SE Labs

INTELLIGENCE-LED TESTING

NETWORK SECURITY APPLIANCE TEST

JANUARY 2018

www.SELabs.uk info@SELabs.uk @SELabsUK www.facebook.com/selabsuk blog.selabs.uk

SE Labs tested a variety of network security appliances from a range of well-known vendors in an effort to judge which were the most effective.

Each product was exposed to the same threats, which were a mixture of targeted attacks using well-established techniques and public email and web-based threats that were found to be live on the internet at the time of the test.

The results indicate how effectively the products were at detecting and/or protecting against those threats in real time.

MANAGEMENT
Director Simon Edwards
Operations Director Marc Briggs
Office Manager Magdalena Jurenko
Technical Lead Stefan Dumitrascu

TESTING TEAM
Thomas Bean
Dimitar Dobrev
Liam Fisher
Gia Gorbold
Pooja Jain
Ivan Merazchiev
Jon Thompson
Jake Warren
Stephen Withey

IT SUPPORT
Danny King-Smith
Chris Short

PUBLICATION
Steve Haines
Colin Mackleworth

Website www.SELabs.uk
Twitter @SELabsUK
Email info@SELabs.uk
Facebook www.facebook.com/selabsuk
Blog blog.selabs.uk
Phone 0203 875 5000
Post ONE Croydon, London, CR0 0XT

SE Labs is BS EN ISO 9001 : 2015 certified for The Provision of IT Security Product Testing.

SE Labs Ltd is a member of the Anti-Malware Testing Standards Organization (AMTSO)

AMTSO Standard public pilot reference:
https://www.amtso.org/se-labs-test-reviews-public-pilot/

CONTENTS

Document version 1.0 Written 12th June 2018

INTRODUCTION

What's the difference between SE Labs and a cyber-criminal?

As we prepared this report for publication we were also getting ready to present at BT's internal security conference Snoopcon. We had been asked to talk about security products and how they might not do what you assume they will.

Reports like this provide an interesting insight into how security products actually work. Marketing messages will inevitably claim world-beating levels of effectiveness, while basic tests might well support these selling points. But when you actually hack target systems through security appliances you sometimes get a very different picture.

Some vendors will support the view that testing using a full attack chain (from a malicious URL pushing an exploit, which in turn delivers a payload that finally provides us with remote access to the system) is the right way to test. Others may point out that the threats we are using don't exactly exist in the real world of criminality because we created them in the lab and are not using them to break into systems worldwide.

We think that is a weak argument. If we can obtain access to certain popular, inexpensive tools online and create threats then these (or variants extremely close to them) are just as likely to exist in the 'real world' of the bad guys as in a legitimate, independent test lab. Not only that, but we don't keep creating new threats until we break in, which is what the criminals (and penetration testers) do. We create a set and, without bias, expose all of the tested products to these threats.

But in some ways we have evolved from being anti-malware testers to being penetration testers, because we don't just scan malware, execute scripts or visit URLs. Once we gain access to a target we perform the same tasks as a criminal would do: escalating privileges, stealing password hashes and installing keyloggers. The only difference between us and the bad guys is that we're hacking our own systems and helping the security vendors plug the gaps.

Executive Summary

Product names
It is good practice to stay up to date with the latest versions of your chosen network security appliance.

This means updating its range of available updates and updating its operating system firmware. We made best efforts to ensure that each appliance tested was running the very latest operating system and updates available to demonstrate the best possible outcome.

For specific operating system and updates details, see **Appendix C: Product versions** on page 15.

EXECUTIVE SUMMARY

Products tested	Protection Accuracy Rating (%)	Legitimate Accuracy Rating (%)	Total Accuracy Rating (%)
Symantec Advanced Threat Protection	89%	100%	96%
Fortinet FortiGate	92%	97%	95%
Palo Alto Networks PA200	38%	96%	77%
Cisco Snort	-13%	99%	63%

■ **Products highlighted in green were the most accurate, scoring 85 per cent or more for Total Accuracy. Those in yellow scored less than 85 but 75 or more. Products shown in red scored less than 75 per cent.**

For exact percentages, see **1. Total Accuracy Ratings** on page 6.

