Independent Tests of
Anti-Virus Software

Enhanced Real-World Test
Advanced Threat Protection - Consumer

Targeted Attacks – Exploits and Fileless Threats

TEST PERIOD:	AUGUST-NOVEMBER 2019
LANGUAGE:	ENGLISH
LAST REVISION:	16TH DECEMBER 2019

WWW.AV-COMPARATIVES.ORG

Contents

Introduction

Advanced Persistent Threat (APT) is a term commonly used to describe a targeted cyber-attack that employs a complex set of methods and techniques to penetrate information system(s). Different aims of such attacks could be stealing / substituting / damaging confidential information, or establishing sabotage capabilities, the last of which could lead to financial and reputational damage of the targeted organisations. Such attacks are very purposeful, and usually involve highly specialized tools. The tools employed include heavily obfuscated malicious code, the malicious use of benign system tools, and non-file-based malicious code.

In our "Enhanced Real-World Test", we use hacking and penetration techniques that allow attackers to access internal computer systems. These attacks can be broken down into Lockheed Martin's Cybersecurity Kill Chain, and seven distinct phases - each with unique IOCs (Indicators of Compromise) for the victims. All our tests use a subset of the TTP (Tactics, Techniques, Procedures) listed in the MITRE ATT&CK framework[1]. A false alarm test is also included in the report.

The tests use a range of techniques and resources, mimicking malware used in the real world. Some examples of these are given here. We make use of system programs, in an attempt to bypass signature-based detection. Popular scripting languages (JavaScript, batch files, PowerShell, Visual Basic scripts, etc.) are used. The tests involve both staged and non-staged malware samples, and deploy obfuscation and/or encryption of malicious code before execution (Base64, AES). Different C2 channels are used to connect to the attacker (HTTP, HTTPS, TCP). Use is made of known exploit frameworks (Metasploit Framework, Meterpreter, PowerShell Empire, Puppy, etc.).

To represent the targeted system, we use fully patched 64-bit Windows 10 systems, each with a different AV product installed. In the enterprise test, the target user has a standard user account. In the consumer test, an admin account is targeted. For this reason and others (e.g. possibly different settings), the results of the Consumer Test should not be compared with those of the Enterprise Test.

Once the payload is executed by the victim, a Command and Control Channel (C2) to the attacker's system is opened. For this to happen, a listener has to be running on the attacker's side. For example, this could be a Metasploit Listener on a Kali Linux system. Using the C2 channel, the attacker has full access to the compromised system. The functionality and stability of this established access is verified in each test-case.

The test consists of 15 different attacks. In future tests, we plan to provide additional, more granular information, complexity and coverage in the public report. This test currently focuses on protection, not on detection. It is carried out completely manually.

AV Main-Test-Series vendors were given the opportunity to opt out of this test before the public test started, which is why not all vendors are included in this test.

[1] https://attack.mitre.org/matrices/enterprise/windows/

Scope of the test

The Enhanced Real-World Test looks at how well the tested products protect against very specific targeted attack methods. It does not consider the overall security provided by each program, or how well it protects the system against malware downloaded from the Internet or introduced via USB devices.

It should be considered as an addition to the Real-World Protection Test and Malware Protection Test, not a replacement for either of these. Consequently, readers should also consider the results of other tests in our Main-Test Series when evaluating the overall protection provided by any individual product. This test focuses on whether the security products protect against specific attack/exploitation techniques used in APTs. Readers who are concerned about such attacks should consider the consumer products participating in this test, whose vendors were confident of their ability to protect against these threats in the test. We expect more vendors to participate in next year's Enhanced Real-World Test.

Differences between the "MITRE test" and our "Enhanced Real-World Test"

Whilst our Enhanced Real-World Test makes use of elements of the MITRE ATT&CK framework, it is a very different sort of test from the "MITRE test". The "MITRE test" evaluates EDR (Endpoint Detection and Response) systems in situations where the respective vendors actively monitor the attack being performed in real time, sometimes also referred as "red and blue team testing". The emphasis is very much on detecting and logging attack processes (visibility), alerting administrators, and providing data to assist with manual threat-hunting and threat-countering measures.

