360 Degree Assessment & Certification

Q1 2020

MRG Effitas
EFFICACY ASSESSMENT & ASSURANCE

MRG Effitas Ltd.
Efficacy Assessment & Assurance

MRG Effitas

Contents

MRG Effitas is a world-leading, independent IT security efficacy testing & assurance company. We are trusted by antimalware vendors across the world.

TEL:
+44 (0)20 3239 9289

EMAIL:
contact@mrg-effitas.com

TWITTER:
@mrgeffitas

Introduction

MRG Effitas is a world leader independent IT research company having a core focus on AV efficacy assessments both in the traditional “Real World” malware detection capabilities and in the anti–financial fraud area.

The methodology employed in this test maps closely to Real World practice representing the valid threads endangering anyone using Windows computers. This evaluation is aimed to help users choosing the most suitable security application.

This Programme is called “360 Assessment & Certification” since it tests the security applications capabilities with a full spectrum of attack vectors. In the 360 Assessment, trojans, backdoors, spyware, financial malware, ransomware and “other” malicious applications are all used. Alongside the usual In-The-Wild (ITW) file-based attacks, evaluation also contains scenarios where fileless cases and exploitation techniques are applied.

Besides the malicious attacks, in order to evaluate the practical accuracy of AV products, they were exposed to potentially unwanted applications (PUA or Greyware) and clean files (FP) as well.

Additionally, besides security capabilities tests, our assessment measured the footprint each security software on a computer’s performance.

Executive Summary

This Certification Programme is designed to serve as a reflection of product efficacy based on what we have previously termed “metrics that matter”.

Based on decades of experience in IT security, our previous tests, and being one of the world’s largest supplier of early-life malicious binaries and URLs, we know that all endpoints can be infected, regardless of the security solutions employed. The question is not ‘if’ but ‘when’ a malicious binary hits the system.

A security product’s ability to block initial infection (although critical in most cases) is not the only metric that matters. Measuring the time taken to detect malicious files or actions, is another metric that can be crucial in evaluation. An additional key factor is the point in time when the fact of the infection and any associated malicious behaviour are detected.

When conducting these tests, we tried to simulate normal user behaviour. We are aware that a “Real World” test cannot be conducted by a team of professionals inside a lab, because we understand how certain types of malware work, how malware attacks are conducted and how such attacks could be prevented. Simulating normal user behaviour means that we paid special attention to all alerts given by security applications. A pass was given only when alerts were straightforward and clearly suggested that some malicious action should be blocked.

With these in mind, it is very important to note that the best choice for an average user is to keep things as simple as possible and not to overwhelm the non-tech savvy with cryptic pop-ups, alerts or questions.

Out of twelve tested security products, eleven managed to meet the specification to attain our Q1 2020 360 Degree Certification, these being:

- Avast Business Antivirus
- Avira Antivirus Pro
- Bitdefender Endpoint Security
- CrowdStrike Falcon Protect Sensor
- ESET Endpoint Security
- F-Secure Computer Protection Premium
- Kaspersky Small Office Security
- McAfee Endpoint Security
- Microsoft Windows Defender
- Symantec Endpoint Protection Cloud
- Sophos Intercept X
- Trend Micro Security

Certification

In order to attain a quarterly MRG Effitas 360 Degree Level 1 certification, a security application must completely protect the system from initial infection either by automatically blocking every ITW sample, or by blocking them based on their behaviour, prior to any malicious actions. (PUA, FP, Exploit/Fileless, and performance tests are not part of the certification.)

Level 2 certification is given if the application blocks or detects any initially missed malware in at least 98% of all cases on the 24-hour retest, while the initially missed test cases are less than 10%. If a ransomware/wiper successfully runs and files are not available anymore, Level 2 certification is lost.

Under the MRG Effitas 360 Degree Assessment & Certification, the following products were certified for Q1 2020.

Certified (Level 1):

- Bitdefender Endpoint Security
- ESET Endpoint Security
- Symantec Endpoint Protection Cloud

Certified (Level 2):

- Avast Business Antivirus
- Avira Antivirus Pro
- CrowdStrike Falcon Protect
- F-Secure Computer Protection Premium
- Kaspersky Small Office Security
- McAfee Endpoint Security
- Microsoft Windows Defender
- Sophos Intercept X
- Trend Micro Security

The Purpose of this Report

Since its inception in 2009, MRG Effitas has strived to differentiate itself from traditional testing houses by having its primary focus on providing “efficacy assessments” and not just performing “tests”.

Traditionally, testing of security software has been aimed at measuring a product’s ability to detect malware. Testing has evolved rapidly over the last couple of years, as most labs, under the direction of AMTSO (of which MRG Effitas is a member) has been striving to conduct “Real World” testing, based on standardised guidelines. More information about the compliance status of this test can be found on the AMTSO website.

https://www.amtso.org/amtso-ls1-tp020

Although there is no absolute definition of this kind of testing, loosely speaking, it involves the introduction of malware to an endpoint through a realistic entry point, such as downloading the sample using a browser or getting it from a USB memory stick. Real world testing mostly involves “dynamic testing” (i.e. the malware is executed and then the ability of the security product to block the malware is measured).

Whilst these types of tests are useful, yielding valid and meaningful data, MRG Effitas wanted to merge standalone tests and also go the extra mile by measuring the time security products take to detect infections and remediate the endpoint.

To make testing more akin to real world scenarios, no manual scanning was conducted. Instead, the system was retested exactly 24 hours after the system was compromised, thereby giving security applications the opportunity to detect infections on restart.

As we have stated in our previous test reports, most malware has one primary objective, and that is to make money for the cybercriminals.

Measuring initial detection rates and the time taken to detect active malware is important, particularly in today’s threat landscape with the mix of malware that is prevalent.

As we have repeated in our previous financial malware test reports, the longer a cybercriminal can run their malware on a system, the greater the opportunity is for them to be able to capture private user information, including banking passwords and social media credentials, etc.

There has been a recent increase in the prevalence of targeted ransomware, which, once active on the system, holds the user at ransom to decrypt system data or unlock the system.

For these types of malware, initial detection is of the utmost importance, since the vast majority of security solutions will be unable to remediate the problem of an encrypted system. In incident response scenarios, it is usually advised to purchase bitcoin right away, should the experts be unable to retrieve the encrypted files.

In providing these quarterly certifications, the MRG Effitas 360 Assessment & Certification Programme is the de facto standard by which security vendors, financial institutions and other corporations can attain the most rigorous and accurate determination of a product’s efficacy against the full spectrum of malware that is prevalent during the period.

Tests Employed

In this assessment (Q1 2020), we ran the following tests:

In the Wild 360 / Full Spectrum Test

Most of the malicious URLs used in this test were compromised legitimate websites, serving malware. We believe that such URLs pose the greatest danger to users, as this is the place where they least expect to get infected, and any URL based protection fails on them. Some URLs originate from our honeypots, or in case of ransomware and financial malware in particular, we used URLs from newly discovered distribution sites.

Malware delivered by URLs used in this test can be considered as zero-day in the true meaning of the phrase. This posed a significant challenge to the participant products.

~10% of the threats used in this test were introduced to the system via internal webmail sites. We have witnessed many SMBs being infected through internal webmails and lack of spam filtering. Downloading malware attachments from internal webmail sites bypass the URL blocking features of the products, and this happens in-the-wild.

During the In the Wild 360 / Full Spectrum test, 398 live ITW samples were used. The stimulus load comprised the following: 111 trojans, 70 backdoors, 70 financial malware samples, 8 ransomware samples, and 101 spyware, 19 malicious document, 3 spam emails, 11 malicious script files and 5 others.

PUA / Adware Test

The PUA samples used in this test are deceptive, or potentially unwanted applications (PUA), that are not malicious, but are generally considered unsuitable for most home or business networks. They usually contain adware, installs toolbars or have other unclear objectives. They may also contribute to consuming computing resources or network bandwidth. PUAs can be deceptive, harmful, hoax, show aggressive popups and misleading or scaring the user. They may provide some unconventional ways of uninstalling the application, maybe retain some of their components on the device without the user's consent. We mainly use a filtered version of AppEsteem's feed, as they developed deceptor requirements as part of a cross-industry effort of many of the world's leading security companies and represent a minimum bar that all apps and services must meet to avoid being titled deceptive.

AppEsteem, as a member of the AMTSO group is dedicated to help protecting consumers from harassing and objectionable material, and to enable security companies to restrict access to such actions. MRG Effitas, as a member of the AMTSO group, is also dedicated to protecting these thoughts.

In the PUA/Adware part we tested the products against 20 PUAs.

Exploit/Fileless Test

The main purpose of this test is to see how security products protect against a specific exploitation technique. In order to measure this, we developed test cases that simulate the corresponding exploit and post-exploitation techniques only.

Drive-by download exploits are the biggest threats for an enterprise environment, since no user interaction is needed to start the chain of infection on a victim machine. Outdated browsers and Office environments are widespread in enterprise environments, due to compatibility issues or the lack of proper patch management process.

We were testing the products' abilities to avoid any exposure to adversaries, to interrupt malicious payload delivery before performing malicious actions. We focus explicitly on each product's ability to mitigate each attack technique. The results are not intended to evaluate the complete efficacy of the products, but rather the products' anti-exploit and anti-post-exploit features in isolation.

During this test we used 10 different exploitation techniques. The detailed description can be found in the 'Appendix'.

False Positive Test

Perfect blocking of malicious content is only part of the story from a practical point of view for any decent AV product. In many cases all malware blocking is a result of a very aggressive filter which can block non-malicious legitimate applications as well prohibiting everyday work by blocking legitimate, perhaps newly developed in-house software.

In order to test this feature, we tested the security applications against completely clean, recently created applications.

False positive assessment consisted of 1186 clean and legitimate application samples. The selection has been focused on applications, frequently found in enterprise environments (drivers, media editors, developer tools, etc.)

Performance Test

A security product's usefulness does not depend on protection level solely, but also on its resource footprint and its effect of the overall operating system performance.

