



NEXT GENERATION FIREWALL (NGFW)
TEST METHODOLOGY V4.0

To receive a licensed copy or report misuse,
Please contact NSS Labs at: +1 (760) 270-9852
or advisor@nsslabs.com

© 2010 NSS Labs, Inc. All rights reserved. No part of this publication may be reproduced, photocopied, stored on a retrieval system, or transmitted without the express written consent of the authors.

All rights reserved. No part of this publication may be reproduced, photocopied, stored on a retrieval system, or transmitted without the express written consent of the authors. This report shall be treated at all times as a confidential and proprietary report for internal use only.

Please note that access to or use of this Methodology is conditioned on the following:

1. The information in this Methodology is subject to change by NSS Labs without notice.
2. The information in this Methodology is believed by NSS Labs to be accurate and reliable, but is not guaranteed. All use of and reliance on this Methodology are at your sole risk. NSS Labs is not liable or responsible for any damages, losses or expenses arising from any error or omission in this Report.
3. NO WARRANTIES, EXPRESS OR IMPLIED ARE GIVEN BY THE NSS LABS. ALL IMPLIED WARRANTIES, INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT ARE DISCLAIMED AND EXCLUDED BY NSS LABS. IN NO EVENT SHALL NSS LABS BE LIABLE FOR ANY CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, OR FOR ANY LOSS OF PROFIT, REVENUE, DATA, COMPUTER PROGRAMS, OR OTHER ASSETS, EVEN IF ADVISED OF THE POSSIBILITY THEREOF.
4. This Methodology does not constitute an endorsement, recommendation or guarantee of any of the products (hardware or software) tested or the hardware and software used in testing the products. The testing does not guarantee that there are no errors or defects in the products, or that the products will meet your expectations, requirements, needs or specifications, or that they will operate without interruption.
5. This Methodology does not imply any endorsement, sponsorship, affiliation or verification by or with any companies mentioned in this report. For PCI-related reports, this does not constitute an endorsement by the PCI Security Standards Council.
6. All trademarks, service marks, and trade names used in this Report are the trademarks, service marks, and trade names of their respective owners, and no endorsement of, sponsorship of, affiliation with, or involvement in, any of the testing, this Report or NSS Labs is implied, nor should it be inferred.

CONTACT INFORMATION

NSS Labs

P.O. Box 130573
Carlsbad, CA 92013 USA
+1 (760) 270-9852
info@nsslabs.com
www.nsslabs.com

CONTENTS

1	<i>Introduction</i>	1
1.1	The Need for the Next Generation Firewall (NGFW).....	1
1.2	What is an NGFW?	1
1.3	When is an NGFW not an NGFW?	2
1.4	The Need for Testing.....	2
1.5	About This Test Methodology and Report	3
1.6	Inclusion Criteria	4
1.7	About NSS Labs.....	4
2	<i>Product Comparisons</i>	5
2.1	Security Effectiveness.....	5
2.2	Resistance to Evasion	5
2.3	Performance	6
2.4	Stability	6
2.5	Total Cost of Ownership.....	6
3	<i>Product Guidance</i>	8
3.1	Recommend	8
3.2	Neutral.....	8
3.3	Caution	8
4	<i>Methodology Elements Overview</i>	9
5	<i>Security Effectiveness</i>	10
5.1	Firewall Policy Enforcement.....	10
5.2	Application Control	11
5.3	User/Group ID Aware Policies	11
5.4	Intrusion Prevention Policies	12
5.5	Evasion	15
6	<i>Firewall Performance</i>	20
6.1	Raw Packet Processing Performance (UDP Traffic).....	20
6.2	Maximum Capacity.....	21
6.3	Behavior Of The State Engine Under Load.....	23
6.4	HTTP Capacity With No Transaction Delays.....	24
6.5	“Real World” Traffic.....	25
6.6	Latency & Response Times.....	26
7	<i>Stability & Reliability</i>	28
	<i>Appendix A: Firewall Test Environment</i>	32
	<i>Appendix B: Special Thanks</i>	33

1 INTRODUCTION

1.1 THE NEED FOR THE NEXT GENERATION FIREWALL (NGFW)

Firewall technology has been around for at least 25 years, and undergone several stages of development; from early packet and circuit firewalls to application layer and dynamic packet firewalls. Across these stages, the goal has continued to be to provide a protective barrier between internal and external networks, while allowing for productive communications to pass from one side to the other. With the emergence of new web applications and security threats, firewalls are again evolving.

Whereas in the past we could say with a reasonable degree of certainty that application X runs over TCP port 552, and web traffic (and web traffic *alone*) runs over TCP port 80, this is no longer true today. Add to that, the rise of Web 2.0 and the proliferation of applications which bypass traditional firewall controls by tunneling over HTTP and HTTPS, and it becomes apparent that additional security controls (based upon the application vs. the port) must be added to firewalls. This means that relying on port and protocol combinations to define network applications is no longer enough. Firewalls need to be capable of performing deep packet inspection of all packets, on all ports and over all protocols in order to determine which applications are running over which ports.

NSS Labs' research indicates that over the past 18 months, the sophistication and strategic capabilities of cybercriminals has outstripped the pace of advancement within information security products. In addition to traditional remote attacks against servers, cybercriminals are increasingly waging highly targeted campaigns against desktop client applications. These campaigns include the use of encrypted websites (such as Gmail), social networking sites, advertising networks, and a long list of compromised websites. The Wall Street Journal, the New York Times and ESPN were all found to have been (inadvertently) dishing up exploits to their clients. As such, users need not venture into a "dark corner" of the Internet to be exploited.

Some high profile examples of desktop clients being the primary attack vector are the Operation Aurora attack against Google and the numerous variants of the Zeus attack against financial institutions. Further, compromised systems often communicate back to command and control servers via ports 80 (HTTP) 443 (HTTPS), or DNS (53) since those ports are most likely not blocked by traditional firewalls, which define security policies in terms of IP Addresses, ports, protocols and services.

1.2 WHAT IS AN NGFW?

As Firewalls which will be deployed at critical choke-points in the network, the stability and reliability of an NGFW is imperative. Therefore prime directive of any NGFW is that it must be as stable, as reliable, as fast, and as flexible as the existing firewall that it is replacing.

In addition, an NGFW must provide granular control based upon applications, not just ports. This capability is needed to re-establish a secure perimeter where unwanted applications are not able to tunnel over HTTP/s. As such, granular application control is a requirement of NGFW since it enables the administrator to define security policies based upon applications vs. ports. For example, the administrator could block all Skype traffic while allowing Twitter and Facebook. More advanced offerings will inspect and expose specific functionality within web applications; enabling administrators to allow users to read Facebook status, but not update; or to read **and** update status, while blocking Facebook applications.

Also important is the ability to identify users and groups and apply security policy based on identity. Where possible, this should be achieved via direct integration with existing enterprise authentication systems (such as Active Directory) without the need for custom server-side software. This allows the administrator to create even more granular policies. For example, it would be possible to restrict the use of social media applications such as Facebook and Twitter to the marketing department, whilst prohibiting use elsewhere in the company during working hours. Combining this with our previous example, it would be possible to allow Facebook and Twitter status updates from the marketing department only, whilst read-only operation would be permitted throughout the company. Facebook applications would be blocked for all users.

Intrusion Prevention Systems (IPS) have become standard security devices in almost all sizes of enterprise. And enterprises are looking to consolidate IPS capabilities within the NGFW. Therefore an NGFW must apply full-strength IPS functionality such that it is as capable of identifying and blocking exploits as the existing IPS that it is replacing.

In order to enhance the blocking decisions of the NGFW, some form of external intelligence should be incorporated. This can include the use of so-called reputation services and black/white lists to perform early blocking of traffic without the need for resource-hungry deep packet inspection.

Based on the needs identified in the previous section, the following capabilities are considered essential as part of a NGFW device:

- Traditional “first generation firewall” including:
 - Basic packet filtering
 - Stateful multi-layer inspection
 - NAT
 - VPN
 - Highly Stable
 - High Availability
- Application awareness/control
- User/group control
- Integrated IPS
- Ability to operate at layer 3 (“traditional”) or layer 2 (“bump in the wire”)
- External intelligence to enhance blocking decisions (i.e. “reputation services”)

1.3 WHEN IS AN NGFW NOT AN NGFW?

When it is a UTM. NSS believes that the enterprise customer is not yet ready to install broad-spectrum UTM devices at the edge of the corporate network. Functions such as web filtering, anti-spam and anti-malware are better reserved for dedicated security gateway devices in larger deployments.

