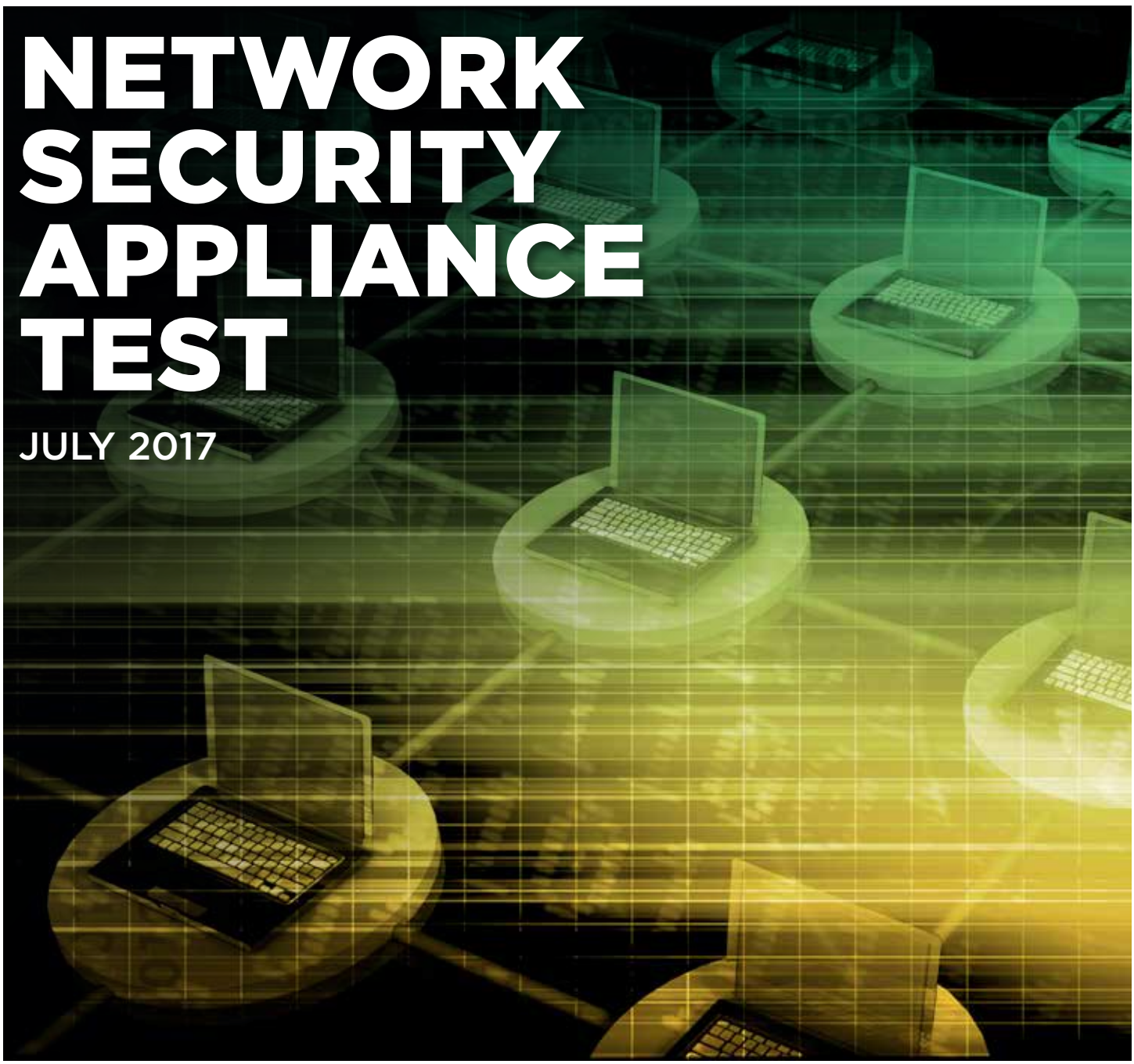
SE Labs

INTELLIGENCE-LED TESTING

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

NETWORK SECURITY APPLIANCE TEST

JULY 2017



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.

