ANALYST BRIEF – May 2013

Correlation Of Detection Failures

THE CHALLENGE FOR LAYERED SECURITY

Author – Stefan Frei, Ph.D.

Overview

Over the past eighteen months, NSS Labs has tested the security effectiveness of typical defense technologies, such as next generation firewall (NGFW), intrusion prevention systems (IPS), and endpoint protection (EPP – also referred to as antivirus/malware detection). A comparison of exploit block performances within and across these defense technologies has revealed a significant correlation of failures to detect exploits. Such detection failures present a serious challenge to the security industry as they allow an attacker to bypass several layers of defense using only a small set of exploits.

In multiple independent group tests carried out at the NSS testing facilities in Austin TX, NSS engineers determined the ability of 37 security products from 24 different security vendors to block exploits in real world attack scenarios. The 1,711 exploits used in these tests target 816 software products from 208 different software vendors, thereby covering 21% of all vulnerabilities published against these software products in the last 10 years.

None of the 37 tested security devices managed to detect all exploits and only 3% of 606 unique combinations of two security products managed to detect all exploits. Further, there is a large diversity in the security performance between individual security products or combinations of security products.

The analysis of these test results documents a significant correlation of failures to detect exploits between security products. The number of exploits found to bypass multiple security devices, as well as the number of security devices simultaneously bypassed by these exploits, is significantly higher than the common expectation, or than the predictions of those risk models that ignore the effects of this correlation.

This can lead security professionals to overestimate the combined security effect of deploying multiple different protection technologies. This significant correlation of detection failures indicates that deploying multiple products within a security category (such as IPS), or even multiple products across multiple categories (such as EPP behind IPS behind NGFW), does not always provide the “defense in depth” that we are led to believe exists from studying vendor claims for the efficacy of their products. This is because most vendors use the same sources of threat intelligence and the same vulnerability research feeds as each other, and this means that they will, more often than not, have the same deficiencies in their coverage.

Layered security, e.g. the deployment of layers of different security technologies, is beneficial when looking to secure the enterprise. However the choice of security devices to be combined is key to realizing substantial security gains and offsetting the increase in complexity, management, and cost.

Naïve risk modeling that ignores correlation, however, will result in a basic lack of understanding of the scope of exploits currently in common use that are able to bypass multiple security products.

The identification and analysis of exploits that escape detection by the majority of the security devices/products in a group test is globally relevant, as these exploits present a significant challenge to the security industry. This analysis shows that, while it is helpful to adopt a layered approach to security, the real key to effective protection against threats lies in an organization's choice of protection technologies to be combined.

NSS Labs Findings

- There is only limited breach prevention available. Not one of the 37 tested security devices managed to detect all of the exploits, and only 3% of the 606 unique security product combinations were able to detect all of the exploits.
- Security performance varies considerably between individual security products, or between combinations of security products. A comparison of the combined block performance of 606 unique pairs of security products revealed the performances to be wide ranging.
- The significant correlation of failures to detect exploits over a wide range of security devices particularly impacts the layered security approach, since the enterprise is inclined to overestimate the security effect of combining multiple protection technologies.
- There are some exploits targeting relevant software that are able to bypass detection by the majority of security devices or combinations of security devices.
- The number of exploits that were able to bypass multiple security devices, as well as the number of security devices that were bypassed by these exploits, is significantly higher than is the prediction for risk models that ignore correlation.
- The average joint failure rate for IPS **and** NGFW is 0.8%, down from an average single-device failure rate of 5.8%.
- The average joint failure rate for multiple enterprise endpoint protection (EPP) products is 26.0%, down from an average single-product failure rate of 45.4%.
- No combination of two security devices in the NGFW 2012 group test would detect all exploits (see Figure 9).
- While it is helpful to adopt a layered approach to security, the real key to effective protection against threats lies in an organization's choice of protection technologies to be combined.

NSS Labs Recommendations

- Enterprises should focus on the effectiveness of specific combinations of devices at blocking specific exploits rather than simply layering randomly in the hope that multiple devices equal higher protection.
- Organizations should assume they are already breached. Prevention should be paired with both breach detection and security information and event management (SIEM) to enable the prompt detection of successful security breaches.
- Security professionals should take into account the effects of correlation when modeling risk. Naïve risk models that ignore correlation of detection failures are severely underestimating the risk of successful compromise.
- Enterprise should prioritize patch management programs to minimize the effects of correlation of failure across multiple security devices.

Table of Contents

Table of Figures

Analysis

Constantly evolving attack methodologies, advanced malware, and innovative evasion techniques present an ongoing challenge to enterprise security. The enterprise has responded to these security threats by deploying multiple layers of diverse security technologies.

NSS testing has discovered a significant correlation of detection failures across a range of diverse security products; this correlation can reduce, or even eliminate, any potential gain in effectiveness from the combination of multiple security technologies and products. Data from NSS real-world testing is used to quantify the potential of malware to successfully evade detection of multiple defense layers or defense technologies.

Comparative analysis of the security products is based upon data gathered during product testing at the NSS facility in Austin, Texas. Testing is performed in accordance with NSS methodologies, which are publicly available on the NSS web site prior to the tests. During these tests, NSS collects data on a range of properties, such as security effectiveness, performance, enterprise management capabilities, and total cost of ownership.

This brief correlates results from the security effectiveness tests, with specific focus on correlating the measured exploit block performance across multiple tested devices/products in the NGFW, IDS, and EPP categories. The data used is derived from NSS group tests conducted over the course of the past 18 months.