NSS will consider UTM devices for inclusion in its NGFW testing program only when it is possible to completely disable all functions other than those listed above.

1.4 THE NEED FOR TESTING

Firewalls are the cornerstone of network security. However, over the past few years we have witnessed the marginalization of the firewall due applications designed to bypass access control by tunneling over HTTP/HTTPS as well web applications which are “permitted” due limitations of traditional firewalls ability to control web applications. Simultaneously, our research indicates that cyber-criminals have evolved from phishing and web-based malware to client exploits. As such, the threats traversing port 80 are greater than ever – a trend that is accelerating.

Correspondingly, vendors have begun to market evolving technologies known as “Next Generation Firewalls”, based on nomenclature coined by Gartner. As a result, the team at NSS Labs decided to investigate the level to which different vendors are delivering ‘next generation’ capabilities, and what the trade-offs are. As part of this research, we are conducting a group test to provide the industry with a current scientific baseline of NGFW effectiveness.

1.5 ABOUT THIS TEST METHODOLOGY AND REPORT

NSS Labs’ test reports are designed to address the challenges faced by IT professionals in selecting and managing security products. The scope of this particular report is focused on:

- Security effectiveness
- Performance
- Stability
- Total Cost of Ownership (TCO)

Live Exploit Testing: NSS Labs’ security effectiveness testing leverages deep expertise of our engineers utilizing multiple commercial, open source and proprietary tools as appropriate. With over 1,200 live exploits. This is the industry’s most comprehensive test to date. Most notable, all of the live exploits and payloads in our test have been validated in our lab such that:

- a reverse shell is returned
- a bind shell is opened on the target allowing the attacker to execute arbitrary commands
- a malicious payload is installed
- a system is rendered unresponsive
- etc.

This test goes far beyond replaying PCAPs or pressing the button on a test tool. In short, our engineers triggered vulnerabilities for the purpose of validating that an exploit was able to pass through the device under test.

Performance: NGFW devices exhibit an inverse correlation between security effectiveness and performance. The more deep packet inspection is performed, the longer it takes to forward packets. Furthermore, it is important to consider a real-world mix of traffic that a device will encounter. NSS Labs utilizes a range of traffic types and mixes.

Tuning: Security engineers tune an IPS to ensure its protection coverage matches the needs of the environment where it is being placed. This strategy works well for datacenters and DMZs. However, protecting desktops is a whole different matter. In surveying enterprises, we found most enterprises do not strictly control the desktop and that in larger enterprises it is safe to assume that pretty much anything can be running. As such, enterprises are expecting IPS and NGFW vendors to provide maximum security for desktop client applications with their *recommended* policies. Further, research indicates that enterprises are not ready to replace their dedicated IPS solutions in the datacenter. Simple deduction therefore tells us that intrusion prevention functionality within an NGFW needs to protect desktop clients – with optimal protection pre-defined via a vendor recommended policy.

Additional Reports and Documents

This group test methodology is based on years of testing experience and summarizes thousands of individual test cases. Given the overwhelming breadth and depth of this methodology, not all data could be represented comparatively. Please reference the following additional documents for more information:

- Individual Product Reports contain detailed results from the extensive testing by product configuration
- Exposure Reports provide a searchable database of exposed vulnerabilities not protected by NGFW and IPS products (listed by CVE) to help organizations identify and mitigate specific risks to assets
- Product reports and analysis at www.nsslabs.com provide more in-depth coverage of required functionality, including management and stability in addition to security and performance tested here

1.6 INCLUSION CRITERIA

In order to garner the greatest participation, and allay any potential concerns of bias, we invited all leading firewall vendors, and those claiming NGFW capabilities, to submit products at no cost. Thus vendors with major market share, as well as challengers with new technology and less of a track record were included.

The NGFW should be supplied as a single appliance, where possible (cluster controller solutions are also acceptable), with the appropriate number of physical interfaces capable of achieving the required level of connectivity and performance (minimum of one in-line port pair per Gigabit of throughput, or one in-line 10Gbps port pair per 10Gbps of throughput).

Firewall products should be implemented as in-line Layer 3 (routing) devices. Multiple separate 1Gbps or 10Gbps connections will be made from the external to internal switches via the Device Under Test (DUT), subject to a minimum of one in-line port pair per Gigabit of throughput. Thus, an 8 Gbps device with only four port pairs will be limited to 4 Gbps. The minimum number of port pairs will be connected to support the claimed maximum bandwidth of the DUT (thus an 8 Gbps DUT with ten port pairs will have eight 1Gbps connections tested).

Once installed in the test lab, the DUT will be configured for the use-case appropriate to the target deployment (corporate network perimeter). The DUT should also be configured to block all traffic when resources are exhausted or when traffic cannot be analyzed for any reason.

1.7 ABOUT NSS LABS

NSS Labs performs expert, independent security product evaluations and certifications to assist IT teams in selecting and managing the right security products for their environment. We operate the largest security and performance lab in the world. Our test reports and analysis are highly regarded by information security professionals for their rigor, depth, and integrity, and are used to validate purchasing decisions in global organizations.

2 PRODUCT COMPARISONS

2.1 SECURITY EFFECTIVENESS

Security products are growing increasingly complex and vendors are responding by simplifying the user interface and security policy selection to meet the usability needs of a broadening user base. Indeed, many organizations accept and deploy the vendor pre-defined recommended settings, understanding these to be the best recommendations from the vendor.

2.1.1 EFFECTIVENESS BY ATTACK VECTOR

Exploits can be initiated either locally by the target (client) or remotely by the attacker. Since 2007, we have seen a dramatic rise in the number of client-side exploits, as these can be easily launched by an unsuspecting user who visits an infected website. NGFW products will be tested using these types of attacks.

NSS utilizes the following definitions:

- **Attacker Initiated:** The threat/exploit is executed by the attacker remotely against a vulnerable application and/or operating system.
- **Target Initiated:** The threat/exploit is initiated by the vulnerable target. The attacker has little or no control as to when the target user or application will execute the threat.

2.1.2 EFFECTIVENESS BY DISCLOSURE DATE

The old attacks are still relevant and must be protected against. Different vendors take different approaches to adding coverage once a vulnerability is disclosed. The result is varying levels of protection for vulnerabilities.

2.2 RESISTANCE TO EVASION

Evasion techniques are means of disguising and modifying attacks in order to avoid detection and blocking by security products. Missing a type of evasion means a hacker can use an entire class of exploits to circumvent the NGFW, rendering it virtually useless. The techniques used in this test have been widely known for years and should be considered minimum requirements for the NGFW product category.

Providing exploit protection results without factoring evasion in can be misleading since the more types of evasion that are missed—IP Fragmentation, TCP Segmentation, RPC Fragmentation, URL Obfuscation, HTML Obfuscation and FTP Evasion—the worse the situation. It is better to miss five evasion techniques in one category (say FTP evasion) than one in each category. Furthermore, missing either IP Fragmentation or TCP Segmentation is worse than missing any of the other three types of obfuscation since they impact ALL exploits.

2.2.1 IMPACT OF EVASION – RECOMMENDED POLICY

IP Fragmentation and TCP Segmentation evasions are by far the worst since, if an attacker can avoid detection by fragmenting IP Packets or segmenting TCP Streams, an NGFW will be completely blind to ALL attacks. In addition, ready-made tools are available to help attackers with these evasion techniques. Nearly as bad are RPC Fragmentation, URL Obfuscation HTML Obfuscation and FTP / Telnet Evasion since those represent some of the most popular applications on a network.

Thus, missing a single evasion technique opens wider holes for attackers to get through and vendors should rectify such omissions immediately. Any financially-motivated hacker with basic skills will know

how to take advantage of these weaknesses, and simple toolkits exist to assist them. Further analysis of evasion techniques is provided in the Exposure Report, including more advanced evasions such as JavaScript evasions and PDF Evasions.

2.3 PERFORMANCE

NSS Labs collects extensive performance metrics during this test, according to our established methodology. The volumes of data produced by these tests are designed to capture maximum capacities or “the edge of performance” that may be obtainable for a given metric. In addition, our real-world traffic mix testing methods enable us to more accurately estimate the performance users can expect in their environments.

Beyond overall throughput of the device, connection dynamics can play an important role in sizing a security device that will not unduly impede the performance of a system or an application.

2.4 STABILITY

Long term stability is particularly important for a firewall, where failure can produce network outages. NSS Labs tests verify the stability of the DUT along with its ability to maintain security effectiveness while under normal load and while passing malicious traffic.