■ **The appliances were mainly effective at handling prevalent web threats aimed at the general public…**
All products were capable of blocking attacks such as those used by cyber criminals to attack Windows PCs and install ransomware and other threats.

■ **… and targeted attacks were also detected and blocked well**
Most of the products were very competent at blocking more targeted, exploit-based attacks. These types of attacks are challenging for endpoint security solutions so having them caught on the network has great value. **Cisco Snort** was notably weaker in this part of the test.

■ **But email attacks were successful against two of the three products tested.**
While appliances from **Symantec** and **Fortinet** defended against the majority of email threats, those from **Palo Alto Networks** and **Cisco's** open source offering were much less successful.

■ **Which products were the most effective?**
Symantec's and **Fortinet's** appliances stopped the most threats and, because they only blocked a small amount of legitimate traffic, they win AAA awards. **Palo Alto** achieved a C grade and **Cisco Snort** failed to score well enough for an award.

Simon Edwards, SE Labs, 12th June 2018

1. Total Accuracy Ratings

Judging the effectiveness of a security product is a subtle art, and many factors are at play when assessing how well it performs. To make things easier, we've combined all the different results from this report into one easy-to-understand graph.

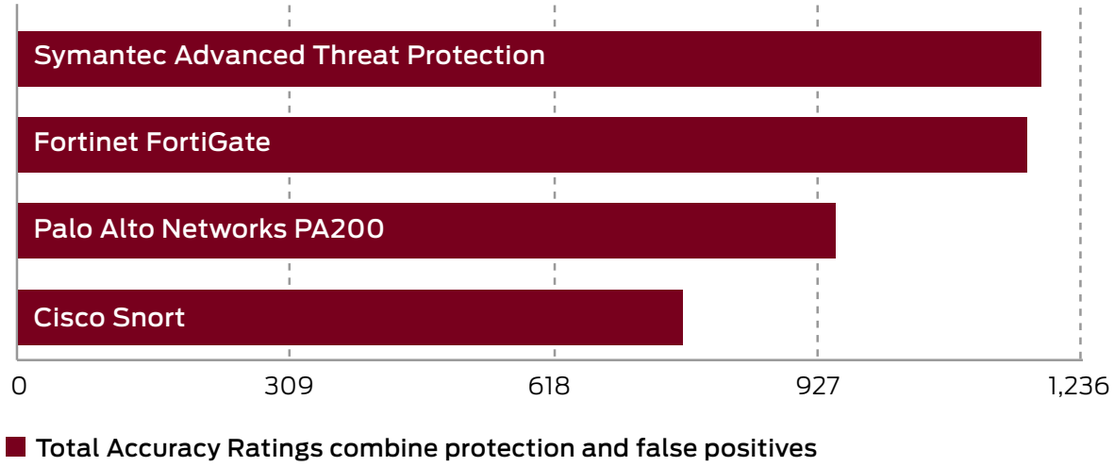
The graph below takes into account not only each product's ability to detect and protect against threats, but also its handling of non-malicious objects such as web addresses (URLs) and applications.

Not all protections, or detections for that matter, are equal. A product might completely block a URL, which prevents the threat completely before it can even start its intended series of malicious events. Alternatively, the product might allow a web-based exploit through one time but block subsequent similar threats. It might also allow the malware to download onto the target but block further threats the malware attempts to download. We take these outcomes into account when attributing points that form final ratings.

For example, a product that completely blocks a threat is rated more highly than one which allows a threat to run for a while before eventually evicting it. Products that allow all malware infections, or that block popular legitimate applications, are penalised heavily.

Categorising how a product handles legitimate objects is complex, and you can find out how we do it in **5. Legitimate Software Ratings** on page 10.

