was supplemented with [deep packet inspection](#). Now every vendor has a different way of talking about their ability to look into packet streams moving through the gateway, but fundamentally they're not all that different.

Our main contention is that application awareness (building policies and rules based on how users interact with applications) isn't something that fits well into the existing firewall architecture. Why? Basically, the current technology (stateful + deep packet inspection) is still focused on ports and protocols. Yes, there are some things (like bolting an IPS onto the firewall) that can provide rudimentary application support, but ultimately we believe the existing firewall architecture is on its last legs.

*Ultimately we believe
the existing firewall
architecture is on
its last legs.*

Packet Processing Evolves

So what is the difference between what we see now and what we need? Basically it's about the number of steps to enforce an application-oriented rule. Current technology can identify the application, but then needs to map it to the existing hierarchy of ports and protocols. Although this all happens behind the scenes, doing all this mapping in real time at gigabit speeds is *very* resource intensive. Clearly it's possible to throw hardware at the problem, and at lower speeds that's fine. But this approach won't work forever.

The long term answer is a *brain transplant* for the firewall, and we are seeing numerous companies adopting a new architecture based not on ports and protocols, but on specific applications and identities instead. So once the application is identified, rules can be applied directly *to the application*, or *to the user/group for that application*. State is now managed for the specific application (or user/group). No mapping, no performance hit. Identifying that application is not a one time activity either, since policies and state must be continually evaluated since application activity can change at any time (ah, the wonders of the web). For instance, you may allow a user to check their GMail account, but not do GTalk. The session information remains the same (since GTalk can be launched from within GMail), thus the firewall needs to constantly apply application policies to each packet that traverses the firewall. Right, this requires serious horsepower.

Again, at lower speeds it'll be hard to decipher which architecture a specific vendor is using, but turn on a bunch of application rules and crank up the bandwidth, and old architectures will grind to a stop. The only way to figure it out for your specific traffic is to actually test it, but that's getting a bit ahead of ourselves. We'll talk about that later, when we discuss the selection process.

Application Profiles

For a long time, security research was the purview of the anti-virus vendors, vulnerability management folks, and IDS/IPS guys. They had to worry about these 'signatures', which were basically profiles of bad things. Their devices enforce policies by looking for bad stuff: a typical negative security model.

The new firewall architecture enables rules to look only for the good applications, and to block everything else. A positive security model makes a lot more sense strategically. We cannot continue looking for, identifying, and enumerating bad stuff because there is an infinite amount of it, but the number of good things that are specifically authorized is much more manageable. We should mention this does overlap a bit with typical IPS behavior (in terms of blocking stuff that isn't good), and clearly there will be increasing rationalization of these functions on the perimeter gateway.

In order to make this architecture work, the application profiles (to recognize application #1 vs. application #2) must be correct. If you thought bad IPS rules wreak havoc (false positives, blocked traffic, & general chaos), wait until you implement a screwy firewall application profile. So as we have mentioned numerous times in the [Network Security Operations Quant series on Managing Firewalls](#), testing these profiles and rules multiple times before deployment is critical.

If this sounds a bit like application white listing on the endpoint, you're be right. So all of the potential issues we see in endpoint white listing — including impacting the user experience, breaking key applications, and ensuring a strong process to manage exceptions apply here. Not that this can or should stop the adoption of the technology (as it shouldn't impact endpoint white listing either), but it still creates a perception issue and must be handled carefully. Again, that's why testing application oriented policies is so important. You want the application aware firewall to be invisible to users — until the violate policy, that is.

It also means firewall vendors need to make a significant and ongoing investment in application research, because many of these applications are deliberately difficult to identify. With a variety of port hopping and obfuscation techniques used even by the good guys (to enhance performance mostly, but also to easily traverse firewalls), digging deeply into a vendor's application research capabilities will be a big part of choosing between devices.

We also expect open interfaces from the vendors to allow enterprise customers to build their own application profiles. As much as we'd like to think all our applications are all web-friendly and stuff, not so much. So in order to truly support all applications, customers will need to be able to build and test their own profiles.

Identity Integration

Take everything we just said about applications and apply it to identity. Just as we need to be able to identify applications and apply certain rules to these behaviors, we need to apply rules to specific users and groups as well. That means integration with the dominant identity stores (Active Directory, LDAP, RADIUS, etc.) becomes very important.

Do you *really* need real-time identity sync? Probably not. Obviously if your organization has lots of moves/adds/changes and they need to be reflected by real-time access control, then the sync window should be minutes rather than hours. But for most organizations, a couple hours should suffice, if not a daily batch job. Just keep in mind that syncing with the firewall is unlikely to be the bottleneck in your identity management process. Most organizations have a significant lag (a day or more) between when a personnel change happens and when it filters through to the directories and other application access control technologies.

Management Evolution

As we have already described, thinking in terms of applications and users — rather than ports and protocols — can add significantly to the complexity of setting up and maintaining the rule base. So enterprise firewalls leveraging this new architecture need to bring forward enhanced management capabilities. Slick application awareness features are useless if you cannot configure them. That means built-in policy checking/testing capabilities, better audit and reporting, and preferably a mechanism to check which rules are useful based on *real traffic*, not a simulation.

A cottage industry has emerged to provide enterprise firewall management, particularly focused on auditing and providing a workflow for configuration and rule changes. But let's be clear: if the firewall vendors didn't suck at management, there would be no market for those tools. So a key aspect of looking at these updated firewalls is to make sure the management capabilities will make things easier for you, rather than harder.

Scalability

The reality of the firewall market remains that most of the propaganda pushed by the firewall vendors revolves around speeds and feeds. Of course, in the hands of savvy marketers in mature markets, it seems less than 10gbps magically becomes 40gbps, 20gbps becomes 100gbps, and software on an industry-standard blade becomes a purpose-built appliance. No wonder buying anything in security remains such a confusing and agonizing endeavor.

In a market dominated by what we lovingly call “bit haulers” (networking companies), everything gets back to throughput and performance. And to be clear, throughput is important — especially depending on how you want to deploy and which security capabilities you want to implement. But you also need to be very wary of the religious connotations of any speeds and feeds discussion, and able to wade through the cesspool without getting lost, to determine the best fitting firewall for your environment.