The DUT is required to remain operational and stable throughout these tests, and to block 100 per cent of previously blocked traffic, raising an alert for each. If any non-allowed traffic passes successfully - caused by either the volume of traffic or the DUT failing open for any reason - this will result in a FAIL.

2.5 TOTAL COST OF OWNERSHIP

NGFW implementations can be complex projects with several factors affecting the overall cost of deployment, maintenance and upkeep. All of these should be considered over the course of the useful life of the solution.

- Product Purchase – the cost of acquisition
- Product Maintenance – the fees paid to the vendor
- Installation – the time required to take the device out of the box, configure it, put it into the network, apply updates and patches, initial tuning, and set up desired logging and reporting.
- Upkeep – the time required to apply periodic updates and patches from vendors, including hardware, software, and protection (signature/filter/rules) updates.
- False Positives – the time required to eliminate false alarms and false positives.

2.5.1 LABOR PER PRODUCT (IN HOURS)

2.5.2 PURCHASE PRICE AND TOTAL COST OF OWNERSHIP

- *Year One TCO was determined by multiplying the Labor Rate (\$75 per hour fully loaded) x (Installation + Upkeep + Tuning) and then adding the Purchase Price + Maintenance.*
- *Year Two TCO was determined by multiplying the Labor Rate (\$75 per hour fully loaded) x (Upkeep + Tuning) and then adding Year One TCO.*
- *Year Three TCO was determined by multiplying the Labor Rate (\$75per hour fully loaded x (Upkeep + Tuning) and then adding Year Two TCO.*

2.5.3 VALUE: COST PER MBPS AND EXPLOIT BLOCKED – TUNED POLICY

2.5.4 THREE-YEAR TOTAL COST OF OWNERSHIP PER PROTECTED MBPS

- *Price/Mbps-Protected = Three-Year TCO/(Protection x Throughput).*

This formula discounts the throughput based upon the effectiveness of the device at blocking attacks in order to provide a weighted value.

2.5.5 COMPARISON BY PERFORMANCE CATEGORY

3 PRODUCT GUIDANCE

NSS Labs issues summary product guidance based on evaluation criteria that is important to information security professionals. The evaluation criteria are weighted as follows:

1. **Security effectiveness** - The primary reason for buying a next-generation firewall is to achieve a high percentage coverage of common threats
2. **Resistance to evasion** - Failure in any evasion class permits attackers to circumvent protection
3. **Performance** – Correctly sizing a firewall is essential
4. **Stability** - Long term stability is particularly important for an in-line device, where failure can produce network outages
5. **Simplicity of management** - In particular, how difficult is it to configure the highest degree of protection
6. **Value** – Customers should seek low TCO and high effectiveness and performance rankings

Products are listed in rank order according to their guidance rating.

3.1 RECOMMEND

A Recommend rating from NSS Labs indicates that a product has performed well and deserves strong consideration. Only the top technical products earn a Recommend rating from NSS Labs—regardless of market share, company size, or brand recognition.

3.2 NEUTRAL

A Neutral rating from NSS Labs indicates that a product has performed reasonably well and should continue to be used if it is the incumbent within an organization. Products that earn a Neutral rating from NSS Labs deserve consideration during the purchasing process.

3.3 CAUTION

A Caution rating from NSS Labs indicates that a product has performed poorly. Organizations using one of these products should review their security posture and other threat mitigation factors, including possible alternative configurations and replacement. Products that earn a Caution rating from NSS Labs should not be short-listed or renewed.

4 METHODOLOGY ELEMENTS OVERVIEW

The NSS NGFW test methodology is designed to stress and evaluate an NGFW device in three main areas: *security effectiveness, performance and stability*.

The following table lists the individual tests NSS Labs performed on each of the products. Direct references are provided to NSS Labs Test IDs from Sections 5 through 7 of our full NGFW methodology. The detailed results per product are available separately.

All testing results as specified in this methodology will be reported in the final test document.

5 SECURITY EFFECTIVENESS

This section verifies that the DUT is capable of enforcing a specified security policy effectively. NSS Labs' NGFW certification is conducted by incrementally building upon a baseline configuration (simple routing with no policy restrictions and no content inspection) to a complex real world multiple zone configuration supporting many addressing modes, policies, applications, and inspection engines.

At each level of complexity, test traffic is passed across the DUT to ensure that only specified traffic is allowed and the rest is denied, and that appropriate log entries are recorded.

The DUT must support stateful firewalling either by managing state tables to prevent "traffic leakage" or as a stateful proxy. The ability to manage firewall policy across multiple interfaces/zones is a required. At a minimum, the DUT must provide a "trusted" internal interface, an "untrusted" external/Internet interface, and one or more DMZ interfaces. In addition, a dedicated management interface is preferred.

5.1 FIREWALL POLICY ENFORCEMENT

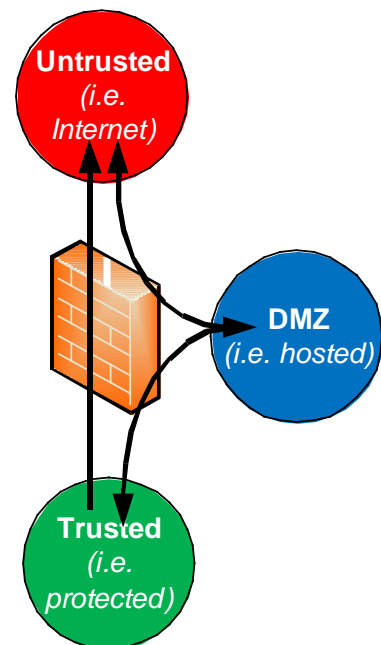
Policies are rules that are configured on a firewall to permit or deny access from one network resource to another based on identifying criteria such as: source address, destination address, and service. A term typically used to define the demarcation point of a network where policy is applied is a *demilitarized zone* (DMZ). Policies are typically written to permit or deny network traffic from one or more of the following zones:

- **Untrusted** – This is typically an external network and is considered to be an unknown and non-secure. An example of an untrusted network would be the Internet.
- **DMZ** – This is a network that is being *isolated* by the firewall restricting network traffic to and from hosts contained within the isolated network.
- **Trusted** – This is typically an internal network; a network that is considered secure and protected.

The NSS Labs Firewall certification tests performance and the ability to enforce policy between the following:

- Trusted to Untrusted
- Untrusted to DMZ
- Trusted to DMZ

Note: Firewalls must provide at a minimum one DMZ interface in order to provide a DMZ or "transition point" between untrusted and trusted networks



5.1.1 BASELINE POLICY

Routed configuration with an "allow all" policy

5.1.2 SIMPLE POLICIES

Simple outbound and inbound policies allowing basic browsing and e-mail access for internal clients and no external access

5.1.3 COMPLEX POLICIES

Complex outbound and inbound policies consisting of many rules, objects, and services.

5.1.4 STATIC NAT (NETWORK ADDRESS TRANSLATION)

Inbound Network Address Translation (NAT) to DMZ using fixed IP address translation with one-to-one mapping.

5.1.5 DYNAMIC/HIDE NAT (NETWORK ADDRESS TRANSLATION)

Outbound Network Address Translation (NAT) (from Internal to External) where all outbound traffic “hides” behind the IP Address of the External Interface of the Firewall utilizing a pool of high ports to manage multiple connections.

5.1.6 SYN FLOOD PROTECTION

The DUT is expected to protect against SYN Floods

5.1.7 ADDRESS SPOOFING

The DUT must prevent Address Spoofing by enforcing rules that define which IP Addresses & Networks reside behind each of the DUT’s interfaces.

5.1.8 SESSION HIJACKING

The DUT is expected to prevent Session Hijacking

Note that vendors should provide tools/scripts to permit the creation of thousands of rules automatically.

5.2 APPLICATION CONTROL

Complex outbound and inbound policies consisting of many rules, objects and applications, verifying that the DUT is capable of correctly determining the correct application from deep packet inspection (regardless of port/protocol used) and taking the appropriate action.

- Popular Social Networking Websites (Web Applications)
- Instant Messaging
- Skype and other VoIP
- Torrents
- Other TBD

For each application, NSS Labs will test the NGFW’s ability to perform the following functions

5.2.1 BLOCK

The Firewall should be able to correctly identify the application and block it

5.2.2 BLOCK SPECIFIC ACTION (DEPENDS ON THE APPLICATION)

For example, with Instant Messaging, the NGFW should allow text communications while blocking file transfers.

5.3 USER/GROUP ID AWARE POLICIES

Complex outbound and inbound policies consisting of many rules, objects and applications, verifying that the DUT is capable of correctly determining the correct user/group ID from deep packet inspection and taking the appropriate action.