TOTAL ACCURACY RATINGS			
Product	Total Accuracy Rating	Total Accuracy (%)	Award
Symantec Advanced Threat Protection	1,192	96%	AAA
Fortinet FortiGate	1,177	95%	AAA
Palo Alto Networks PA200	953	77%	C
Cisco Snort	776	63%	



SE Labs

Network Security Appliance Test Awards

The following products win SE Labs awards:

- Symantec Advanced Threat Protection
- Fortinet FortiGate

- Palo Alto Networks PA200

2. Protection Ratings

The results below indicate how effectively the products dealt with threats. Points are earned for detecting the threat and for either blocking or neutralising it.

Detected (+1)
If the product detects the threat with any degree of useful information, we award it one point.

Blocked (+2)
Threats that are disallowed from even starting their malicious activities are blocked. Blocking products score two points.

Neutralised (+1)
Products that kill all running malicious processes 'neutralise' the threat and win one point.

Complete remediation (+1)
If, in addition to neutralising a threat, the product removes all significant traces of the attack, it gains an additional one point.

Compromised (-5)
If the threat compromises the system, the product loses five points. This loss may be reduced to four points if it manages to detect the threat (see Detected, above), as this at least alerts the user, who may now take steps to secure the system.

Rating calculations
We calculate the protection ratings using the following formula:

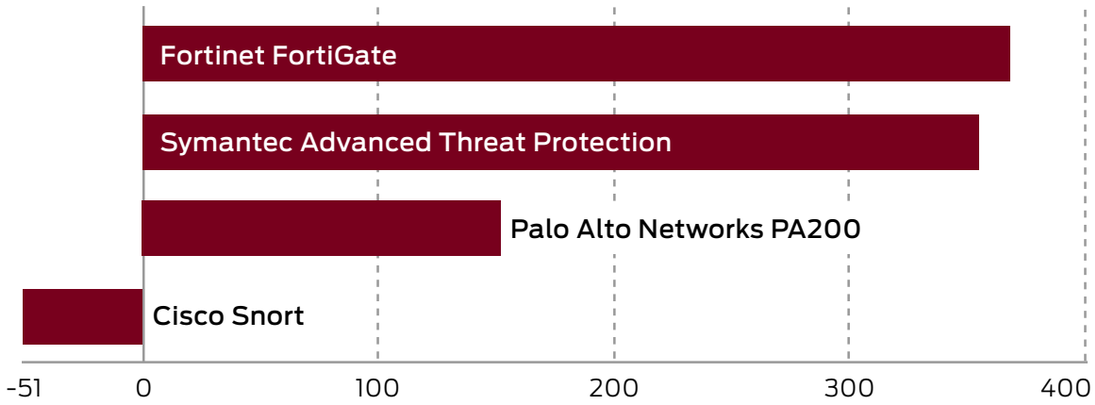
Protection rating =
(1x number of Detected) +
(2x number of Blocked) +
(1x number of Neutralised) +
(1x number of Complete remediation) +
(-5x number of Compromised)

The 'Complete remediation' number relates to cases of neutralisation in which all significant traces of the attack were removed from the target. Such traces should not exist if the threat was 'Blocked' and so Blocked results imply Complete remediation.

These ratings are based on our opinion of how important these different outcomes are. You may have a different view on how seriously you treat a 'Compromise' or 'Neutralisation without complete remediation'. If you want to create your own rating system, you can use the raw data from **4. Protection Details** on page 9 to roll your own set of personalised ratings.

PROTECTION RATINGS		
Product	Protection Rating	Protection Accuracy (%)
Fortinet FortiGate	369	92%
Symantec Advanced Threat Protection	356	89%
Palo Alto Networks PA200	153	38%
Cisco Snort	-51	-13%

Average: 51.5%



Protection Ratings are weighted to show that how products handle threats can be subtler than just 'win' or 'lose'.