Here are a few things to consider:

- **Top Speed:** Most vendors want to talk about the peak throughput of their devices. In fact many pricing models are based on this number — which is mostly useless. You see, a 100gbps firewall under the right circumstances can process 100gbps. But turn anything on — like more than two filtering rules, a few application policies, and/or identity integration, and you'll be lucky to get a fraction of the specified throughput. So it's far more important to understand *your requirements*, which will then give you a feel for the real-world top speed you need. And testing is your chance to confirm the device can keep up.
- **Proprietary vs. industry-standard hardware:** Two camps exist in the enterprise firewall market: those who spin their own chips and those who don't. The chip folks have all these cool pictures that show how their proprietary chips enable all sorts of cool things. On the other hand, the guys who focus on software tell stories about how they take advantage of cool hardware technologies in industry-standard chips (read: Intel processors). This is mostly just religious/PR banter, and not very relevant to your decision process. The fact is, you are buying an enterprise firewall, which needs to be a perimeter gateway **solution**. How it's packaged and who makes the chips shouldn't really matter to you. *The real*

*It's far more important to understand **your requirements**, which will then give you a feel for the real-world top speed you need.*

question is whether the device will provide the services you need at the speed you require. There is no place for religion in buying security devices.

- **UTM:** Many of the players in this space talk about their ability to add capabilities such as IDS/IPS and content security to their devices to reduce complexity. Again, if you are buying a firewall, buy a firewall. In an enterprise deployment, turning on these additional capabilities may kill the performance of the firewall, which kind of defeats the purpose of buying an evolved firewall. That said, there are clearly use cases where the simplicity offered by UTM is a consideration (especially smaller/branch offices) and having that capability can swing the decision. We will discuss these issues later in the paper. First and foremost, make sure you can meet your *firewall* requirements, and keep in mind that additional UTM features may not be important to the enterprise firewall decision.
- **Networking functions:** A major part of the firewall's role is to be a *traffic cop* for both ingress and egress traffic passing through it. So it's important that your device can run at the speeds required for the use case. If the plan is to deploy it in the data center to segment credit card data, then playing nice with the switching infrastructure (VLANs, etc.) is key. If the device is to be deployed on the perimeter, how well it plays with the IP addressing environment (network address translation) and perhaps bandwidth rate-limiting capabilities (by destination or application) are important. Are these features that will make or break your decision? Probably not, but if your network is a mess (you are free to call it 'special' or 'unique'), then good interoperability with the network vendor is important, and may drive you toward security devices offered by your primary network vendor.

So it's critical that in the initial stage of procurement you be very clear about what you are buying and why. If it's a firewall, that's great. If you need some firewall capabilities plus other stuff, that's great too. But figure this out early in the process, because it shapes the way you make this decision.

Product Line Consistency

Given the significant consolidation that has happened in the network security business over the past 5 years, another aspect of the technical architecture is product line consistency. By that, we mean the degree to which devices within a vendor's product line offer the same capabilities and user experience. In an enterprise rollout you'll likely deploy a range of different-sized devices, depending on location and which capabilities each deployment requires.

Successfully managing these devices requires enforcing a consistent policy across the enterprise.

Usually we don't much care about the underlying guts and code base these devices use, because we buy *solutions* to problems. But we do have to understand and ask whether the same capabilities are available up and down the product line, from the small boxes that go in branches to the big box sitting at HQ. Why? Because successfully managing these devices requires enforcing a consistent policy across the enterprise, and that's hard if you have different devices with different capabilities and management requirements.

We also need to mention the *v-word*: virtualization. A lot of the vendors (especially the ones praying to the software god) offer their firewalls as virtual appliances. If you can get past the idea that the anchor of your secure perimeter could be abstracted and run on top of a hypervisor, this opens up a variety of deployment alternatives. But again, you need to

ensure that a consistent policy can be implemented, the user experience is the same, and ultimately all the relevant capabilities from the appliances are also available in the VM version.

As we've learned through the Network Security Operations Quant research, there is a significant cost to operating an enterprise firewall environment, which means you must look to streamline operations when buying new devices. Consistency is one of the keys to making your environment more efficient.

Embedded Firewalls

Speaking of consistency, we also see a number of firewalls that run not on a traditional appliance, dedicated device, or VM — but instead embedded on another device. This might be a WAN optimization box which lets you do everything from a single device in the branch office, a network switch to provide more granular segmentation internally, or even a server (although it's always a bad idea to make your servers Internet-visible). The same deal applies here as on a vendor's own dedicated hardware. Can you manage the firewall policy on an enterprise-wide basis? Do you have all the same capabilities? And even more important, what are the performance characteristics of the device with the firewall capabilities active and fully configured? It's very interesting to think about integrated WAN optimizers with a built-in firewall, but not if the firewall rules add latency to the connection — that would just be silly.

Trust, but Verify

What all this discussion really boils down to is the need to test the device as you'll be using it *before* you buy. It makes no difference what a product testing lab says about throughput.

Based on how you'll use the device, what rules and capabilities you'll implement (especially relative to application awareness), and what size device you deploy, your real performance may be slower or faster than the spec. The only way to figure that out is to actually run a proof of concept to discover the performance characteristics. Again, we'll discuss this in great detail when we look at the selection process, but it needs to be mentioned repeatedly because most enterprises make the mistake of figuring "a firewall is a firewall" and believing performance metrics provided by marketing folks.

*What all this discussion really boils down to is the need to test the device as you'll be using it **before** you buy.*

Deployment Considerations

Depending on requirements and the use cases for the device, there many different ways to deploy enterprise firewalls. Do this wrong and you end up with either too many or too few boxes, single points of failure, suboptimal network access, and/or crappy application performance.