5.3.1 USERS DEFINED VIA NGFW INTEGRATION WITH ACTIVE DIRECTORY

5.3.2 USERS DEFINED IN NGFW DB

Should integration with Active Directory be unavailable, the firewall is expected to allow local creation of Users / Groups. In addition, the ability to import thousands of Users and dozens of Groups is highly desirable.

The following table is an example of Users & Groups + Firewall and Application Control Policies that will be defined.

Users	Application
David (Sales Person)	Salesforce.com
Jay (DB Administrator)	MySQL DB + SSH
Jeff (Operations)	ERP
Pam (Controller)	Accounting software
Richard (VP of Marketing)	ALL
Scott (Auditor)	Accounting software (Read Only)

Groups	Applications
Accounting	Accounting software
Consultant	ERP
Executive	ALL
IT	SSH
Operations	ERP
Sales	Salesforce.com

5.4 INTRUSION PREVENTION POLICIES

Policies consisting of threat protection signatures, verifying that the DUT is capable of correctly blocking malicious traffic based on a comparison of packet/session contents against signatures/filters/protocol decoders.

This section verifies that the NGFW is capable of detecting and blocking a wide range of common exploits accurately, while remaining resistant to false positives. All tests are performed initially with no background network load. The tests are then repeated under varying levels and mixes of background traffic to ensure that the results do not vary when handling normal network traffic.

The latest signature pack is acquired from the vendor, and the NGFW is deployed with the pre-defined recommended security policy. NSS Labs considers it unacceptable for a product of this nature to be sold without a *recommended* policy. No custom signatures are permitted in the testing. All signatures used must be available to the general public at the time of testing.

Although intrusion detection systems operate in detection-only mode, an NGFW is required to block and log exploit attempts and hostile traffic.

5.4.1 DETECTION ENGINE

While it is not possible to validate the entire signature set of any IPS / NGFW, the NSS Labs testing provides a demonstration of effectiveness for the NGFW to protect vulnerable assets from targeted threats and exploitation. This asset/target and threat based approach forms the basis from which NGFW security effectiveness is measured. (See NSS Labs' white paper *Intrusion Prevention Security Effectiveness*, available at www.nsslabs.com).

The NSS Labs threat and attack suite contains thousands of publically-available exploits (including multiple variants of each exploit) from which groups of exploits are carefully selected to test based on appropriate usage. Each exploit has been validated to impact the target vulnerable host(s). Based on the impact of the threat against the target, the following metrics are reported:

5.4.1.1 SYSTEM EXPOSURE

Attacks resulting in remote system compromise and the ability of the attacker to execute arbitrary system-level commands. Most exploits in this class that are "weaponized" will provide the attacker with a fully interactive remote shell on the target client or server.

5.4.1.2 SERVICE EXPOSURE

Attacks resulting in an individual service compromise but not arbitrary system-level command execution. Typical attacks in this category include service specific attacks such as SQL injection that enable the attacker to execute arbitrary SQL commands within the database service. These attacks are somewhat isolated to the service and do not immediately result in full system-level access to the operating system and all services. However, using additional localized system attacks it may be possible for the attacker to go from the service level to the system level.

5.4.1.3 SYSTEM OR SERVICE FAULT

Attacks resulting in a system- or service-level fault that crashes the targeted service or application and requires administrative action to restart the service or reboot the system. These attacks do not enable the attacker to execute arbitrary commands. However, the resulting impact to the business could be severe given that the attacker could crash the protected system or service.

5.4.2 THREAT VECTOR

Threats and exploits can be initiated by either the target or the attacker targeting either local or remote vulnerabilities. As a result, NSS Labs categorizes threats and exploits into the following matrix:

	Network	Local
Attacker	RPC Exploit	Root Kit
Target	Browser Exploit	Trojan

*Example exploits included above for reference purposes.

5.4.2.1 ATTACKER INITIATED

The threat/exploit is executed by the attacker remotely against a vulnerable application and/or operating system.

5.4.2.2 TARGET INITIATED

The threat/exploit is initiated by the vulnerable target. The attacker has little or no control as to when the target user or application will execute the threat.

5.4.2.3 NETWORK

Threat/exploits that are initiated as a result of network communication.

5.4.2.4 LOCAL

Local execution that requires existing access to the target (not applicable to NGFW).

Protective ratings are reported in raw percentages of mitigated attacks and their resulting impact: system, service, fault, reconnaissance. Although a system or service exploit may be partially mitigated by the IPS, the service could have crashed because of the residual communications resulting in a fault impact on the service or operating system.

5.4.3 TARGET TYPE

The following list of web target types is represented in NSS Labs' live exploit test. Protection capabilities are indicated as percentages.

Web Server	Web Browser
ActiveX	JavaScript
Browser Plug-ins/Add-ons	

5.4.4 COVERAGE BY RESULT

The following results of exploitation are represented in NSS Labs' live exploit test. Protection capabilities are indicated as percentages.

Arbitrary Code Execution	Cross-Site Script
Buffer Overflow	Directory Traversal
Code Injection	Privilege Escalation

5.4.5 COVERAGE BY VENDOR

The following list of vendors is represented in NSS Labs' live exploit test. Protection capabilities are indicated as percentages.

- | | |
|---|---|
| <ul style="list-style-type: none"> • 3Com • Alt-N • Apple • Avast • BitDefender • CA • Citrix • EMC • GNU • HP • IPSwitch • Kaspersky • lighttpd • Macromedia • Mailenable • Mercury • MIT | <ul style="list-style-type: none"> • Adobe • Apache • Atrium • BEA • Borland • Cisco • ClamAV • Facebook • Google • IBM • ISC • LanDesk • Linux • MacroVision • McAfee • Microsoft • Mozilla |
|---|---|

- Mplayer
- MySQL
- Novell
- OpenLDAP
- OpenSSH
- Oracle
- Panda
- Samba
- Snort
- SpamAssassin
- Sun Microsystems
- Trend Micro
- UltraVNC
- VideoLan
- WinAmp
- Winzip
- Multiple Vendors
- NOD32
- Nullsoft
- OpenOffice
- OpenSSL
- Other Misc
- RealNetworks
- SAP
- Sophos
- Squid
- Symantec
- Trillian
- Veritas
- VMWare
- WinFTP
- Yahoo

5.5 EVASION

Cyber-criminals can modify basic attacks to evade detection in a number of ways. If an NGFW fails to detect a single form of evasion, any exploit can pass through the device, rendering it ineffective. NSS Labs verifies that the NGFW is capable of detecting and blocking basic exploits when subjected to varying common evasion techniques. Further, the DUT is expected to successfully “decoded” the evasion to provide an accurate alert relating to the original exploit, rather than alerting purely on anomalous traffic detected as a result of the evasion technique itself.

UNMODIFIED EXPLOIT VALIDATION

A number of common exploits are executed across the NGFW to ensure that they are detected in their unmodified state. These will be chosen from a suite of older/common basic exploits for which NSS Labs is certain that all vendors will have signatures. None of the exploits that were used in Section 5.4 will be used as evasion baselines. This ensures that vendors are not provided with any information on the content of any part of the main NSS Labs exploit library in advance of the test.

5.5.1 PACKET FRAGMENTATION

These tests determine the effectiveness of the fragment reassembly mechanism of the NGFW.

- Ordered 8 byte fragments
- Ordered 24 byte fragments
- Out of order 8 byte fragments
- Ordered 8 byte fragments, duplicate last packet
- Out of order 8 byte fragments, duplicate last packet
- Ordered 8 byte fragments, reorder fragments in reverse
- Ordered 16 byte fragments, fragment overlap (favor new)
- Ordered 16 byte fragments, fragment overlap (favor old)
- Out of order 8 byte fragments, interleaved duplicate packets scheduled for later delivery

It is a requirement of the test that the NGFW submitted should have all IP fragmentation reassembly options enabled by default in the shipping product.

5.5.2 STREAM SEGMENTATION

These tests determine the effectiveness of the stream reassembly mechanism of the NGFW.