In 2012, NSS tested:

- 16 IPS devices from 10 vendors. (IPS 2012)
- 8 NGFW devices from 8 vendors. (NGFW 2012)
- 13 EPP products from 13 vendors. (EPP 2012)

In 2013, NSS tested (up to the date of this writing):

- 9 NGFW devices from 8 vendors (NGFW 2013)

During these group tests, NSS recorded the security performance of 37 unique security products (13 endpoint protection products and 24 network level protection products) from 24 different security vendors.

This brief correlates those results to determine the joint failure rate of 606 unique combinations of security products (note that some of the devices participated in more than one test.)

NSS Exploit Portfolio

An exploit is a piece of software, a piece of data, or a sequence of commands that takes advantage of a security vulnerability in software or hardware, in order to cause the target system to behave in an unintended or unanticipated manner. Exploits allow an attacker to gain control or to escalate privileges on the targeted system, or to render the target unusable through a denial-of-service (DoS) attack.

For the purpose of security testing, NSS uses a current portfolio of exploits that is maintained through the findings of NSS research (reverse engineering), analysis of live malware, and collaboration with the security industry. NSS exploit testing leverages the expertise of NSS engineers, who make use of multiple commercial, open source, and proprietary tools.

With tens of thousands of live exploits, NSS has the industry's most comprehensive test harness. A subset of the exploit library is selected for each round of testing, and all of the live exploits and payloads in the tests have been validated in the NSS 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.

During the tests, NSS engineers trigger vulnerabilities to determine whether an exploit is able to pass through the device under test (DUT).

For the tests referred to in this analysis, NSS used a representative sample of 1,486 exploits for the NGFW 2012 tests and IPS 2012 tests, 1,711 exploits for the NGFW 2013 tests, and a sample of 43 recent endpoint exploits for the EPP 2012 tests. The DUT is required to block and log exploitation attempts and hostile traffic.

Exploit And Vulnerability Classification

The data supporting this analysis is combined with publicly available information from the National Vulnerability Database (NVD)[1], the U.S. government repository of standards-based vulnerability management data. The NVD represents all vulnerability disclosures that have a common vulnerabilities and exposures (CVE) identifier[2].

CVE is a security industry standard that is used to uniquely identify and correlate vulnerabilities. By using CVE identifiers, the vulnerability information can easily be correlated to the respective security patches, exploit availability, or corresponding signatures in protection technologies, such as IPS or anti-virus engines.

The criticality of the vulnerabilities being exploited is rated based on the common vulnerability scoring system (CVSS)[3]. The CVSS base metric assigns a numeric value between 0 and 10 to vulnerabilities, according to criticality, with higher scores indicating greater criticality:

- High criticality – vulnerability has a CVSS base score of 7.0 - 10.0.
- Medium criticality – vulnerability has a CVSS base score of 4.0 - 6.9.
- Low criticality – vulnerability has a CVSS base score of 0.0 - 3.9.

The 1,711 exploits used in the recent tests targeted 816 software products from 208 different vendors. These 208 vendors were affected by 20,230 (43%) of all vulnerabilities published by the NVD since 2002. 24% of these vulnerabilities and 60% of the NSS exploits are classified as highly critical, with a CVSS score ≥ 8.

[1] NVD National Vulnerability Database - http://nvd.nist.gov

[2] CVE Common Vulnerabilities and Exposure - http://cve.mitre.org

[3] CVSS Vulnerability Scoring System - http://www.first.org/cvss

The average CVSS score of the exploits used in the tests is 8.2 – thus NSS tested on average 21% of the highly critical vulnerabilities affecting these 208 vendors in the last 10 years, as seen in Figure 1.

Targeted Vendors: 208	All Risks	Highly Critical	
Vulnerabilities since 2002 (CVE)	20,230	4,913	24%
NSS exploits used (only exploits with CVE)	1,675	1,008	60%
Exploit sampling rate	**8.3%**	**20.5%**	

Figure 1 - Representation of the exploits used in recent tests (highly critical with a CVSS score ≥8).

Figure 2 shows the distribution of these exploits by year of publication. Throughout the testing, NSS found significant numbers of older exploits that were still not detected by multiple security devices.

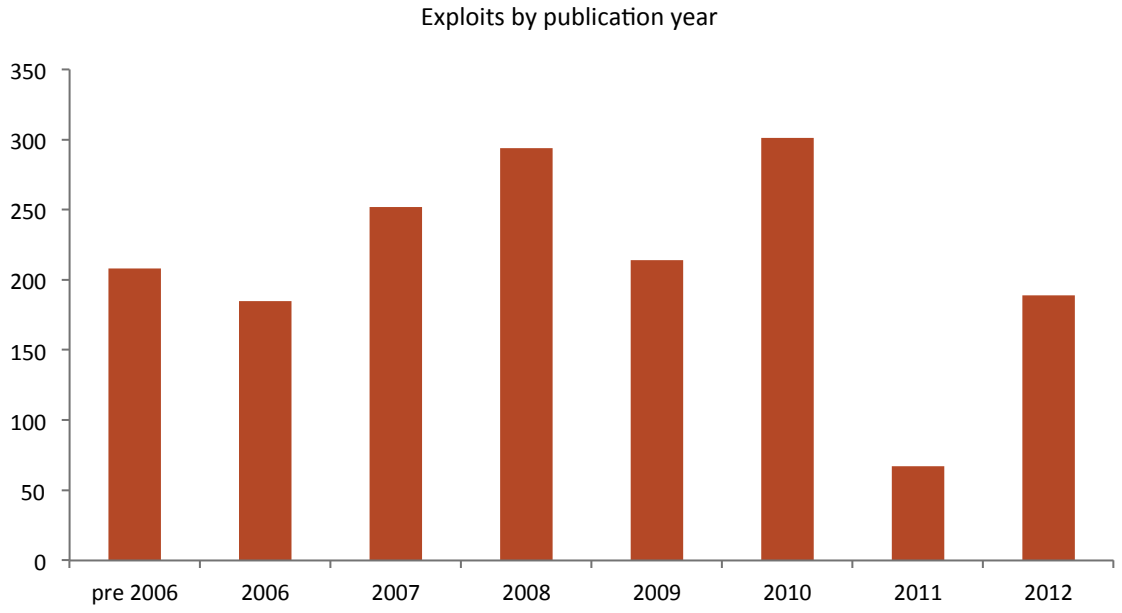

Figure 2 - Distribution of exploits by year of publication.