CONTENTS

Document version 1. 0. Written 12th July 2017

SIMON EDWARDS
Director

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

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

TESTING TEAM
Thomas Bean
Dimitar Dobrev
Gia Gorbold
Alexandru Statie
Jon Thompson
Jake Warren
Stephen Withey

IT SUPPORT
Danny King-Smith
Chris Short

PUBLICATION
Steve Haines
Colin Mackleworth

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

INTRODUCTION

Using layers of security is a well-known concept designed to reduce the chances of an attacker succeeding in breaching a network. If one layer fails, others exist to mitigate the threat. In this report we explore the effectiveness of network appliances designed to detect and block attacks against endpoint systems.

The systems we have tested here are popular appliances designed to sit between your endpoints and the internet router. They are designed to detect, and often protect against, threats coming in from the internet or passing through the local network. Their role is to stop threats before they reach the endpoints. If they fail to stop a threat, they might learn that an attack has happened and generate an alert, while subsequently blocking future, similar attacks.

In some cases an appliance will take information it considers suspicious and send it to a cloud-based service for further analysis. In this way it might allow a threat through the first time, explore it more deeply using the cloud service and send back information to the appliance so that it will block that same (or similar) attack in future. It's a little like an immune system.

As immune systems adapt to protect against known threats, so threats adapt in an arms race to defeat protection mechanisms. This report includes our first public set of network security appliance results. Future reports will keep you updated as to how well the industry competes with the bad guys in the real world.

If you spot a detail in this report that you don't understand, or would like to discuss, please contact us via Twitter or Facebook.

SE Labs uses current threat intelligence to make our tests as realistic as possible. To learn more about how we test, how we define 'threat intelligence' and how we use it to improve our tests, please visit our website and follow us on Twitter.

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 19.

Products tested

PRODUCT	PROTECTED ACCURACY RATING	LEGITIMATE ACCURACY RATING	TOTAL ACCURACY RATING
Symantec Advanced Threat Protection	100%	97%	99%
Fortinet FortiGate	100%	95%	98%
Palo Alto Networks PA200	82%	100%	91%
Cisco Snort	81%	87%	84%

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 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**

All 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.

- **False positives were not an issue for most products**

With the exception of Cisco Snort, the products did not generate significant numbers of false positives. Palo Alto Networks' appliance made no errors at all in this part of the test.

- **Which products were the most effective?**

Symantec's and Fortinet's appliances stopped the most threats, while not blocking significant amounts of legitimate traffic. They both win AAA awards.