- Ordered 1 byte segments, interleaved duplicate segments with invalid TCP checksums
- Ordered 1 byte segments, interleaved duplicate segments with null TCP control flags
- Ordered 1 byte segments, interleaved duplicate segments with requests to resync sequence numbers mid-stream
- Ordered 1 byte segments, duplicate last packet
- Ordered 2 byte segments, segment overlap (favor new)
- Ordered 1 byte segments, interleaved duplicate segments with out-of-window sequence numbers
- Out of order 1 byte segments
- Out of order 1 byte segments, interleaved duplicate segments with faked retransmits
- Ordered 1 byte segments, segment overlap (favor new)
- Out of order 1 byte segments, PAWS elimination (interleaved duplicate segments with older TCP timestamp options)
- Ordered 16 byte segments, segment overlap (favor new (Unix))

It is a requirement of the test that the DUT submitted should have all TCP stream reassembly options enabled by default in the shipping product.

5.5.3 RPC FRAGMENTATION

Both Sun/ONC RPC and MS-RPC allow the sending application to fragment requests, and all MS-RPC services have a built-in fragmentation reassembly mechanism.

An attacker can transmit the BIND followed by a single request fragmented over a hundred actual requests with small fragments of the malicious payload. Alternatively, the attacker could transmit both the BIND and request fragments in one large TCP segment, thus foiling any signatures which use a simple size check.

Immunitysec's CANVAS test tool combines large writes with many tiny MS-RPC fragments and provides up to ten levels of fragmentation. These tests determine the effectiveness of the RPC reassembly mechanism of the NGFW:

- One-byte fragmentation (ONC)
- Two-byte fragmentation (ONC)
- All fragments, including Last Fragment (LF) will be sent in one TCP segment (ONC)
- All fragments except Last Fragment (LF) will be sent in one TCP segment. LF will be sent in separate TCP segment (ONC)
- One RPC fragment will be sent per TCP segment (ONC)
- One LF split over more than one TCP segment (in this case, no RPC fragmentation is performed (ONC))
- Canvas Reference Implementation Level 1 (MS)
- Canvas Reference Implementation Level 2 (MS)
- Canvas Reference Implementation Level 3 (MS)
- Canvas Reference Implementation Level 4 (MS)
- Canvas Reference Implementation Level 5 (MS)
- Canvas Reference Implementation Level 6 (MS)
- Canvas Reference Implementation Level 7 (MS)
- Canvas Reference Implementation Level 8 (MS)
- Canvas Reference Implementation Level 9 (MS)
- Canvas Reference Implementation Level 10 (MS)

5.5.4 URL OBFUSCATION

Random URL encoding techniques are employed to transform simple URLs which are often used in pattern-matching signatures to apparently meaningless strings of escape sequences and expanded path characters using a combination of the following techniques:

- Escape encoding (% encoding)
- Microsoft %u encoding
- Path character transformations and expansions (/./ , //, \)

These techniques are combined in various ways for each URL tested, ranging from minimal transformation, to extreme (every character transformed). All transformed URLs are verified to ensure they still function as expected after transformation.

- URL encoding - Level 1 (minimal)
- URL encoding - Level 2
- URL encoding - Level 3
- URL encoding - Level 4
- URL encoding - Level 5
- URL encoding - Level 6
- URL encoding - Level 7
- URL encoding - Level 8 (extreme)
- Premature URL ending
- Long URL
- Fake parameter
- TAB separation
- Case sensitivity
- Windows \ delimiter
- Session splicing

5.5.5 HTML OBFUSCATION

Recognizing malicious HTML documents is becoming increasingly important when protecting the enterprise. Malicious HTML documents exploit flaws in common web browsers, browser plug-ins, and add-ons to gain control of the client system and silently install malware such as Trojans, rootkits, and key loggers.

Therefore, it is becoming increasingly important that security products charged with protecting end systems must correctly interpret HTML documents. Many security products use simple pattern matching systems with very little semantic or syntactic understanding of the data they are analyzing. This leaves them vulnerable to evasion through use of redundant, but equivalent, alternative representations of malicious documents.

This test suite uses a number of malicious HTML documents which are transferred from server to client through the DUT. Each malicious HTML document is served with a different form of obfuscation, as follows:

- UTF-16 character set encoding (big-endian)
- UTF-16 character set encoding (little-endian)
- UTF-32 character set encoding (big-endian)
- UTF-32 character set encoding (little-endian)
- UTF-7 character set encoding
- Chunked encoding (random chunk size)
- Chunked encoding (fixed 8 byte chunk size)
- Chunked encoding (chaffing / arbitrary numbers inserted between chunks)

- Compression (Deflate)
- Compression (Gzip)
- Combination: UTF-7 encoding + Gzip compression + chunked encoding (random chunk sizes)

The UTF-16 character set specifies a 2-byte sequence for most characters and a 4-byte sequence for the others (a small percentage). Recoding an HTML document in UTF-16 significantly changes its appearance. A document that contains just the ASCII subset of characters will appear to have a null byte between every one of the original characters. There are also two different forms of the UTF-16 encoding depending on whether the null high byte comes first (big-endian) or second (little-endian) - this test uses big-endian byte ordering

The UTF-32 character set specifies a 4-byte sequence. Like the UTF-16 character set encoding there are two variations -big-endian and little-endian and this test case uses big-endian byte ordering.

The UTF-7 character set encodes most ASCII characters as themselves. However, in addition to recoding non-English characters as other encodings do, it also recodes many punctuation symbols, including many of the symbols that are important to the HTML specification. Therefore, recoding an HTML document in UTF-7 significantly changes its appearance

Chunked encoding allows the server to break a document into smaller chunks and transmit them individually. The server needs only to specify the size of each chunk before it is transmitted and then indicate when the last chunk has been transmitted. Since chunked encoding intersperses arbitrary numbers (chunk sizes) with the elements of the original document, it can be used to greatly change the appearance of the original document as observed "on the wire". In addition, the server can choose to break the document into chunks at arbitrary points. This makes it difficult for simple pattern matching systems to reliably identify the original HTML document from the raw data on the network.

Per RFC 2616, the HTTP protocol allows the client to request and the server to use several compression methods. These compression methods not only improve performance in many circumstances, they completely change the characteristic size and appearance of HTML documents. Furthermore, small changes in the original document, can greatly change the final appearance of the compressed document. This property of these algorithms could be used to obfuscate hostile content for the purpose of evading detection. The deflate compression method is a Lempel-Ziv coding (LZ77), specified in RFC 1951. The gzip compression method is specified in RFC 1952.

For each of the above, it is verified that a standard Web browser (such as Internet Explorer) is capable of rendering the results of the evasion.

5.5.6 FTP EVASION

When attempting FTP exploits, it is possible to evade some IDS/NIPS products by inserting additional spaces and telnet control sequences in FTP commands.

These tests insert a range of valid telnet control sequences that can be parsed and handled by IIS FTP server and wu-ftpd, and which also conform to Section 2.3 of RFC 959. Control opcodes are inserted at random, ranging from minimal insertion (only one pair of opcodes), to extreme (opcodes between every character in the FTP command):

- Inserting spaces in FTP command lines
- Inserting non-text Telnet opcodes - Level 1 (minimal)
- Inserting non-text Telnet opcodes - Level 2
- Inserting non-text Telnet opcodes - Level 3
- Inserting non-text Telnet opcodes - Level 4

- Inserting non-text Telnet opcodes - Level 5
- Inserting non-text Telnet opcodes - Level 6
- Inserting non-text Telnet opcodes - Level 7
- Inserting non-text Telnet opcodes - Level 8 (extreme)

6 FIREWALL PERFORMANCE

This section measures the performance of the DUT using various traffic conditions that provide metrics for real world performance. Individual implementations will vary based on usage, however these quantitative metrics provide a gauge as to whether a particular DUT is appropriate for a given environment.

The DUT will be deployed with NAT enabled and the policy defined for test 5.1.3. This represents the most likely real-world deployment configuration.

Multiple separate 1Gbps connections will be made from the external to internal switches via the DUT, subject to a minimum of one in-line port pair per Gigabit of throughput. Thus an 8Gbps device with only four port pairs will be limited to 4Gbps. The minimum number of port pairs will be connected to support the claimed maximum bandwidth of the DUT. Thus an 8 Gbps device with ten port pairs will be deployed with eight 1Gbps connections. One in-line connection will pass from the trusted to external network, and the remainder will pass from internal to DMZ.

Basic traffic – both permitted and non-allowed – is passed through the DUT to verify the policy is operational. Once that has been established, increasing levels of varying types of background traffic are generated through the DUT in order to determine the point at which the DUT begins to drop packets.

All tests are repeated with background traffic levels of 25%, 50%, 75% and 100% of the maximum throughput of the device, and for each type of background traffic the maximum load the DUT can sustain before it begins to drop packets is determined.

Any device which permits malicious traffic to pass through the DUT will fail the overall test immediately.