Exploit Detection Failures

The NSS test results were mapped using the software program Maltego. The relationship mapping capabilities of this program allowed researchers to see that there were significantly varying exploit block performances by device and that there were a considerable number of exploits jointly missed by multiple devices. Figure 3 illustrates the results for 11 devices in the IPS 2012 test.

Figure 3 illustrates the significantly varying block performances of 11 of the tested devices. The devices under test are presented as green bubbles, with the size of each bubble indicating the number of exploits that went undetected by that device. The size of the exploit bubbles is in proportion to the number of devices that did not detect a given exploit. The orange bubbles represent exploits that went undetected by one security device, and the brown bubbles represent exploits that went undetected by multiple devices.

Exploits that are able to bypass several security devices in parallel are considered critical, since such exploits would allow an attacker to either bypass several layers of defense on a given target, or successfully attack a large number of targets, regardless of the types of security devices deployed. Figure 3 depicts a large number of such exploits.

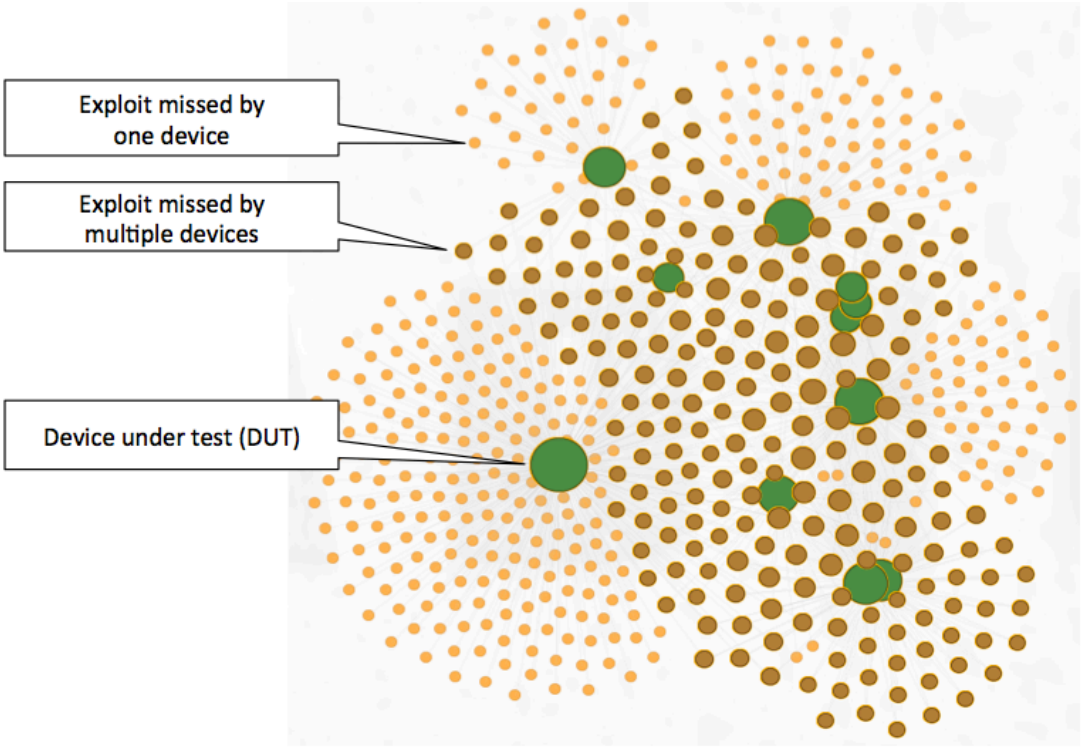

Figure 3 - Visualization of test result.

Green bubbles indicate devices tested with size proportional to number of missed exploits. Brown bubbles indicate jointly missed exploits, while orange bubbles indicate exploits missed by a single device.

Figure 4 provides a summary of the group tests results from EPP and in-line network security devices tested in 2012 and Q1 2013. During the IPS 2012 group test, it can be seen that 1,486 exploits were tested against 16 devices from 10 different security vendors. The average number of undetected exploits per device for the IPS 2012 group test is 82.8, which results in an average failure rate of 5.6% per device.

Group Test			Total Unique Exploits			Exploits by DUT		Failure
	Vendors	Devices	Tested	Missed	%	Missed	Std.Dev	Rate
2012 EPP	13	13	**43**	39	90.7%	**19.5**	5.5	**45.4%**
2012 IPS	10	16	**1,486**	716	48.2%	**82.8**	80.2	**5.6%**
2012 NGFW	8	8	**1,486**	752	50.6%	**132.4**	93.9	**8.9%**
2013 NGFW	8	9	**1,711**	392	22.9%	**67.7**	35.1	**4.0%**

Figure 4 - Summary of group test results.

Figure 5 depicts the number of exploits that were undetected by any security device. The table shows that the poorest performing device in the IPS 2012 group test missed 334 exploits (D1), while the best performing device missed 16 exploits (D16).

During the IPS 2012 group test, 51.8% of the exploits were detected by all devices, and 48.2% (712 exploits) managed to bypass at least one security device.

Group Test	Exploits missed by device under test (DUT)															
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
EPP 2012	28	28	25	23	22	19	19	18	18	18	16	11	9			
IPS 2012	334	166	135	132	111	78	74	74	60	51	25	17	17	17	17	16
NGFW 2012	307	235	168	131	90	81	31	16								
NGFW 2013	155	93	67	67	55	51	45	45	31							

Figure 5 - Number of exploits missed for each device in the four group tests.

These results highlight several key considerations:

- The measured dispersion of exploit block performance among the devices in a group test is significant. The number of exploits that were undetected by the best performing device versus the number of exploits that were undetected by the worst performing device differs by a factor of 3 to 20, between the EPP 2012 and IPS 2012 group tests, respectively.
- The high average failure rate of 45.4% in the EPP 2012 group test is significantly higher than other protection mechanisms.
- A considerable number of exploits (between 22.9% and 90.7%) managed to bypass at least one of the security devices tested.

The high failure rate of the EPP 2012 tests indicates that anti-virus/anti-malware products are inadequate protection against current exploits. The EPP 2012 group test used exploits that were current at the time of testing and that targeted mainstream products typically found on endpoint systems. These included popular web browsers, and plug-ins such as Adobe Flash and Java.

Given the average failure rate of 45.4%, strong correlation between the products tested is to be expected. Figure 5 clearly depicts multiple exploits that go undetected by several of the security products tested.

On the face of it, the average failure rates of between 4.0% and 8.9% of the network protection group tests (IPS and NGFW) appear considerably lower than the failure rate of the EPP group test. However, given the larger number of exploits used in the IPS and NGFW tests, these "low" failure rates still translate into a considerable average number of missed exploits, between 67 and 132 per security device.

On the other hand, a lower failure rate potentially makes it less likely to find correlation, that is, to identify individual exploits that bypass detection by multiple devices. However, in order to maximize protection, enterprises need to focus on the correlation of exploits bypassing multiple devices, and of the software targeted by these exploits.

Targeted Software Vendors

The exploits used in these group tests target 816 products from 208 software vendors. These products account for 20,230 (43%) of all vulnerabilities published by the NVD since 2002. The 33 NGFW and IPS security devices that were tested did not detect exploits targeting 639 products from 164 software vendors.

Figure 6 shows the top 20 targeted software vendors, ranked by the number of security devices that failed to detect exploits against the given vendor. Figure 6 reveals that 27 of the 33 devices under test failed to detect all exploits against software from Apple, resulting in 54 unique exploits bypassing detection. The least effective device missed 17 exploits (D14) and the most effective devices missed 0 exploits (D8, D13, D21, D23, D28, D29).

				NGFW 2013									NGFW 2012								IPS 2012																			
#	Vendor	Exploits missed	By # devices	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31	D32	D33
1	Microsoft	408	33	7	19	27	20	14	13	13	19	29	99	84	41	37	89	12	7	24	9	74	37	37	7	6	17	49	102	50	17	17	7	7	7	20
2	ca	39	31		1	1	1	1	1	1	1	7	12	5	5	2	9	1	1	6	2	3	5	1	1	1	1	4	11		6	6	1	1	1	1
3	Symantec	24	28	2	1	4	1	3	1	1	4	6	2	3	2	3	10			3		6	2	3	1			3	3	4	3	3	1	1	1	3
4	Apple	54	27	1	7	4	2	1	5	5		7	4	14	3		17	2	2	1	1	10	3		1		6	8	15	8			1	1	1	1
5	IBM	35	25	1		5	3	2			2	9	4	8	5	1	15	1		4	1	4	5	1		2	1	5	6	2	2	2				2
6	Adobe	25	24		7	5	1		3	3		1	5	7	3	1	1	4	1			1	1	1	1		2	2	10	6			1	1	1	
7	Oracle	55	24	5	2	3	2	6	3	3	2	13		19	5	1	20	1		3	1	17	3	1			3	1	12	16						2
8	Squid	11	24					1	1	1	2	3		4	1	6	2			1	1	1	1	6	1			4	6	1	2	2	1	1	1	2
9	Mozilla	36	23	1	3	3	2	2			3	6	8	8	4	2	2	2		1	1	3	4	2		1	2	2	23	1						
10	SUN	22	22	1	1	2	3	1	3	3				8	3	1	6	1		5		2	3	1				2	7	3	4	4				2
11	Cisco	7	22		1	1	1	2			1		1	1	3	2	3	1		1		3	3	2			1	2	3	3	1	1				2
12	Linux	13	22			5	1	3	1	1	8			8	8	8	4			2	1	2	8	8		1		8	8	6	3	3				2
13	OpenBSD	3	19		1		1	2			1	1		1	2	2	2	1		1		1	2	2			1	1	2	2						2
14	FreeBSD	4	19		1		1	2			1			1	2	2	2	1		1		1	2	2		1	1	1	2	2						2
15	Trendmicro	10	19				1	1	1	1		4		2	1		4		1	1		1	1		1			1	2	1			1	1	1	
16	Novell	33	18		1	3	1					6	2	7	1	1	11			3		2	1				5	2	7	3	1	1				
17	HP	31	18	1	1	1	6	1	3	3	1	7	1		2		16			2	1	3	1						5							1
18	Samba	11	17				2		1	1	1	2		1	3		7			2					1				1	1	2	2				1
19	Videolan	8	17	1	2	1	1	1	1	1	1	2	1	1			1	1				1					1		6	1					
20	mit	8	15	1		1	1				2	1	1	2	2					4			2					1	1		4	4			2

Figure 6 - Top 15 most targeted software vendors by number of undetected exploits.

