SE Labs

INTELLIGENCE-LED TESTING

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

NETWORK SECURITY APPLIANCE TEST

DECEMBER 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 4th December 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
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

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)

While every effort is made to ensure the accuracy of the information published in this document, no guarantee is expressed or implied and SE Labs Ltd does not accept liability for any loss or damage that may arise from any errors or omissions.

INTRODUCTION

There have been so many publicised data breaches in 2017 that we don't even have enough space on this page to provide a basic summary. In many cases a business network was breached. Business networks comprise endpoints (usually Windows PCs), servers, Point of Sale computers and a range of other devices.

One approach to compromising a business is to hack an endpoint (PC) and then to use it as a platform from which to launch further attacks into the network. For example, rather than going straight for a company's main servers why not trick a user into infecting his/ her computer with malware? We can then scan and infect the entire network, stealing information, causing damage and generally behaving in ways contrary to the business' best interests.

There is some really good endpoint software available, as we see in our regular Endpoint Protection tests, but nothing is perfect and any extra layers of security are welcome. If one layer fails, others exist to mitigate the threat. In this report we explore the effectiveness of network appliances designed to detect and protect against 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.

There are no guarantees that technology will always protect you from attackers, but our results show that adding layers of security is an effective way to improve your prospects when facing general and more targeted attacks.

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

EXECUTIVE SUMMARY			
Product	**Protection Accuracy Rating**	**Legitimate Accuracy Rating**	**Total Accuracy Rating**
Fortinet FortiGate	96%	95%	95%
Symantec Advanced Threat Protection	78%	99%	92%
Palo Alto Networks PA200	73%	95%	88%
Cisco Snort	-25%	100%	59%

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**
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.

- **False positives were not an issue for most products**
With the exception of **Fortinet's** appliance, the products did not generate significant numbers of false positives. **Cisco Snort** made no errors at all in this part of the test.

- **Which products were the most effective?**
Fortinet's appliances stopped the most threats and, despite blocking some legitimate traffic, still managed to win an AAA award. **Symantec Advanced Threat Protection** came a close second, winning a AA award.

Simon Edwards, SE Labs, 4th December 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.

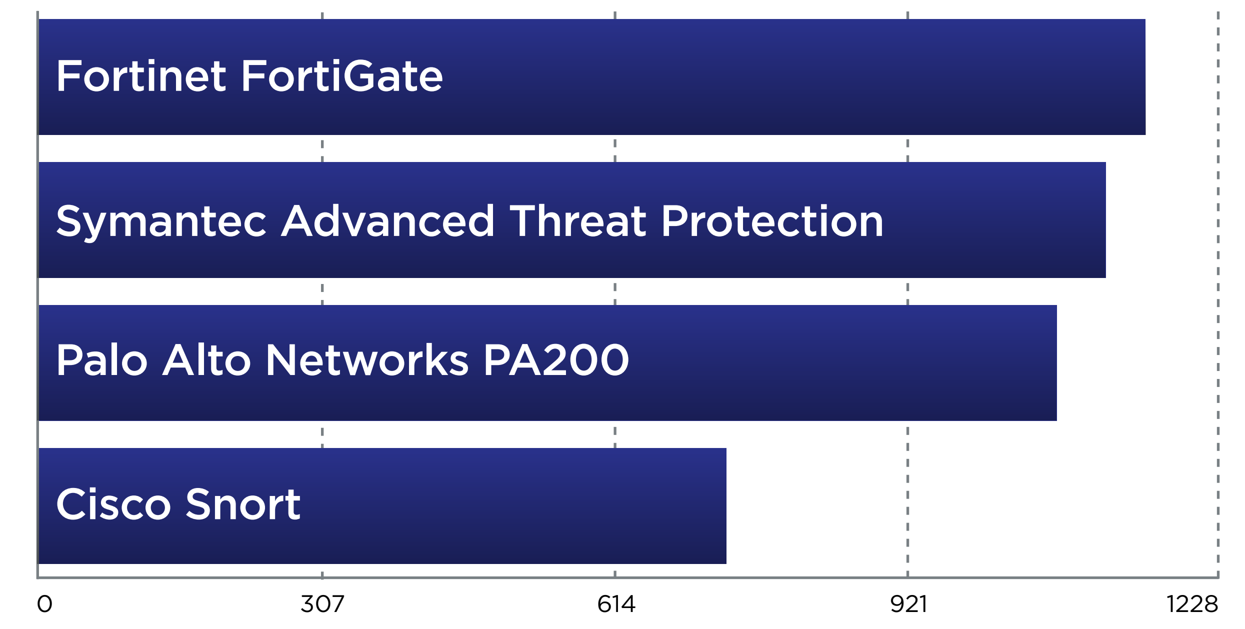
The graph below takes into account not only each product's ability to detect and protect against threats, but also its handling of non-malicious objects such as web addresses (URLs) and applications.