6.1 RAW PACKET PROCESSING PERFORMANCE (UDP TRAFFIC)

This test uses UDP packets of varying sizes generated by Spirent SmartBits traffic generation tools.

A constant stream of the appropriate packet size - with variable source IP addresses and ports transmitting to a single fixed IP address/port - is transmitted bi-directionally through each port pair of the DUT.

Each packet contains dummy data, and is targeted at a valid port on a valid IP address on the target subnet. The percentage load and packets per second (pps) figures across each in-line port pair are verified by the Adtech network monitoring tool before each test begins. Multiple tests are run and averages taken where necessary. Each test is repeated with traffic loads of 25%, 50%, 75% and 100% of the maximum throughput of the DUT, and the percentage of attacks detected and blocked is recorded at each load level. Maximum throughput with zero packet loss is also recorded.

This traffic does not attempt to simulate any form of “real world” network condition. No TCP sessions are created during this test, and there is very little for the detection engine to do in the way of protocol analysis (although each vendor will be required to write a signature to detect the test packets to ensure that they are being passed through the detection engine and not “fast-tracked” from the inbound to outbound port).

The aim of this test is purely to determine the raw packet processing capability of each in-line port pair of the DUT, and its effectiveness at passing “useless” packets quickly in order to pass potential attack packets to the detection engine.

6.1.1 128 BYTE PACKETS

Maximum 842,000 Packets Per Second per Gigabit of traffic. Repeated with traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT.

6.1.2 256 BYTE PACKETS

Maximum 452,000 Packets Per Second per Gigabit of traffic. Repeated with traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT.

6.1.3 512 BYTE PACKETS

Maximum 235,000 Packets Per Second per Gigabit of traffic. Repeated with traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. This test provides a reasonable indication of the ability of a device to process packets from the wire on an “average” network.

6.1.4 1024 BYTE PACKETS

Maximum 120,000 Packets Per Second per Gigabit of traffic. Repeated with traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT.

6.1.5 1514 BYTE PACKETS

Maximum 82,000 Packets Per Second per Gigabit of traffic. Repeated with traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. This test has been included mainly to demonstrate how easy it is to achieve good results using large packets – beware of test results that only quote performance figures using similar packet sizes.

6.2 MAXIMUM CAPACITY

The aim of these tests is to stress the DUT and determine how it copes with large numbers of TCP connections per second, application layer transactions per second, and concurrent open connections. All packets contain valid payload and address data, and these tests provide an excellent representation of a live network at various connection/transaction rates.

Note that in all tests, the following critical “breaking points” - where the final measurements are taken - are used:

- **Excessive concurrent TCP connections** - latency within the DUT is causing unacceptable increase in open connections on the server-side
- **Excessive response time for HTTP transactions/SMTP sessions** - latency within the DUT is causing excessive delays and increased response time to client
- **Unsuccessful HTTP transactions/SMTP sessions** - normally there should be zero unsuccessful transactions. Once these appear, it is an indication that excessive latency within the DUT is causing connections to time out

The use of multiple BreakingPoint appliances allows us to create true “real world” traffic at multi-Gigabit speeds as a background load for our tests.

6.2.1 THEORETICAL MAXIMUM CONCURRENT TCP CONNECTIONS

This test is designed to determine the maximum concurrent TCP connections of the DUT with no data passing across the connections. This type of traffic would not typically be found on a normal network, but it provides the means to determine the maximum possible concurrent connections figure.

A maximum of 7.5 million Layer 4 TCP sessions are opened across the DUT. Each session is opened normally and then held open for the duration of the test as additional sessions are added up to the maximum possible. Load is increased until no more connections can be established, and this number is recorded.

6.2.2 THEORETICAL MAXIMUM CONCURRENT TCP CONNECTIONS WITH DATA

This test is identical to 6.2.1 except that once the maximum number of concurrent connections have been established, 1GB of data is transmitted across them in 21KB segments. This ensures that the DUT is capable of passing data across the connections once they have been established.

6.2.3 MAXIMUM CONCURRENT STATEFUL TCP CONNECTIONS

This test is identical to 6.2.1, but is designed to verify the maximum concurrent TCP connections on which the vendor claims the DUT can maintain state.

A legitimate HTTP session is opened and the first packet of a two-packet exploit is transmitted. As the number of open connections approaches the maximum claimed by the vendor, the initial HTTP session is then completed with the second half of the exploit and the session is closed. If the sensor is still maintaining state on the original session, the exploit will be recorded. If the state tables have been exhausted, the exploit string will be seen as a non-stateful attack, and will thus be ignored.

Both halves of the exploit are required to trigger an alert - a product will fail the test if it fails to generate an alert after the second packet is transmitted, or if it raises an alert on either half of the exploit on its own.

6.2.4 MAXIMUM TCP CONNECTIONS PER SECOND

This test is designed to determine the maximum TCP connection rate of the DUT with one byte of data passing across the connections. This type of traffic would not typically be found on a normal network, but it provides the means to determine the maximum possible TCP connection rate.

A maximum of 2,500,000 connections per second are generated across the DUT, ramped slowly to determine the exact point of failure. Each session is opened normally, one byte of data passed through the DUT, and then the session is closed immediately. Load is increased until one or more of the breaking points defined earlier is reached.

6.2.5 MAXIMUM HTTP CONNECTIONS PER SECOND

Given that the NGFW is designed specifically to identify applications that typically tunnel over port 80, HTTP Connections Per Second is a key metric. This test is designed to determine the maximum TCP connection rate of the DUT with a 1 byte HTTP response size. The response size defines the number of bytes contained in the body, excluding any bytes associated with the HTTP header. A 1 byte response size is designed to provide a theoretical maximum HTTP connections per second rate.

Client and server are using HTTP 1.0 without keep alive, and the client will open a TCP connection, send one HTTP request, and close the connection. This ensures that all TCP connections are closed immediately the request is satisfied, thus any concurrent TCP connections will be caused purely as a result of latency within the DUT. Load is increased until one or more of the breaking points defined earlier is reached.

6.2.6 MAXIMUM HTTP TRANSACTIONS PER SECOND

This test is designed to determine the maximum HTTP transaction rate of the DUT with a 1 byte HTTP response size. The object size defines the number of bytes contained in the body, excluding any bytes associated with the HTTP header. A 1 byte response size is designed to provide a theoretical maximum connections per second rate.

Client and server are using HTTP 1.1 with persistence, and the client will open a TCP connection, send ten HTTP requests, and close the connection. This ensures that TCP connections remain open until all ten HTTP transactions are complete, thus eliminating the maximum connection per second rate as a bottleneck (one TCP connection = 10 HTTP transactions). Load is increased until one or more of the breaking points defined earlier is reached.

6.2.7 MAXIMUM THROUGHPUT

Given that the NGFW is designed specifically to identify applications that typically tunnel over port 80, HTTP Maximum Throughput is a key metric. This test is designed to determine the maximum throughput of the DUT using a low HTTP transaction rate with large response sizes to provide maximum bandwidth from minimum TCP connections per second or HTTP transactions per second.

Client and server are using HTTP 1.1 with persistence, and the client will open a TCP connection, send ten HTTP requests with large response sizes, and close the connection. This ensures that TCP connections remain open until all ten HTTP transactions are complete, thus eliminating the maximum connection per second rate as a bottleneck (one TCP connection = 10 HTTP transactions). Load is increased until one or more of the breaking points defined earlier is reached.

6.3 BEHAVIOR OF THE STATE ENGINE UNDER LOAD

This test determines whether the DUT is capable of preserving state across a large number of open connections over an extended time period.

At various points throughout the test (including after the maximum has been reached), it is confirmed that the DUT is still capable of inspecting and blocking traffic which is in violation of the currently-applied security policy, whilst confirming that legitimate traffic is not blocked (perhaps as a result of state tables filling up). The NGFW needs to be able to apply policy decisions effectively based on inspected traffic at all load levels.

6.3.1 ATTACK DETECTION/BLOCKING - NORMAL LOAD

This test determines if the DUT is able to detect and block policy violations as the number of open sessions reaches 75 per cent of the maximum determined in Test 6.2.1.

6.3.2 PASS LEGITIMATE TRAFFIC - NORMAL LOAD

This test ensures that the DUT continues to pass legitimate traffic as the number of open sessions reaches 75 per cent of the maximum determined in Test 6.2.1.

6.3.3 STATE PRESERVATION - NORMAL LOAD

This test determines if the sensor maintains the state of pre-existing sessions as the number of open sessions reaches 75% of the maximum determined in Section 6.2.1.