Simon Edwards, SE Labs, 12th July 2017

1. TOTAL ACCURACY RATINGS

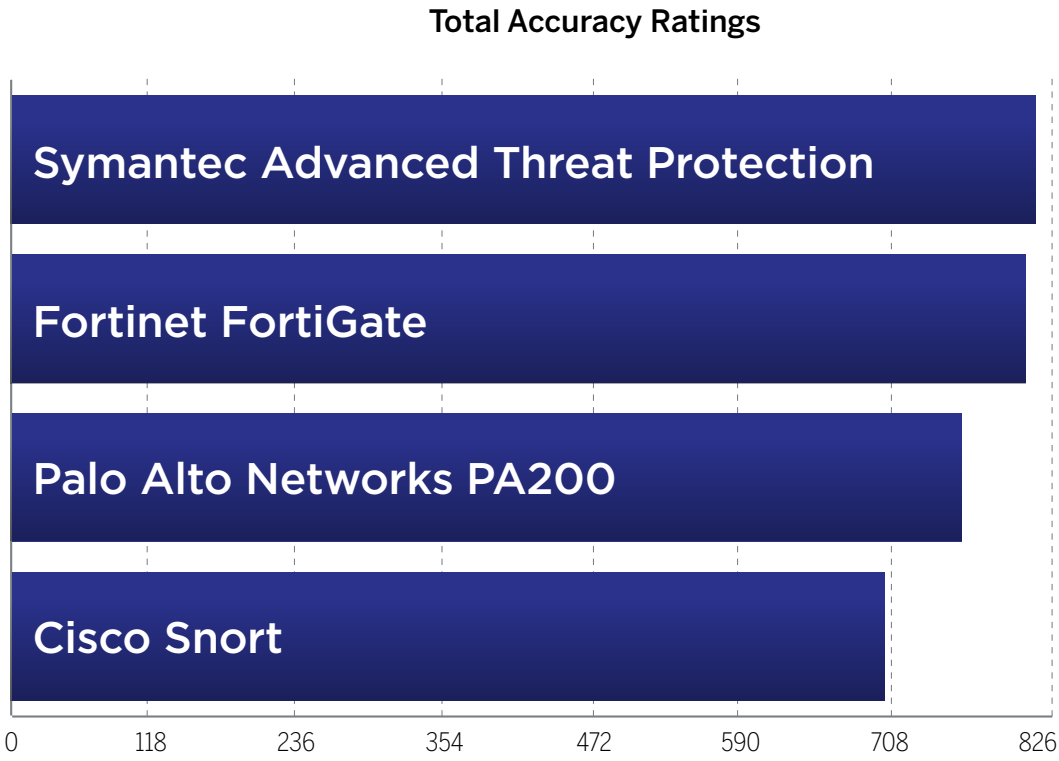
Judging the effectiveness of a security product is a subtle art, and many factors are at play when assessing how well it performs. To make things easier, we've combined all the different results from this report into one easy-to-understand graph.

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 12.



Total Accuracy Ratings combine protection and false positives.

Awards

The following products win SE Labs awards:

- Symantec Advanced Threat Protection
- Fortinet FortiGate

- Palo Alto Networks PA200

- Cisco Snort

TOTAL ACCURACY RATINGS

Product	Total Accuracy Rating	Total Accuracy (%)	Award
Symantec Advanced Threat Protection	814	99%	AAA
Fortinet FortiGate	806	98%	AAA
Palo Alto Networks PA200	755	91%	AA
Cisco Snort	693.5	84%	B

2. PROTECTION RATINGS

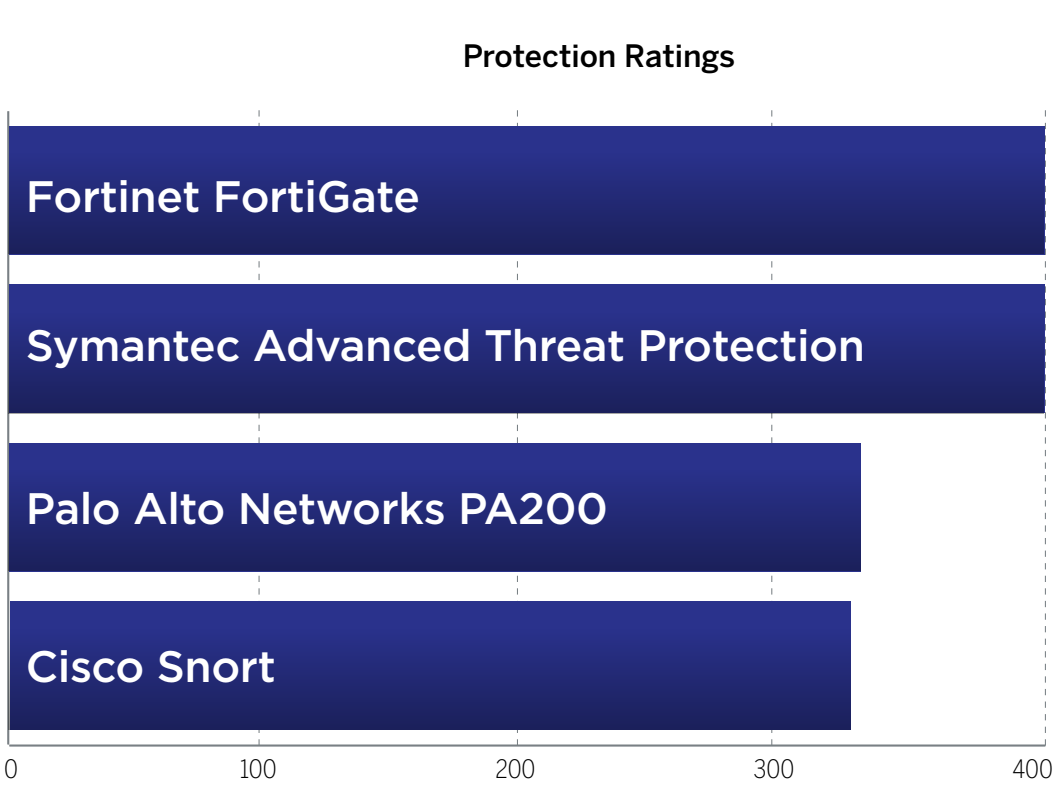
The results below indicate how effectively the products dealt with threats. Points are earned for detecting the threat and for either blocking or neutralising it.