We could talk about all sorts of different models and use fancy names like *tiered*, *mesh*, *peer to peer*, and the like for them — but fortunately the situation isn't really that complicated. To choose the most appropriate architecture you must answer a few questions:

- **Public or private network?** Are your remote locations all connected via private connections such as MPLS or managed IP services, or via public Internet services leveraging site-to-site VPN tunnels?
- **How much is avoiding downtime worth?** This fairly simple question will drive both network architecture and perimeter device selection. You can implement high availability architectures to minimize the likelihood of downtime, but the additional cost is generally significant.
- **What egress filtering/protection do you need?** Obviously you want to provide web and email filtering on outbound traffic. Depending on bandwidth availability and cost, it may make sense to haul remote traffic back to a central location to be processed by large (existing) content security gateways. But for bandwidth-constrained sites, it may make more sense to do web/email filtering locally (using a UTM box), with the understanding that filtering at the smaller sites might be less sophisticated.
- **Who controls gateway policy?** Depending on the size of your organization, there may be different policies for different geographies, business units, locations, etc. Some enterprise firewall management consoles support this kind of granular policy distribution, but you need to figure out who will control policy, and use this to guide how you deploy the boxes.

Remember the technical architecture post where we pointed out the importance of consistency? A consistent feature set on devices up and down a vendor's product line provides a lot of flexibility in how you can deploy — this enables you to select equipment based on the throughput requirement rather than feature set. This is also preferable because application architectures and requirements change, and support for all features on branch equipment (even if you don't initially expect to use them) can save on having to deploy new equipment later. But future-proofing may be outweighed by economic reality.

Bandwidth Matters

We most frequently see firewalls architectures implemented in either two or three tiers. Central sites (geographic HQ) get big honking firewalls deployed in a high-availability cluster to ensure resilience and throughput — especially if they provide higher-level application and/or UTM features. Distribution locations, if they exist, are typically connected to the central site

via a private IP network. These tend to be major cities with good bandwidth. With plentiful bandwidth, most organizations tend to centralize egress filtering (yes, hauling outbound traffic back to the central site) to minimize the control points.

With smaller locations like stores, or in emerging countries with expensive private network options, it may make more economic sense to use public IP services (commodity Internet access) with site-to-site VPN between the locations. In this case centralizing egress filtering probably doesn't make sense, so local firewalls generally must do the filtering as well.

Regardless of the egress filtering strategy you should have a consistent set of ingress policies in place, which usually means (almost) no traffic originating from the Internet is accepted: a [default deny security posture](#). Most organizations leverage hosting providers for web apps, which enable tight rules on the enterprise perimeter for inbound traffic. Likewise, allowing inbound Internet traffic to a small location usually doesn't make sense, since those small sites shouldn't be directly serving up data. Unless you are cool with tellers running their Internet-based side businesses on your network.

High Availability Clusters

Downtime is generally a bad thing — end users can get very grumpy when they can't manage their fantasy football teams during the work day — so you should investigate the hardware resilience features of firewall devices. Things like hot swappable drives and power supplies, redundant backplanes, multiple network connections, redundant memory, etc. Obviously the more redundancy built into the box, the more it will cost, but you already knew that.

Another option is to deploy a high availability cluster. Basically, this means you've got two (or more) boxes sharing a single configuration, allowing automated and transparent load balancing between them to provide scalable performance and to ride out any equipment failures. If a box fails, its peer(s) transparently pick up the slack. But to be clear, load balancing tends to be deployed to address scalability issues, *not* high availability.

High availability and clustering used to be different capabilities (and on some older firewall architectures, still are). But given the state of the hardware and maturity of the space, the terminology has evolved to active/active (the high availability cluster mentioned above where all boxes in the cluster process traffic to address scalability issues) and active/passive (some boxes are normally hot spares, offering no load balancing to ensure a box is always available to protect the perimeter). Bandwidth/scalability requirements tend to drive whether multiple gateways are active.

Internal Deployment

We have mostly discussed the perimeter gateway use case. But there is another scenario, where the firewall is deployed within the data center or at distribution points in the network to provide network segmentation and filtering. This is a bit different than managing inbound/outbound traffic at the perimeter, and largely driven by network architecture. The bandwidth requirements for internal devices are intense — typically 40-100gbps — and here downtime is definitely a no-no, so provision these devices accordingly and bring your checkbook.

Migration

The final issue we'll tackle in relation to deployment is getting old boxes out and new ones in. Depending on the size of the environment, it may not be feasible to do a flash cutover. So the more the new vendor can assist in the migration, the better. Fortunately the market is mature enough that many vendors can import a competitors' rule sets, which facilitates switchovers.

But don't forget that a firewall migration is normally a great opportunity to revisit the firewall rule base and clear out the crap. As we discussed in the Network Security Ops Quant research, you should revisit policies and rules systematically — hopefully a couple times a year — but we are realists. Having to update rules for new capabilities within new gear provides both the means and motive to kill some of those stale firewall rules.

Firewall Management

During procurement it's very easy to focus on shiny objects and blinking lights. By that we mean getting enamored with speeds, feeds, and features — to the exclusion of what you will do with the device once it's deployed. Without focusing on management *during procurement*, you could miss a key requirement — or even worse, sign yourself up to a virtual lifetime of inefficiency and wasted time struggling to manage the security perimeter.

To be clear, most of the base management capabilities of firewall devices are subpar. In fact, a cottage industry of firewall management tools has emerged to address the gaps in these built-in capabilities. Unfortunately that doesn't surprise us, because vendors tend to focus on managing their *devices*, rather than protecting the perimeter. There is a huge difference, and if you have more than 15-20 firewalls to worry about, you need to be very sensitive to how the rule base is built, distributed, and maintained.

*Vendors tend to focus on managing **their** devices, rather than protecting the perimeter.*

What to Manage?

Let's start by making a list of the things you tend to need to manage. It's pretty straightforward and includes (but isn't limited to): ports, protocols, users/groups, applications, network access, network segmentation, and VPN access. You need to understand whether the rules will apply at all times or only at certain times, and whether they apply to all users or just certain groups of users. You'll need to think about what behaviors are acceptable within specific applications as well — especially web-based apps. We talk about building these rule sets in detail in our [Network Security Operations Quant](#) research.

Once we have lists of things to be managed, and some buy-in of what the rules need to be (yes, that involves building consensus among business users, tech colleagues, legal, and lots of other folks there to make you miserable), you can configure the rule base and distribute to the boxes. Another key question is where you will manage the policy — or really at how many levels. You'll likely have some corporate policies driven from HQ which can't be messed with by local admins. You can also opt for some level of regional administration, so part of the rule base reflects corporate policy, but local administrators have the ability to add rules to deal with local issues.

