

Independent Tests of Anti-Virus Software



Advanced Threat Protection - Consumer Enhanced Real-World Test - Targeted Attacks

TEST PERIOD: AUTUMN 2023

LAST REVISION: 15TH NOVEMBER 2023

WWW.AV-COMPARATIVES.ORG

Contents

INTRODUCTION	3
TEST PROCEDURE	5
TESTED PRODUCTS	6
TEST RESULTS	7
AWARD LEVELS REACHED	9
TEST CASES EMPLOYED	10
ABOUT THIS TEST	12
COPYRIGHT AND DISCLAIMER	14

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 specialised tools. The tools used are partly free and partly commercial, partly their payloads are based on non-evasive techniques such as using standard Windows APIs, and partly their payloads are based on evasive techniques such as direct syscalls, indirect syscalls, user-mode unhooking, shellcode obfuscation, API hashing, hardware breakpoints, etc.

In our Advanced Threat Protection Test, we use Tactics, Techniques and Procedures (TTPs) that reflect the strategies attackers use to infiltrate a network with malware. These multifaceted attacks can be classified using Lockheed Martin's Cybersecurity Kill Chain, which divides them into seven distinct phases, each marked by its own unique Indicators of Compromise (IOCs). Our testing approach is heavily influenced by a subset of the TTPs found in the respected MITRE ATT&CK® framework. To reinforce the authenticity and reliability of our findings, a false alarm test is integrated into our report. Our tests are designed to simulate real-world scenarios as closely as possible, using a variety of techniques and resources that mimic the malware found in real-world cyber-attacks. We use system programs designed to evade signature-based detection, while also exploiting the versatility of popular scripting languages such as JavaScript, batch files, PowerShell and Visual Basic scripts. Our tests intricately interweave both staged and non-staged malware samples, cleverly using obfuscation and encryption strategies such as Base64, XOR and AES to disguise malicious code before it executes. We use a range of C2 channels to communicate with the attacker, including HTTP, HTTPS and TCP. In addition, our arsenal includes a variety of well-known exploit frameworks such as the Metasploit Framework, PowerShell Empire and several other commercial tools. This holistic and complex approach ensures that our tests remain at the forefront of cybersecurity evaluation and reflect the ever-evolving threat landscape.

To represent the targeted hosts, we use fully patched 64-bit Windows 10 systems, each with a different AV product installed. In the consumer test, an admin account is targeted, although every POC is executed using only a standard-user account, with medium integrity. Windows User Account Control is enabled and set to the default level in both tests. With regard to vendors whose products were tested in both the Consumer and Enterprise ATP Tests, please note that the products and their settings may differ. Hence, 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. If a stable C2 connection is made, the system is considered to be compromised. The test consists of 15 different attacks. It focuses on protection, not on detection, and is carried out entirely manually. Whilst the testing procedure is necessarily complex, we have used a fairly simple description of it in this report.

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.

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 and shared network drives.

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.

In the ATP test, we focus on crafting and testing different kinds of C2 malware POCs, based on different adversary tactics and techniques. We use a variety of delivery scenarios to include the possible adversary strategies. The goal of the ATP Test is to demonstrate the prevention capabilities of the respective products. To accomplish this, we use different POCs, all of which try to open a stable C2 channel after execution, thus simulating a successful initial compromise. In cases where a POC was not prevented and the attacker was able to open a stable C2 session, the target PC was considered to be compromised. The test does not check across different stages of an attack (which is done in our EPR¹ test).

¹ <https://www.av-comparatives.org/enterprise/testmethod/endpoint-prevention-response-tests/>

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.

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 recognise 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 way². 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 can include e.g. file uploads/downloads, screenshots, keylogging, Windows shell (GUI), and webcam snapshots. We expect attacks to be blocked regardless of where/how they are hosted and where from/how they are executed. If an attack is detected only under very specific circumstances, we would say the product does not provide effective protection.