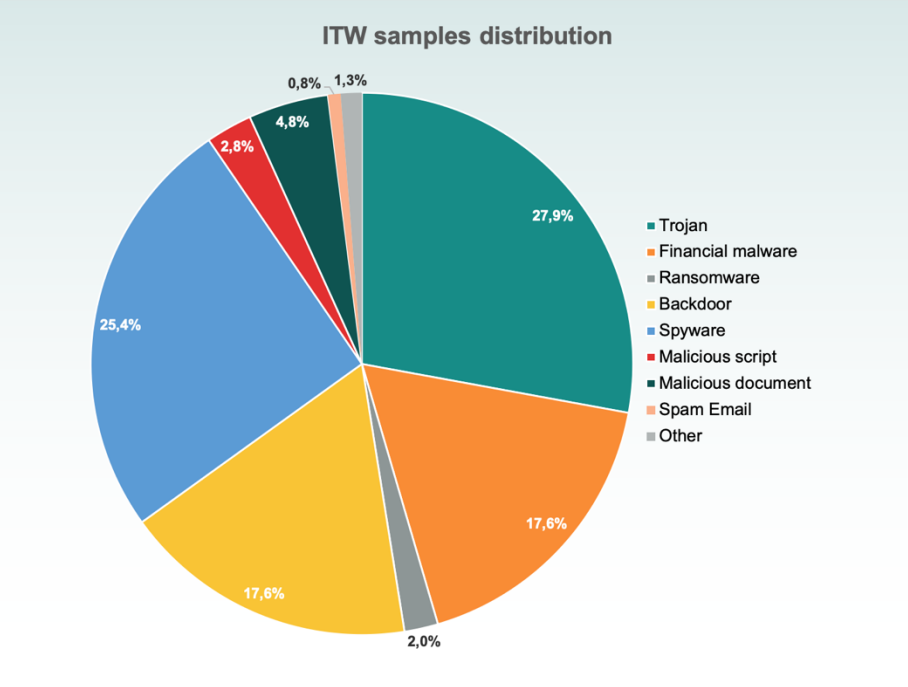
In order to assess the products' influence on the operating system, we tested several performance factors on a physical machine and combined the results, based on a scoring approach. Detailed information can be found in the 'Appendix'.

In every test case, (except for the performance test) our testing environment supports the execution of VM-aware malware, this is the reason why we were able to use more sophisticated threats which normally would not run on Virtual Machines.

Security Applications Tested

- Avast Business Antivirus 20.1.2581
- Avira Antivirus Pro 1.0.23.8081
- Bitdefender Endpoint Security 6.6.17.243
- CrowdStrike Falcon Protect Sensor 5.28.11009.0
- Microsoft Windows Defender 4.18.1911.3
- ESET Endpoint Security 7.1.2045.5
- F-Secure Computer Protection Premium 20.1
- Kaspersky Small Office Security 19.0.0.1088(l)
- McAfee Endpoint Security 10.7.0.1285
- Sophos Intercept X 2.0.16
- Symantec Endpoint Protection Cloud 22.19.8.65
- Trend Micro Security 6.7.1185/14.2.1108

Malware sample types used to conduct the tests

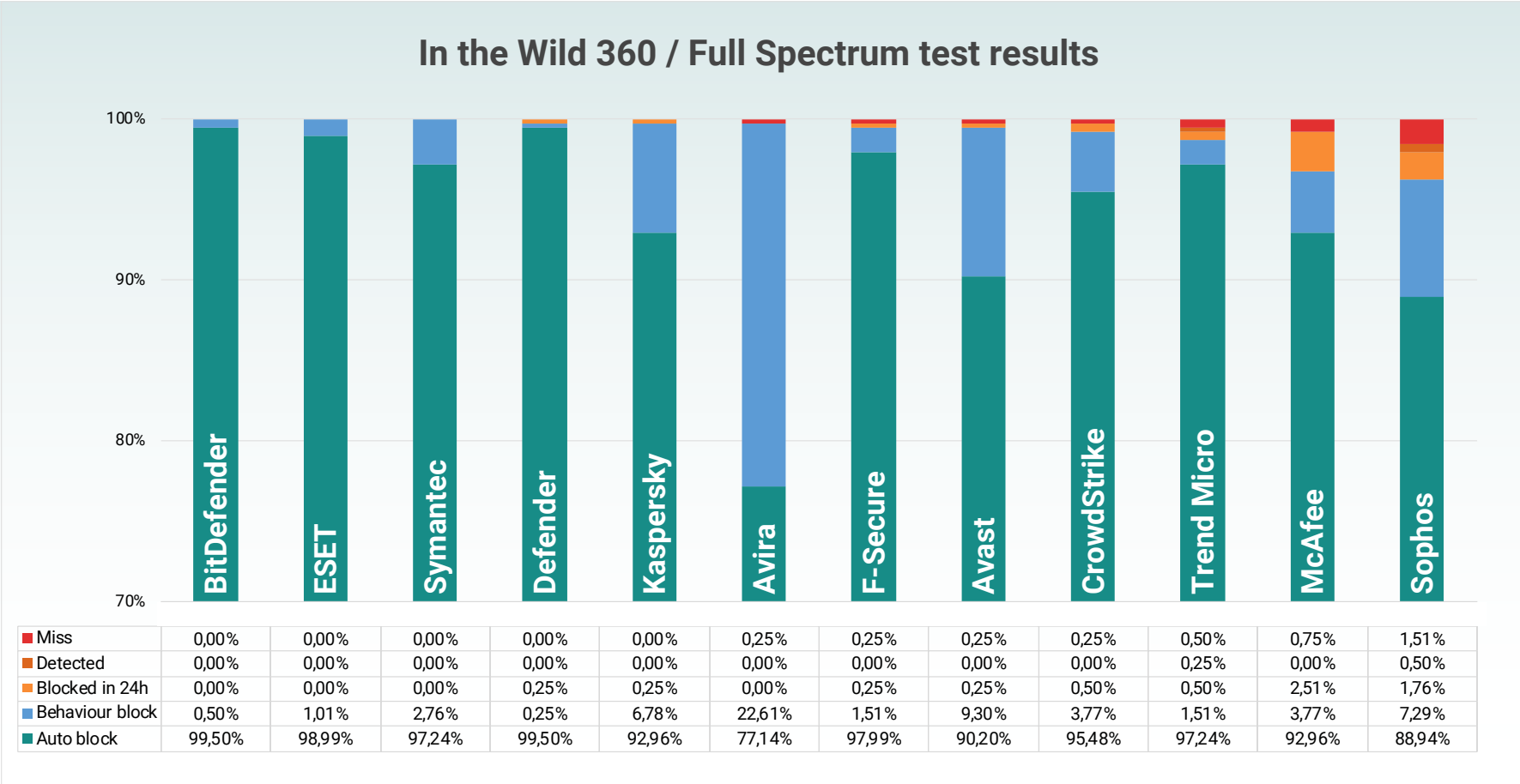


Test Results

The tables below show the results of testing under the MRG Effitas 360 Q1 2020 Assessment Programme.

Q1 2020 In the Wild 360 / Full Spectrum test results

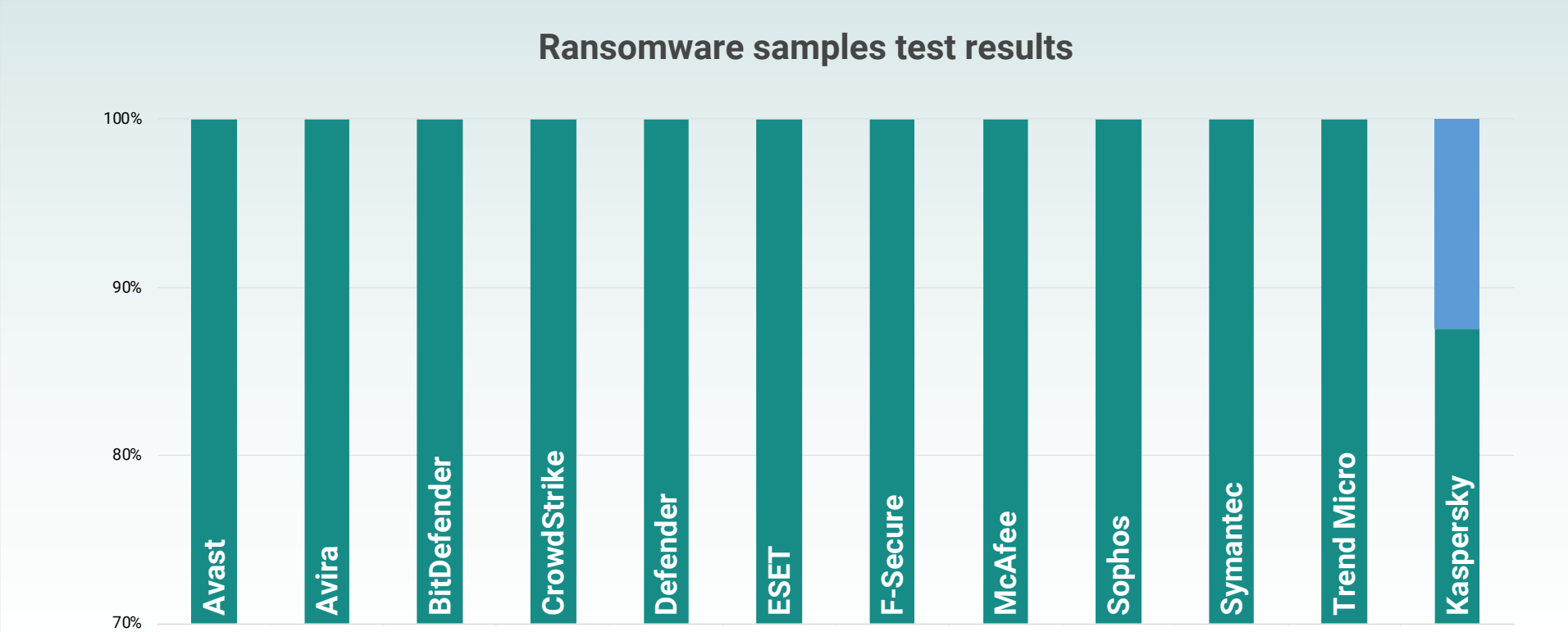
The table below shows the detection rates of the security products for 398 ITW samples. This table is sorted by smallest number of missed samples.