- **Detected (+1)**
If the product detected 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 allow the initial attack stage to succeed but blocks the full attack.

- **Compromised (-5)**
If the threat compromised 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.



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

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.

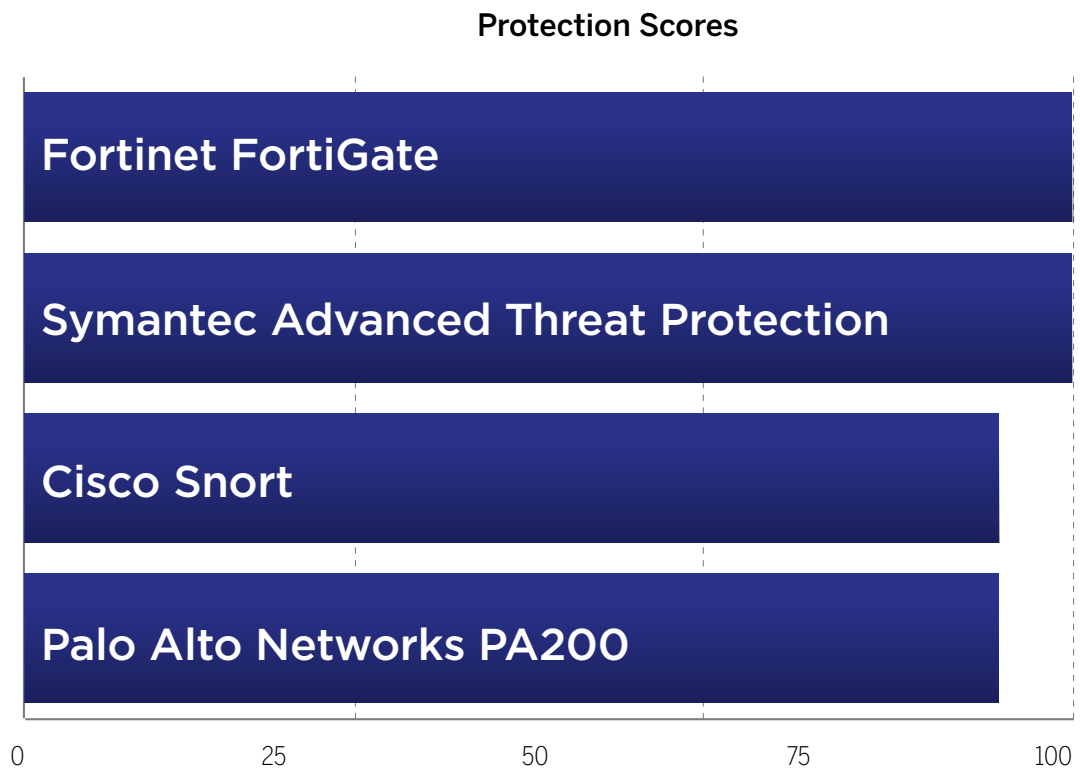
These ratings are simple and based on our opinion of how important these different outcomes are. You may have a different view on how seriously you treat a 'Compromise' or 'Neutralisation without complete remediation'. If you want to create your own rating system, you can use the raw data from 4. Protection Details on page 11 to roll your own set of personalised ratings.