For the "MITRE test", vendors set their products to "log-only" mode, in order to find out as much as possible about the attack chain. Such tests very definitely have their uses and provide valuable data. However, protecting individual systems against infection, and thus system/data damage, is not the principle aim in such a test. We also note that MITRE does not provide a final scoring or ranking system; rather, it simply provides raw data for analysis.

Our Enhanced Real-World Test, on the other hand, aims to determine how well a security product protects the system on which it is installed in everyday use. The critical question is whether the product protects the system against the attack, whereby it is not important which protection component blocks the attack, or at which stage the attack is stopped, provided the system is not compromised (this sort of granularity might be added in future Enhanced Real-World Test for informational purposes). We also consider false alarms in our test.

Test procedure

Scripts such as VBS, JS or MS Office macros can execute and install a file-less backdoor on victims' systems and create a control channel (C2) to the attacker, who is usually in a different physical location, and maybe even in a different country. Apart from these well-known scenarios, it is possible to deliver file-less malware using exploits, remote calls (PSexec, wmic), task scheduler, registry entries, Arduino hardware (USB RubberDucky) and WMI calls. This can be done with built-in Windows tools like PowerShell. These methods load the actual malware directly from the Internet into the target system's memory, and continue to expand further into the local area network with native OS tools. They may even become persistent on machines in this way. This test does not make use of portable executable (PE) malware.

Fileless attacks

In the field of malware there are many (possibly overlapping) classification categories, and amongst other things a distinction can be made between file-based and fileless malware. Since 2017, a significant increase in fileless threats has been recorded. One reason for this is the fact that such attacks have proved very successful from the attackers' point of view. One factor in their effectiveness is the fact that fileless threats operate only in the memory of the compromised system, making it harder for security solutions to recognize them. It is important that fileless threats are recognised by consumer security programs as well as by business products, for the reasons given below.

Attack vectors and targets

In penetration tests, we see that certain attack vectors may not yet be well covered by security programs, and many popular AV products still provide insufficient protection. Some business security products are now making improvements in this area, and providing better protection in some scenarios. As mentioned above, we believe that consumer products also need to improve their protection against such malicious attacks; non-business users can be, and are, attacked in the same way. Anyone can be targeted, for a variety of reasons, including "doxing" (publishing confidential personal information) as an act of revenge. Attacking the home computers of businesspeople is also an obvious route into accessing their company data.

Attack methods

In the Enhanced Real-World Test, we also include several different command-line stacks, CMD/PS commands, which can download malware from the network directly into RAM (staged) or base64 encoded calls. These methods completely avoid disk access, which is (usually) well guarded by security products. We sometimes use simple concealment measures, or change the method of the stager call as well. Once the malware has loaded its 2nd stage, an http/https connection to the attacker will be established. This inside-out mechanism has the advantage of establishing a C2 channel to the attacker that is beyond the protection measures of the majority of NAT and firewall products. Once the C2 tunnel has been established, the attacker can use all known control mechanisms of the common C2 products (Meterpreter, PowerShell Empire, etc.). These include e.g. file uploads/downloads, screenshots, keylogging, Windows shell (GUI), and webcam snapshots. All the tools used are freely available. Their source code is open and created for research purposes. However, the bad guys often abuse these tools for criminal purposes.

False Positive (False Alarm) Test

A security product that blocks 100% of malicious attacks, but also blocks legitimate (non-malicious) actions, can be hugely disruptive. Consequently, we conduct a false-positives test as part of the Enhanced Real-World Test, to check whether the tested products are able to distinguish malicious from non-malicious actions. Otherwise a security product could easily block 100% of malicious attacks that e.g. use email attachments, scripts and macros, simply by blocking such functions. For many users, this could make it impossible to carry out their normal daily tasks. Consequently, false-positive scores are taken into account in the product's test score.

We also note that warning the user against e.g. opening harmless email attachments can lead to a "boy who cried wolf" scenario. Users who encounter a number of unnecessary warnings will sooner or later assume that all warnings are false alarms, and thus ignore a genuine warning when it comes along.