	BitDefender	ESET	Symantec	Defender	Kaspersky	Avira	F-Secure	Avast	CrowdStrike	Trend Micro	McAfee	Sophos
Miss	0,00%	0,00%	0,00%	0,00%	0,00%	0,25%	0,25%	0,25%	0,25%	0,50%	0,75%	1,51%
Detected	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,25%	0,00%	0,50%
Blocked in 24h	0,00%	0,00%	0,00%	0,25%	0,25%	0,00%	0,25%	0,25%	0,50%	0,50%	2,51%	1,76%
Behaviour block	0,50%	1,01%	2,76%	0,25%	6,78%	22,61%	1,51%	9,30%	3,77%	1,51%	3,77%	7,29%
Auto block	99,50%	98,99%	97,24%	99,50%	92,96%	77,14%	97,99%	90,20%	95,48%	97,24%	92,96%	88,94%

Ransomware samples test results

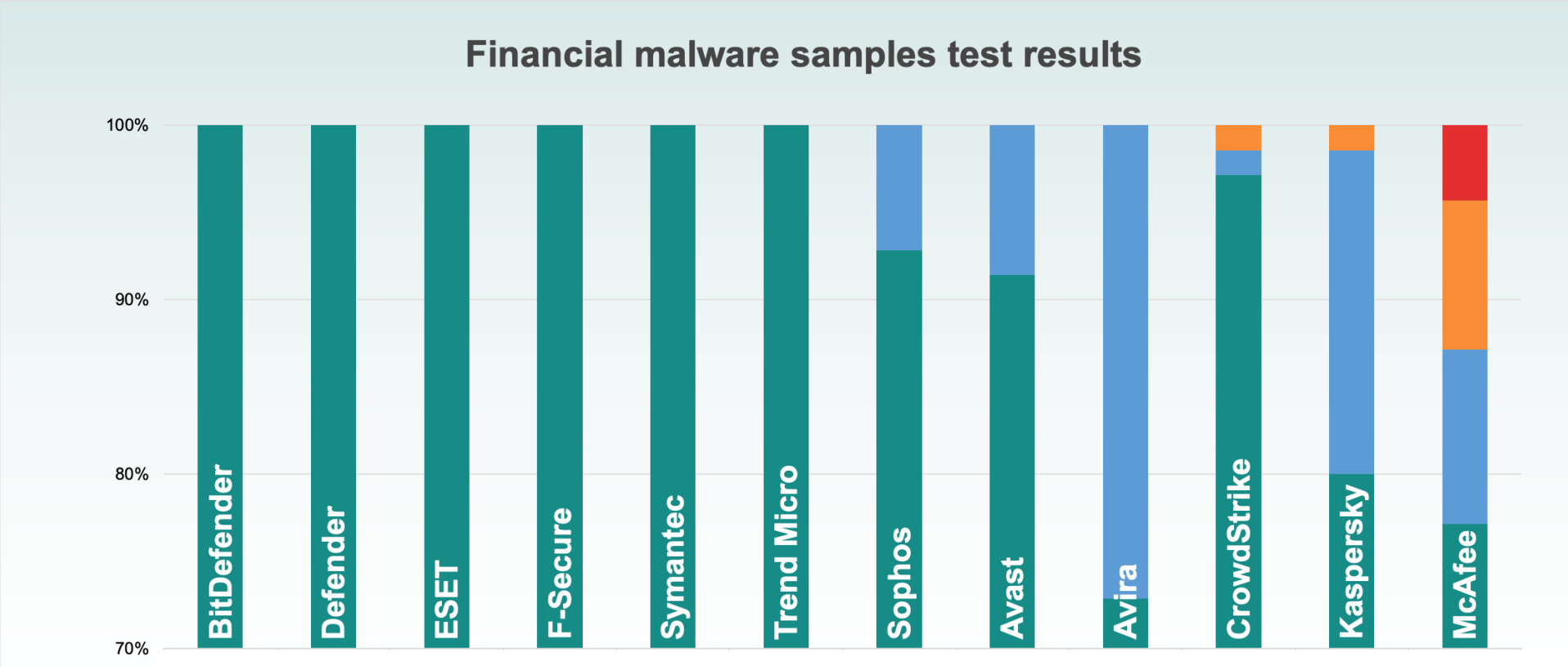
The table below shows the detection rates of the security products for 8 ransomware samples. This table is sorted by smallest number of missed samples.



	Avast	Avira	BitDefender	CrowdStrike	Defender	ESET	F-Secure	McAfee	Sophos	Symantec	Trend Micro	Kaspersky
Miss	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Blocked in 24h	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Behaviour block	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	12,50%
Auto block	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	87,50%

Financial malware samples test results

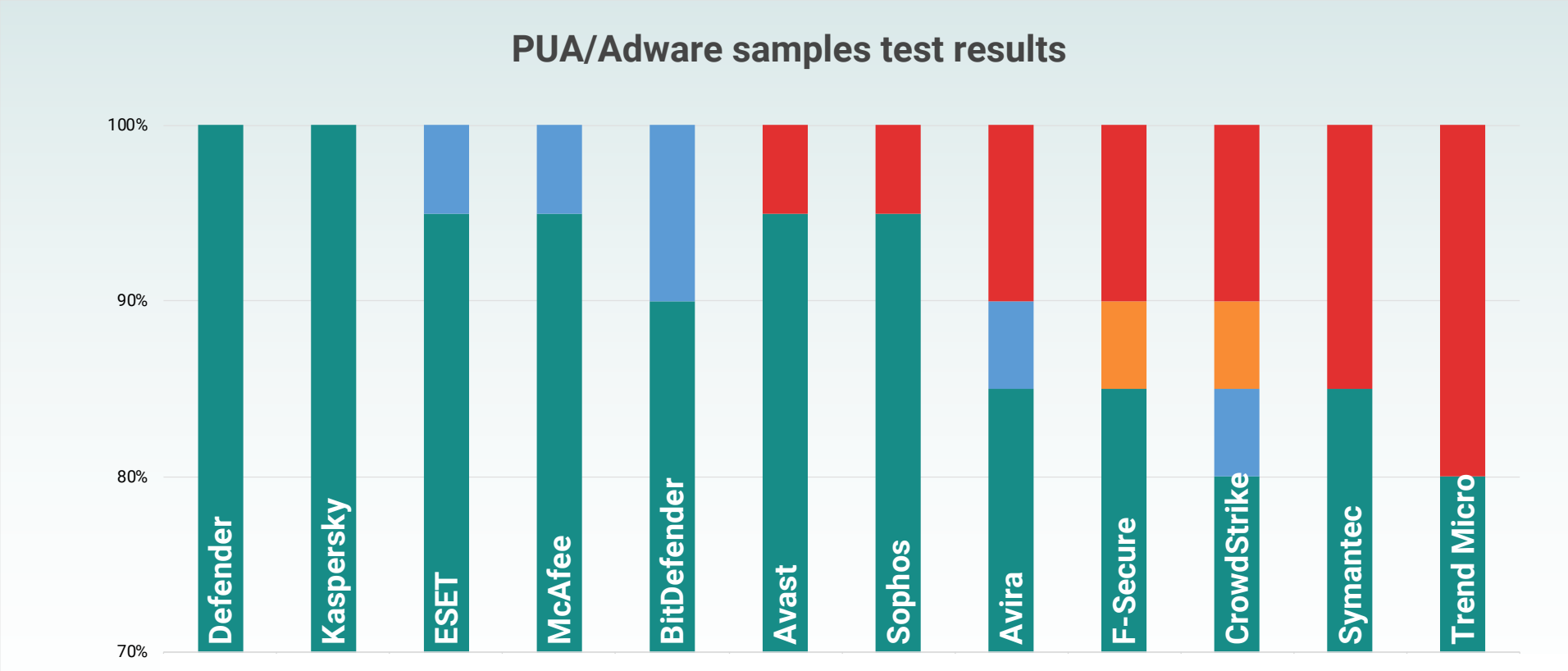
The table below shows the detection rates of the security products for 70 financial malware samples. This table is sorted by smallest number of missed samples.



	BitDefender	Defender	ESET	F-Secure	Symantec	Trend Micro	Sophos	Avast	Avira	CrowdStrike	Kaspersky	McAfee
Miss	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	4,29%
Blocked in 24h	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	1,43%	1,43%	8,57%
Behaviour block	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	7,14%	8,57%	27,14%	1,43%	18,57%	10,00%
Auto block	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	92,86%	91,43%	72,86%	97,14%	80,00%	77,14%

PUA/adware samples test results

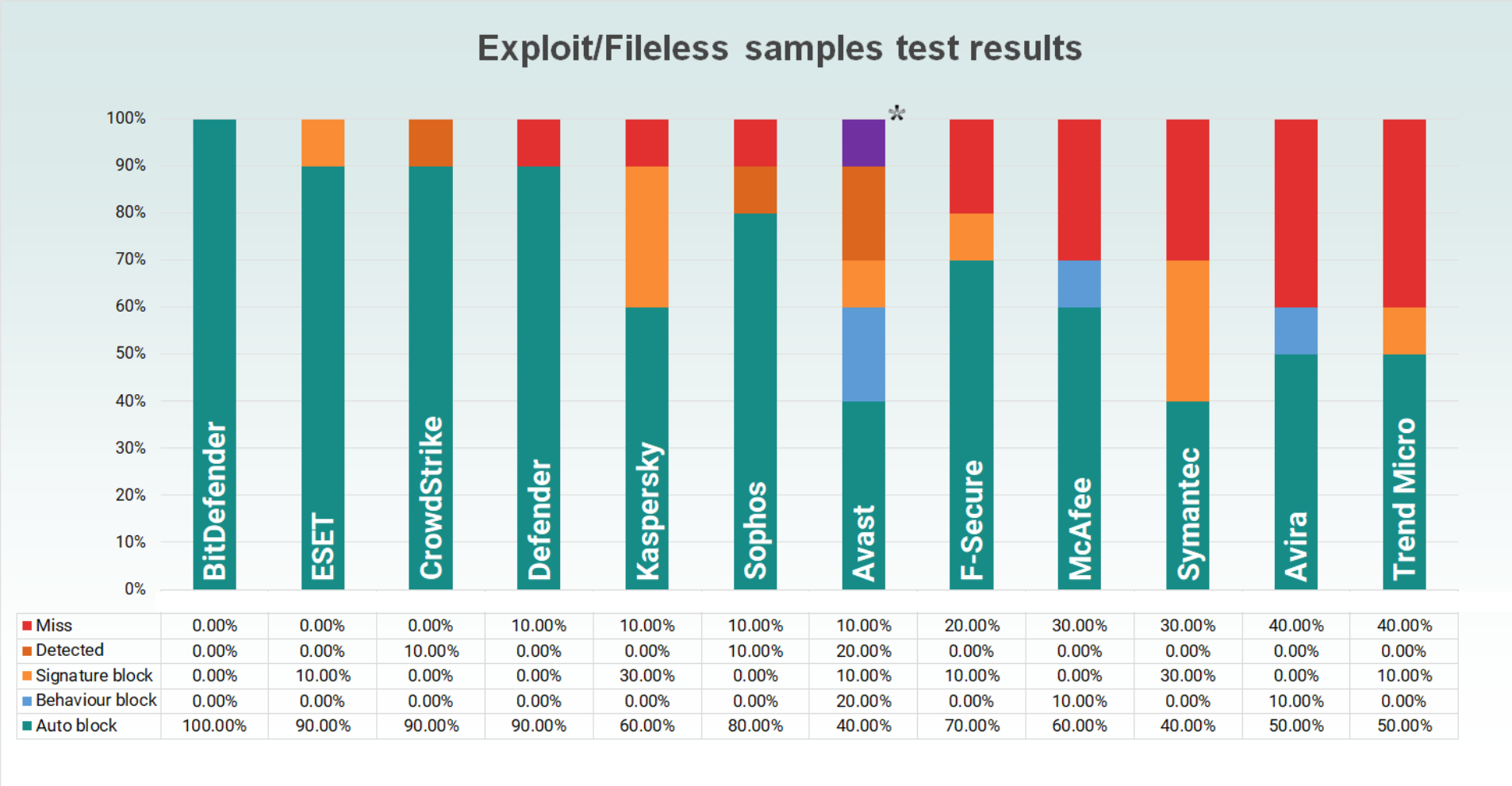
The table below shows the detection rates of the security products for 20 PUA/Adware samples. This table is sorted by smallest number of missed samples.