PROTECTION RATINGS		
Product	**Protection Rating**	**Protection Accuracy (%)**
Fortinet FortiGate	400	100%
Symantec Advanced Threat Protection	400	100%
Palo Alto Networks PA200	329	82%
Cisco Snort	325	81%

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.



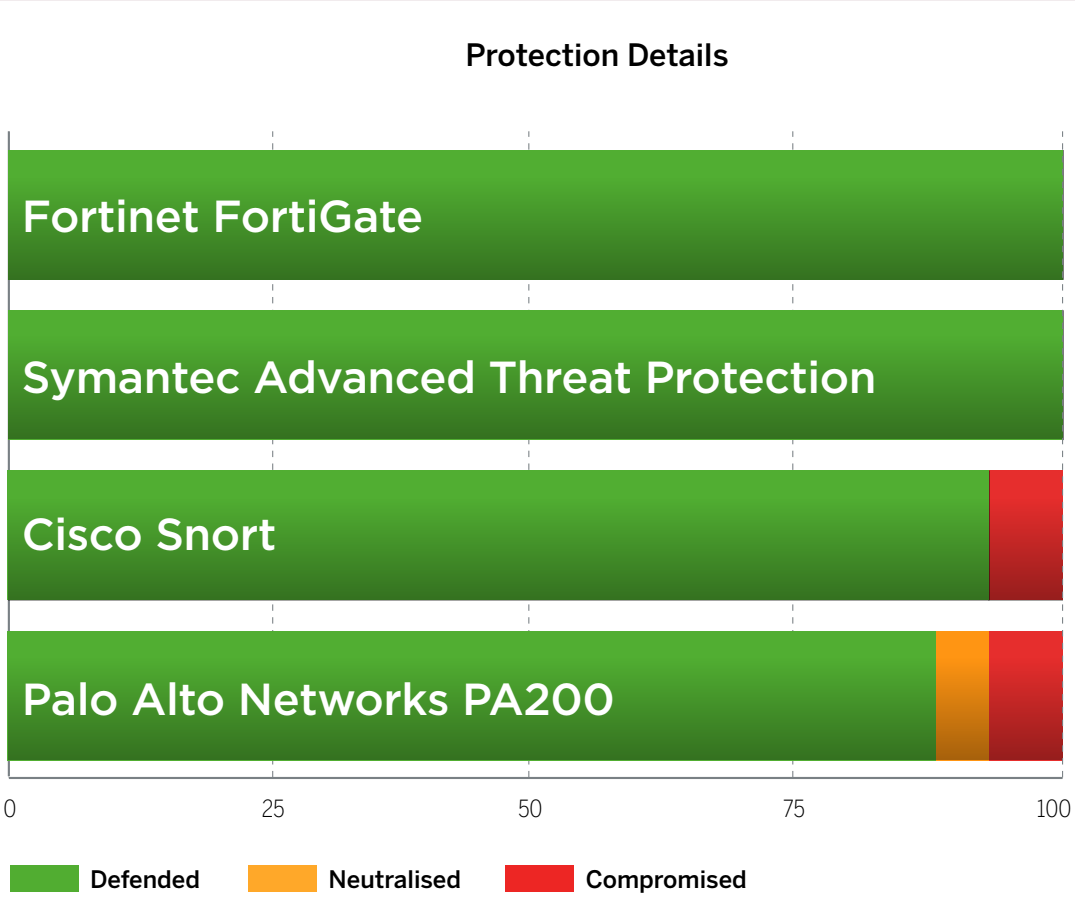
Protection Scores are a simple count of how many times a product protected the system.