Given the sheer number of options available to manage an enterprise firewall environment, don't forget to consider:

- **Role-based access control:** Make sure you can define different classes of administrators. Some manage the enterprise policy, others just manage their local devices. You also need to pay attention to separation of duties, driven by the firewall change management workflow. Keep in mind the requirement for some level of privileged user monitoring to keep everyone honest (and also to pass those pesky audits) and to provide an audit trail for any changes.

- **Multi-domain administration:** As the perimeter gets more complicated, we see a lot of focus on technologies to allow differing rule bases on the firewalls. This doesn't just provision for different administrators needing access to different functions on the devices, but supports totally different policies running on a single device. Large enterprises with multiple operating units tend to have this requirement, as each department may have unique requirements which require different policy — even when they share the same underlying network. Implementing numerous security *zones* (or domains) allows different policies to be enforced for different logical networks. Ultimately corporate headquarters bears responsibility for the integrity of the entire perimeter, so you'll need a management environment that can effectively map to the way your business operates.
- **Virtual firewalls:** Since everything eventually gets virtualized, why not the firewall? We aren't talking about running the firewall in a general purpose virtual machine again, but instead about having multiple *virtual* firewalls running on the same device. Depending on network segmentation and load balancing requirements, it may make sense to deploy totally separate rule sets within a single device. This is an emerging requirement but worth investigating, because supporting virtual firewalls isn't easy with traditional hardware architectures, particularly at the performance levels typically required for a larger organization.

Checking the Policy

People with experience managing firewalls know all about the pain of a deploying a faulty rule. To avoid that pain and learn from our mistakes, it's critical to *test* rules before they go live. That means the management tools must be able to tell you how a new rule or rule change will impact the rest of the rule base. For example, if you insert a rule at one point in the tree, does it bypass rules in other places? First and foremost, you want to ensure that any change doesn't violate your policies or create a gaping hole in the perimeter. That is job #1.

Also important is rule efficiency. Most organizations have firewall rule bases resembling old closets. Lots of stuff in there, and no one is quite sure why you keep this stuff or which rules still apply. So having the ability to check rule hits (how many times the rule was triggered) helps ensure all your rules remain relevant. It's helpful to have a utility to help optimize

the rule base. The rules tend to be checked sequentially for each incoming packet, so make sure you have the most frequently used rules early, so your expensive devices can work smarter rather than harder and provide some scalability headroom.

Blind devotion to a policy tool is dangerous.

But blind devotion to a policy tool is dangerous too. Remember, these tools *simulate* the policies and impact of new rules and updates. Don't mistake simulation for reality — we strongly recommend confirming changes with actual *tests*. Perhaps not every change, but periodically pen testing your own perimeter will

make sure you didn't miss anything, and minimize surprises. And we know you don't like surprises.

Reporting

As interesting as managing the rule base is, at some point you'll need to prove that you are doing the right thing. That means a set of reports substantiating the controls in place. You'll want to be able to schedule specific times to get this report, as well as how to receive it (web link, PDF, etc.). You should be able to run reports about attacks, traffic dynamics, user activity, etc. You'll also need the ability to dig into the event logs to perform forensic analysis, if you don't send those

events to a SIEM or Log Management device. Don't neglect report customization capabilities either. You know the auditor or your own internal teams will want a custom report — even if the firewall includes thousands built-in — so an environment for quickly and painlessly building your own *ad hoc* reports helps.

Finally, you'll need a set of compliance specific reports — unless you are one of the 10 companies unconcerned with regulatory oversight and still in operation. Most vendors have a series of reports customized to the big regulations (PCI, HIPAA, SoX, NERC CIP, etc.). Again, make sure you can customize these reports, but ultimately the vendor should be doing most of the legwork to map rules to specific regulations.

Other Management Considerations

- **Integration:** Since we're pretty sure you use more than just a firewall, integrating with other IT and security management systems remains a requirement. On the inbound side, you'll need to pull data from the identity store for user/group data and possibly the CMDB (configuration management database) for asset and application data. From an outbound perspective; sending data to a SIEM/Log Management product or service is the most critical requirement; to support centralized activity monitoring, reporting, and forensics; while being able to interface directly with a trouble ticket system to manage requests helps with operational workflow.
- **Workflow:** Speaking of workflow, organizations should have some type of defined authorization process for implementing new rules and changes. Both common sense and compliance guidelines dictate this, but unfortunately it's not a particular strength of device management offerings from firewall vendors. This is really where the third-party firewall management tools are gaining traction.
- **Heterogeneous Firewalls:** This is another area where most vendors' device management offerings are weak. They don't want to help you use competitors' boxes, so they tend to ignore the need to manage a heterogeneous firewall environment. This is another area where third-party management tools are doing well, and as organizations continue acquiring each other this requirement will remain.
- **Outsourcing:** Many organizations are also outsourcing either the monitoring or actual management of their firewalls, so the management capability must offer some kind of interface for the internal team. That may involve a web portal provided by the service provider or some kind of integration. But given the drive toward managed security services, it makes sense to at least ask the vendors whether and how their management consoles can support a managed environment.

Advanced Features

Since this report is largely about the movement toward application aware firewalls, it makes sense to dig a bit deeper into the technology that will make this happen and the major uses for these capabilities. With an understanding of what to look for, you should be in a better position to judge whether a vendor's application awareness capabilities will match your requirements.

Application Visibility

As we discussed, visibility is one of the key use cases for application aware firewalls. What exactly does that mean and what should you look for? We'll break this up into the following buckets:

- **Eye Candy:** Most security folks don't care about fancy charts and graphs, but senior management loves them. What CFO doesn't turn to jello at the first sign of a colorful pie chart? The ability to see application usage and traffic, and who is consuming bandwidth over a long period over time, provides huge value in understanding *normal* behavior on your network. Look for granularity and flexibility in these application-oriented visuals. Top 10 lists are a given, but be sure you can slice the data the ways you need — or at least export to a tool that can. Having the data is nice; being able to use it is better.