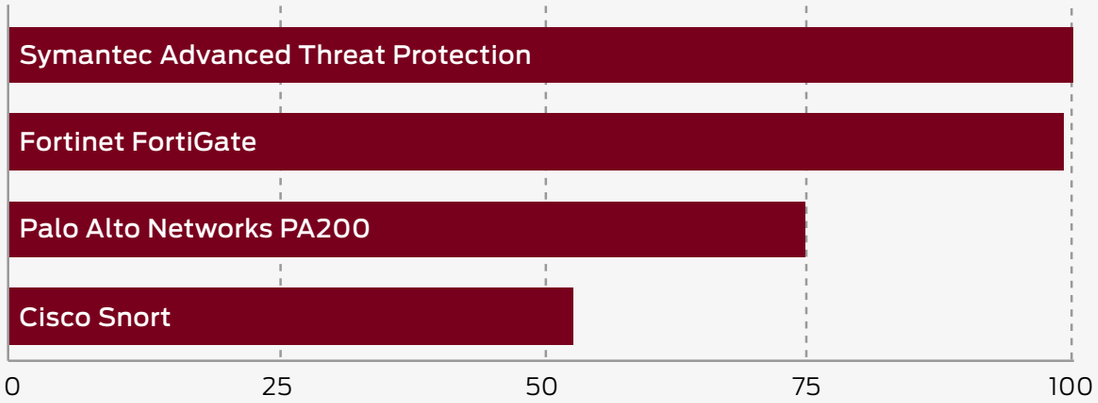
3. Protection Scores

This graph shows the overall level of protection, making no distinction between neutralised and blocked incidents.

For each product we add Blocked and Neutralised cases together to make one simple tally.

PROTECTION SCORES	
Product	Protection Score
Symantec Advanced Threat Protection	100
Fortinet FortiGate	99
Palo Alto Networks PA200	75
Cisco Snort	53

■ Protection Scores are a simple count of how many times a product protected the system.

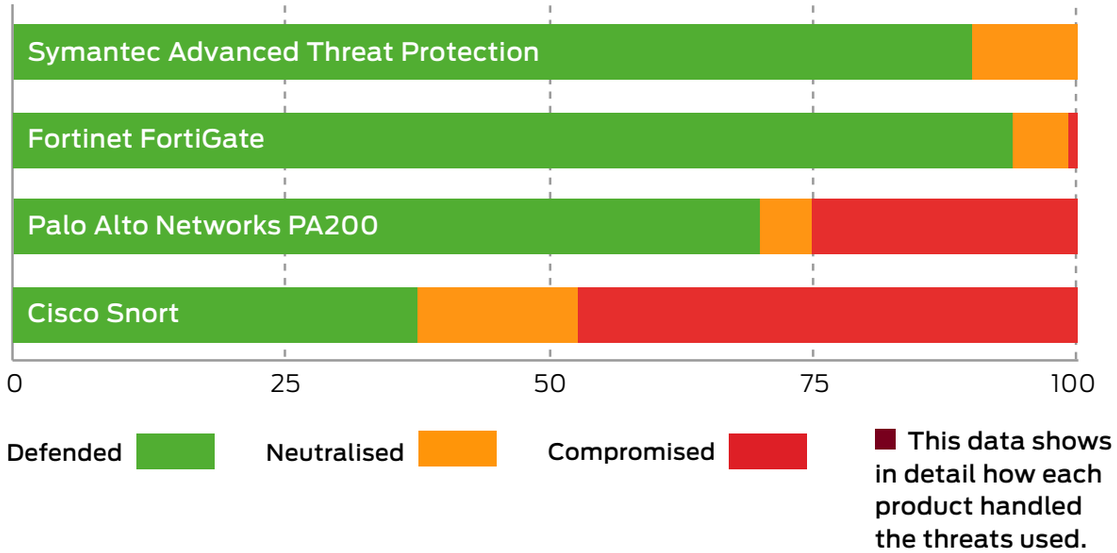


4. Protection Details

These results break down how each product handled threats into some detail. You can see how many detected a threat and the levels of protection provided.

Products sometimes detect more threats than they protect against. This can happen when they recognise an element of the threat but are not equipped to stop it. Products can also provide protection even if they don't detect certain threats. Some threats abort on detecting specific protection software.