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



Total Accuracy Ratings combine protection and false positives.

Awards

The following products win SE Labs awards:

- Fortinet FortiGate
- Symantec Advanced Threat Protection
- Palo Alto Networks PA200

TOTAL ACCURACY RATINGS			
Product	**Total Accuracy Rating**	**Total Accuracy (%)**	**Award**
Fortinet FortiGate	1171	95%	AAA
Symantec Advanced Threat Protection	1129	92%	AA
Palo Alto Networks PA200	1077	88%	A
Cisco Snort	728	59%	

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.

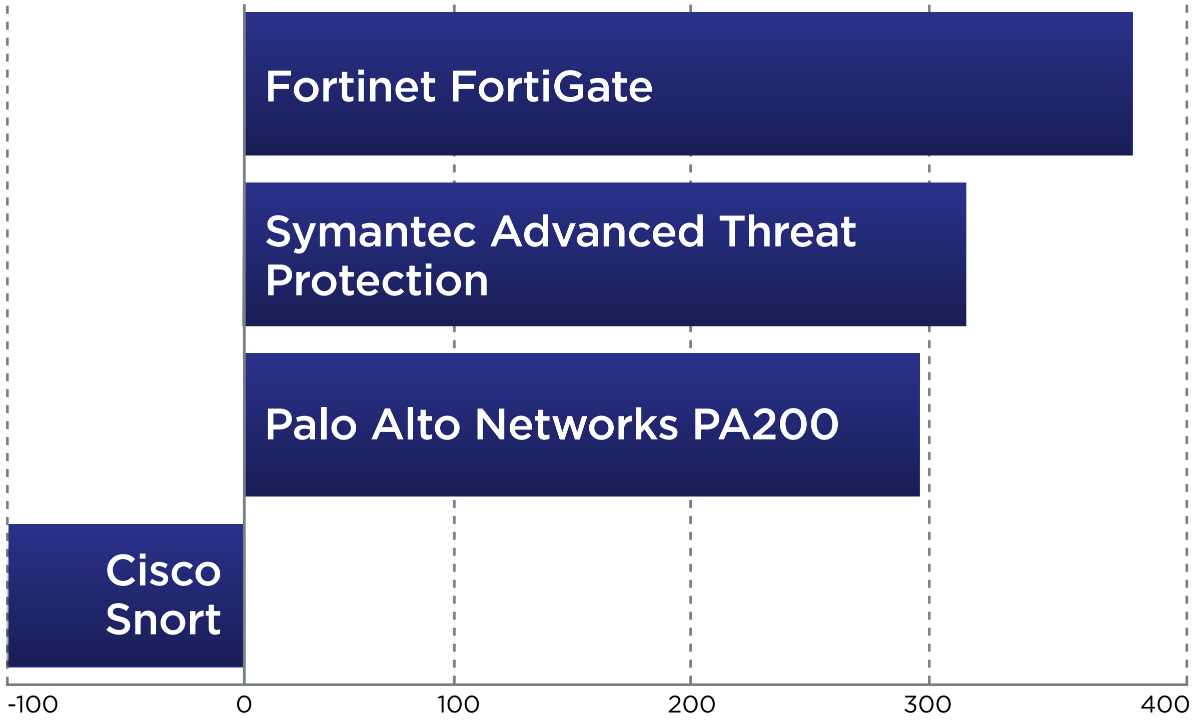
- **Detected (+1)**
If the product detected the threat with any degree of useful information, we award it one point.