A legitimate HTTP session is opened and the first packet of a two-packet exploit is transmitted. As the number of open connections approaches the maximum, the initial HTTP session is then completed with the second half of the exploit and the session is closed. If the sensor is still maintaining state on the original session, the exploit will be recorded. If the state tables have been exhausted, the exploit string will be seen as a non-stateful attack, and will thus be ignored. Both halves of the exploit are

required to trigger an alert. A product will fail the test if it fails to generate an alert after the second packet is transmitted or if it raises an alert on either half of the exploit on its own.

6.3.4 STATE PRESERVATION - MAXIMUM EXCEEDED

This test determines if the sensor maintains the state of pre-existing sessions as the number of open sessions exceed the maximum determined in Test 6.2.1. Method of execution is identical to Test 6.3.3.

6.3.5 DROP LEGITIMATE TRAFFIC - MAXIMUM EXCEEDED

This test ensures that the sensor continues to drop all traffic as the number of open sessions exceed the maximum determined in Test 6.2.1.

Note: If a DUT allows traffic to “leak” due to the way it expires old connections, the result will be an automatic fail for the entire test.

6.4 HTTP CAPACITY WITH NO TRANSACTION DELAYS

Given that the NGFW is designed specifically to identify applications that typically tunnel over port 80, the ability to handle heavy loads of HTTP traffic is a key metric. The aim of these tests is to stress the HTTP detection engine and determine how the DUT copes with detecting and blocking security policy violations under network loads of varying average packet size and varying connections per second. By creating genuine session-based traffic with varying session lengths, the sensor is forced to track valid TCP sessions, thus ensuring a higher workload than for simple packet-based background traffic. This provides a test environment that is as close to “real world” as it is possible to achieve in a lab environment, while ensuring absolute accuracy and repeatability.

Each transaction consists of a single HTTP GET request and there are no transaction delays (i.e. the Web server responds immediately to all requests). All packets contain valid payload (a mix of binary and ASCII objects) and address data, and this test provides an excellent representation of a live network (albeit one biased towards HTTP traffic) at various network loads.

6.4.1 2,500 CONNECTIONS PER SECOND

Max 2,500 new connections per second per Gigabit of traffic with a 44KB HTTP response size - average packet size 900 bytes - maximum 140,000 packets per second per Gigabit of traffic. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. With relatively low connection rates and large packet sizes, all sensors should be capable of performing well throughout this test.

6.4.2 5,000 CONNECTIONS PER SECOND

Max 5,000 new connections per second per Gigabit of traffic with a 21KB HTTP response size - average packet size 670 bytes - maximum 185,000 packets per second per Gigabit of traffic. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. With average connection rates and average packet sizes, this is a good approximation of a real-world production network, and all sensors should be capable of performing well throughout this test.

6.4.3 10,000 CONNECTIONS PER SECOND

Max 10,000 new connections per second per Gigabit of traffic with a 10KB HTTP response size - average packet size 550 bytes - maximum 225,000 packets per second per Gigabit of traffic. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of

DUT. With average packet sizes coupled with very high connection rates this represents a very heavily used production network and is a strenuous test for any sensor.

6.4.4 20,000 CONNECTIONS PER SECOND

Max 20,000 new connections per second per Gigabit of traffic with a 4.5KB HTTP response size - average packet size 420 bytes - maximum 300,000 packets per second per Gigabit of traffic. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. With small packet sizes and extremely high connection rates this is an extreme test for any sensor.

6.4.5 40,000 CONNECTIONS PER SECOND

Max 40,000 new connections per second per Gigabit of traffic with a 1.7KB HTTP response size - average packet size 270 bytes - maximum 445,000 packets per second per Gigabit of traffic. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. With small packet sizes and extremely high connection rates this is an extreme test for any sensor.

6.5 "REAL WORLD" TRAFFIC

The aim of this test is to simulate a "real world" environment by introducing additional protocols, applications and real content while still maintaining a precisely repeatable and consistent background traffic load (something rarely seen in a real world environment).

The result is a background traffic load that is closer to what may be found on a heavily-utilized "normal" production network. Load is increased to the maximum throughput level as previously determined to ensure that the same level can be achieved with real world traffic. Average response times per protocol are also recorded.

6.5.1 "REAL WORLD" PROTOCOL MIX (PERIMETER)

Traffic is generated across the DUT comprising the following protocol mix typical of that seen by a perimeter security device:

- | | | |
|-----------------------|-----|--------|
| • HTTP text | 33% | |
| • SMTP | 18% | |
| • HTTP Images (<50k) | 14% | FTP 8% |
| • DNS | 6% | |
| • HTTP Video | 4% | |
| • HTTP Audio | 4% | |
| • HTTP Images (>300k) | 4% | |
| • SSH | 4% | |
| • AIM | 3% | |
| • SIP/RTP | 1% | |
| • BitTorrent | 1% | |

For this test and 0, HTTP traffic comprises genuine transactions and Web pages from real Web sites such as Google, Yahoo, MSN, NSS Labs, etc. including small (<50KB) and large (>300KB) Jpeg images. Also included as part of the HTTP traffic is genuine QuickTime movie content and MP3 files, taking the total HTTP traffic of all types to approximately 65% of the overall load. SMTP traffic comprises real e-mail messages of varying lengths (with and without attachments) from the NSS Labs mail server.

Maximum 6000 connections per second per Gigabit of traffic - 220,000 packets per second per Gigabit of traffic - average packet size of 550 bytes. Repeated with background traffic loads of 25%, 50%, 75% and 100% of maximum throughput of DUT. Maximum of 5,000 open connections during the test.

With lower connection rates, average packets sizes, and a common protocol mix comprising protocols which all require inspection by the NGFW Deep Inspection / IPS Engine, this is a good approximation of a heavily-used production network. All sensors should be capable of performing well throughout this test (and 0).

6.5.2 “REAL-WORLD” PROTOCOL MIX (CORE)

Traffic is generated across the IPS comprising the following protocol mix typical of that seen by a network core security device:

- HTTP text 24%
- SMB File transfer 14%
- HTTP Images (<50 kb) 12%
- SMTP 12%
- PostgreSQL 10%
- DNS 6%
- DCERPC 4%
- FTP 3%
- SMB NULL 3%
- HTTP Video 2%
- HTTP Audio 2%
- HTTP Images (>300 kb) 2%
- AIM 2%
- SIP/RTP 1%
- NFS 1%
- SSH 1%
- RTSP 1%

Maximum 5000 connections per second per Gigabit of traffic - 270,000 packets per second per Gigabit of traffic - average packet size of 440 bytes. Repeated with background traffic loads of 25%, 50%, 75%, and 100% of maximum throughput of IPS. Maximum of 6,000 open connections during the test.

6.6 LATENCY & RESPONSE TIMES

The aim of the latency and user response time tests is to determine the effect the Firewall has on the traffic passing through it under various load conditions.

Should a device impose a high degree of latency on the packets passing through it, a network or security administrator would need to think carefully about how many devices could be installed in a single data path before user response times became unacceptable or the combination of devices caused excessive timeouts.

This test uses UDP packets of varying sizes generated by BreakingPoint appliances to determine raw packet latency. The BreakingPoint software runs through several iterations of the test, varying the traffic load through multiple in-line port pairs bi-directionally from 25% to 100% of the maximum DUT throughput with zero packet loss with 512 byte packets as determined in test 6.1.3.

This is repeated for a range of packet sizes (128, 256, 512, 1024 and 1514 bytes) of UDP traffic with variable IP addresses and ports. At each iteration of the test, the test equipment records the number of packets dropped, together with average and maximum latency, measured in microseconds.

This test provides an indication of how much the DUT affects the traffic flow through it. This data is particularly useful for network administrators who need to gauge the effect of any form of in-line device which is likely to be placed at critical points within the corporate network.

Note that at any packet size NSS considers 1ms to be the acceptable limit for a typical perimeter deployment, and 300 μ s to be the acceptable limit for the network core.

6.6.1 LATENCY - UDP

Test traffic is passed across the infrastructure switches and through all in-line port pair of the DUT simultaneously (the latency of the basic infrastructure is known and is constant throughout the tests).

The packet loss and average latency (μ S) are recorded at each packet size (128, 256, 512, 1024 and 1514 bytes) and each load level from 25% to 100% of the maximum throughput with zero packet loss with 512 byte packets as previously determined in test 6.1.3.

6.6.2 APPLICATION RESPONSE TIME – HTTP

Test traffic is passed across the infrastructure switches and through all in-line port pair of the DUT simultaneously (the latency of the basic infrastructure is known and is constant throughout the tests).