- **Alerting:** Trending capabilities for application traffic analysis allow you to set alerts to fire when abnormal behavior appears. Given the infinite attack surface we must protect, any help you can get pinpointing and prioritizing investigative resources increases efficiency. Be sure you have sufficient knobs and dials to set appropriate alerts. You'd like to be able to alert on applications, user/group behavior in specific applications, and possibly even packet payloads (through regular expression type analysis), and any combination therein. Obviously the more flexibility you have in setting application alerts and tightening thresholds, the better you'll be able to reduce the noise. This sounds very similar to managing an IDS, but we'll get to that later. Also make sure setting lots of application rules won't kill performance. Dropped packets are a lousy trade-off for application alerts.

The more flexibility you have in setting application alerts and tightening thresholds, the better you'll be able to reduce the noise.

the noise

One challenge of using a traditional firewall is the interface. Unless the user experience has been rebuilt around an application context (what folks are doing), it still feels like everything is ports and protocols (*how* they are doing it). Clearly the further you can abstract network behavior to application behavior, the more applicable and understandable your rules will be.

Application Blocking

Visibility is the first step, but you also want to be able to block certain applications, users, and content activities. We told you this was very similar to the IPS concept — the difference is in how detection works. The IDS/IPS uses a negative security model (matching patterns to identify bad stuff) to fire rules, while application aware firewalls use a positive security model — they determine what application traffic is authorized and block everything else.

*Just because you **can** block doesn't mean you **should**.*

Extending this IPS discussion a bit, we see most organizations use blocking on only a small minority of the rules/signatures on the box, usually less than 10%. This is for obvious reasons — primarily because they frown on blocking legitimate traffic — and gets back to a fundamental tenet of IPS which also applies to application aware firewalls. *Just because you **can** block doesn't mean you **should**.* Of course, a positive security model means you are defining what is acceptable and blocking everything else, but be careful here. Most security organizations aren't in the loop

on everything that is happening (we know — quite a shocker), so you may inadvertently stymie a new/updated application because the firewall doesn't allow it. To be clear, from a security standpoint that's a **great** thing. You want to be able to vet each application before it goes live, but politically that might not work out very well (for you). You'll need to gauge your own ability to get away with this.

Aside from the IPS analogy, there is also a very clear endpoint white listing analogy to blocking application traffic. One of the issues with application white listing on the endpoints is the challenge of getting applications classified correctly and providing a clear workflow mechanism to deal with exceptions. The same issues apply to application blocking on an application aware firewall. First you need to ensure the application profiles are accurate and up to date. Second, you need a process to allow traffic to be accepted, balancing the need to protect infrastructure and information against responsiveness to business needs.

Yeah, this is non-trivial, which is why blocking is done on a fraction of application traffic.

Overlap with Existing Web Security

Think about the increasing functionality of your operating system or your office suite. Basically, the big behemoth squashed a whole bunch of third party utilities that added value by bundling such capabilities into each new release. The same thing is happening here.

If you look at the typical capabilities of your web application filter, there isn't a lot that can't be done by an application aware firewall. Visibility? Check. Employee control/management? Check. URL blocking, heuristics, script analysis, AV? Check, check, check, check. The standalone web filter is an endangered species — which, given the complexity of the perimeter, isn't a bad thing. Simplifying is good. Moreover, a lot of folks are doing web filtering in the cloud now, so the movement from on-premises web filters was under way anyway. Of course, no entrenched device gets replaced overnight, but the long slide towards standalone web filter oblivion has begun.

As you look at application aware firewalls, you may be able to displace an existing device (or eliminate the maintenance renewal) to justify the cost of the new gear. Clearly going after the web filtering budget makes sense, and the more expense neutral you can make any purchase, the better.

What about web application firewalls? To date, they have been more differentiated from application aware firewalls, with less clear overlap. The WAF's ability to profile and learn about application behavior — in terms of parameter validation, session management, flow analysis, etc. — aren't available on application aware firewalls. For now. But let's be clear: it's not a technical issue. Most of the vendors moving towards these new firewalls also offer web app firewalls. Why build everything into one box if you can charge twice?

Sure, that's cynical, but it's the way things work. Over time, we do expect web application firewall capabilities to be added to application aware firewalls, but that's more of a 3-year scenario, and doesn't mean WAFs will go away entirely. Within a large organization, the WAF may be under the control of the web app team, because the rules are directly related to application functionality rather than security. In this case, there is little impetus for integration/convergence of the devices. But again, this isn't a technical issue — it's a cultural one.

Bot Detection

As law enforcement got much better at tracking attackers, the bad guys adapted by hiding behind armies of compromised machines. Better known as zombies or bots, these devices (nominally controlled by consumers) send spam, perform reconnaissance, and launch other attacks. Due to their sophisticated command and control structures, it's very difficult to map out these bot networks, and attacks can be launched from anywhere at any time.

So how do we deal with this new kind of attacker on the enterprise firewall?

- **Reputation:** Reputation analysis was originally created to help fight spam, and is rapidly being adopted in the broader network security context. We know some proportion of the devices out there are doing bad things, and we know many of those IP addresses. Yes, they are likely compromised devices (as opposed to owned by bad folks specifically for nefarious purposes) but regardless, they are doing bad things. You can check a reputation service in real time and either block or take other actions on traffic originating from recognized bad actors. This is primarily a black list, though some companies track 'good' IPs as well, which allows them to take a cautious stance on devices not known to be either good or bad.
- **Traffic Analysis:** Another technique we are starting to see on enterprise firewalls is traffic analysis. Network behavioral analysis didn't really make it as a standalone capability, but tracking network flows across the firewall (with origin, destination, and protocol information) allows you to build a baseline of acceptable traffic patterns and highlight abnormal activity. You can also set alerts on specific traffic patterns associated with command and control (bot) communications, and as such use a firewall as an IDS/IPS.

Are these two capabilities critical right now? Given the prevalence of other mechanisms to detect these attacks — such as flow analysis through SIEM and pattern matching via network IDS — this is a nice-to-have capability. But we expect a lot of these capabilities to centralize on application aware firewalls, positioning these devices as *the* perimeter security gateway. As such, we expect them to become more common over the next 2 years, and in the process make the bot detection specialists acquisition targets.

Content Inspection

It's funny, but many vendors use the term 'DLP' to describe how they analyze content within the firewall. To be clear, firewall vendors are *not* performing Data Leak Prevention. [Not the way we define it, anyway.](#) At best, it's content analysis

a bit more sophisticated than regular expression scanning. There are no capabilities to protect data at rest or in use, and their algorithms for deep content analysis are immature — when they exist at all.