- **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



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

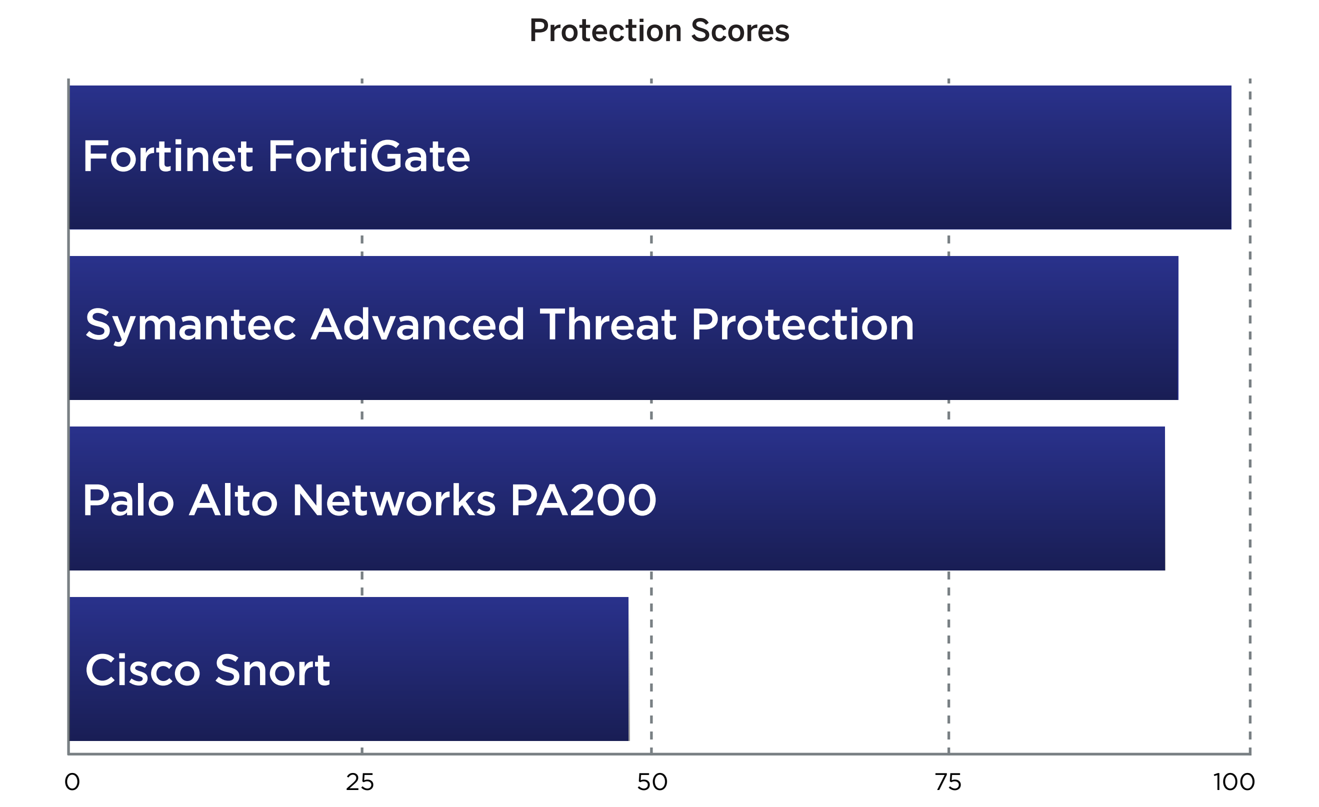
Product	Protection Rating	Protection Accuracy (%)
Fortinet FortiGate	385	96%
Symantec Advanced Threat Protection	313	78%
Palo Alto Networks PA200	293	73%
Cisco Snort	-100	-25%

Average: 56%

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.



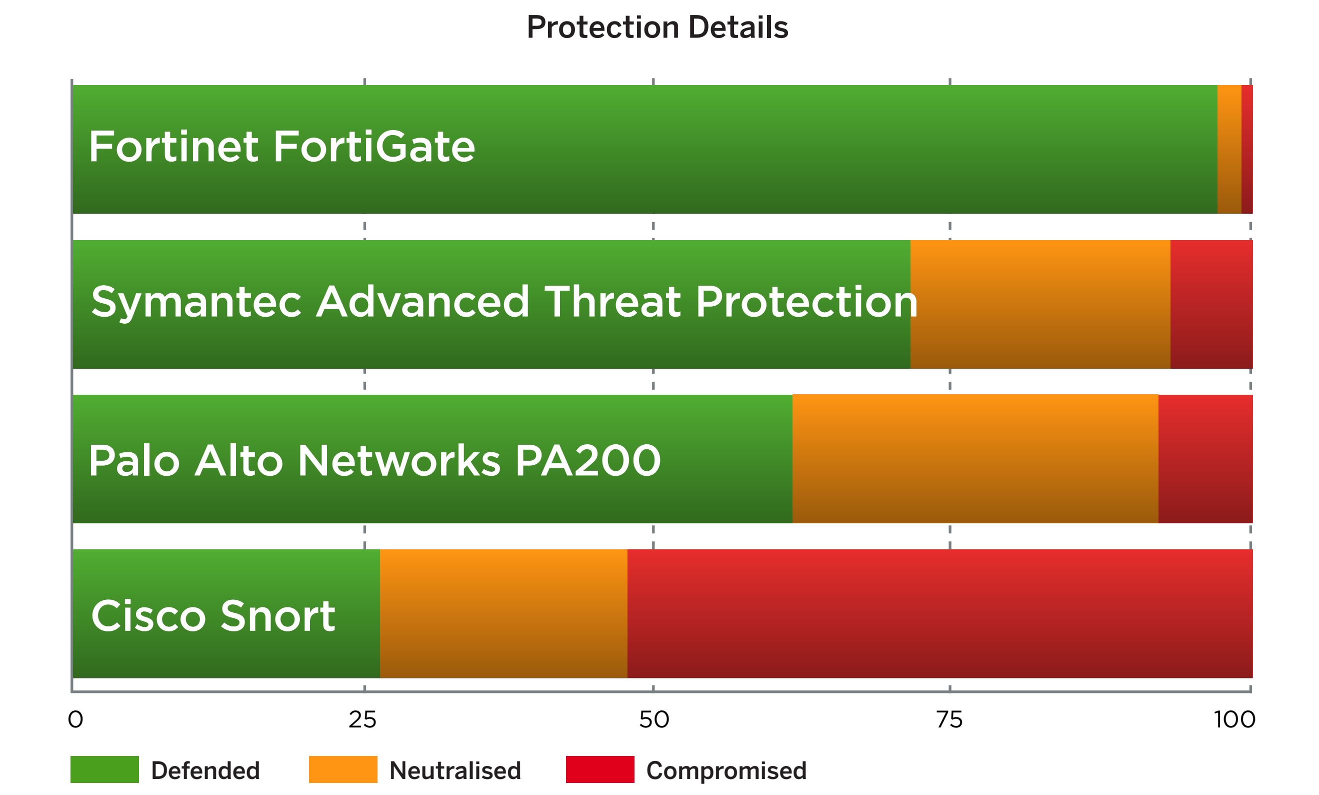
Protection Scores are a simple count of how many times a product protected the system.

PROTECTION SCORES	
Product	**Protection Score**
Fortinet FortiGate	98
Symantec Advanced Threat Protection	93
Palo Alto Networks PA200	92
Cisco Snort	47

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.

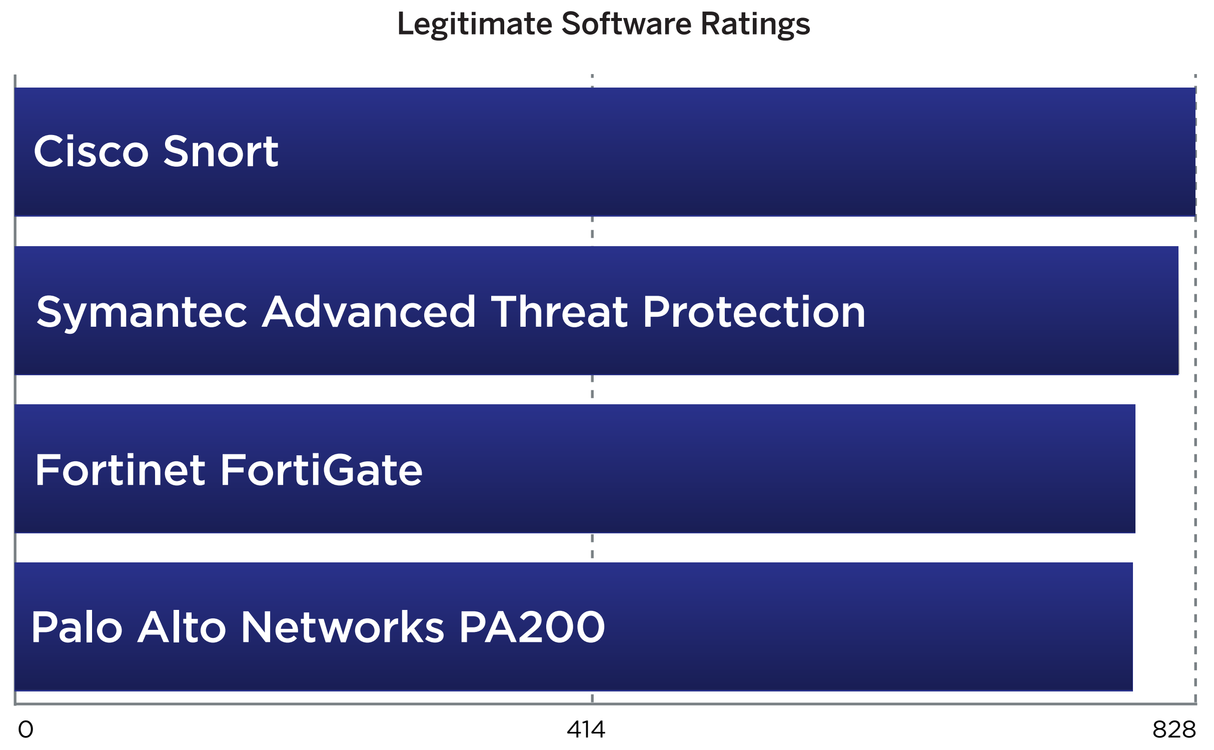
PROTECTION DETAILS					
Product	**Detected**	**Blocked**	**Neutralised**	**Compromised**	**Protected**
FortiGate	99	97	2	1	99
Symantec Advanced Threat Protection	92	71	22	7	93
PA200	92	61	31	8	92
Snort	56	26	21	53	47

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 (%)**
Cisco Snort	828	100%
Symantec Advanced Threat Protection	816	99%
Fortinet FortiGate	786	95%
Palo Alto Networks PA200	784	95%

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)**
Cisco Snort	100	0
Symantec Advanced Threat Protection	98	2
Fortinet FortiGate	97	3
Palo Alto Networks PA200	97	3

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	53
High impact	24
Medium impact	12
Low impact	6
Very low impact	5
Grand total	**100**

6. CONCLUSIONS

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.

Fortinet FortiGate protected against all of the public attacks and only missed two targeted attacks. Similarly, it detected and blocked outright all of the threats. It blocked three legitimate objects.

Symantec Advanced Threat Protection protected against a good number of the public email threats, malware downloads and targeted attacks. It blocked 93 per cent of the threats and was also accurate when handling legitimate objects, blocking only two. It achieves an overall total accuracy rating of 92 per cent, which puts it in second place in this test.

Palo Alto Networks PA200 was strong when handling targeted attacks and email threats but was less effective against web-based malware. It also blocked three legitimate objects so its overall total accuracy rating is (slightly above average) 88 per cent.

Cisco Snort detected more threats than it blocked. It detected just 56 per cent of the threats and stopped 47 per cent. **Snort** was the strongest product when it came to handling legitimate objects, blocking none of them. This is a vast improvement over its performance in our previous test, in which **Snort** was the product most prone to false positives.

Fortinet's appliance wins an AAA award for its strong overall performance. **Symantec's** achieves a solid AA, while **Palo Alto Networks**' product managed 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 10th July 2017 to 31st August 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	FortiGate	v.5.4.5, build 1138 (GA)
Palo Alto Networks	PA200	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 Drive-by	Web-Download	Targeted Attack	Protected (Total)
Fortinet FortiGate	23	25	50	98
Symantec Advanced Threat Protection	25	20	48	93
Palo Alto Networks PA200	24	24	44	92
Cisco Snort	17	11	19	47