Independent Tests of Anti-Virus Software

Advanced Threat Protection - Consumer

Enhanced Real-World Test - Targeted Attacks

TEST PERIOD: SEPTEMBER-NOVEMBER 2020
LANGUAGE: ENGLISH
LAST REVISION: 4TH DECEMBER 2020

WWW.AV-COMPARATIVES.ORG

Content

Introduction

"Advanced persistent threat" 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 Advanced Threat Protection Test (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. It currently focuses on protection, not on detection, and is carried out completely manually. Whilst the testing procedure is necessarily complex, we have used a fairly simple description of it in this report. This is in accordance with reader feedback, and we hope that it will make it comprehensible to a wider audience.

AV Consumer Main-Test-Series vendors were given the opportunity to opt out of this test before the test started, which is why not all vendors are included in this test. Some vendors are continuing to perfect their products before joining this advanced test. We congratulate all those vendors who took part in the test, even those whose products did not get the best scores, as they are striving to make their software better.

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

Scope of the test

The Advanced Threat Protection (ATP) 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 advanced persistent threats. Readers who are concerned about such attacks should consider the products participating in this test, whose vendors were confident of their ability to protect against these threats in the test.

Differences between the MITRE ATT&CK® Test and our ATP Test

Whilst our Advanced Threat Protection Test makes use of elements of the ATT&CK framework, it is a very different sort of test from the ATT&CK Test. The ATT&CK Test principally evaluates enterprise security products with investigative and response capabilities in situations where the respective vendors actively monitor the attack being performed in real time. This is 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 ATT&CK 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 ATT&CK Test does not provide a final scoring or ranking system; rather, it provides raw data for analysis.

Our ATP 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. 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 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 year's test does not make use of portable executable (PE) malware. However, as the nature of advanced persistent threats continues to evolve, we may introduce one or two samples of these in the future if appropriate.

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 Advanced Threat Protection 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 second 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 Advanced Threat Protection 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 Advanced Threat Protection Test. These are the vendors who were confident enough in the protection capabilities of their products[2] against targeted 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	20.8
AVG	Free Antivirus	20.8
Bitdefender	Internet Security	25.0
ESET	Internet Security	13.2 - 14.0
F-Secure	Safe	17.8
Kaspersky	Internet Security	21.1
Vipre	Advanced Security	11.0

All consumer products were tested with default settings.

[2] Information about additional third-party engines/signatures used inside the products: Vipre use the Bitdefender engine. F-Secure use the Avira engine. AVG is a rebranded version of Avast.

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	**11**
AVG	🛡	✓	✓	✓	✓	🛡	✓	✗	✗	✓	✗	✓	✗	✓	✓	N	**11**
Bitdefender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	N	**13**
ESET	✓	🛡	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	N	**13**
F-Secure	✓	🛡	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✓	N	**6**
Kaspersky	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	**14**
Vipre	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	N	**10**

Key

✓	Threat blocked, no C2 session, system protected	1 point
🛡	No alert shown, but no C2 session established, system protected	1 point
✗	Threat not blocked, C2 session established	0 points
✓ (grey)	Protection result invalid, as also non-malicious scripts/functions were blocked	N/A

In our opinion, the goal of every AV/EPP/EDR system should be to detect and prevent attacks or other malware as soon as possible. In other words, if the attack is detected before, at or soon after execution, thus preventing e.g. 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 blocked certain legitimate functions (e.g. email attachments or scripts) would be categorised only as "Tested".

Observations on consumer products

In this section, we report some additional information which could be of interest to readers.

Detection/Blocking stages

Pre-execution (PRE): when the threat has not been run, and is inactive on the system.
On-execution (ON): immediately after the threat has been run.
Post-execution (POST): after the threat has been run, and its actions have been recognised.

	Test scenarios														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Avast	ON	POST	PRE	POST	PRE	ON	POST	-	-	ON	-	POST	-	POST	ON
AVG	ON	POST	PRE	POST	PRE	ON	POST	-	-	ON	-	POST	-	POST	ON
Bitdefender	PRE	PRE	PRE	POST	PRE	ON	PRE	POST	PRE	-	PRE	ON	-	PRE	PRE
ESET	ON	ON	ON	ON	ON	ON	POST	ON	-	-	PRE	PRE	PRE	PRE	ON
F-Secure	ON	ON	-	POST	-	-	-	-	PRE	-	-	-	-	ON	ON
Kaspersky	POST	PRE	PRE	PRE	PRE	-	ON	POST	ON	POST	PRE	PRE	POST	ON	ON
Vipre	PRE	-	PRE	-	PRE	ON	PRE	POST	PRE	-	PRE	-	-	PRE	PRE

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

Bitdefender: Many detections occurred before the threat was executed, due to heuristics for malicious scripts.

ESET: In one case, there was no alert, but also no stable C2-session. Most of the malicious email attachments were detected before the attachments were saved to disk.

F-Secure: Detections were mostly on PowerShell exploit code.

Kaspersky: About half of the attacks were blocked before the threat was executed, due to heuristics for malicious scripts, and most of the other attacks were blocked post-execution by the behaviour-blocker.

Vipre: Many detections occurred before the threat was executed, due to heuristics for malicious scripts. One case (#12) was detected (and file on disk was deleted), but the threat continued to run in memory, and the attack continued without problems.

All the tested vendors continuously implement improvements in the product, so it is to be expected that many of the missed attacks used in the test are covered by now.