So we are pretty skeptical about the level of real content inspection you can get from a firewall. If you are just looking to make sure Social Security numbers and account IDs don't leave the perimeter through email or web traffic, a sophisticated firewall can do that. But don't expect to protect your intellectual property with sophisticated analysis and content matching algorithms. When firewall vendors start bragging about 'DLP', feel free to snicker and challenge them regarding *exactly* what kinds of content they can detect and how.

That said, clearly there are opportunities for better integration between real DLP solutions and the enterprise firewall, which can provide an additional layer of enforcement. We also expect to see maturation of inspection algorithms available on firewalls, which could supplant the current DLP network gateways (the ones looking to control data in motion) — particularly in smaller locations where multiple devices are problematic.

Vulnerability Integration

One of the more interesting integrations we see on the horizon is the ability for a web application scanner or service to find an issue and set a blocking rule directly on the web application firewall. This is not a long-term fix but does buy time to investigating a potential application flaw, and provides breathing room to choose the most appropriate remediation approach. Some vendors refer to this as *virtual patching*. Whatever it's called, we think it's interesting. So we expect the same kind of capability to show up on general purpose enterprise firewalls as well.

You'd expect the vulnerability scanning vendors to lead the way on this integration, but regardless, it will make for an interesting capability on the application aware firewall. Especially if you broaden your thinking beyond general network/system scanners. A database scan would likely yield some interesting holes that could be addressed with an application blocking rule at the firewall, no? There are numerous intriguing possibilities, and of course there is always a risk of over-automating ([SkyNet](#), anyone?), but the additional capabilities are likely worth the complexity risk.

To UTM or not to UTM?

Given how much time we've spent discussing application awareness and how these new capabilities pretty much stomp all over existing security products like IDS/IPS and web filters, does that mean standalone network security devices go away? Should you just quietly accept that unified threat management (UTM) is the way to go because the enterprise firewall provides multiple functions? Not exactly.

First let's talk about the rise of UTM, even in the enterprise. The drive towards UTM started with smaller businesses, where using a single device for firewall, IDS/IPS, anti-spam, web filtering, gateway AV, and other functions reduced complexity and cost — and thus made a lot of sense. But over time as device performance increased, it became feasible even for enterprises to consolidate functions into a single device. This doesn't mean many enterprises tried this, but they had the option.

So why hasn't the large enterprise embraced UTM? It comes down to predictable factors we see impacting enterprise technology adoption in general:

- **Branding:** UTM was perceived as a SMB technology, so many enterprise snobs didn't want anything to do with it. Why pay \$2,500 for a box when you can pay \$50,000 to make a statement about being one of the big boys? Of course, notwithstanding the category name, every vendor brought a *multi-function security gateway* to market. They realize 'UTM' could be a liability so they use different names for people who don't want to use the same gear as the great unwashed.
- **Performance Perception:** Again, given the SMB heritage of UTM, enterprise network security players could easily paint UTM as low-performance, and customers believed them. To be clear, the UTM-centric vendors didn't help by pushing their boxes into use cases where they couldn't succeed, thus demonstrating they weren't always suitable. If you try to do high-speed firewall, IDS/IPS, and anti-spam with thousands of rules, all in the same box, it's not going to work well. Hell, even standalone devices use load balancing techniques to scale performance, but the enterprise perception was that UTM doesn't scale. And we all know that perception is reality in the buyer's mind.
- **Single Point of Failure:** If the box goes down you are completely dead in the water, right? Well, yes since firewalls should fail closed. Many enterprises remain unwilling to put all their eggs in one basket, even with high availability configurations and the like. As fans of layered security we don't blame folks for thinking this way, but recognize that you can deploy a set of multi-function gateways in a high availability configuration to address the issue. Furthermore, there is no difference whether you are talking about a UTM or a stand-alone firewall. If it goes down you are out of business. But when you are looking for excuses *not* to do something, you can always find at least one.
- **Specialization:** The complexity of large enterprise environments demands lots of resources, and the resources tend to be specialized in the operation of one specific device. So you'll have a firewall jockey, an IDS/IPS guru, and an anti-spam queen. If you have all those capabilities in a single box, what does that do for the job security of all three? To be

clear every UTM device supports role-based management so administrators can have control only over the functions in their area, but it's easier for security folks to justify their existence if they have a dedicated box/function to manage. Yes, this boils down to politics, but we all know political machinations have killed more than a handful of emerging technologies.

- **Pricing:** There is no reason you can't get a multi-function security device and use it as a single-purpose device. You can get a UTM and run it as a plain firewall. Really. But to date, the enterprise pricing of these UTM devices made that unattractive for most organizations. Again, a clear case of vendors not helping themselves. So we'd like to see more of a smorgasbord pricing model, where you buy the modules you need. Yes, some of the vendors (especially those selling software on commodity hardware) are there. But their inclination is to nickel and dime customers, charging too much for each module, so enterprises start to lose the idea that multi-function devices actually save money.

Ultimately, these factors will not stop the multi-function security device juggernaut from continuing to collapse more functions into the perimeter gateway. Vendors changed the branding to avoid the term 'UTM', but it's the same thing. The devices have increased performance with new chips and updated architectures. And even the political stuff works out over time due to economic pressure to increase operational efficiency.

So the conclusion we draw is that consolidation of network security functions is inevitable, even in the large enterprise.

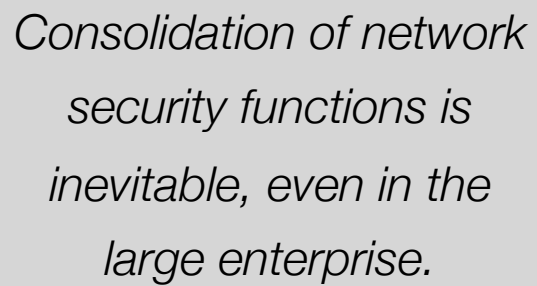
But we aren't religious about UTM vs. standalone devices.

All we care about is seeing the right set of security controls implemented in the most effective way to protect critical information. We don't expect standalone IDS/IPS devices to go away any time soon. And much of the content filtering (email and web) is moving to cloud-based services. We believe this is a very positive trend. These new abilities of the enterprise firewall give us more *flexibility*.

