SE LABS

EPS
PROTECTION

Security Evaluation Test Report

[redacted] with

[redacted]

JANUARY 2025

SE LABS ® was commissioned to compare two threat protection products, [redacted] and [redacted], both installed on the standard [redacted] endpoint configuration. The goal was to judge their respective effectiveness at detecting and protecting against advanced targeted attacks and prevalent known malware attacks.

Each configuration was exposed to the same threats, which were a mixture of targeted attacks using well-established techniques; the same targeted attacks disguised with a variety of evasion methods; and public web-based threats that were found 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 1st January 2025

SE LABS

Management
Chief Executive Officer Simon Edwards
Chief Operations Officer Marc Briggs
Chief Human Resources Officer Magdalena Jurenko
Chief Technical Officer Stefan Dumitrascu

Testing Team
Nikki Albesa
Thomas Bean
Solandra Brewster
Jarred Earlington
Gia Gorbold
Anila Johny
Erica Marotta
Jeremiah Morgan
Julian Owusu-Abrokwa
Joseph Pike
Georgios Sakatzidi
Enejda Torba
Dimitrios Tsarouchas
Stephen Withey

Marketing
Sara Claridge
Janice Sheridan

Publication
Rahat Hussain
Colin Mackleworth

IT Support
Danny King-Smith
Chris Short

Website selabs.uk
Email info@SELabs.uk
LinkedIn www.linkedin.com/company/se-labs/
Blog blog.selabs.uk
Post SE Labs Ltd, 55A High Street, Wimbledon, SW19 5BA, UK

SE Labs is ISO/IEC 27001 : 2022 certified and BS EN ISO 9001 : 2015 certified for The Provision of IT Security Product Testing.

SE Labs is a member of the Microsoft Virus Initiative (MVI); the Anti-Malware Testing Standards Organization (AMTSO); the Association of anti Virus Asia Researchers (AVAR); and NetSecOPEN.

CEO
Simon Edwards

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

Are all breaches as important as each other?

And how many missed detections are OK?

What do computer breach detection products have in common with household bleach? They claim to protect against most known threats, killing 99.9 per cent of all known germs/ malware.

Consider that estimates place the total mass of all bacteria in the world as being larger than the combined mass of all other living plants and animals. That 0.1 per cent of missed known germs seems rather large now, and the protection afforded by a $2 bottle of bleach seems scarily pathetic. It gets worse when you consider the possible danger posed by unknown threats.

Similarly, the number of malware samples known to one or more researchers has been charted regularly as being in an exponential explosion for years. If an endpoint product managed to stop just short of 100 per cent should we be impressed, considering that 0.1 per cent still represents a massive problem? And again, what about the unknown threat?

We've all seen security reports in which products achieve fantastic results. Maybe Endpoint X scored 99 per cent, while Endpoint Y scored just 50 per cent. Clearly you should invest in X and not Y.

But when products' test results look good but are clustered together (as they are in this report), how do you decide which is best for your organisation? Cost? Manageability? Existing skills in the IT department? All of these are important factors and it's possible to secure an endpoint with a less than stellar product if the configurations used are sufficiently hardened. A good team can turn an average product into something more effective than a standard installation of a 'good' product.

In this test we have an interesting set of results. Both products perform very well, but one missed four threats and, in three of those cases, the attacker was able to gain complete control of the endpoint – to the point where it was even possible to achieve a more powerful level of access than that enjoyed by an administrator.

In terms of percentages, there's little to choose between the two products. But it only takes one successful attack to compromise a network, and in this relatively small test we managed four compromises, three of which would have been devastating to the business' security.

Executive Summary

Product Names

It is good practice to stay up to date with the latest version of your chosen endpoint security product. We made best efforts to ensure that each product tested was the very latest version running with the most recent updates to give the best possible outcome.

- **The endpoints were effective at handling general threats from cyber criminals...**

All products were capable of handling public web-based threats such as those used by criminals to attack Windows PCs and install ransomware. The same applies to more targeted file-based malware attacks.

- **... but targeted attacks posed more of a challenge**

[redacted] product was very effective at stopping all targeted attacks, while [redacted] handled all but four. Of these, three comprised a thorough and complete compromise of the targeted system.

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

The two products generated no false positive results, meaning that no legitimate applications were mis-classified or blocked.

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

Both products were capable, but [redacted] blocked all threats, while [redacted] stopped all but four. Both generated no false positives. Both are classed by SE Labs as being AAA grade options.

Products Tested	Protection Accuracy Rating (%)	Legitimate Accuracy Rating (%)	Total Accuracy Rating (%)
[redacted] with [redacted]	100%	100%	100%
[redacted] with [redacted]	99%	100%	100%

- **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 **1.3 Total Accuracy Ratings** on page 8.

Security Evaluation Test Award

The following product wins the SE Labs award:

SE LABS
AAA
January 2025
Security Evaluation Award

[redacted] with [redacted]

[redacted] with [redacted]

Threat Responses

Full Attack Chain: Testing Every Layer of Detection and Protection

Attackers start from a certain point and don't stop until they have either achieved their goal or have reached the end of their resources (which could be a deadline or the limit of their abilities). This means that, in a test, the tester needs to begin the attack from a realistic first position, such as sending a phishing email or setting up an infected website, and moving through many of the likely steps leading to actually stealing data or causing some other form of damage to the network.

If the test starts too far into the attack chain, such as executing malware on an endpoint, then many products will be denied opportunities to use the full extent of their protection and detection abilities. If the test concludes before any 'useful' damage or theft has been achieved, then similarly the product may be denied a chance to demonstrate its abilities in behavioural detection and so on.

Attack Stages

The illustration (below) shows typical stages of an attack. In a test, each of these should be attempted to determine the security solution's effectiveness. This test's results record detection and protection for each of these stages.

We measure how a product responds to the first stages of the attack with a detection and/ or protection rating. Sometimes products allow threats to run yet still detect them. Other times they might allow the threat to run briefly before neutralising it. Ideally, they detect and block the threat before it has a chance to run. Products may delete threats or automatically contain them in a 'quarantine' or other safe holding mechanism for later analysis.

Should the initial attack phase succeed, we then measure post-exploitation stages, which are represented by steps two through to seven below. We broadly categorise these stages as: Access (step 2); Action (step 3); Escalation (step 4); and Post-Escalation (steps 5-6).

In figure 1. you can see a typical attack running from start to end, through various 'hacking' activities. This can be classified as a fully successful breach.

In figure 2. a product or service has interfered with the attack, allowing it to succeed only as far as stage 3, after which it was detected and neutralised. The attacker was unable to progress through stages 4 onwards.

It is possible for an attack to run in a different order with, for example, the attacker attempting to connect to other systems without needing to escalate privileges. However, it is common for password theft (see step 5) to occur before using stolen credentials to move further through the network.

How Hackers Progress

Figure 1. A typical attack starts with an initial contact and progresses through various stages, including reconnaissance, stealing data and causing damage.

Figure 2. This attack was initially successful but only able to progress as far as the reconnaissance phase.

1. Protection and Legitimate Handling Accuracy

1.1 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 aren't equipped to stop it. Products can also provide protection even if they don't detect certain threats. Some threats abort on detecting specific endpoint protection software.

Product	Detected	Blocked	Completely Remediated	Compromised	Protected
[redacted] with [redacted]	125	125	93	0	125
[redacted] with [redacted]	121	121	74	4	121

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

1.2 Attack Types

The table below shows how each configuration protected against the different types of attacks used in the test. The following labels for each group of attack types is used:

TA HTTP: Targeted attacks using shellcode injection techniques, delivered over the HTTP (web) protocol.
TA HTTP (extended): As TA HTTP, but using an executable compression technique for evasion purposes.
TA Email: Targeted attacks, including Macro-enabled and malicious Microsoft Office documents and other script-based 'file-less' attacks, all delivered via email.

Attack Type	[redacted] with [redacted]	[redacted] with [redacted]
Web Download	25	25
TA HTTP	20	20
TA Email	20	16
Web Download (extended – method 1)	23	23
Web Download (extended – method 2)	17	17
TA HTTP (extended)	20	20

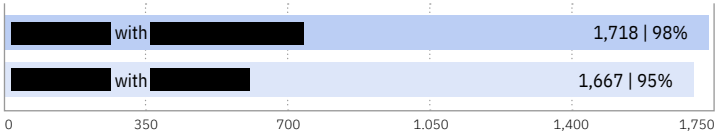
Web Download: Prevalent, non-targeted attacks that potentially affect any internet users visiting websites with the Google Chrome web browser.
Web Download (extended – method 1): As Web Download, but each file was altered using an executable compression technique for evasion purposes.
Web Download (extended – method 2): As Web Download (extended – method 1), but using a different executable compression technique.

1.3. Total Accuracy Ratings

Judging the effectiveness of an endpoint 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 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 stops the threat before it can even start its intended series of malicious events. Alternatively, the product might allow a web-based exploit to execute but prevent it from downloading any further code to the target.



● **Total Accuracy Ratings combine protection and false positives**

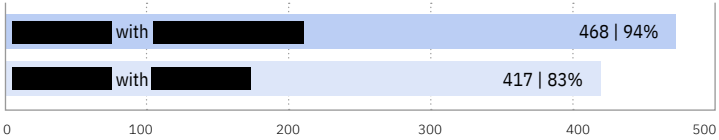
In another case malware might run on the target for a short while before its behaviour is detected and its code is deleted or moved to a safe 'quarantine' area for future analysis. We take these outcomes into account when attributing points that form final ratings.

For example, a product that completely blocks a threat is rated more highly than one that 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 **Legitimate Accuracy Ratings** on page 9.

1.4 Protection Accuracy

To understand how we calculate these ratings, see **Appendix A: Protection Ratings** on page 11.



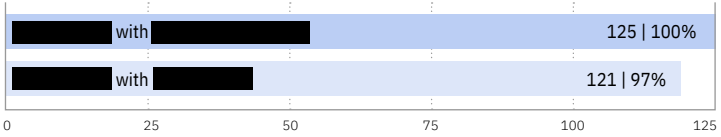
● **Protection Ratings are weighted to show that how products handle threats can be subtler than just 'win' or 'lose'.**

Average 88.5%

1.5 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.

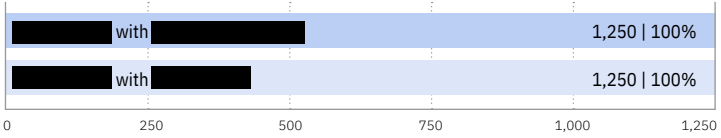


● **Protection Scores are a simple count of how many times a product protected the system.**

1.6 Legitimate Accuracy 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.



● **Legitimate Accuracy Ratings can indicate how well a vendor has tuned its detection engine.**

To understand how we calculate these ratings, see **Accuracy Ratings** on page 13.

2. Conclusion

Attacks in this test included targeted attacks and attacks that affected the general public at the time of testing. In order to test how effective the different products and configurations were at detecting and protecting against the threats, we used the attacks in a number of ways. These varied from attacking systems using threats as found in the public domain and as generated by tools, through to using varying evasion techniques to challenge the security products to detect threats that were disguised as legitimate applications or scrambled to appear differently to known threats.

The products themselves were [redacted] and [redacted] (otherwise known as [redacted]. Both products claim to detect and protect from known and unknown threats. These products were tested alongside [redacted] standard endpoint configuration, which includes other security products and measures.

The results in this report illustrate how effectively each of the two products performed against known and unknown threats when deployed according to [redacted] usual processes.

Both products handled the known public threats perfectly and also detected and protected against these same threats when each was disguised using two different techniques. These techniques scramble the code so that regular signature-based detections would likely fail. It is reasonable to conclude that the behavioural detection abilities of the products work well with prevalent threats that are both known and unknown.

File-based targeted attacks were also no match for the two configurations, and the evasion techniques used in the test were insufficient to bypass detection. However, there was a weakness with [redacted] when certain 'file-less' targeted attacks were introduced via email. One fifth of the email-bourn targeted attacks were successful against [redacted] with three of the four attacks being able to achieve a thorough and undetected compromised. [redacted] detected and blocked all of these attacks.

If this was a public test we would rate both products as AAA grade.

Appendices

Appendix A: 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.

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.

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

Persistent Neutralisation (-2) This result occurs when a product continually blocks a persistent threat from achieving its aim, while not removing it from the system.

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.

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 **1.1 Protection Details** on page 7 to roll your own set of personalised ratings.

Targeted Attack Scoring

The following scores apply only to targeted attacks and are cumulative, ranging from -1 to -5.

Access (-1) If any command that yields information about the target system is successful this score is applied. Examples of successful commands include listing current running processes, exploring the file system and so on. If the first command is attempted and the session is terminated by the product without the command being successful the score of Neutralised (see above) will be applied.

Action (-1) If the attacker is able to exfiltrate a document from the target's Desktop of the currently logged in user then an 'action' has been successfully taken.

Escalation (-2) The attacker attempts to escalate privileges to NT Authority/System. If successful, an additional two points are deducted.

Post-Escalation Action (-1) After escalation the attacker attempts actions that rely on escalated privileges. These include attempting to steal credentials, modifying the file system and recording keystrokes. If any of these actions are successful then a further penalty of one point deduction is applied.

Appendix B: Legitimate 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.

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**

	None (allowed)	Click to Allow (default allow)	Click to Allow/ Block (no recommendation)	Click to Block (default block)	None (blocked)	
Safe	2	1.5	1			A
Unknown	2	1	0.5	0	-0.5	B
Not Classified	2	0.5	0	-0.5	-1	C
Suspicious	0.5	0	-0.5	-1	-1.5	D
Unwanted	0	-0.5	1	-1.5	-2	E
Malicious				2	-2	F
	1	2	3	4	5	

Legitimate Software Prevalence Rating Modifiers

Very High Impact	5
High Impact	4
Medium Impact	3
Low Impact	2
Very Low Impact	1

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.

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 Tranco.com's global traffic ranking system.

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 **Legitimate Accuracy Ratings** on page 9.

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 Interaction Ratings

Product	None (allowed)	None (blocked)	Click to Block (default block)
[redacted] with [redacted]	125	0	0
[redacted] with [redacted]	125	0	0

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

Legitimate Software Category Frequency

Prevalence Rating	Frequency
Very High Impact	125
High Impact	0
Medium Impact	0
Low Impact	0
Very Low Impact	0

Appendix C: Terms Used

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 D: FAQs

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.

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

- The products and configurations for this test were selected by [redacted]
- The test was sponsored by [redacted]
- The test was conducted between 18th March and 24th April 2019.
- All products had full internet access and were confirmed to have access to any required or recommended back-end systems.
- Threats and legitimate objects were independently located and verified by SE Labs.
- Targeted attacks were generated, verified and selected by SE Labs. They were created and managed by well-known, publicly and freely available tools. The choice of exploits and other techniques used was advised by public information about ongoing attacks. One notable source was the 2018 Data Breach Investigations Report from Verizon.
- SE Labs conducted this endpoint security testing on virtual systems.

SE LABS

SE Labs Report Disclaimer

1. The information contained in this report is subject to change and revision by SE Labs without notice.
2. SE Labs is under no obligation to update this report at any time.
3. SE Labs believes that the information contained within this report is accurate and reliable at the time of its publication, which can be found at the bottom of the contents page, but SE Labs does not guarantee this in any way.
4. All use of and any reliance on this report, or any information contained within this report, is solely at your own risk. SE Labs shall not be liable or responsible for any loss of profit (whether incurred directly or indirectly), any loss of goodwill or business reputation, any loss of data suffered, pure economic loss, cost of procurement of substitute goods or services, or other intangible loss, or any indirect, incidental, special or consequential loss, costs, damages, charges or expenses or exemplary damages arising his report in any way whatsoever.
5. The contents of this report does not constitute a recommendation, guarantee, endorsement or otherwise of any of the products listed, mentioned or tested.
6. The testing and subsequent results do not guarantee that there are no errors in the products, or that you will achieve the same or similar results. SE Labs does not guarantee in any way that the products will meet your expectations, requirements, specifications or needs.
7. Any trade marks, trade names, logos or images used in this report are the trade marks, trade names, logos or images of their respective owners.
8. The contents of this report are provided on an “AS IS” basis and accordingly SE Labs does not make any express or implied warranty or representation concerning its accuracy or completeness.