	Defender	Kaspersky	ESET	McAfee	BitDefender	Avast	Sophos	Avira	F-Secure	CrowdStrike	Symantec	Trend Micro
Miss	0,00%	0,00%	0,00%	0,00%	0,00%	5,00%	5,00%	10,00%	10,00%	10,00%	15,00%	20,00%
Blocked in 24h	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	5,00%	5,00%	0,00%	0,00%
Behaviour block	0,00%	0,00%	5,00%	5,00%	10,00%	0,00%	0,00%	5,00%	0,00%	5,00%	0,00%	0,00%
Auto block	100,00%	100,00%	95,00%	95,00%	90,00%	95,00%	95,00%	85,00%	85,00%	80,00%	85,00%	80,00%

Exploit/fileless samples test results

The table below shows the initial detection rates of the security products for 10 exploit/fileless test. This table is sorted by smallest number of missed attack vectors.

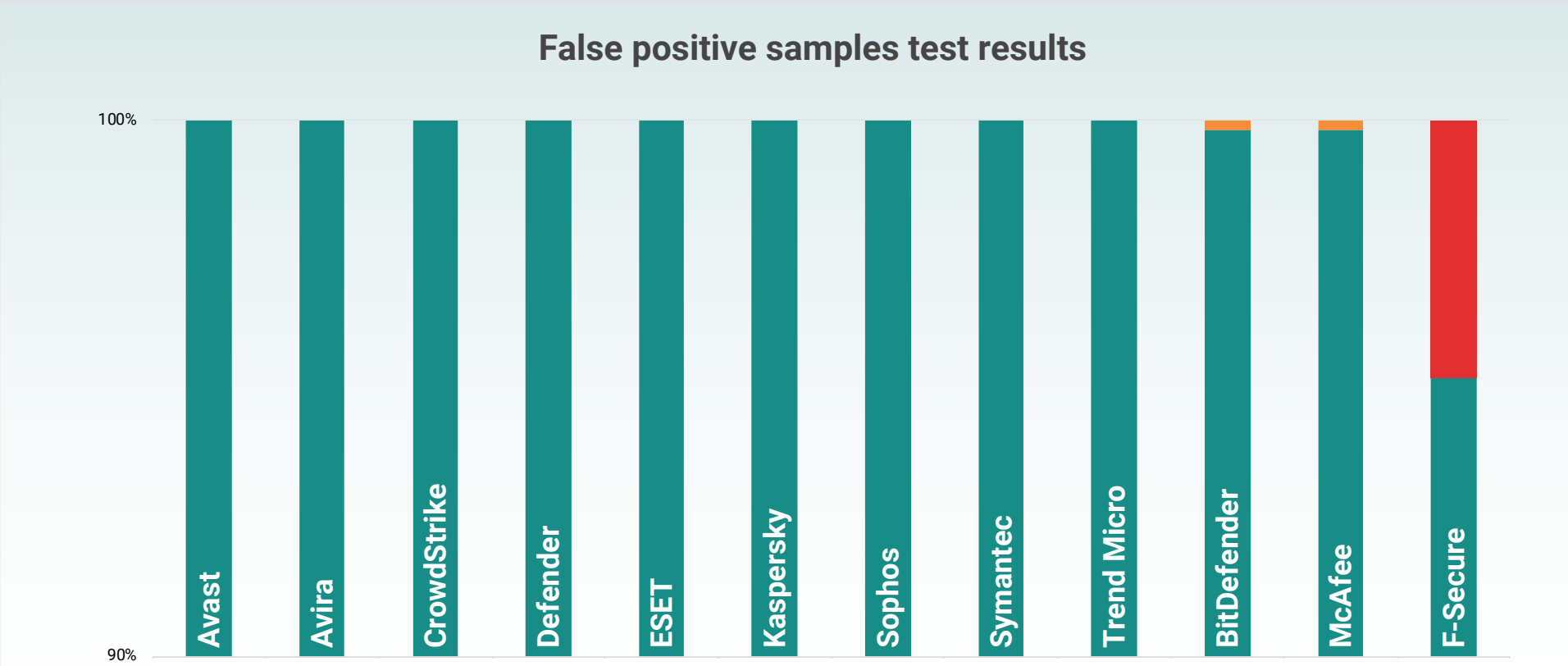


	BitDefender	ESET	CrowdStrike	Defender	Kaspersky	Sophos	Avast	F-Secure	McAfee	Symantec	Avira	Trend Micro
Miss	0.00%	0.00%	0.00%	10.00%	10.00%	10.00%	10.00%	20.00%	30.00%	30.00%	40.00%	40.00%
Detected	0.00%	0.00%	10.00%	0.00%	0.00%	10.00%	20.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Signature block	0.00%	10.00%	0.00%	0.00%	30.00%	0.00%	10.00%	10.00%	0.00%	30.00%	0.00%	10.00%
Behaviour block	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%	0.00%	10.00%	0.00%	10.00%	0.00%
Auto block	100.00%	90.00%	90.00%	90.00%	60.00%	80.00%	40.00%	70.00%	60.00%	40.00%	50.00%	50.00%

*Missed sample #008 (False positive case) indicated with * sign and purple colour is disputed by Avast because in their point of view, executing any binary file via an Office macro should be considered as malicious action.*

False positive samples test results

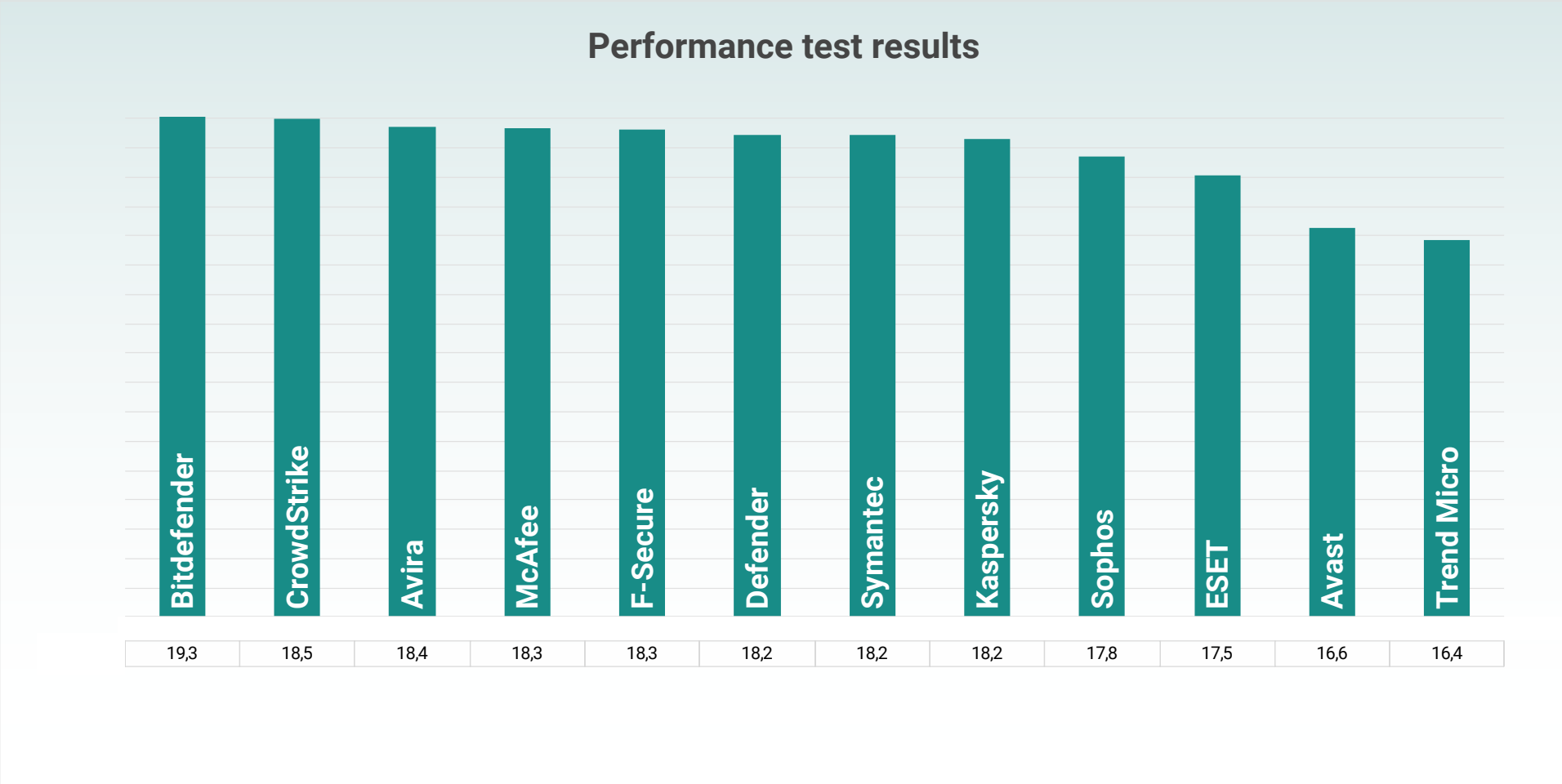
The table below shows the initial detection rates of the security products for 1186 false positive samples. This table is sorted by smallest number of false positive sample blocks.