Tested Products

The following vendors participated in the Enhanced Real-World Test. These are the vendors whose products scored well in the internal pre-test, and who were confident enough in the protection capabilities of their products against file-less attacks to take part in this public test. All other vendors in the Consumer Main-Test Series opted out of the test.

Vendor	Product	Version
Avast	Free Antivirus	19.8
AVG	Free Antivirus	19.8
Bitdefender	Internet Security	24.0
ESET	Internet Security	13.0
F-Secure	Safe	17.7
Kaspersky	Internet Security	20.0

All consumer products were tested with default settings.

Test Results

Below are the results for the 15 attacks used in this test:

	Test scenarios																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	FPs	Score
Avast	✓	✓	✓	✓	✓	🛡	✓	✓	🛡	✓	✓	✓	✓	🛡	✗	N	14
AVG	✓	✓	✓	✓	✓	🛡	✓	✓	🛡	✓	✓	✓	✓	🛡	✗	N	14
Bitdefender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	N	13
ESET	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	15
F-Secure	✓	✓	✓	✓	✓	✓	✓	✓	🛡	✓	✓	✓	✓	✗	✗	N	13
Kaspersky	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	15

Key

✓	Threat detected, no C2 session, system protected	1 point
🛡	No alert shown, but no C2 session established, system protected	1 point
✗	Threat not detected, C2 session established	0 points

In our opinion, the goal of every AV/EPP/EDR system should be to detect and prevent APTs or other malware as soon as possible. In other words, if the APT is detected before, at or soon after execution, thus preventing the opening of a Command and Control Channel, there is no need to prevent post-exploitation activities. A good burglar alarm should go off when somebody breaks into your house, not wait until they start stealing things.

A product that blocks certain functions, e.g. email attachments or scripts, would be categorised only as "Tested". However, none of the tested products exhibited any such behaviour in the false-alarm/functionality-blocking scenarios used in this particular test.

If a user-dependent alert were shown, we would award half a point. However, there were no such cases in this specific test.

Observations on consumer products

In this section, we report any additional information of interest to readers.

Avast, AVG: In three cases, there was no alert, but also no stable C2-session.

Bitdefender: Almost all detections occurred on-access, i.e. before the threat was executed.

ESET, Kaspersky: All threats were blocked: most of them were blocked during execution, and some few ones before the threat was executed (on-access).

F-Secure: All detections occurred during execution of the threats. In one case, there was no alert, but also no stable C2-session.

Award levels reached

From our experience, we know that many consumer AV programs do not provide effective protection against the types of threat used in this test. For this reason, a consumer security app that detects even 5 out of 15 threats is worthy of an award for "Advanced Threat Protection" (ATP). Precise criteria for awards in this test are given below:

	Blocked Threats (out of 15)			
	0-4	5-8	9-12	13-15
No false alarms/functionality blocking	TESTED	STANDARD	ADVANCED	ADVANCED+
False alarms/functionality blocking seen	TESTED	TESTED	TESTED	TESTED

AWARDS[2] (based on protection rates and false alarms)	PRODUCTS
AV comparatives ADVANCED+ ★★★ ATP Consumer 2019	✓ Avast ✓ AVG ✓ Bitdefender ✓ ESET ✓ F-Secure ✓ Kaspersky
AV comparatives ADVANCED ★★ ATP Consumer 2019	-
AV comparatives STANDARD ★ ATP Consumer 2019	-
AV comparatives TESTED ATP Consumer 2019	-

[2] ATP stands for „Advanced Threat Protection"

Test cases employed

We used five different Initial Access Phases, distributed among the 15 test cases (e.g. 3 testcases came via email/spear-phishing attachment).