PROTECTION SCORES	
Product	**Protection Score**
Fortinet FortiGate	100
Symantec Advanced Threat Protection	100
Cisco Snort	93
Palo Alto Networks PA200	93

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.



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

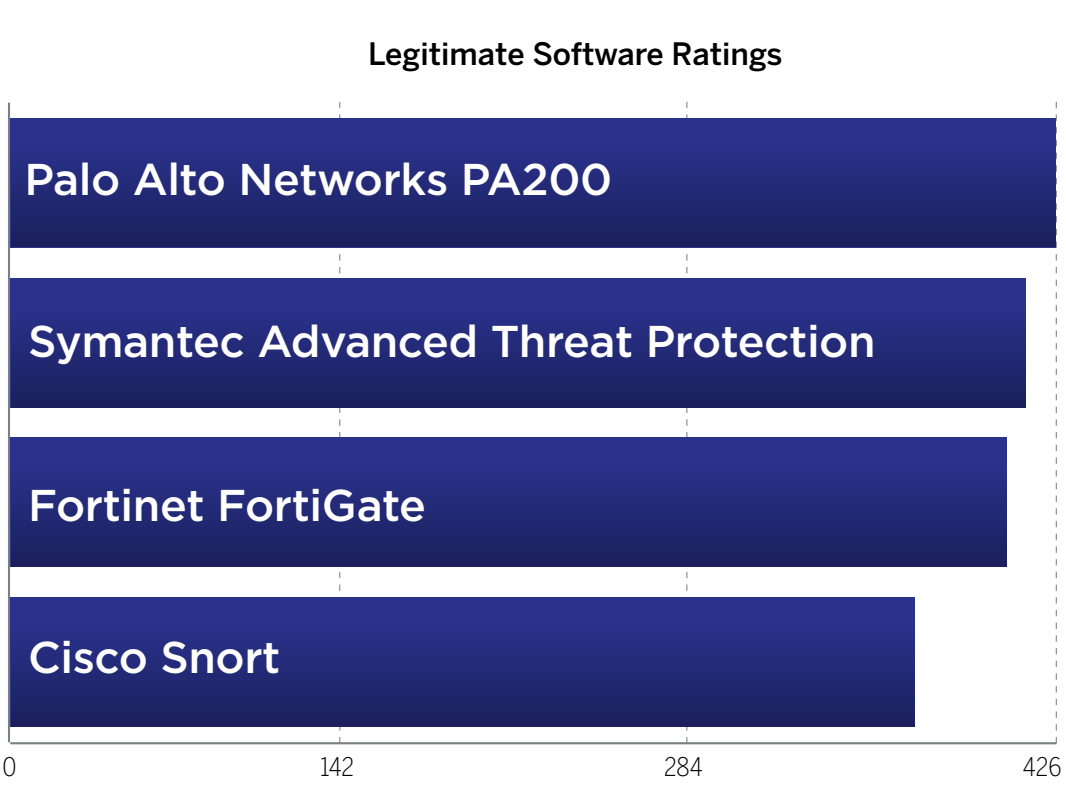
PROTECTION DETAILS					
Product	**Detected**	**Blocked**	**Neutralised**	**Compromised**	**Protected**
Fortinet FortiGate	100	100	0	0	**100**
Symantec Advanced Threat Protection	100	100	0	0	**100**
Cisco Snort	81	93	0	7	**93**
Palo Alto Networks PA200	95	88	5	7	**93**

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 15.



Legitimate software ratings can indicate how well a vendor has tuned its detection engine.

LEGITIMATE SOFTWARE RATINGS		
Product	**Legitimate Accuracy Rating**	**Legitimate Accuracy (%)**
Palo Alto Networks PA200	426	100%
Symantec Advanced Threat Protection	414	97%
Fortinet FortiGate	406	95%
Cisco Snort	368.5	87%

5.1 Interaction ratings

It's crucial that security 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 false positives 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 a product'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 product 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.