	Avast	Avira	CrowdStrike	Defender	ESET	Kaspersky	Sophos	Symantec	Trend Micro	BitDefender	McAfee	F-Secure
False block	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	4,81%
Detected	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,17%	0,17%	0,00%
Allowed to run	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	99,83%	99,83%	95,19%

Performance test results

This table is sorted from highest to lowest score where the highest score denotes the lowest impact on the system.



Scoring details can be found in the 'Appendix'.

Detailed results of the Performance test

The table below shows the detailed results of the performance test of the security products. This table is sorted alphabetically.

	Windows 10 Base	Avast	Avira	Bitdefender	CrowdStrike	Defender	Eset	F-Secure	Kaspersky	McAfee	Sophos	Symantec	Trend Micro
Install time (s)	n/a	98,0	150,0	252,0	158,0	n/a	41,0	115,0	79,0	219,0	381,0	162,0	475,0
Bootup time (s)	32,2	40,0	40,0	42,1	40,9	32,9	41,3	36,7	36,5	58,5	53,5	44,9	51,6
Firefox startup time (s)	1,3	1,9	1,7	1,6	1,8	1,3	1,5	1,4	1,7	1,7	1,8	1,7	2,8
10 minutes of idling													
CPU usage (%)	0,3	1,7	0,9	1,0	0,7	0,7	0,8	5,8	0,9	0,7	2,1	0,9	3,5
Memory usage (Mb)	0 (Reference)	126,4	61,8	347,3	60,2	102,0	48,2	300,9	233,8	311,6	422,2	213,7	283,7
Physical disk usage (%)	0,4	1,6	0,8	0,9	0,8	1,0	1,1	1,2	1,3	0,8	1,7	1,0	1,7
Network interface usage (B/s)	208,2	792,1	446,9	370,4	458,2	627,3	754,7	422,8	643,3	189,9	651,4	697,1	793,6
Security software update													
Time (s)	n/a	42,7	38,0	120,3	n/a	37,7	22,7	41,0	62,0	142,7	81,3	35,0	30,3
CPU usage (%)	n/a	35,6	42,5	32,9	n/a	44,5	44,1	37,9	30,1	33,8	28,0	42,3	39,5
Memory usage (Mb)	n/a	369,7	524,2	530,3	n/a	118,8	153,2	500,9	300,3	394,3	518,8	333,7	432,2
Physical disk usage (%)	n/a	29,5	19,3	11,6	n/a	14,2	93,5	22,6	8,2	5,8	9,9	17,0	15,3
Network interface usage (B/s)	n/a	99366,7	18359,1	23970,5	n/a	166029,0	65760,2	71033,4	23236,1	183684,3	17502,8	85874,8	185177,9
Security software scanning - C:\													
Time (s)	n/a	269,3	432,7	99,3	n/a	353,7	108,7	73,7	113,7	1206,3	326,0	127,0	790,3
CPU usage (%)	n/a	20,9	28,8	33,2	n/a	90,9	24,5	80,0	58,4	96,2	29,3	83,6	33,7
Memory usage (Mb)	n/a	634,8	513,6	691,0	n/a	381,7	245,3	524,2	373,6	837,7	802,6	717,0	685,0
Physical disk usage (%)	n/a	31,5	17,6	21,4	n/a	30,6	31,6	28,6	28,0	6,0	9,4	15,4	11,3
Network interface usage (B/s)	n/a	2643,5	6330,1	1478,5	n/a	1009,1	766,0	5132,5	5011,5	943,2	1261,8	1200,8	4080,3
Security software size on disk													
Just after install (Mb)	n/a	1180,6	872,8	989,4	33,2	n/a	1082,3	604,4	824,1	738,6	1696,6	599,2	650,6
After usage (at least: 30 mins idling, 3 scans, 3 updates)	n/a	1167,1	675,6	1037,7	38,3	288,4	1148,7	791,4	1085,6	1289,1	1740,6	649,7	785,5

Understanding Grade of Pass

Level 1

All threats detected on first exposure or via behaviour protection.

- Bitdefender Endpoint Security
- ESET Endpoint Security
- Symantec Endpoint Protection Cloud

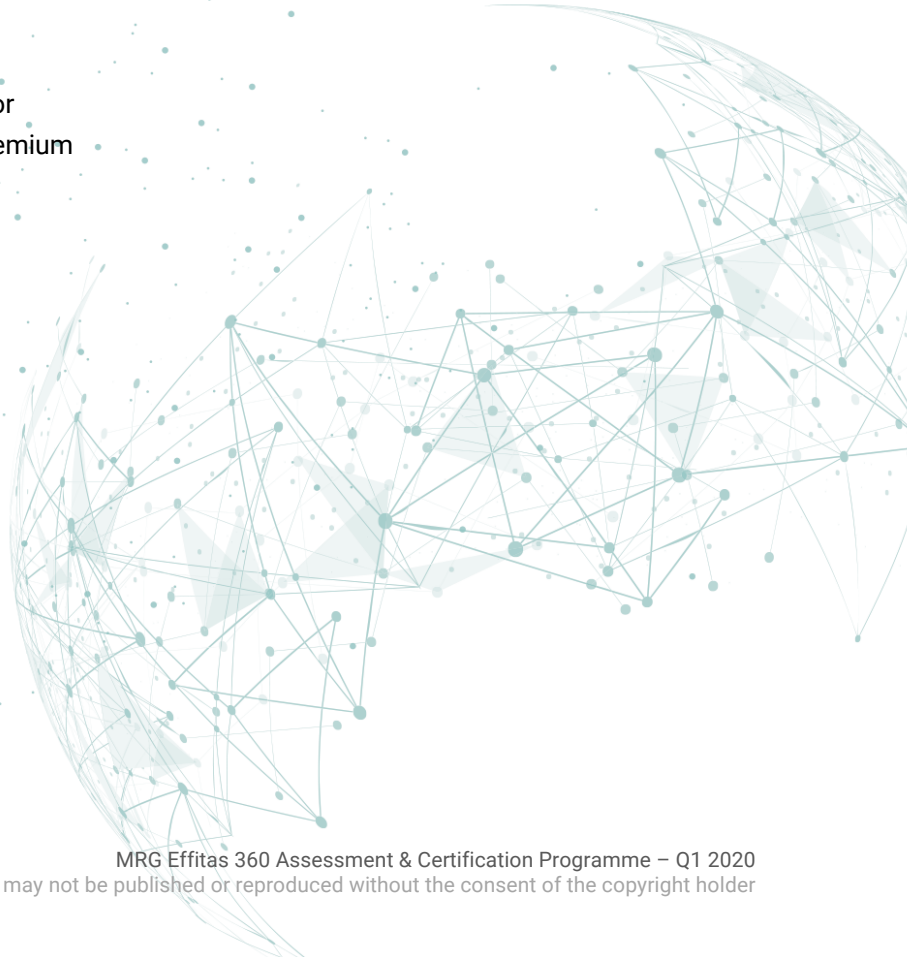
Level 2

At least 98% of the threats detected and neutralised / system remediated before or on the first rescan while the initially missed test cases are less than 10%.

- Avast Business Antivirus
- Avira Antivirus Pro
- CrowdStrike Falcon Protect Sensor
- F-Secure Computer Protection Premium
- Kaspersky Small Office Security
- McAfee Endpoint Security
- Microsoft Windows Defender
- Sophos Intercept X
- Trend Micro Security

Failed

Security product failed to detect all infections or at least 98% of them and remediate the system during the test procedure.



Appendix 1

Methodology used in the “In the Wild 360 / Full Spectrum” and the PUA tests

1. Windows 10 64-bit operating system was installed on a hardened virtual machine, all updates are applied, and third-party applications installed and updated.
2. An image of the operating system was be created.
3. A clone of the imaged systems was made for each of the security applications used in the test.
4. An individual security application was installed using default settings on each of the systems created in (3) and then, where applicable, updated. If the vendor provided a non-default setting, this setting was checked whether it was realistic. If yes, the changes were documented, applied, and added in the report in an appendix (if any).
5. A clone of the system as at the end of (4) was created.
6. Each live URL test was conducted by the following procedure.
    a. Downloading a single binary executable (or document, script, etc.) from its native URL using Google Chrome to the Downloads folder and then executing the binary from the browser.
    b. Either the security application blocked the URL where the malicious binary was located.
        i. Or the security application detected and blocked the malicious binary whilst it was being downloaded to the desktop.
        ii. Or the security application detected the malicious binary when it was executed according to the following criteria: It identified the binary as being malicious and either automatically blocked it or postponed its execution and warned the user that the file was malicious and awaited user input.
7. The system under test was deemed to have been infected if the security application failed to detect or block the binary at any stage in (6) and allowed it to be executed.
8. The test case was retested 24 hours after the initial test if the security application failed to detect or block the malicious binary.
9. Tests are conducted with all systems having internet access.
10. As no user-initiated scans was involved in this test, applications rely on various technologies to detect, block and remediate threats. Some of these technologies were: URL blacklist, reputation, signature, machine learning, heuristics, behaviour etc.

Methodology used in the false positive test

1. Windows 10 64-bit operating system was installed on a hardened virtual machine, all updates are applied, and third-party applications installed and updated.
2. An image of the operating system was be created.
3. A clone of the imaged systems was made for each of the security applications used in the test.
4. An individual security application was installed using default settings on each of the systems created in (3) and then, where applicable, updated. If the vendor provided a non-default setting, this setting was checked whether it was realistic. If yes, the changes were documented, applied, and added in the report in an appendix (if any).
5. A clone of the system as at the end of (4) was created.
6. Each false positive test case was conducted by the following procedure.
    a. Copying the binary executable from an external drive to the Desktop
    b. Executing the binary.
7. The test case is marked as a False Positive block if the security application detects or blocks the binary at any stage in (6).
8. The test case was retested 24 hours after the initial test if the security application blocked the binary.
9. Tests are conducted with all systems having internet access.

Methodology used in the exploit/fileless test – in-the-wild exploits

1. One default install Windows 10 virtual machine endpoint is created. The default HTTP/HTTPS proxy is configured to point to a proxy running on a different machine. SSL/TLS traffic is not intercepted on the proxy, and optionally, AV's have been configured to skip the proxy.
2. The security of the OS is weakened by the following actions:
    a. Microsoft Defender is disabled (except in case of Microsoft Defender)
    b. Internet Explorer SmartScreen is disabled (except in case of Microsoft Defender)
3. The following vulnerable software is installed:
    a. Java 1.7.0.17
    b. Adobe Reader 9.3.0
    c. Flash Player 15.0.0.152 or Flash Player 16.0.0.287 in a small number of cases
    d. Silverlight 5.1.10411.0
    e. Internet Explorer 11
    f. Firefox 31.0

g. Chrome 38.0.2125.101

These version numbers were specified with the following two requirements:

- The highest number of in-the-wild exploits should be able to exploit this specific version, thus increasing the coverage of the tests.
- The version must currently be popular among users.
- Windows Update is disabled.