PROTECTION DETAILS					
Product	Detected	Blocked	Neutralised	Compromised	Protected
Symantec Advanced Threat Protection	89	90	10	0	100
Fortinet FortiGate	99	94	5	1	99
Palo Alto Networks PA200	79	70	5	25	75
Cisco Snort	53	38	15	47	53



■ This data shows in detail how each product handled the threats used.

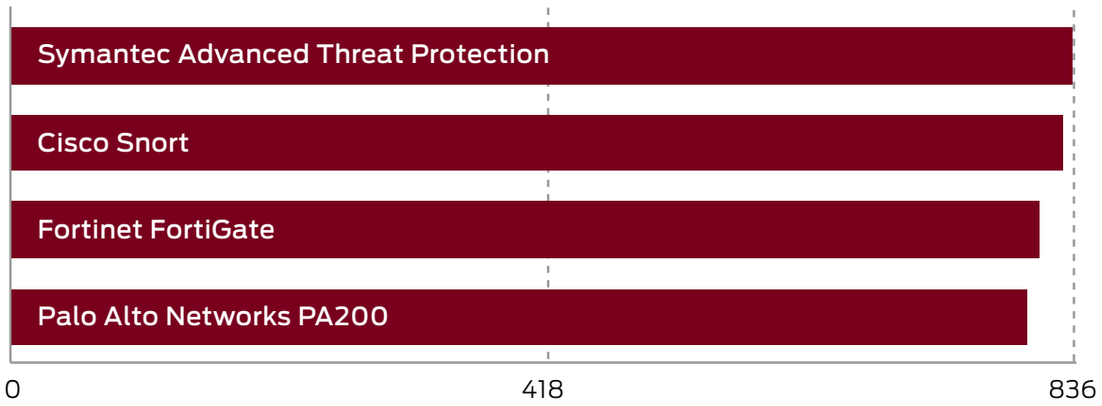
5. Legitimate Software Ratings

These ratings indicate how accurately the products classify legitimate applications and URLs, while also taking into account the interactions that each product has with the user. Ideally a product will either not classify a legitimate object or will classify it as safe. In neither case should it bother the user.

We also take into account the prevalence (popularity) of the applications and websites used in this part of the test, applying stricter penalties for when products misclassify very popular software and sites.

To understand how we calculate these ratings, see **5.3 Accuracy Ratings** on page 12.

LEGITIMATE SOFTWARE RATINGS		
Product	**Legitimate accuracy rating**	**Legitimate accuracy (%)**
Symantec Advanced Threat Protection	836	100%
Cisco Snort	827	99%
Fortinet FortiGate	808	97%
Palo Alto Networks PA200	800	96%



Legitimate Software Ratings can indicate how well a vendor has tuned its detection engine.

5.1 Interaction Ratings

It's crucial that anti-malware endpoint products not only stop – or at least detect – threats, but that they allow legitimate applications to install and run without misclassifying them as malware. Such an error is known as a 'false positive' (FP).

In reality, genuine FPs are quite rare in testing. In our experience it is unusual for a legitimate application to be classified as 'malware'. More often it will be classified as 'unknown', 'suspicious' or 'unwanted' (or terms that mean much the same thing).

We use a subtle system of rating an endpoint's approach to legitimate objects, which takes into account how it classifies the application and how it presents that information to the user. Sometimes the endpoint software will pass the buck and demand that the user decide if the application is safe or not. In such cases the product may make a recommendation to allow or block. In other cases, the product will make no recommendation, which is possibly even less helpful.

If a product allows an application to install and run with no user interaction, or with simply a brief notification that the application is likely to be safe, it has achieved an optimum result. Anything else is a Non-Optimal Classification/Action (NOCA). We think that measuring NOCAs is more useful than counting the rarer FPs.