That's right: *We still believe (strongly) in defense in depth*. So having an IDS/IPS sitting behind an application aware firewall isn't a bad thing. Attacks change every day and sometimes it's best to look for a specific issue. Let's use a battle analogy

— if we have a sniper (in the form of IDS/IPS) sitting behind the moat (firewall) looking for a certain individual (the new attack), there is nothing wrong with that. If we want to provision some perimeter security in the cloud, and have a cleaner stream of traffic hitting the network, that's all good. If you want to maintain separate devices at HQ and larger regional locations, while integrating functions in small offices and branches, or maybe even running network security in a virtual machine, you can.

And that's really the point. For a long time, we security folks have been building security architectures based on *what the devices could do*, not what's appropriate (or necessary) to protect information assets. Having the ability to provision the security you need where you need it is exactly what we've been looking for. All these technologies remain relevant. Even if enterprises fully embrace application awareness on the enterprise firewall — and they will — there will still be plenty of boxes at your perimeter. So don't go eBaying your 19" racks quite yet. They'll still be full for a while...



Consolidation of network security functions is inevitable, even in the large enterprise.

Selection Process

Now that we've been through the drivers for evolved, application-aware firewalls, and a lot of the technology enabling them, how does the selection process need to evolve to keep pace? As with most of our research at Securosis, we favor map out a *very* detailed process, and leave you to decide which steps make sense in your situation. So we don't expect every organization to go through every step in this process. Figure out which are appropriate for your organization and use those.

To be clear, buying an enterprise firewall usually involves calling up your reseller and getting the paperwork for the renewal. But given that these firewalls imply new application policies and perhaps a different deployment architecture, some work must be done during selection to get things right.

Define Needs

The key here is to understand which applications you want to control, and how much you will consider collapsing functionality (IDS/IPS, web filtering, UTM) into the enterprise firewall. A few steps to consider here are:

- **Create an oversight committee:** We hate the term 'committee' too, but the reality is that an application aware firewall impacts activities across several groups. Clearly this is not just all about the security team, but also the network team and the application teams as well — at minimum, you will need to profile the applications traversing the firewall. So it's best to get someone from each of these teams (to whatever degree they exist in your organization) on the committee. Ensure they understand the objectives for the new enterprise firewall, and make sure it's clear how their operations will change.
- **Define the applications to control:** Which applications do you need to control? You may not actually know this until you install one of these devices and see what visibility they provide into applications traversing the firewall. We'll discuss phasing in your deployment, but you need to understand what degree of granularity you need from a blocking standpoint, as that drives some aspects of selection.
- **Determine management requirements:** The deployment scenario will drive these. Do you need the console to manage the policies? To generate reports? For dashboards? The degree to which you need management help (if you have a third party tool, the answer should be: not much) will define a set of management requirements.
- **Product versus managed service:** Do you plan to use a managed service for either managing or monitoring the enterprise firewall? Have you selected a provider? The provider might define your short list before you even start.

By the end of this phase you should have identified key stakeholders, convened a selection team, prioritized the applications to control, and determined management requirements.

Formalize Requirements

This phase can be performed by a smaller team working under the mandate of the oversight committee. Here the generic needs determined in phase 1 are translated into specific technical features, and any additional requirements are considered. You can always refine these requirements as you proceed through the selection process and get a better feel for how the products work — and how effective and flexible they are at blocking applications.

At the conclusion of this stage you will develop a formal RFI (Request For Information) to release to vendors, and a rough RFP (Request For Proposals) to clean up and issue formally in the evaluation phase.

Evaluate Products

Increasingly we see firewall vendors starting to talk about application awareness, new architectures, and very similar feature sets. The following steps should minimize your risk and help you feel confident in your final decision:

- **Issue the RFI:** Larger organizations should issue an RFI through established channels and contact a few leading enterprise firewall vendors directly. Though in reality virtually all the firewall players sell through the security channel, so it's likely you will end up going through a VAR.
- **Define the short list:** Before bringing anyone in, match any materials from the vendor or other sources to your RFI and draft RFP. Your goal is to build a short list of 3 products *which can satisfy most of your needs*. You should also use outside research sources and product comparisons. Understand that you'll likely need to compromise at some point in the process, as it's unlikely any vendor can meet every requirement.
- **Dog and pony show:** Instead of generic presentations and demonstrations, ask the vendors to walk you through how they protect the specific applications you are worried about. This involves not just the technical capabilities of the box, but also the research capabilities to profile applications and keep those profiles relevant and accurate. This is **critical**, because the vendors are very good at showing cool eye candy and presenting a long list of generic supported applications. Don't expect a full response to your draft RFP — these meetings are to help you understand how each vendor can solve your *specific* use cases, and to finalize your requirements.
- **Finalize and issue your RFP:** At this point you should completely understand your specific requirements, and issue a final formal RFP.
- **Assess RFP responses and start proof of concept (PoC):** Review the RFP results and drop anyone who doesn't meet your hard requirements. Then bring in any remaining products for in-house testing. Given that it's not advisable to pop holes in your perimeter while learning how to manage these devices, we suggest a layered approach to test them.
 - **Test Ingress:** First test your ingress connection by installing the new firewall **outside** the existing perimeter gateway. Migrate your policies over, let the box run for a while, and see what it's blocking and what it's not.
 - **Test Egress:** Then move the firewall **inside** the perimeter gateway, so it's in position to do egress filtering on all your traffic. We suggest you monitor the traffic for a while to understand what is happening, and then define egress filtering policies.

Understand that you need to devote resources to each PoC, and testing ingress separately from egress adds time to the process. But it's not feasible to leave the perimeter unprotected while you figure out what works, so this approach gives

you that protection and the ability to run the devices in pseudo-production mode. In fact, we see many organizations actually deploying new firewalls in this manner (behind or in front of the existing devices) to reduce the risk of embracing the new technology.

Selection and Deployment

- **Select, negotiate, and buy:** Finish testing, take the results to the full selection committee, and begin negotiating with your top two choices, assuming more than one meets your needs. We understand this takes more time, but you should be able to walk away from one of the vendors if they won't play on pricing, terms, or conditions.
- **Implementation planning:** Congratulations, you've selected a product, navigated the procurement process, and made a sales rep happy. But now the next stage begins: the last phase of selection is planning the deployment. That means making sure of little details, lining up resources, locking in an install schedule, and even figuring out the logistics of getting devices to (and installed at) the right locations.