4. From this point, a number of different snapshots are created from the virtual machine, each with different endpoint protection products and one with none. This procedure ensures that the base system is exactly the same in all test systems. The following endpoint security suites, with the following configuration, are defined for this test:
   a. No additional protection, this snapshot is used to infect the OS and to verify the exploit replay.
   b. Vendor A
   c. Vendor B
   d. …

   The endpoint systems are installed with default configuration, potentially unwanted software removal is enabled, and if it was an option during install, cloud/community participation is enabled. The management servers (if needed) are installed onto a different server. The purpose of management servers is to centrally administer, update and analyse logs in an enterprise environment. Installing the management server on a different server is highly recommended by vendors, so it does not interfere with the testing, machine resources are not used by the management server, etc.
5. Two sources of exploits are used during the test. One in-the-wild exploit kits, and one from publicly available open-source exploit frameworks (e.g. Metasploit). In spite of other "real world protection tests", no binary downloads (e.g. exe) were tested. ActiveX, VBscript based downloaders are out of scope in the exploit test section.
6. The virtual machine is reverted to a clean state and traffic was replayed by the proxy server. The replay meant that the browser is used as before, but instead of the original webservers, the proxy server answers the requests based on the recorded traffic. In this replay, other traffic is allowed, which means that unmatched requests (previously not recorded) are answered as without the proxy. When the "replayed exploit" is able to infect the OS, the exploit traffic is marked as a source for the tests. This method guarantees that exactly the same traffic will be seen by the endpoint protection systems, even if the original exploit kit goes down during the tests. Although this might be axiomatic, it is important to note that no exploit traffic test case was deleted after this step of the test. All tests are included in the final results. In the case of HTTPS traffic, the original site is contacted, without replaying.
7. After new exploit traffic is approved, the endpoint protection systems are tested, in a random order. Before the exploit site is tested, it is verified that the endpoint protection had been updated to the latest version with the latest signatures and that every cloud connection is working. If there is a need to restart the system, it is restarted. In the proxy setup, unmatched requests are allowed to pass through. No VPN is used during the test. When user interaction is needed from the endpoint protection (e.g. site visit not recommended, etc.), the block/deny action is chosen. When user interaction is needed from Windows, we chose the run/allow options, except for UAC. No other processes are running on the system, except the Process Monitor from Sysinternals and Wireshark (both installed to non-default directories and modified not to be detected by default tools).
8. After navigating to the exploit site, the system is monitored to check for new processes, loaded DLLs or C&C traffic.

9. After an endpoint protection suite is tested, a new endpoint protection is randomly selected for the test until all endpoint protection products had been tested.
10. The process goes back to step 7. until all exploit site test cases are reached.
11. If the exploitation had been successful and considered 'Missed', the following actions could had been taken.
    - Download a file from victim machine
    - Upload a file to the victim machine
    - Execute a command on the victim machine

Detailed description of the Exploit / Fileless cases.

Test case 001

Koadic/wmic

Koadic is a framework using VBScript stagers for increased stealth and limited footprint. In this test case, a Koadic connectback payload is instantiated using a wmic command.

In case the exploitation was successful, as a proof of that working session has been establised, the following actions has been carried out through the connection.

- A directory list is queried
- A file is uploaded to the victim
- A file is downloaded
- A shell command is executed

The test case is flagged as MISSED if exploitation was successful and test machine had been successfully controlled via the new session.

References:

https://github.com/zerosum0x0/koadic

Test case 002

Koadic/mshta

Koadic is a framework using VBScript stagers for increased stealth and limited footprint. In this test case, a Koadic connectback payload is instantiated using a malicious Windows help .hta document.

In case the exploitation was successful, as a proof of that working session has been establised, the following actions has been carried out through the connection.

- A directory list is queried
- A file is uploaded to the victim

- A file is downloaded
- A shell command is executed

The test case is flagged as MISSED if exploitation was successful and test machine had been successfully controlled via the new session.
References:
https://github.com/zerosum0x0/koadic

Test case 003

Koadic/regsvr32

Koadic is a framework using VBScript stagers for increased stealth and limited footprint. In this test case, a Koadic connectback payload is instantiated using a regsvr32 remote object load call.
In case the exploitation was successful, as a proof of that working session has been establised, the following actions has been carried out through the connection.

- A directory list is queried
- A file is uploaded to the victim
- A file is downloaded
- A shell command is executed

The test case is flagged as MISSED if exploitation was successful and test machine had been successfully controlled via the new session.
References:
https://github.com/zerosum0x0/koadic

Test case 004

EMPIRE/.net compilation job

In this test case, we use the Empire PowerShell framework to create a crafted .net build job XML to spawn an Empire connectback shell upon a compilation process.
In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.
References:

https://github.com/EmpireProject/Empire/

Test case 005

EMPIRE/MSHTA

In this test case, we use the Empire PowerShell framework to create malicious Windows help document to spawn an Empire connectback shell upon a compilation process.

In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.

References:

https://github.com/EmpireProject/Empire/

Test case 006

Foxit reader Use After Free + Empire

In this test case, we use the Foxit Reader v9.0.1.1049 exploit (foxit_reader_uaf) to start the exploit chain. After successfully exploiting the vulnerability an Empire (PowerShell) stager is executed.

In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.

Exploited application: Foxit Reader v9.0.1.1049 OS version: Windows 7

CVE:

CVE-2018-9948
CVE-2018-9958

The exploit

Foxit Reader v9.0.1.1049 and earlier are affected by use-after-free and uninitialized memory vulnerabilities that can be used to gain code execution. This module uses Uint32Array uninitialized memory and text annotation use-after-free vulnerabilities to call WinExec with a share file path to download and execute the specified exe. The module has been tested against Foxit Reader v9.0.1.1049 running on Windows 7 x64 and Windows 10 Pro x64 Build 17134. Windows 10 Enterprise needs to have insecure logons enabled for the exploit to work as expected.

References:

https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-9948
https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-9958
https://www.rapid7.com/db/modules/exploit/windows/fileformat/foxit_reader_uaf
https://www.powershellempire.com/
https://github.com/EmpireProject/Empire.

Test case 007

Firefox version 31.0 exploit with Empire

In this test case, we target Firefox 31.0 with an exploit (CVE-2014-8636, CVE-2015-0802) starting the exploit chain. After successfully exploiting the vulnerability an Empire (PowerShell) stager is executed.

In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.

The exploit

This exploit gains remote code execution on Firefox 31-34 by abusing a bug in the XPConnect component and gaining a reference to the privileged chrome:// window. This exploit requires the user to click anywhere on the page to trigger the vulnerability.

CVE:

CVE-2014-8636
CVE-2015-0802

References:

https://www.rapid7.com/db/modules/exploit/multi/browser/firefox_proxy_prototype
https://www.powershellempire.com/
https://github.com/EmpireProject/Empire

Test case 008

Microsoft Office False Positive test

False positive test: Word document running an Office macro that spawns the existing Windows Calculator. Since this is not a malicious action, expected behaviour from the security product is not to block or detect this test case at all.

Test case 009

MSBuild + Metasploit Meterpreter

In this test case, we target MSBuild starting the exploit chain. Assuming that MSBuild.exe is allowed since this tool is part of the Microsoft .NET Framework, we can invoke it to execute a .xml file as a Visual Studio .NET C# Project descriptor. The well-composed file contains a CSharp code which starts a Meterpreter stager. If code execution is not blocked, as a result, a new Meterpreter session back to MRG-Effitas CnC server will be created.
In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.
References:

https://ired.team/offensive-security/code-execution/using-msbuild-to-execute-shellcode-in-c

Test case 010

Code Injection via NtCreateSection (shellcode: bind shell)

In this test, we used a code injection technique that leverages Native APIs NtCreateSection, NtMapViewOfSection, and RtlCreateUserThread to inject code to a trusted process.
If the code successfully executed, bind shell shellcode is injected to the C:\Windows\System32\exprorer.exe. This payload accepts remote TCP connection and serve them by cmd.exe. Doing this, targeted machine can be controlled from local network.
The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.
In case the exploitation was successful, as a proof of a working session, the following steps are taken.

screenshot has been made
downloading a file
uploading a file

The test case is flagged as MISSED if exploitation was successful and the test machine had been successfully controlled via the new session.
References:

https://ired.team/offensive-security/code-injection-process-injection/ntcreatesection-+-ntmapviewofsection-code-injection

Methodology used in performance test

1. Windows 10 64-bit operating system was installed on a physical machine, all updates are applied, and third-party applications installed and updated.
2. A backup image of the operating system was created.
3. A security application was installed into the OS. Same configuration is used as in the other tests.
4. The following performance metrics were measured:
    a. Install time, starting from downloading the installer binary, finished when the security application is installed, started, and the GUI is working.
    b. Size of the files installed and created by the security application. The size is measured both after the installation, update, scan and after some time passed with normal computer usage.
    c. CPU overhead of the processes and services belonging to the security applications are summed.
    d. Memory footprint (private and shared working set) of the processes and services belonging to the security applications are summed.
    e. Performance impact on the browser load time is measured. The browser should fully load a complex website, from a local network URL or replay proxy.
    f. Average network loading was measured on the interface while the device was idling, during AV update and during system drive scan as well.
    g. Physical Disk usage was measured while the device was idling, during AV update and during system drive scan as well.

Every performance result is the average of three times measurement except for the Firefox start-up time as it was measured twenty times for each vendor.

Performance chart was calculated based on:

- The security product reaching the best result in the category was rewarded with 12 points, the second received 11 points and so on. Once every performance category was measured, the points were summed, and the final calculation was made by dividing the summarized points by the number of tests the product's result could have been measured.