The vendors listed in Figure 6 are almost exclusively mainstream software vendors, leading to the conclusion that these exploits attack software that is regularly used in the enterprise and in private environments. It is worth noting that of the 33 security devices tested, none were successful in detecting all exploits against Microsoft products. The need to adequately protect widely-used software from such attacks is a serious challenge that faces the security industry.

The failures to detect attacks on widely used software products are particularly relevant from the perspective of protection and risk management.

Correlation Of Detection Failures

The information presented thus far indicates substantial correlation between detection failures. To test this hypothesis, NSS examined the number of exploits that were able to bypass detection by several security devices simultaneously. The *Measured Data* table shown in Figure 7 documents the number of unique exploits found to bypass X or more security devices for each NSS group test.

For example, during the NGFW 2013 group test, NSS testing identified and confirmed 18 exploits that bypassed five or more security devices, and eight exploits that bypassed all nine of the security devices tested. These exploits successfully attacked widely used software such as a web browser, a media plug-in, or a popular FTP server.

Measurement Versus Risk Model

To assess the effect of correlation (the extent to which exploits bypass multiple devices), it is necessary to compare the measured data to a model that assumes no correlation between detection failures. Ignoring correlation, it is possible to calculate the probability of finding exploits that bypass X or more security devices, given the average failure rate of the security devices in a given group test.

In terms of probability theory and statistics, the binomial distribution models the number of x successes in a sequence of N independent yes/no experiments, each of which yields success with probability p. Success in this context means an exploit was not detected.

The binomial model $B(x, N, p)$ returns the probability of finding exploits that are simultaneously not detected by x of the N security devices of the group test, given the average failure probability p of the group test.

The *Model (Ignoring Correlation)* table that is shown in Figure 7 lists the number of exploits that bypass x or more security devices for each group test, as predicted by the binomial model.

Measured Data

	Exploits	Devices		Number of exploits bypassing X or more devices (measured)																
Group Test	total	total	max	x=0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EPP 2012	43	13	12	43	39	37	34	33	31	24	16	15	10	9	4	2	0			
IPS 2012	1,486	16	11	1,486	716	289	127	76	47	32	21	9	4	2	1	0	0	0	0	0
NGFW 2012	1,486	8	6	1,486	752	216	58	23	8	2	0	0								
NGFW 2013	1,711	9	9	1,711	392	102	41	23	18	9	8	8	8							

Model (Ignoring Correlation)

	Exploits	Devices		Number of exploits bypassing X or more devices (model)																
Group Test	total	total	max	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EPP 2012	43	13	10	43	43	43	42	39	34	25	16	8	3	1	0	0	0			
IPS 2012	1,486	16	5	1,486	892	331	83	15	2	0	0	0	0	0	0	0	0	0	0	0
NGFW 2012	1,486	8	4	1,486	782	230	42	5	0	0	0	0								
NGFW 2013	1,711	9	7	1,711	521	80	7	0	0	0	0	0	0							

Figure 7 - Number of exploits bypassing X or more security devices, Measured Data and Model (Ignoring Correlation).

A comparison of the measurement data with the prediction ignoring correlation reveals a pattern where the ignoring correlation model consistently underestimates the number of exploits able to bypass a high number of security devices. For example, the model predicted zero exploits bypassing X or more security devices, while NSS tests positively identified multiple exploits that bypass X or more security devices.

- EPP 2012: No exploits for more than 10 devices predicted – found 2 exploits bypassing 12 devices.
- IPS 2012: No exploits for more than 5 devices predicted – found 1 exploit bypassing 11 devices.
- NGFW 2012: No exploits for more than 4 devices predicted – found 2 exploits bypassing 6 devices.
- NGFW 2013: No exploits for more than 3 devices predicted – found 8 exploits bypassing 9 devices.

The difference between the model and the measurement is smallest for the EPP 2012 group test. This is to be expected, given the security devices' average failure rate of 45.4%. When 13 devices are tested and each has an average failure rate of 45.4% (19.5 of 43 exploits missed on average), it is expected that multiple exploits will be identified that have simultaneously gone undetected by multiple devices.

For the other group tests, the outcome is distinctly different. For example, during the NGFW 2012 group test, the 8 devices have an average failure rate of 8.9% (132.5 of 1,486 exploits missed on average). Since 8 x 132.5 = 1,060 is smaller than the total number of 1,486 exploits tested, it is expected that there will be less exploits that have simultaneously gone undetected by multiple devices.

The effect of the correlation of detection failures is twofold:

(A) The binomial model predicts higher probabilities than measured in the test for very few devices (two devices or less for IPS and NGFW). As a result of the correlation found in the test, a smaller set of unique exploits is required to result in the same average device failure rate as the model ignoring correlation. This effect decreases with the increasing failure rate of the devices.

(B) The model underestimates the probability of exploits bypassing an increasing number of devices. The model prediction quickly approaches close to zero probability for increasing device numbers, while measurements confirm multiple exploits that bypass all or almost all devices.

The observed and increasing difference between prediction (zero exploits) and measurements (positive number of exploits) for the increasing numbers of security devices being bypassed is confirmation of the correlation of detection failures.

The identification and analysis of exploits that escape detection by the majority of the security devices/products in a group test is globally relevant, as these exploits present a significant challenge to the security industry. Multiple security vendors continue to jointly miss numerous exploits, in spite of considerable investment in security products.

The test results show that, regardless of the security products deployed, it remains highly probable that a cybercriminal will be able to successfully penetrate several layers of security of a targeted organization, or successfully attack a large number of different organizations. It should be noted that NSS did not use 0-day exploits in these group tests.

Concerns over complexity, manageability, and cost make it a challenge for enterprises to deploy multiple security devices to create a layered defense. This analysis will focus on the cross-correlation of detection failures for any combination of *two* security products within and across group tests.

Risk Management And Cross-Correlation

The test results demonstrate that failures to detect exploits are strongly correlated, particularly in those group tests with a comparatively low average failure rate. A risk model that ignores correlation is in danger of underestimating the probability of exploits able to bypass multiple security devices. The typical approach to calculating the combined failure rate of security devices A and B put in series is to multiply the devices individual failure rates: $P(A \text{ and } B) = P_A \times P_B$. However, this multiplication of probabilities is only valid if P_A and P_B are not correlated.

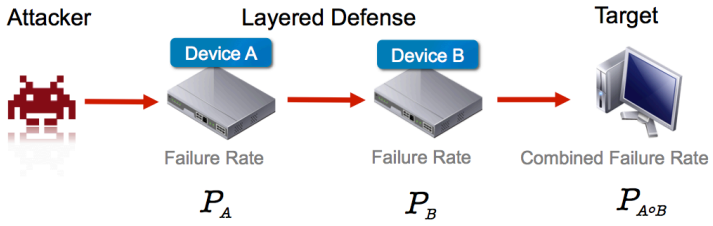

Figure 8 - Combined failure rate of two security devices.

Conversely, detection failures *are* correlated and cannot be considered independent events. For example, during the NGFW 2012 group test, NSS tested 8 different security devices; hence there are a total of 8 x 8 = 64 combinations of two security devices, of which 28 are unique (excluding duplicates and combinations of a device with itself).

Figure 9 illustrates the cross-correlation matrix with the number of jointly missed exploits for any combination of security devices of the NGFW 2012 group test. Diagonal cells indicate the number of exploits missed by a single security device. All other cells indicate the number of exploits jointly missed by the combination of the two security devices given in the column/row header.

The cell that is shaded green identifies the most secure combination and the cell that is shaded red identifies the least secure combination. Figure 9 shows that no combination of two security devices in the NGFW 2012 group test would detect all exploits.

The measured failure rate for each combination of two security devices is calculated by dividing the number of exploits given in each cell by the total number of exploits tested.

#	Security Device	Barracuda F900	Check Point 12600	SuperMassive e10800	Fortinet 3140B	Juniper SRX3600	Palo Alto PA5020	Sourcefire 3D8250	Stonesoft 1301
1	Barracuda F900	307	28	19	8	61	34	3	16
2	Check Point 12600		168	11	5	22	19	4	5
3	SuperMassive e10800			81	6	20	27	6	14
4	Fortinet 3140B				31	17	12	4	6
5	Juniper SRX3600					235	33	11	22
6	Palo Alto PA5020						131	10	18
7	Sourcefire 3D8250							16	2
8	Stonesoft 1301								90

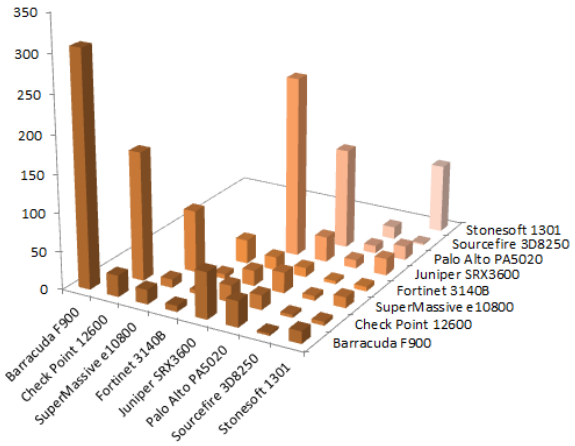

Figure 9. - Joint detection failures for all combinations of two devices of the NGFW test.

Diagonal cells indicate exploits missed by a single device; all other cells indicate the number of exploits jointly missed by both devices.

To assess the benefit of layered security and to quantify the error of ignoring correlation using a naïve risk model, NSS compared the prediction $P(A \text{ and } B) = P_A \times P_B$ of two devices' joint failure rates with their measured joint failure rate for:

- All unique combinations of two devices in a group test (for example, all IPS devices).
- All unique combinations of two security devices between two group tests (for example, all IPS vs. all NGFW devices).

Figure 10 summarizes the results of this analysis for all combinations of unique pairs of security products of the group tests indicated in the first two columns. The predicted and measured failure rates are listed in the final three columns. The "low prediction" column shows the percentage of all combinations of device pairs for which the model underestimates the joint failure rate. The "joint exploits" column shows the number of unique exploits used in the combined group test scenario, "test1/test2".

The first four rows report on combinations of security products within a single group test and the following three rows compare all pairs of security products between the IPS and the two NGFW group tests. Combinations of the EPP versus IPS/NGFW group tests are not listed, since the joint number of exploits tested for these combinations is too low for robust results.

The combined failure rate of layered security is typically found to be considerably higher than the product of the failure rates of the individual layers. It is apparent from this that detection failures are correlated definitively, and they should not be considered independent events.