a) **Trusted Relationship:** “Adversaries may breach or otherwise leverage organizations who have access to intended victims. Access through trusted third party relationship exploits an existing connection that may not be protected or receives less scrutiny than standard mechanisms of gaining access to a network.” (Source: https://attack.mitre.org/techniques/T1199/)

b) **Valid accounts: “**Adversaries may steal the credentials of a specific user or service account using Credential Access techniques or capture credentials earlier in their reconnaissance process through social engineering […].“ (Source: https://attack.mitre.org/techniques/T1078/)

c) **Replication Through Removable Media: “**Adversaries may move onto systems […] by copying malware to removable media […] and renaming it to look like a legitimate file to trick users into executing it on a separate system. […]“ (Source: https://attack.mitre.org/techniques/T1091/)

d) **Spearphishing Attachment:** “Spearphishing attachment is […] employs the use of malware attached to an email. […]” (Source: https://attack.mitre.org/techniques/T1193/)

e) **Spearphishing Link: “**Spearphishing with a link […] employs the use of links to download malware contained in email […]“ (Source: https://attack.mitre.org/techniques/T1192/)

The 15 test scenarios (PowerShell-based file-less attacks and file-based exploits) used in this Enhanced Real-World Test are very briefly described below.

1) This threat is introduced via Trusted Relationship. MSHTA launches an HTML application, which executes a PowerShell command via the Windows Scripting Host. This test case was created with Unicorn.

2) This threat is introduced via Trusted Relationship. A batch file with an encoded PowerShell command gets executed. The PowerShell process injects the payload into memory. This test case was created with Unicorn.

3) This threat is introduced via Trusted Relationship. A Microsoft Word document with a malicious macro starts a PowerShell process which loads the payload into memory. This test case was created with Unicorn.

4) This threat is introduced through Valid Accounts. A VBScript spawns a PowerShell process and executes the payload. This test case was created with Empire.

5) This threat is introduced through Valid Accounts. A Shortcut modification technique is used to generate a backdoor. This test case was created with Empire.

6) This threat is introduced through Valid Accounts. MSHTA launches an HTML application, which executes an obfuscated PowerShell command via the Windows Scripting Host. This test case was created with Empire.

7) This threat is introduced via Removable Media (USB). A batch file executes an encoded payload through the PowerShell engine. This test case was created with Empire.

8) This threat is introduced via Removable Media (USB). A An encoded malicious Microsoft Word document macro starts a PowerShell process and loads the payload into memory. This test case was created with Empire.

9) This threat is introduced via Removable Media (USB). A PowerShell script executes a PowerShell payload into memory. This test case was created with Unicorn.

10) This threat is introduced via Spearphishing Attachment. A VBScript executes an obfuscated payload through the PowerShell engine. This test case was created with Metasploit Meterpreter.

11) This threat is introduced via Spearphishing Attachment. An obfuscated Microsoft Word macro-enabled file starts a PowerShell process which loads the payload into memory. This test case was created with Empire.

12) This threat is introduced via Spearphishing Attachment. A JavaScript executes a C# code via the Windows Scripting Host. This test case was created with SharpShooter.

13) This threat is introduced via Spearphishing Link. A JavaScript executes an obfuscated C# code via the Windows Scripting Host. This test case was created with Metasploit Meterpreter.

14) This threat is introduced via Spearphishing Link. A Microsoft Excel macro-enabled file injects obfuscated C# code into memory. This test case was created with Metasploit Meterpreter.

15) This threat is introduced via Spearphishing Link. A PowerShell script injects an obfuscated PowerShell payload into memory. This test case was created with Metasploit Meterpreter.

False Alarm Test: Various false-alarm scenarios were used in order to see if any product is over-blocking certain actions (e.g. by blocking by policy email attachments, communication, scripts, etc.). None of the tested products showed over-blocking behaviour in the false-alarm test scenarios used.

What is covered by the various testcases?