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[3] (based on protection rates and false alarms)	PRODUCTS
ADVANCED+ ★★★ ATP Consumer 2020	✓ Kaspersky ✓ Bitdefender ✓ ESET
ADVANCED ★★ ATP Consumer 2020	✓ Avast ✓ AVG ✓ Vipre
STANDARD ★ ATP Consumer 2020	✓ F-Secure
TESTED ATP Consumer 2020	-

[3] 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."
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 […]."
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. […]"
d) Phishing: Spearphishing Attachment: "Spearphishing attachment is […] employs the use of malware attached to an email. […]"
e) Phishing: Spearphishing Link: "Spearphishing with a link […] employs the use of links to download malware contained in email […]."

The 15 test scenarios used in this test are very briefly described below:

1) This threat is introduced via Trusted Relationship. MSHTA launches an HTML application, which executes a staged Empire PowerShell payload.

2) This threat is introduced via Trusted Relationship. A PowerShell script containing an AMSI bypass and a PowerShell Empire stager was executed.

3) This threat is introduced via Trusted Relationship. Windows Scripting Host was used to download a PowerShell payload via a integrated Empire PowerShell Stager, combined with an AMSI bypass.

4) This threat is introduced through Valid Accounts. The trusted Windows utility Microsoft Build Engine was used to proxy the execution of an Empire macro payload, which opens a command and control channel.

5) This threat is introduced through Valid Accounts. A VBScript which spawns a PowerShell process and executes an Empire payload has been used.

6) This threat is introduced through Valid Accounts. A batch file was used to execute an obfuscated PowerShell stager, download an obfuscated PoshC2 payload.

7) This threat is introduced via Removable Media (USB). A JavaScript executes an obfuscated PowerShell stager, which downloads and executes a PoshC2 PowerShell payload.

8) This threat is introduced via Removable Media (USB). MSHTA.exe executes a PowerShell stager which launches a base64-encoded PoshC2 staged PowerShell payload.

9) This threat is introduced via Removable Media (USB). A malicious Microsoft Office macro executes a PoshC2 PowerShell payload.

10) This threat is introduced via Spearphishing Attachment. VBScript downloads and executes an XSL PoshC2 payload.

11) This threat is introduced via Spearphishing Attachment. A HTML application downloads and executes an obfuscated PowerShell payload. This test case was created with Metasploit Meterpreter.

12) This threat is introduced via Spearphishing Attachment. VBScript downloads and executes an XSL payload. This test case was created with Metasploit Meterpreter.

13) This threat is introduced via Spearphishing Link. MSHTA.exe downloads and executes an obfuscated XSL payload. This test case was created with Metasploit Meterpreter.

14) This threat is introduced via Spearphishing Link. A JavaScript downloads and executes an obfuscated PowerShell payload. This test case was created with Metasploit Meterpreter.

15) This threat is introduced via Spearphishing Link. MSHTA.exe downloads and executes a PowerShell stager which downloads and executes an encrypted PowerShell Empire staged PowerShell payload, combined with an AMSI bypass.

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.

If during the course of the test, we were to observe products adapting their protection to our test environment, we would use countermeasures to evade these adaptations, to ensure that each product can genuinely detect the attack, as opposed to the test situation.

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. This year, the above 15 testcases cover the items shown in the table below:

Initial Access	Execution	Persistence	Defense Evasion	Discovery	Lateral Movement	Collection	Command and Control	Exfiltration
Replication Through Removable Media	Command and Scripting Interpreter	Boot or Logon Autostart Execution	Obfuscated Files or Information	System Owner/User Discovery	Replication Through Removable Media	Data from Local System	Non-Application Layer Protocol	Exfiltration Over C2 Channel
Trusted Relationship	User Execution	Valid Accounts	Modify Registry	Software Discovery	Internal Spearphishing	Screen Capture	Application Layer Protocol	Automated Exfiltration
Valid Accounts			Signed Binary Proxy Execution	System Information Discovery		Clipboard Data	Data Obfuscation	
Phishing			Template Injection				Encrypted Channel	
			Masquerading				Multi-Stage Channels	
			Valid Accounts				Data Encoding	
			XSL Script Processing				Non-Standard Port	

About this test

The Advanced Threat Protection Test for consumer products is an optional part of the Public Main-Test Series[4]. 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 Consumer Main-Test Series had the opportunity to participate in the Public Advanced Threat Protection Test of 2020 at no additional cost to themselves.

As some of the attack methods used in the test make use of legitimate system programs and techniques, it would be fairly easy for a vendor to stop such attacks e.g. simply by blocking the use of these legitimate processes. However, this would result in the product concerned being marked down for false positives, in the same way that a security program would be marked down for e.g. blocking all unknown executable program files. Likewise, in this test, preventing an attack e.g. by simply blacklisting used servers, files or emails originating from a particular domain name would not be allowed as a means of preventing a targeted attack. Similarly, we do not accept an approach which does not distinguish between malicious and non-malicious processes, but requires e.g. an admin to whitelist ones that should be allowed.

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, precisely the same product and configuration is used for all the tests in the 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 setting 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. After the test, we provide each participating vendor with sufficient data to assist them in understanding any of their missed test cases.

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 fairly 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.

[4] https://www.av-comparatives.org/consumer/

Copyright and Disclaimer

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

AV-Comparatives
(December 2020)