Group Test		Combinations			Joint	Group Failure Rate		Joint Failure Rate		
Test 1	Test 2	Total	Unique	Low Prediction	Exploits	Test 1	Test 2	Model	Measured	Off By
2012 EPP	2012 EPP	13x13	78	91%	43	45.4%	45.4%	20.51%	26.00%	1.3 x
2012 IPS	2012 IPS	16x16	120	96%	1,486	5.6%	5.6%	0.29%	0.77%	2.6 x
2012 NGFW	2012 NGFW	8x8	28	82%	1,486	8.9%	8.9%	0.74%	1.06%	1.4 x
2013 NGFW	2013 NGFW	9x9	36	100%	1,711	4.0%	4.0%	0.15%	0.87%	5.8 x
2012 IPS	2012 NGFW	16x8	128	92%	1,486	5.6%	8.9%	0.50%	1.07%	2.2 x
2012 IPS	2013 NGFW	16x9	144	91%	1,482	5.5%	3.0%	0.16%	0.46%	2.8 x
2012 NGFW	2013 NGFW	8x9	72	83%	1,482	8.8%	3.0%	0.26%	0.64%	2.4 x

Figure 10 - Prediction error of simple risk model vs. measured joint failure rate for any combination of two security products.

As an example, the results of the last row "NGFW 2012 vs. NGFW 2013" in Figure 10 indicate that:

- The NGFW 2012 and NGFW 2013 group tests share 1,482 exploits.
- For 83% of the 72 unique pairs of security products, the model prediction underestimated the measured joint failure rate.
- Averaged over these 72 combinations, the predicted joint failure rate is 0.26%, while measurements found 0.64%, or 2.4 times the predicted value.

Figure 10 summarizes data of the joint failure rate of 606 unique pairs of security products. Only 19 (3%) of these 606 pairs of security products succeed in blocking all exploits. In 554 (91%) of these pairs, the naïve risk model underestimates the joint failure rate, on average, by a factor of 6.1.

Figure 11 reveals that the distribution of the prediction error is skewed with 58 (9.5%) of the pairs exceeding a prediction error by a factor of 10. The maximum error found is a factor of 87 off the prediction.

Such outliers with a high prediction error towards insecurity are especially critical, since risk models based on this prediction severely underestimate the effective failure rate.

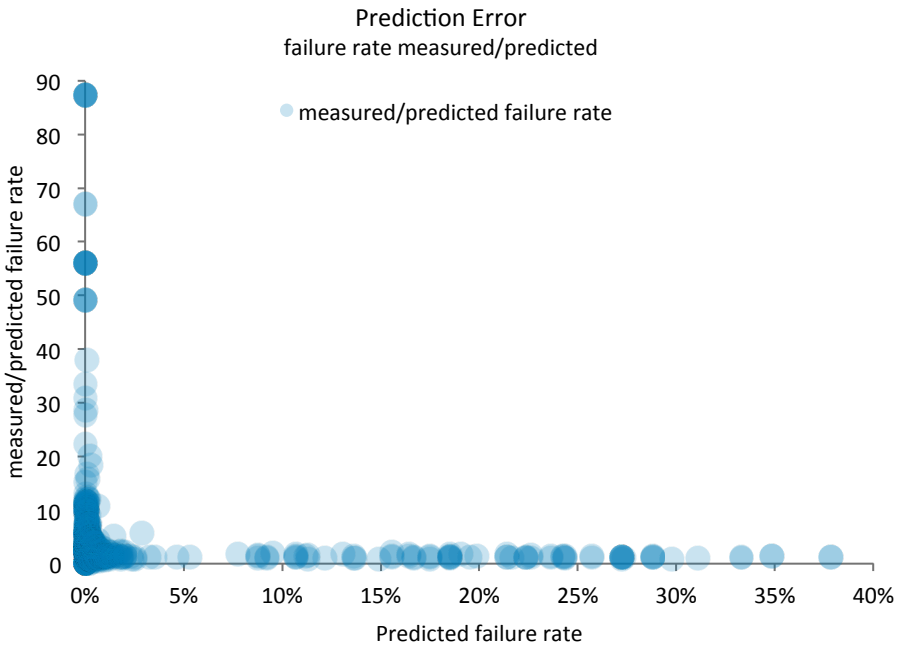

Figure 11 - Distribution of predication error as measured/predicted failure rate.

The Effectiveness Of Layered Security

This analysis examines the effectiveness and the limitations of layered security. The average failure rate of security device pairs is lower than the average failure rate of individual products:

- The average joint failure rate for IPS and NGFW is 0.8%, down from an average single-device failure rate of 5.8%.
- The average joint failure rate for multiple EPP products is 26.0%, down from an average single-product failure rate of 45.4%.

However, only 3% of the 606 pairs of security devices analyzed were able to detect all exploits. Once again, the results show differing reductions in failure rate across all pairs.

Figure 12 compares the single-device failure rate to the joint failure rate, for all 606 combinations of two security devices.

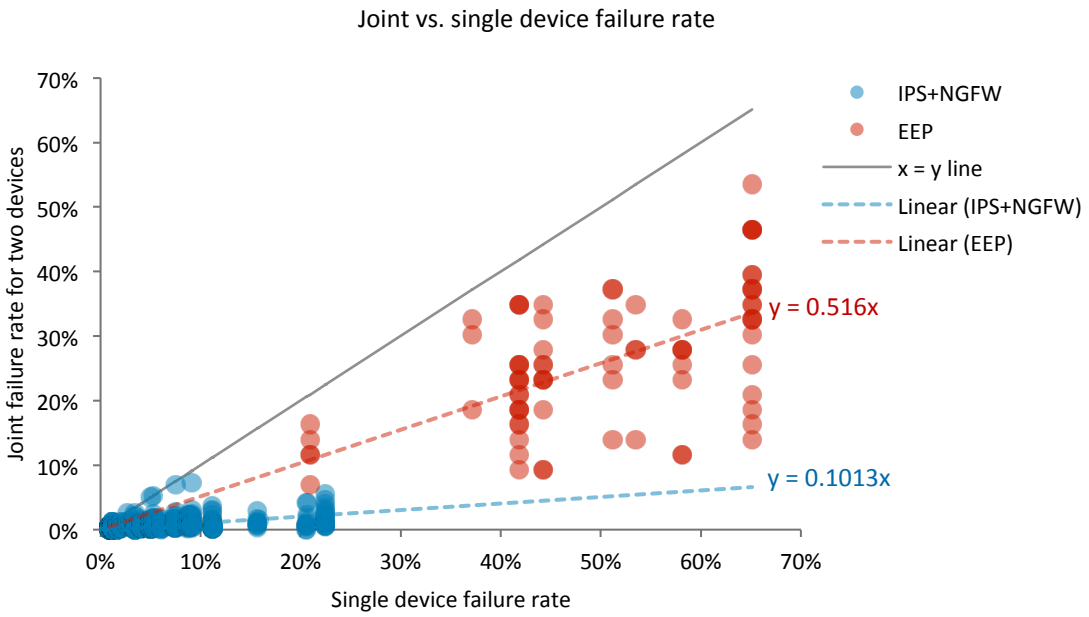

Figure 12 - Reduced failure rate of security device pairs compared to single device failure rate.

The results for IPS+NGFW device pairs are marked in blue and the results for EPP device pairs are marked in red. The diagonal line depicts the unity of x and y. A combination of security products cannot lead to an increase in the joint failure rate, therefore the unity line is an upper limit.

Horizontally, the points are clustered at values representing the failure rate of a single security device, as seen in Figure 5. The vertical spread of the points illustrates the range of the joint failure rate, determined by the choice of security devices being combined.

For example, for a given device (device 1) with a failure rate depicted on the x-axis, the vertical spread indicates the resulting joint failure rate of the pairs (device 1, device 2), depending on the selection of the second security device. This spread shows that the choice of devices to be combined is crucial to realize effective layered security within an organization.

The dashed linear trend lines in Figure 12 demonstrate the impact of layered security on endpoint protection products and on network protection products. The slope for the endpoint protection products is approximately five times larger than the slope for the network protection devices. This is attributed to the combination of the EPP group test's failure rate of 45.4% with the strong correlation of detection failures between these security products.

While the layering of security devices is advantageous, the real key to effective security lies in an organization's choice of protection technologies to be combined. The device combinations close to the unity line in Figure 12 indicate poor combinations of security devices, resulting in little-to-no increase in security.

Reading List

"NGFW Comparative Analysis 2012" NSS Labs, October 17, 2012:
https://www.nsslabs.com/reports/ngfw-comparative-analysis-2012

"2013 Next Generation Firewall Comparative Analysis" NSS Labs, February 26, 2013:
https://www.nsslabs.com/reports/next-generation-firewall-comparative-analysis-2013

IPS Comparative Analysis 2012" NSS Labs, October 3, 2012:
https://www.nsslabs.com/reports/ips-comparative-analysis-2012

"2013 Consumer AV/EPP Comparative Analysis – Phishing Protection" NSS Labs, January 30, 2013:
https://www.nsslabs.com/reports/consumer-avepp-comparative-analysis-phishing-protection

"Consumer AV/EPP Comparative Analysis – Exploit Evasion Defenses" NSS Labs, October 30 2012:
https://www.nsslabs.com/reports/consumer-avepp-comparative-analysis-exploit-evasion-defenses

"Consumer AV/EPP Comparative Analysis – Exploit Protection" NSS Labs, October 24, 2012:
https://www.nsslabs.com/reports/consumer-avepp-comparative-analysis-exploit-protection

"Modeling Evasions In Layered Security" NSS Labs, December 5, 2012:
https://www.nsslabs.com/reports/modeling-evasions-layered-security

Appendix

Calculation Of The Combined Failure Rate

A risk model that ignores correlation is in danger of underestimating the probability of exploits that bypass multiple security devices. In a naïve risk model, one would calculate the combined failure rate *P(A and B)* of two security devices, A and B, put in series as the product of the devices' individual failure rate: $P(A \text{ and } B) = P_A \times P_B$

However, this multiplication of probabilities is only valid if P_A and P_B are statistically independent (for example, if they are two independent events, such as throwing two dice - the probability to get "66" is 1/6 x 1/6 = 1/36).

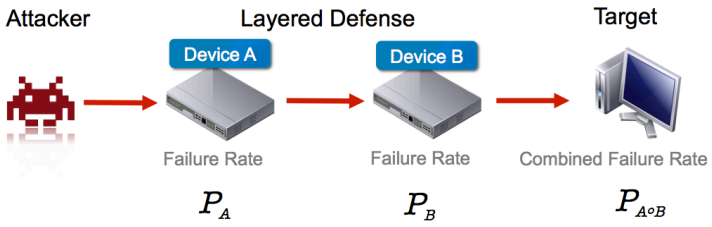

Figure 13 - Combined failure rate of two security devices.

Conversely, detection failures are correlated and they cannot be considered independent events. Therefore the result of the simple formula $P_A \times P_B$ is misleading. Typically, the combined failure rate is considerably higher than the product of the failure rates of the individual layers.

Contact Information

NSS Labs, Inc.
206 Wild Basin Rd
Building A, Suite 200
Austin, TX 78746 USA
(512) 961-5300
info@nsslabs.com
www.nsslabs.com

V 130514a

This analyst brief was produced as part of NSS Labs' independent testing information services. Leading products were tested at no cost to the vendor, and NSS Labs received no vendor funding to produce this analyst brief.