	None (allowed)	Click to allow (default allow)	Click to allow/block (no recommendation)	Click to block (default block)	None (blocked)	
Object is safe	2	1.5	1			A
Object is unknown	2	1	0.5	0	-0.5	B
Object is not classified	2	0.5	0	-0.5	-1	C
Object is suspicious	0.5	0	-0.5	-1	-1.5	D
Object is unwanted	0	-0.5	-1	-1.5	-2	E
Object is malicious				-2	-2	F
	1	2	3	4	5	

COUNT OF INTERACTIONS		
Product	None (Allowed)	None (blocked)
Symantec Advanced Threat Protection	100	0
Cisco Snort	99	1
Fortinet FortiGate	98	2
Palo Alto Networks PA200	98	2

■ Products that do not bother users and classify most applications correctly earn more points than those that ask questions and condemn legitimate applications.

5.2 Prevalence Ratings

There is a significant difference between an endpoint product blocking a popular application such as the latest version of Microsoft Word and condemning a rare Iranian dating toolbar for Internet Explorer 6. One is very popular all over the world and its detection as malware (or something less serious but still suspicious) is a big deal. Conversely, the outdated toolbar won't have had a comparably large user base even when it was new. Detecting this application as malware may be wrong, but it is less impactful in the overall scheme of things.

With this in mind, we collected applications of varying popularity and sorted them into five separate categories, as follows:

1. Very high impact
2. High impact
3. Medium impact
4. Low impact
5. Very low impact

Incorrectly handling any legitimate application will invoke penalties, but classifying Microsoft Word as malware and blocking it without any way for the user to override this will bring far greater penalties than doing the same for an ancient niche toolbar. In order to calculate these relative penalties, we assigned each impact category with a rating modifier, as shown in the table above.

LEGITIMATE SOFTWARE CATEGORY FREQUENCY	
Impact Category	Rating Modifier
Very high impact	5
High impact	4
Medium impact	3
Low impact	2
Very low impact	1

Applications were downloaded and installed during the test, but third-party download sites were avoided and original developers' URLs were used where possible. Download sites will sometimes bundle additional components into applications' install files, which may correctly cause anti-malware products to flag adware. We remove adware from the test set because it is often unclear how desirable this type of code is.

The prevalence for each application and URL is estimated using metrics such as third-party download sites and the data from Alexa.com's global traffic ranking system.

5.3 Accuracy Ratings

We calculate legitimate software accuracy ratings by multiplying together the interaction and prevalence ratings for each download and installation:

Accuracy rating = Interaction rating x Prevalence rating

If a product allowed one legitimate, Medium impact application to install with zero interaction with the user, then its Accuracy rating would be calculated like this:

Accuracy rating = 2 x 3 = 6

This same calculation is made for each legitimate application/site in the test and the results are summed and used to populate the graph and table shown under **5. Legitimate Software Ratings** on page 10.

5.4 Distribution of Impact Categories

Endpoint products that were most accurate in handling legitimate objects achieved the highest ratings. If all objects were of the highest prevalence, the maximum possible rating would be 1,000 (100 incidents x (2 interaction rating x 5 prevalence rating)).

In this test there was a range of applications with different levels of prevalence. The table below shows the frequency:

LEGITIMATE SOFTWARE CATEGORY FREQUENCY	
Prevalence Rating	**Frequency**
Very high impact	55
High impact	5
Medium impact	24
Low impact	10
Very low impact	6
GRAND TOTAL	**100**

6. Conclusion

Attacks in this test included infected websites available to the general public that often tried to trick users into installing the malware.

URLs were introduced to the targets directly and, in relevant cases, via email. Infected emails were also included. We also launched targeted attacks in the form of exploit-based attempts to gain remote control of the target systems.

Crucially we attempt to run a full chain of attack, performing malicious actions on systems to which we manage to obtain remote access. This gives products an opportunity to detect important characteristics of an attack that would be missing if we simply obtained remote access but did nothing else.

Symantec Advanced Threat Protection protected against all of the public attacks and all of the malicious emails. It also blocked all of the targeted attacks and allowed all of the legitimate applications and URLs. Because it neutralised some threats its total accuracy rating is 96 per cent.