We hear the groans from small to medium sized businesses who look at this process and think this is a ridiculous amount of detail. Once again, we deliberately created a granular selection process, but you can pare this down to meet your organization's requirements. We wanted to ensure we captured all the gory details some organizations need to go through for a successful procurement. The full process outlined is appropriate for a large enterprise, but a little pruning can make it manageable for small groups. That's the great thing about process: you can change it any way you see fit at no expense.

Conclusion

The enterprise firewall has long been the anchor of the perimeter, blocking the bad stuff and making sure authorized traffic gets to the right places. But over the past five years, the firewall (generically speaking) has not kept pace with changes in the attack space and the way more and more key applications are encapsulated within a few protocols. That is finally changing, and now it is time to revisit your perimeter security architecture, and start evaluating a new generation of enterprise-class firewalls.

The requirements driving this move to evolved firewalls include not just increasing performance and decreasing the complexity of running them for a large enterprise, but additionally to move beyond the traditional ports & protocols model for firewalls to a new reality based on applications and users. This is basically starting over and will require a total re-architecting of the entire class of firewall devices. That said, there are many ways to get to the promised land, and the incumbent firewall vendors are all working on application aware technologies which promise a reasonable migration to this new functionality.

But application awareness isn't the only new capability increasing in importance on the enterprise firewall. The ability to detect and block command and control traffic (typically associated with bot networks), as well as enhanced management enabling multi-domain and virtual implementations of firewall functionality, are increasing in importance as the perimeter gateway evolves.

Each year, your incumbent firewall vendor comes in to collect their rather sizable maintenance renewal. Most organizations grunt and write the check, but we believe it's time to start challenging the incumbents and pushing them to help you understand how they are moving in these new directions, to more effectively defend your network against the types of attacks we see now.

Also take this opportunity to see if there are capabilities (such as web filtering and IDS/IPS) that can be collapsed into a multi-function security gateway. Notice we didn't say the word 'UTM', although that's what we mean. Depending on the performance requirements and the separation of duties required for your organization, the time to start simplifying perimeter security may be upon you.

In many cases, organizations will decide the *status quo* is okay and that the incumbent firewall vendor has the right strategy to keep things moving forward. We have no issue with that conclusion, as long as you ask the right questions and ensure that your devices are evolving to meeting not only today's needs but will also be able to handle tomorrow's. Our hope is that this guide provides the information necessary to evaluate enterprise firewalls both individually and head-to-head, and help you avoid many of the problems that tend to pop up — usually just *after* purchasing a product, when coming to grips with how it *really* works. Good luck with your selection process — we are happy to help if you run into questions during your efforts; feel free to drop us a note at info@securosis.com, and we'll do our best to help out.



About the Analyst

Mike Rothman, Analyst/President

Mike's bold perspectives and irreverent style are invaluable as companies determine effective strategies to grapple with the dynamic security threatscape. Mike specializes in the sexy aspects of security, such as protecting networks and endpoints, security management, and compliance. Mike is one of the most sought-after speakers and commentators in the security business, and brings a deep background in information security. After 20 years in and around security, he's one of the guys who "knows where the bodies are buried" in the space.

Starting his career as a programmer and a networking consultant, Mike joined META Group in 1993 and spearheaded META's initial foray into information security research. Mike left META in 1998 to found SHYM Technology, a pioneer in the PKI software market, and then held VP Marketing roles at CipherTrust and TruSecure — providing experience in marketing, business development, and channel operations for both product and services companies.

After getting fed up with vendor life, Mike started Security Incite in 2006 to provide a voice of reason in an over-hyped yet underwhelming security industry. After taking a short detour as Senior VP, Strategy and CMO at eIQnetworks to chase shiny objects in security and compliance management, Mike joined Securosis with a rejuvenated cynicism about the state of security and what it takes to survive as a security professional.

Mike published *The Pragmatic CSO* <<http://www.pragmaticcso.com/>> in 2007 to introduce technically oriented security professionals to the nuances of what is required to be a senior security professional. He also possesses a very expensive engineering degree in Operations Research and Industrial Engineering from Cornell University. His folks are overjoyed that he uses literally zero percent of his education on a daily basis. He can be reached at mrothman (at) securosis (dot) com.

About Securosis

Securosis, L.L.C. is an independent research and analysis firm dedicated to thought leadership, objectivity, and transparency. Our analysts have all held executive level positions and are dedicated to providing high-value, pragmatic advisory services.

Our services include:

- *Primary research publishing:* We currently release the vast majority of our research for free through our blog, and archive it in our Research Library. Most of these research documents can be sponsored for distribution on an annual basis. All published materials and presentations meet our strict objectivity requirements, and follow our [Totally Transparent Research](#) policy.
- *Research products and strategic advisory services for end users:* Securosis will be introducing a line of research products and inquiry-based subscription services designed to assist end user organizations in accelerating project and program success. Additional advisory projects are also available, including product selection assistance, technology and architecture strategy, education, security management evaluations, and risk assessment.
- *Retainer services for vendors:* Although we will accept briefings from anyone, some vendors opt for a tighter, ongoing relationship. We offer a number of flexible retainer packages. Example services available as part of a retainer package include market and product analysis and strategy, technology guidance, product evaluations, and merger and acquisition assessments. Even with paid clients, we maintain our strict objectivity and confidentiality requirements. More information on our [retainer services](#) (PDF) is available.
- *External speaking and editorial:* Securosis analysts frequently speak at industry events, give online presentations, and write and/or speak for a variety of publications and media.
- *Other expert services:* Securosis analysts are available for other services as well, including Strategic Advisory Days, Strategy Consulting engagements, and Investor Services. These services tend to be customized to meet a client's specific requirements.

Our clients range from stealth startups to some of the best known technology vendors and end users. They include large financial institutions, institutional investors, mid-sized enterprises, and major security vendors.

Additionally, Securosis partners with security testing labs to provide unique product evaluations that combine in-depth technical analysis with high-level product, architecture, and market analysis.