Interaction Ratings

	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	

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

INTERACTION RATINGS		
Product	**None (allowed)**	**None (blocked)**
Palo Alto Networks PA200	50	0
Fortinet FortiGate	49	1
Symantec Advanced Threat Protection	49	1
Cisco Snort	45	5

5.2 Prevalence ratings

There is a significant difference between a product blocking a popular application like 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 being 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 following table.

LEGITIMATE SOFTWARE PREVALENCE RATING MODIFIERS	
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 date 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 12.

5.4 Distribution of impact categories

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	
Prevelance Rating	**Frequency**
Very high impact	32
High impact	9
Medium impact	5
Low impact	3
Very low impact	1
Grand total	**50**

6. CONCLUSIONS

Attacks in this test included infected websites available to the general public, including sites that automatically attack visitors and attempt to infect them without any social engineering or other interaction. Some sites relied on users being fooled into installing the malware.

URLs were introduced to the targets directly and, in relevant cases, via email. We also included targeted attacks, which were 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 web-based threats, malware downloads and targeted attacks. It blocked 100 per cent of the threats and was also accurate when handling legitimate objects, blocking only one. It achieves an overall total accuracy rating of 99 per cent.

Fortinet FortiGate also protected against all of the public and targeted attacks. Similarly, it detected and blocked outright all of the threats. It blocked only one legitimate object but it has a slightly lower total accuracy rating than Symantec because that object that it blocked was more prevalent, which made the Non-Optimal Classification/Action more impactful. It comes a very close second.

Palo Alto Networks PA200 detected most of the threats but blocked fewer. It neutralised five and allowed seven threats through to the target endpoints. However, it was 100 per cent effective at handling legitimate objects so its overall total accuracy rating is a strong 91 per cent.

Cisco Snort detected 81 per cent of the threats and blocked 93 per cent. While that result may look strange, we determined that in 12 cases the threats were blocked but the appliance did not generate any alerts or logs. Without the appliance in place those threats were able to run fully, so the presence of Snort was blocking the threats without appearing to detect them. Users would have been protected but would not have been alerted about the attacks. Snort was the weakest product when it came to handling legitimate objects, blocking five of them.

The appliances from Symantec and Fortinet win AAA awards for their strong overall performance. Palo Alto Networks' product achieved a good AA grade while Cisco's appliance managed a B 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 20th January 2016 and 15th March 2017.
- 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 Metasploit Framework Edition using default settings. The choice of exploits was advised by public information about ongoing attacks. One notable source was the **2016 Data Breach Investigations Report** from Verizon.
- Malicious and legitimate data was provided to partner organisations once the full test was complete.

Q I am a security vendor. How can I include my product in your test?

A Please contact us at info@SELabs.uk. We will be happy to arrange a phone call to discuss our methodology and the suitability of your product for inclusion.

Q I am a security vendor. Does it cost money to have my product tested?

A We do not charge directly for testing products in public tests. We do charge for private tests.

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

A Partner organisations support our tests by paying for access to test data after each test has completed but before publication. Partners can dispute results and use our awards logos for marketing purposes. We do not share data on one partner with other partners. We do not currently partner with organisations that do not engage in our testing.

Q So you don't share threat data with test participants before the test starts?

A No, this would bias the test and make the results unfair and unrealistic.

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 small subsets of data with non-partner participants at our discretion. A small administration fee is applicable.

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		
Vendor	**Product**	**Build**
Cisco	Snort	2.9.8.3 GRE (Build 383)
Fortinet	FortiGuard	5.4.0
Palo Alto Networks	PA200	7.1.7
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 Drive-by**	**Web-Download**	**Targeted Attack**	**Protected (Total)**
Fortinet FortiGate	51	24	25	**100**
Symantec Advanced Threat Protection	51	24	25	**100**
Palo Alto Networks PA200	50	18	25	**93**
Cisco Snort	47	21	25	**93**