The results recorded at each transaction size (44KB, 21KB, 10KB, 4.5KB, and 1.7KB HTTP Responses) load level of 90% of the maximum throughput with zero packet loss as previously determined in test 6.4

- Application Average Response Time: HTTP

7 STABILITY & RELIABILITY

Long term stability is particularly important for an in-line device, where failure can produce network outages. These tests verify the stability of the DUT along with its ability to maintain security effectiveness while under normal load and while passing malicious traffic. Products that are not able to sustain legitimate traffic (or crash) while under hostile attack will not pass.

The DUT is required to remain operational and stable throughout these tests, and to block 100 per cent of previously blocked traffic, raising an alert for each. If any non-allowed traffic passes successfully - caused by either the volume of traffic or the DUT failing open for any reason - this will result in a FAIL.

7.1 BLOCKING UNDER EXTENDED ATTACK

The DUT is exposed to a constant stream of security policy violations over an extended period of time. The device is configured to block and alert, and thus this test provides an indication the effectiveness of both the blocking and alert handling mechanisms.

A continuous stream of security policy violations mixed with legitimate traffic is transmitted through the device at a maximum of 100Mbps (max 50,000 packets per second, average packet sizes in the range of 120-350 bytes) for 8 hours with no additional background traffic. This is not intended as a stress test in terms of traffic load (covered in the previous section) - merely a reliability test in terms of consistency of blocking performance.

The device is expected to remain operational and stable throughout this test, and to block 100 per cent of recognisable violations, raising an alert for each. If any recognisable policy violations are passed - caused by either the volume of traffic or the sensor failing open for any reason - this will result in a FAIL.

7.2 PASSING LEGITIMATE TRAFFIC UNDER EXTENDED ATTACK

This test is identical to 7.1.1, where the external interface of the device is exposed to a constant stream of exploits over an extended period of time.

The device is expected to remain operational and stable throughout this test, and to pass most/all of the legitimate traffic. If an excessive amount of legitimate traffic is blocked throughout this test - caused by either the volume of traffic or the DUT failing for any reason - this will result in a FAIL.

7.3 PROTOCOL FUZZING

This test stresses the protocol stacks of the DUT by exposing it to traffic from various protocol randomizer tools. Several of the tools in this category are based on the ISIC test suite and the BreakingPoint *Stack Scrambler* component.

Traffic load is a maximum of 350Mbps and 60,000 packets per second (average packet size is 690 bytes). Results are presented as a simple PASS/FAIL - the device is expected to remain operational and capable of detecting and blocking exploits throughout the test.

7.4 PROTOCOL MUTATION

This test stresses the protocol stacks of the DUT by exposing it to traffic from various protocol mutation tools. Several of the tools in this category are based on the Mu Security Analyzer and the BreakingPoint *Stack Scrambler* component.

7.5 POWER FAIL

Power to the DUT is cut whilst passing a mixture of legitimate and disallowed traffic. Firewalls should always be configured to fail closed - no traffic should be passed once power has been cut.

7.6 REDUNDANCY

Does the DUT include multiple redundant critical components (fans, power supplies, hard drive, etc.) (YES/NO/OPTION).

7.7 PERSISTENCE OF DATA

The DUT should retain all configuration data, policy data and locally logged data once restored to operation following power failure.

8 MANAGEMENT AND CONFIGURATION COSTS

Organizations should be concerned with the ongoing, amortized cost of operating security products. This section evaluates the costs associated with the purchase, installation, and ongoing management of the IPS.

8.1 EASE-OF-USE

8.1.1 INITIAL SETUP (HOURS)

The initial setup costs include those associated with time required to install and configure the IPS.

8.1.2 TIME REQUIRED FOR UPKEEP (HOURS PER YEAR)

The time required to keep the device in good health. This includes applying patches, OS updates, exporting and archiving logs, generating reports, and backing up the system.

8.1.3 TIME REQUIRED TO TUNE (HOURS PER YEAR)

The time required to keep create, deploy, and maintain the optimal security policy. This includes downloading signatures/filters, defining updated policies based upon those new signatures, and eliminating all false positives.

8.2 EXPECTED COSTS

8.2.1 INITIAL PURCHASE

8.2.2 ONGOING MAINTENANCE & SUPPORT (ANNUAL)

8.2.3 INSTALLATION LABOR COST (@\$75/HR)

8.2.4 MANAGEMENT LABOR COST (PER YEAR @\$75/HR)

8.2.5 TUNING LABOR COST (PER YEAR @\$75/HR)

8.3 TOTAL COST OF OWNERSHIP

8.3.1 YEAR 1

8.3.2 YEAR 2

8.3.3 YEAR 3

8.3.4 THREE-YEAR TOTAL COST OF OWNERSHIP

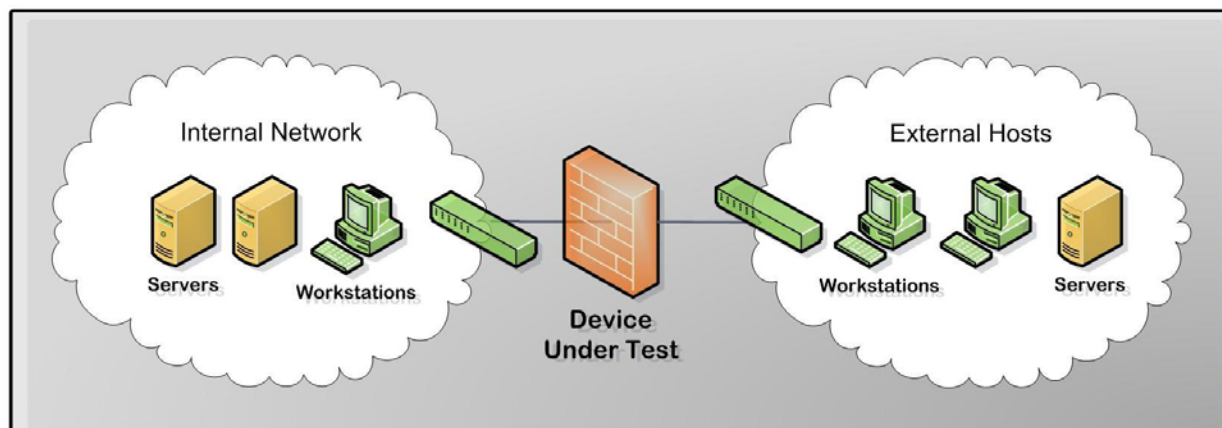
APPENDIX A: FIREWALL TEST ENVIRONMENT

The aim of this procedure is to provide a thorough test of all the main components of a Firewall device in a controlled and repeatable manner and in the most “real world” environment that can be simulated in a test lab.

The NSS Labs test network is a multi-Gigabit infrastructure based around multiple Cisco Catalyst 6500-series switches (these have a mix of fiber and copper Gigabit interfaces). The Firewall will be configured for the use-case appropriate to the target deployment environment.

Traffic generation equipment - such as the BreakingPoint and Spirent Smartbits transmit ports - is connected to both the “internal” network and the “external” network. The “receiving” equipment - such as the BreakingPoint and Spirent Smartbits receive ports - is connected to both the “internal” and “external” network. Attackers are connected to the external network, whilst vulnerable hosts and internet-facing servers (Web, FTP, etc.) are connected to the internal and DMZ networks. This enables testing of product performance for multiple scenarios according to product deployment and usage.

The Firewall is connected between two “gateway” switches - one at the edge of the external network, and one at the edge of the external network.



All “normal” network traffic and background load traffic will therefore be transmitted through the Firewall, from internal to external (responses will flow in the opposite direction) to simulate internal client originating traffic, as well as from external to internal (responses will flow in the opposite direction) to simulate remote client to internal (DMZ) server traffic. Attacker Initiated exploit traffic will flow from remote client to internal (DMZ) server, while Target/Client Initiated exploit traffic will flow in the opposite direction. The same traffic is mirrored to multiple SPAN ports of the external gateway switch, to which Adtech AX/4000 network monitoring devices are connected. The Adtech AX/4000’s monitor the same mirrored traffic to ensure that the total amount of traffic per in-line port pair never exceeds 1Gbps.

The management interface is used to connect the appliance to the management console on a private subnet. This ensures that the Firewall and console can communicate even when the target subnet is subjected to heavy loads, in addition to preventing attacks on the console itself.

APPENDIX B: SPECIAL THANKS

Special thanks go to our test infrastructure partners who provide much of the equipment, software, and support that make this testing possible:

