SE Labs

INTELLIGENCE-LED TESTING

EventTracker EDR

JANUARY 2019

PRIVATE

SE Labs tested **EventTracker's** endpoint security solution, which is designed to detect suspicious activity and remediate threats.

Systems protected by the **EventTracker** endpoint agent were exposed to a mixture of targeted attacks using well-established techniques and public web-based threats that were found to be live on the internet at the time of the test.

The results indicate how effectively **EventTracker's** service was at detecting and/or protecting against those threats in real-time.

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

TESTING TEAM
Thomas Bean
Dimitar Dobrev
Liam Fisher
Gia Gorbold
Pooja Jain
Ivan Merazchiev
Jon Thompson
Dave Togneri
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 is a member of the Microsoft Virus Information Alliance (VIA); the Anti-Malware Testing Standards Organization (AMTSO); and the Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG).

CONTENTS

Document version 1.0 Written 31st January 2019

INTRODUCTION

Assessing next-generation protection

Malware scanning is not enough. You have to hack, too.

The amount of choice when trialling or buying endpoint security is at an all-time high. It has been 36 years since 'anti-virus' first appeared and, in the last five years, the number of companies innovating and selling products designed to keep Windows systems secure has exploded.

And whereas once vendors of these products generally used non-technical terms to market their wares, now computer science has come to the fore. No longer are we offered 'anti-virus' or 'hacker protection' but artificial intelligence-based detection and response solutions. The choice has never been greater, nor has the confusion among potential customers.

While marketing departments appear to have no doubt about the effectiveness of their product, the fact is that without in-depth testing no-one really knows whether or not an Endpoint Detection and Response (EDR) agent can do what it is intended. Internal testing is necessary but inherently biased: 'we test against what we know'. Thorough testing, including the full attack chains presented by threats, is needed to show not only detection and protection rates, but response capabilities.

EventTracker asked SE Labs to conduct an independent test of its EDR agent, running the same tests as are used against some of the world's most established endpoint security solutions available, as well as some of the newer ones.

This report shows EventTracker's performance in this test. The results are directly comparable with the public SE Labs Enterprise Endpoint Protection (Oct – Dec 2018) report, **available here**.

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.

For specific build numbers, see ***Appendix C: Product versions*** on page 13.

EXECUTIVE SUMMARY			
Product Tested	Protection Accuracy Rating (%)	Legitimate Accuracy Rating (%)	Total Accuracy Rating (%)
EventTracker EDR	100%	93%	95%

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.

■ **EventTracker EDR was effective at handling general threats from cyber criminals…**

The agent was capable of handling public web-based threats such as those used by criminals to attack Windows PCs, tricking users into running malicious files or running scripts that download and run malicious files.

■ **… and targeted attacks were prevented in all cases.**

EventTracker EDR was also competent at blocking more targeted, exploit-based attacks. Compared to many well-established anti-malware products it provided superior protection (see ***6. Conclusions*** on page 12 for more details).

■ **False positives were not an issue for EventTracker EDR**

The agent was accurate when assessing legitimate applications and URLs. This is by no means normal in the wider market.

■ **How did the product rate overall?**

EventTracker EDR's excellent performance wins it a AAA award, putting it in the highest class of security products.

1. 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.

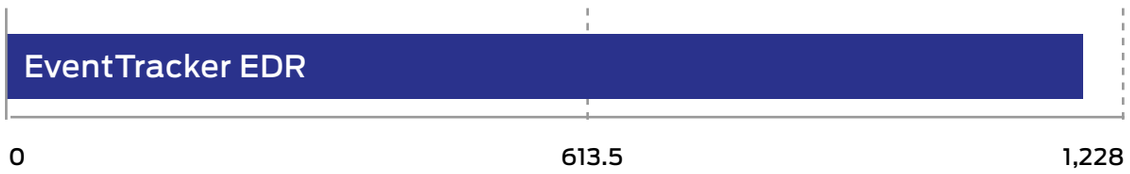
The graph below takes into account not only each product's ability to detect and protect against threats, but also its handling of non-malicious objects such as web addresses (URLs) and applications.

Not all protections, or detections for that matter, are equal. A product might completely block a URL, which 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. 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 ***5. Legitimate Software Ratings*** on page 9.

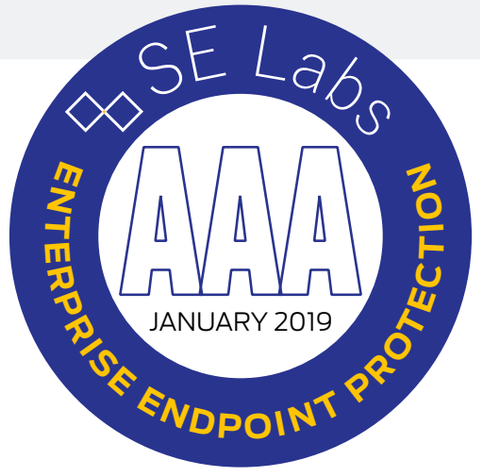
TOTAL ACCURACY RATINGS			
Product	Total Accuracy Rating	Total Accuracy (%)	Award
EventTracker EDR	1,166.5	95%	AAA



Total Accuracy Ratings combine protection and false positives.

Enterprise Endpoint Protection Awards

The following products win SE Labs awards:



EventTracker EDR

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.

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. 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 8 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.

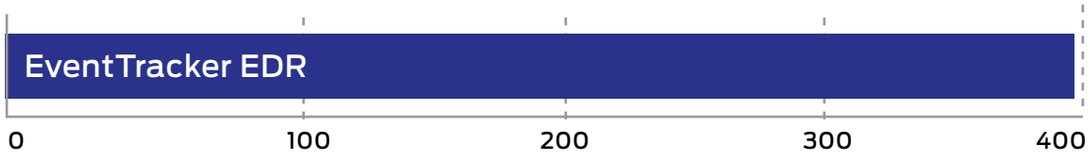
Action (-1)
If the attacker is able to exfiltrate a document from the target's Desktop of the currently logged in user then an 'action' has been successfully taken.

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.

2.1. Protection Ratings

PROTECTION RATINGS		
Product	Protection Accuracy	Protection Accuracy (%)
EventTracker EDR	398	100%



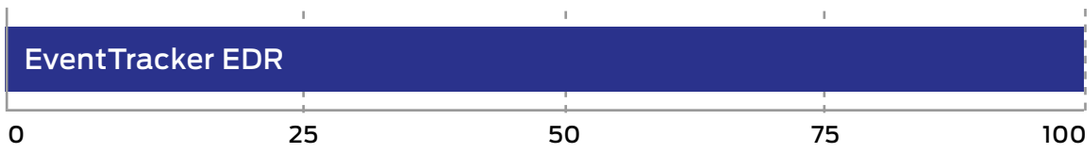
Protection Ratings are weighted to show that how products handle threats can be subtler than just 'win' or 'lose'.

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
EventTracker EDR	100%



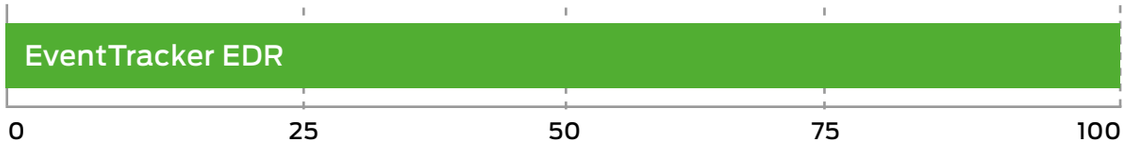
Protection Scores are a simple count of how many times a product protected the system.

4. Protection Details

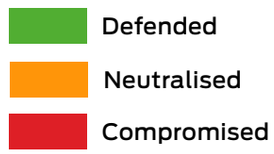
These results break down how each product handled threats into some detail. You can see how many detected a threat and the levels of protection provided.

Products sometimes detect more threats than they protect against. This can happen when they recognise an element of the threat but 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.

PROTECTION DETAILS					
Product	Detected	Blocked	Neutralised	Compromised	Protected
EventTracker EDR	100	100	0	0	100



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



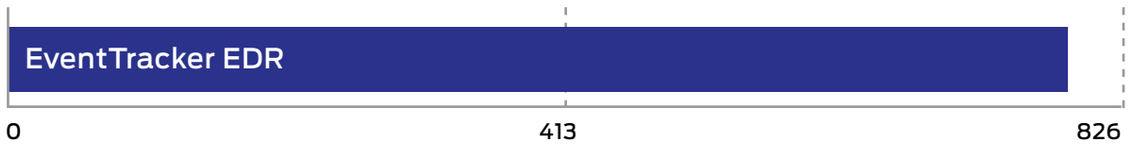
5. Legitimate Software Ratings

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.

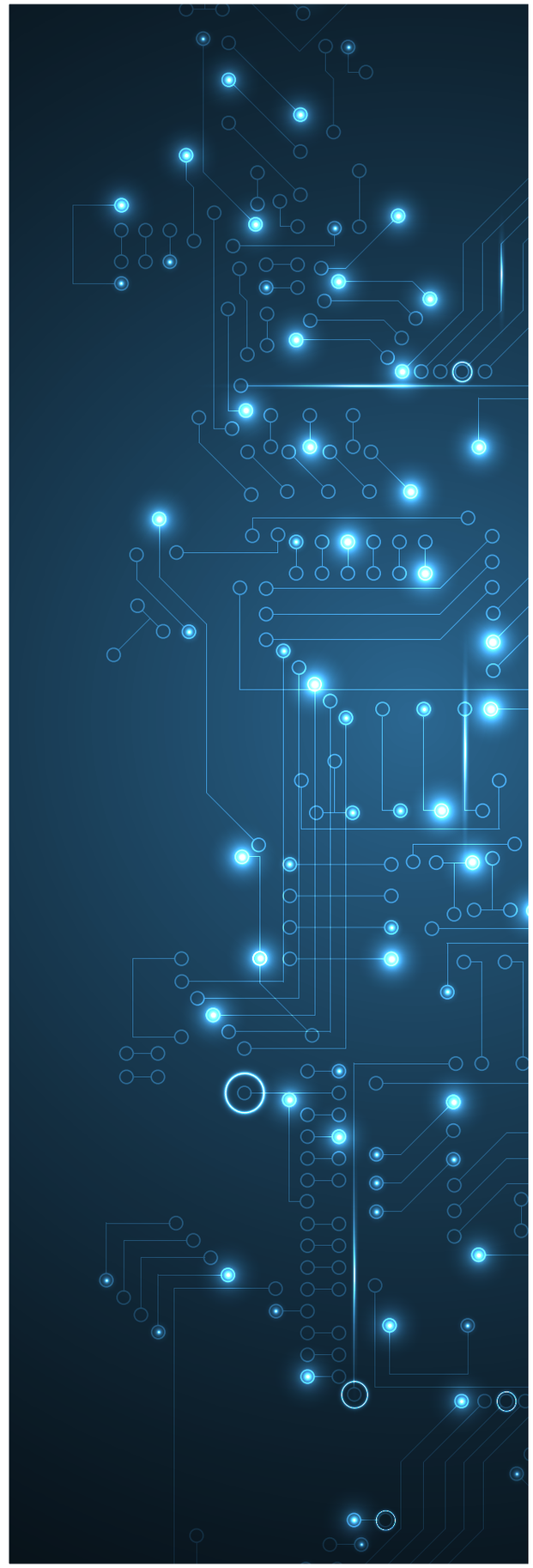
We also take into account the prevalence (popularity) of the applications and websites used in this part of the test, applying stricter penalties for when products misclassify very popular software and sites.

To understand how we calculate these ratings, see ***5.3 Accuracy Ratings*** on page 11.

LEGITIMATE SOFTWARE RATINGS		
Product	Legitimate Accuracy Rating	Legitimate Accuracy (%)
EventTracker EDR	768.5	93%



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	

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)	Click to allow/block (No Recommendation)
EventTracker EDR	91	9

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 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 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 9.

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	54
High Impact	22
Medium Impact	12
Low Impact	8
Very Low Impact	4
GRAND TOTAL	100

6. Conclusions

Attacks in this test included threats that affect the wider public and more closely-targeted individuals and organisations. You could say that we tested the products with 'public' malware and complete hacking attacks.

We introduced the threats in a realistic way, such that threats seen in the wild on websites were downloaded from those same websites, while threats caught spreading through email were delivered to our target systems as emails.

It is expected that most good security products would achieve good levels of protection against the threats used, as they are either prevalent public threats or targeted attacks created using free or inexpensive tools available to the general (but curious) public. We do not create unique malware, as such, although individual targeted attack files will almost certainly be 'unknown', in that their exact likeness won't have been seen by producers of security products.

EventTracker EDR did not disappoint and handled the majority of the targeted attacks well. It protected against 23 out of 25, while also stopping all of the 75 public attacks.

Some products are tuned aggressively to protect at the expensive of legitimate applications and are prone to generating False Positives (FPs) or, as we define them, Non-Optimal Classification/ Actions (NOCAs). FPs are when a product wrongly labels a legitimate application as being malicious, while a NOCA can include overly-sensitive classifications such as "risky", "potentially unwanted" or "suspicious". A NOCA might also block a file or force the user to choose whether or not to allow it. **EventTracker EDR** generated 9 NOCAs, but in such a way that users were alerted but not prevented in achieving their business goals.

EventTracker EDR was tested at the same time as a range of other competing products (some well-established, others newer) during the last few months of 2018. As such it is possible to compare its performance with those listed in the SE Labs Enterprise Endpoint Protection (Oct – Dec 2018) report, **available here**.

EventTracker EDR's performance earned it a AAA rating.

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 test was commissioned by EventTracker.
- The test was conducted between September and November 2018.
- The product was configured according to EventTracker's recommendations.
- Malicious URLs and legitimate applications were independently located and verified by SE Labs.
- Targeted attacks were selected and verified by SE Labs.
- Malicious and legitimate data was provided to EventTracker once the test was complete.
- SE Labs conducted this endpoint security test on physical PCs, not virtual machines.

APPENDIX C: Product Versions

The table below shows the service's name as it was being marketed at the time of the test.

PRODUCT VERSIONS			
Provider	Product Name	Build Version (Start)	Build Version (End)
EventTracker	EDR	9.0	9.0