Our tests use a subset of the TTP (Tactics, Techniques, Procedures) listed in the MITRE ATT&CK framework. In future, we might cover more Techniques and Tactics (such as Privilege Escalation, Credential Access, Lateral Movement and Impact) and provide more details of where the attack is stopped (either as part of this report, or in a separate test report). This year, the above 15 testcases cover the items shown in the table below:

Initial Access	Execution	Persistence	Defense Evasion	*Discovery*	*Collection*	*Command and Control*	*Exfiltration*
Replication Through Removable Media	Mshta	Shortcut Modification	Mshta	*System Information Discovery*	*Data from Local System*	*Commonly Used Port*	*Automated Exfiltration*
Spearphishing Attachment	PowerShell		Masquerading			*Data Encoding*	*Data compressed*
Spearphishing Link	Scripting		Obfuscated Files or Information			*Data Obfuscation*	*Exfiltration Over Command and Control Channel*
Trusted Relationship			Scripting			*Multi-Stage Channels*	
Valid Accounts			Template Injection			*Uncommonly Used Port*	

About this test

The Enhanced Real-World Test for consumer products is a new and optional part of the Public Main-Test Series. We congratulate those vendors who chose to take part. They have obviously worked hard on their products, and are using these public third-party tests as independent verification that their products do what they claim.

The complex nature of the test means that automation is not possible, and it has to be performed entirely manually, making it cost-intensive to run. However, vendors in the Main-Test Series (both Consumer and Enterprise) had the opportunity to participate in the Public Enhanced Real-World Test of 2019 at no additional cost to themselves.

In our Consumer Main-Test Series, products are tested with their default settings. In the Business Main-Test Series, vendors are allowed to configure the products as they see fit – as is common practice with business security products in the real world. However, with each of the two Main Test Series, precisely the same product and configuration is used for all the tests in the respective series. If we did not insist on this, a vendor could turn up protection settings or activate features in order to score highly in the Real-World and Malware Protection Tests, but turn them down/deactivate them for the Performance and False Positive Tests, in order to appear faster and less error-prone. In real life, users can only have one configuration at once, so they should be able to see if high protection scores mean slower system performance, or lower false-positive scores mean reduced protection.

Some vendors asked for precise details of the day and time the test would be performed, so that they could monitor the attacks in real time and interact with their products when they thought it beneficial. Because the aim of the test is to measure protection capabilities, rather than analyse the attack methods, we did not provide any vendors with any advance information about when the test would be performed. In real life, attackers do not tell their victims when they are going to attack, so products must provide protection all the time. We also had requests from vendors regarding the attack methods to be used in the test. Again, because the test is about protection rather than analysis/visibility, we did not divulge specific details of the attack methods.

We did however invite all the vendors in both Main-Test Series to take part in an internal pre-test, which demonstrated broad guidelines for how the test would be performed, and invited vendors to provide feedback on how it might be improved. Each vendor was privately provided with the results for their respective product. As a result of the feedback we received, we implemented some changes in the test methodology, where we felt that this was in the genuine interests of users and helped to promote cybersecurity in general.

The test is very challenging, but at the same time it also reflects realistic scenarios. We have had positive feedback from many vendors' technical departments. Penetration testers see the real capabilities of products in their tests every day. Our comparison test tries to create a level playing-field that allows us to compare the protection capabilities of the different products against such attacks. This lets users see how well they are protected, and allows vendors, where necessary, to improve their products in the future. To get an overall picture of the protection capabilities of any of the tested products, readers should look at the results of the other tests in the Main-Test Series too.

Copyright and Disclaimer

This publication is Copyright © 2019 by AV-Comparatives®. Any use of the results, etc. in whole or in part, is ONLY permitted after the explicit written agreement of the management board of AV-Comparatives prior to any publication. AV-Comparatives and its testers cannot be held liable for any damage or loss, which might occur as result of, or in connection with, the use of the information provided in this paper. We take every possible care to ensure the correctness of the basic data, but a liability for the correctness of the test results cannot be taken by any representative of AV-Comparatives. We do not give any guarantee of the correctness, completeness, or suitability for a specific purpose of any of the information/content provided at any given time. No one else involved in creating, producing or delivering test results shall be liable for any indirect, special or consequential damage, or loss of profits, arising out of, or related to, the use or inability to use, the services provided by the website, test documents or any related data.

For more information about AV-Comparatives and the testing methodologies, please visit our website.

AV-Comparatives
(December 2019)