Physical machine specification

- OS: Windows 10 x64
- CPU: Intel Core i5
- Memory: 8GB
- Storage: 100GB SSD

Hardened virtual machine specification

- OS: Windows 10 x64
- CPU: 2 core processor
- Memory: 4GB
- Storage: 100GB SSD

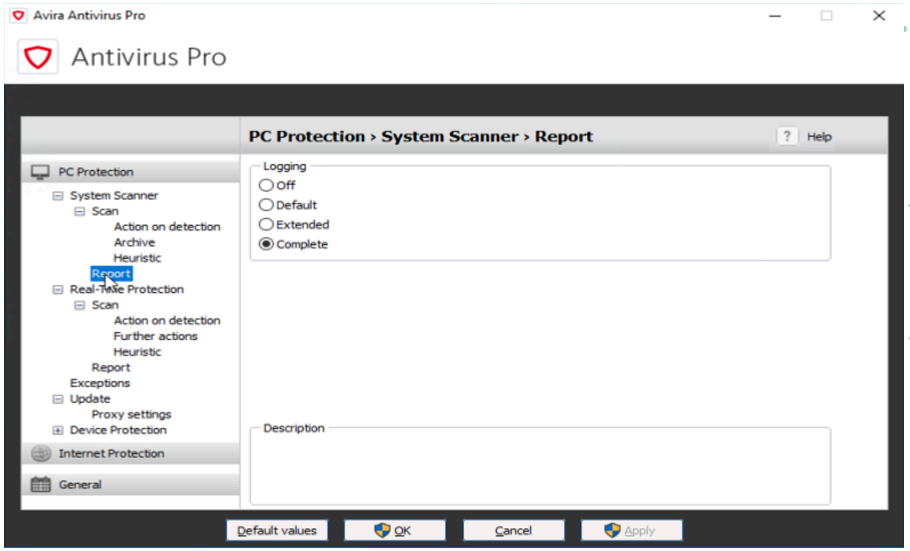
Appendix 2

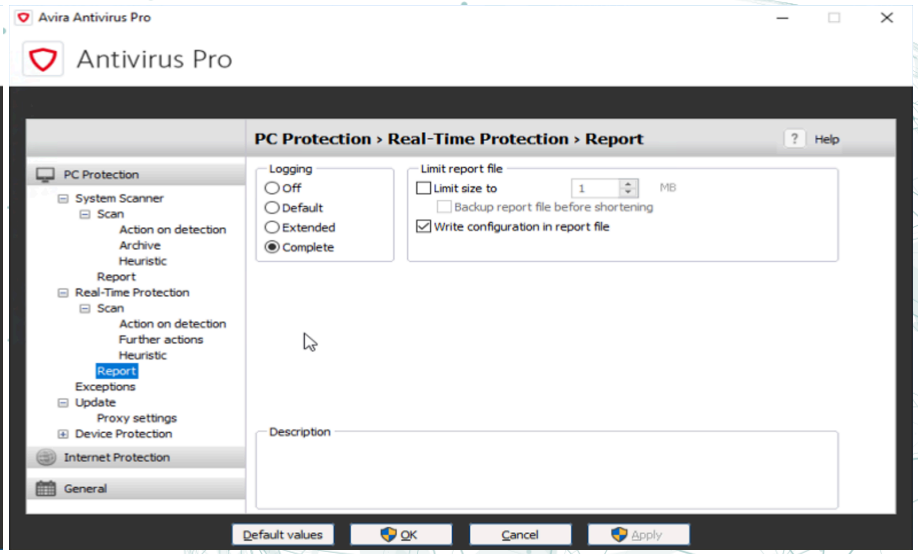
Non-default endpoint protection configurations

Endpoint protection software was running on custom configuration if suggested by the vendor.

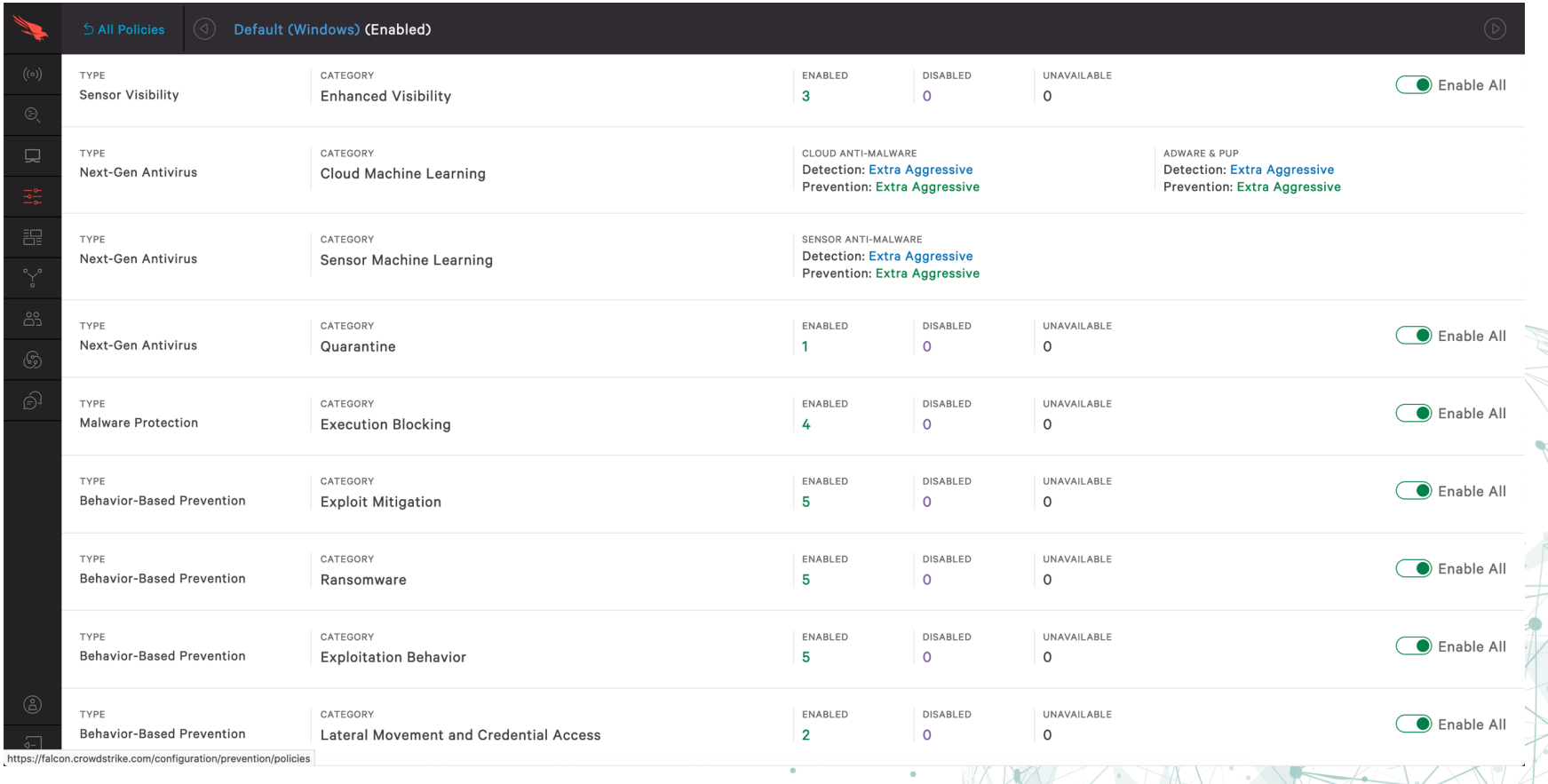
- **Avast Business Antivirus**
  Detailed logging was enabled via configuration file

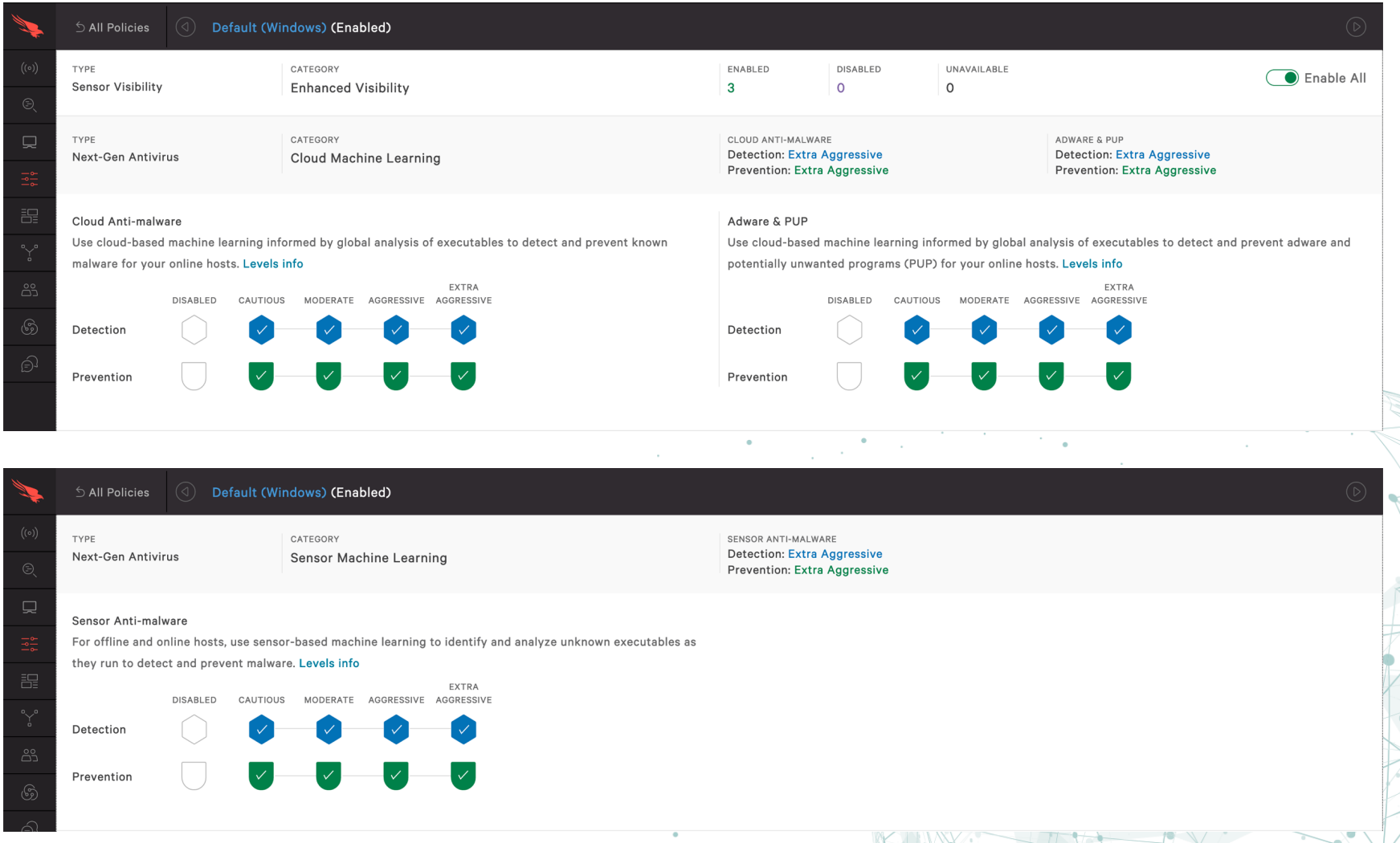
- **Avira Antivirus Pro**
  Log level was set to 'Complete' instead of 'Default' in 'System Scanner' and in 'Real-Time Protection'