Fortinet FortiGate protected against all of the public email threats and malware downloads from the web and managed to handle all but one of the targeted attacks. It was also accurate when handling legitimate objects, blocking only two. It achieves an overall total accuracy rating of 95 per cent, which puts it in second place in this test.

Palo Alto Networks PA200 was strong when handling targeted attacks but was less effective against web-based malware and missed many of the email threats. It also blocked two legitimate objects so its overall total accuracy rating is below average at 52 per cent.

Cisco Snort detected more threats than it blocked. It detected just 53 per cent of the threats and stopped the same number, but 15 of those were neutralised as opposed to being blocked. **Snort** was strong when handling legitimate objects, blocking just one of them.

S**ymantec's** and F**ortinet's** appliances win AAA awards for their strong overall performance. **Palo Alto Networks**' product managed a C grade, which is considerably lower than in the last test, where it achieved an A grade. **Cisco's** appliance did not score well enough to win an award.

Appendices

APPENDIX A: Terms Used

TERM	MEANING
Compromised	The attack succeeded, resulting in malware running unhindered on the target. In the case of a targeted attack, the attacker was able to take remote control of the system and carry out a variety of tasks without hindrance.
Blocked	The attack was prevented from making any changes to the target.
False Positive	When a security product misclassifies a legitimate application or website as being malicious, it generates a 'false positive'.
Neutralised	The exploit or malware payload ran on the target but was subsequently removed.
Complete Remediation	If a security product removes all significant traces of an attack it has achieved complete remediation.
Target	The test system that is protected by a security product.
Threat	A program or sequence of interactions with the target that is designed to take some level of unauthorised control of that target.
Update	Security vendors provide information to their products in an effort to keep abreast of the latest threats. These updates may be downloaded in bulk as one or more files, or requested individually and live over the internet.

APPENDIX B: FAQs

A **full methodology** for this test is available from our website.

- The products chosen for this test were selected by SE Labs.
- The test was not sponsored. This means that no security vendor has control over the report's content or its publication.
- The test was conducted between 17th October 2017 and 4th January 2018.
- Malicious URLs and legitimate applications and URLs were independently located and verified by SE Labs.
- Targeted attacks were selected and verified by SE Labs. They were created and managed by a variety of publicly-available tools including Metasploit Framework Edition. The choice of attack techniques was advised by public information about ongoing attacks. One notable source was the **2018 Data Breach Investigations Report from Verizon**
- Malicious and legitimate data was provided to partner organisations once the full test was complete.

Q What is a partner organisation? Can I become one to gain access to the threat data used in your tests?

A Partner organisations benefit from our consultancy services after a test has been run. Partners may gain access to low-level data that can be useful in product improvement initiatives and have permission to use award logos, where appropriate, for marketing purposes. We do not share data on one partner with other partners. We do not partner with organisations that do not engage in our testing.

Q I am a security vendor and you tested my product without permission. May I access the threat data to verify that your results are accurate?

A We are willing to share a certain level of test data with non-partner participants for free. The intention is to provide sufficient data to demonstrate that the results are accurate. For more in-depth data suitable for product improvement purposes we recommend becoming a partner.

APPENDIX C: Product Versions

A product's update mechanism may upgrade the software to a new version automatically so the version used at the start of the test may be different to that used at the end.

PRODUCT VERSIONS		
Provider	Product Name	Build Service
Cisco	Snort	2.9.5
Fortinet	FortiGuard	5.4.5, build 1138 (GA)
Palo Alto	Networks	8.0.3
Symantec	Advanced Threat protection	2.3.0-233

APPENDIX D: Attack Types

The table below shows how each product protected against the different types of attacks used in the test.

ATTACK TYPES				
Product	Web-Download	E-mail	Targeted Attack	Protected (Total)
Symantec Advanced Threat Protection	50	25	25	100
Fortinet FortiGate	50	24	25	99
Palo Alto Networks PA200	45	5	25	25
Cisco Snort	42	4	7	53