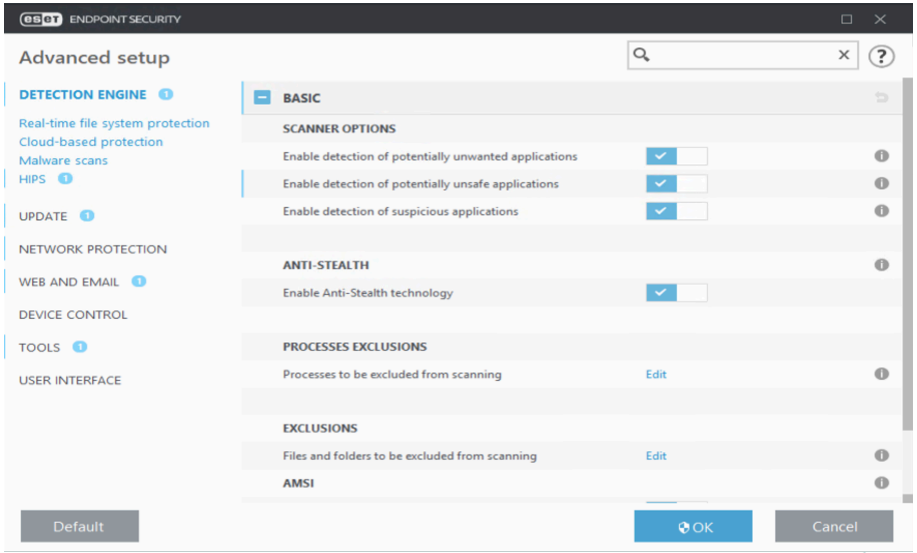


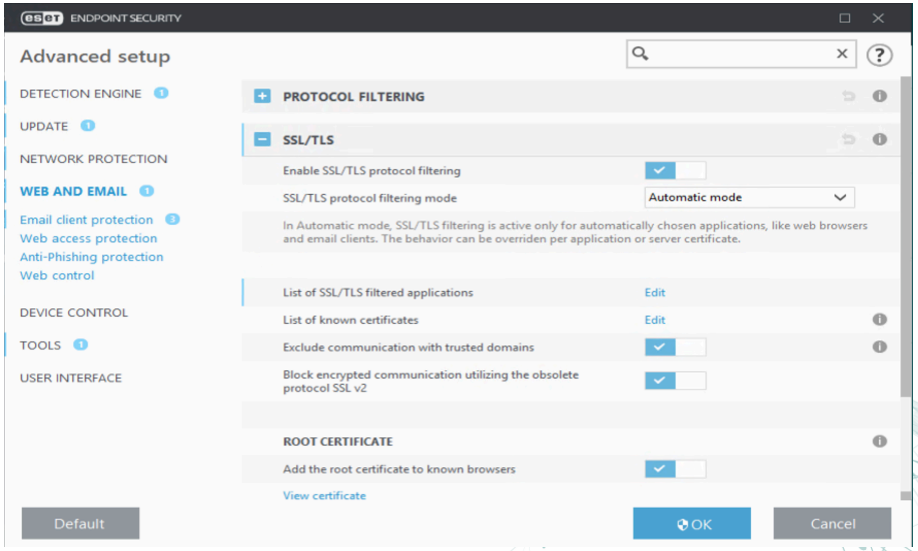
- **CrowdStrike Falcon Protect**
  Cloud Anti-Malware, Sensor Machine Learning Anti-Malware and Adware & PUA detection and prevention levels are set to Extra Aggressive.





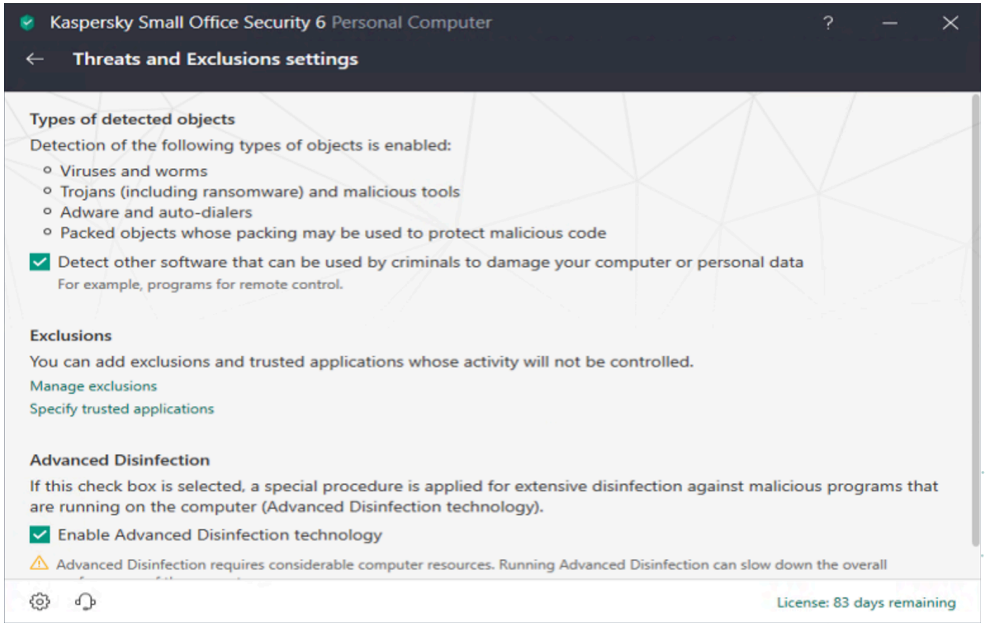
- **ESET Endpoint Security**
  Detection of ‘Potentially unwanted applications’ and ‘Potentially unsafe applications’ were turned on among with ‘SSL/TLS protocol filtering’.

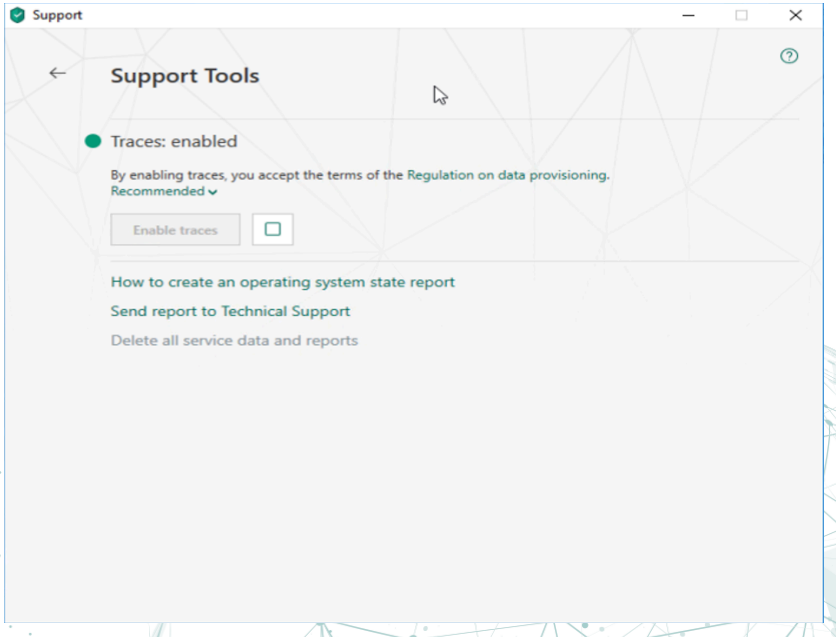




- **Kaspersky Small Office Security**

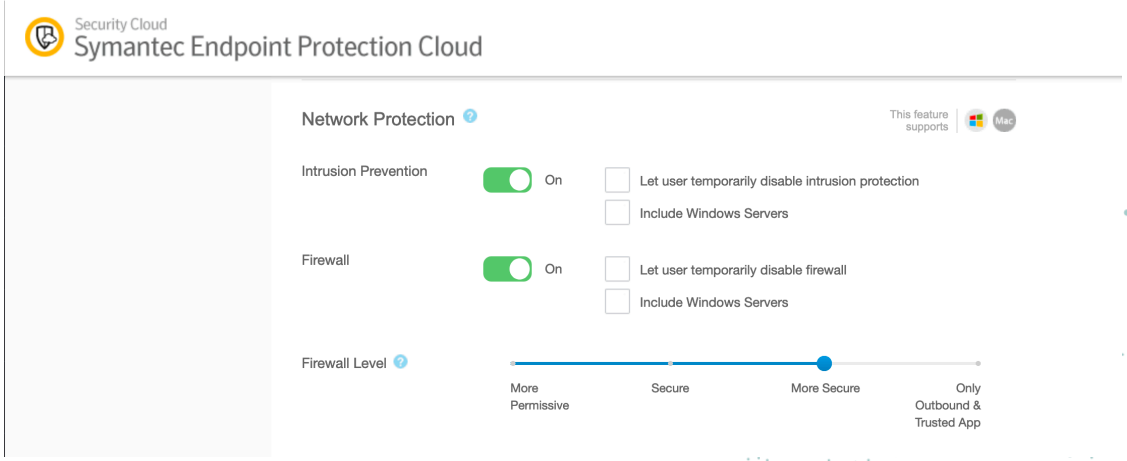
  'Check the URL against the database of URLs containing legitimate applications that can be used by criminals to damager your computer or personal data' and 'enable traces' were turned on.





- **Microsoft Windows Defender**
  Microsoft Defender ATP endpoint detection and response capabilities were turned on including ASR rules.

- **Symantec Endpoint Protection**
  Firewall level was set to ‘More secure’ from ‘Secure’.



Default endpoint protection configurations

- **Bitdefender Endpoint Security**
- **F-Secure Computer Protection Premium**
- **McAfee Endpoint Security**
- **Sophos Intercept X**
- **Trend Micro Security**

Version History

Nr.	Modify date	Comment
1.0	20.05.2020	Published
1.1	15.06.2020	Sophos Intercept X Certification level corrected
1.2	07.09.2020	McAfee Endpoint Security level corrected