² <https://www.bleepingcomputer.com/news/security/google-youtubers-accounts-hijacked-with-cookie-stealing-malware/>

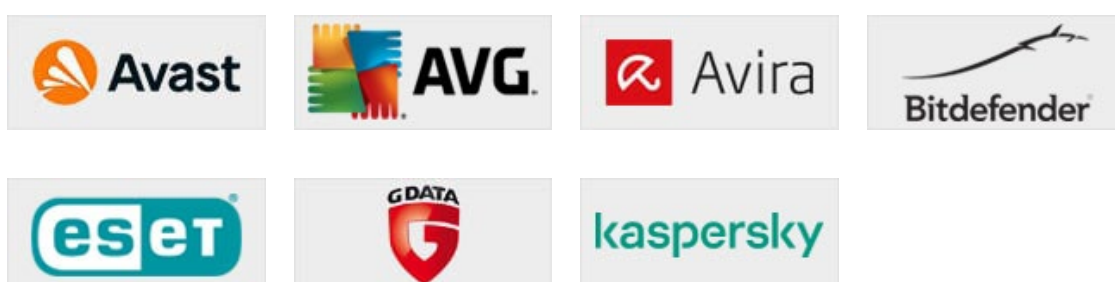
False Positive (False Alarm) Test

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 products³ 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	23.9
AVG	Free Antivirus	23.9
Avira	Prime	1.1
Bitdefender	Internet Security	27.0
ESET	Internet Security	16.2
G Data	Total Security	25.5
Kaspersky	Standard	21.14

All consumer products were tested with default settings.

³ Information about additional third-party engines/signatures used inside the products: **G Data** uses the Bitdefender 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
Avira	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓	✗	N	8
Bitdefender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	N	14
ESET	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	N	13
G Data	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	N	12
Kaspersky	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	N	13

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
⚠️	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/prevented 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 as soon as someone breaks into your home. It should not wait until they start stealing.

The intention of the test is to focus on early detection and prevention, specifically intercepting threats before they progress to post-exploitation stages. The scenarios deliberately excluded certain post-exploitation actions in order to assess the efficacy of hindering Command and Control channels promptly, aiming to neutralize threats at an early stage. The absence of post-exploitation activities does not diminish the significance of early detection, as preventing the establishment of a C2 channel disrupts the cyber kill chain and safeguards against subsequent malicious actions. The inclusion of more damaging actions could skew the evaluation towards post-exploitation capabilities, rather than assessing the system's ability to proactively thwart threats in their early phases.

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 (static).

On-execution (ON): immediately after the threat has been run (dynamic).

Post-execution (POST): after the threat has been run, and its actions have been recognised (in-memory).

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

Avast, AVG: Detections occurred mostly pre- or on-execution, with two post-execution.

Avira: Detections occurred mostly on-execution, with one post-execution.

Bitdefender: Detections occurred either pre- or on-execution.

ESET: Detections occurred mostly pre- or on-execution, with two post-execution.

G Data: Detections occurred mostly pre- or on-execution, with one post-execution.





Kaspersky: Detections occurred mostly on-execution, with five post-execution.

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 now covered.

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
<i>No false alarms/functionality blocking</i>	TESTED	STANDARD	ADVANCED	ADVANCED+
<i>False alarms/functionality blocking seen</i>	TESTED	TESTED	TESTED	TESTED

AWARDS (based on protection rates and false alarms)	PRODUCTS
	<ul style="list-style-type: none"> ✓ Bitdefender ✓ ESET ✓ Kaspersky
	<ul style="list-style-type: none"> ✓ G Data ✓ Avast ✓ AVG
	<ul style="list-style-type: none"> ✓ Avira
	-

*) Products that blocked certain legitimate functions (e.g. email attachments or scripts) were downgraded to "Tested".

Test cases employed

We used five different [Initial Access Phases](#), distributed among the 15 test cases.

- 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 through Replication Trough Removable Media. A malicious binary has been created that executes shellcode to establish a meterpreter C2 channel. The binary was placed in a ZIP container followed by an ISO container.
- 2) This threat is introduced via Replication Trough Removable Media. A malicious JavaScript was crafted, which executes shellcode via DLL Side-Loading to open a meterpreter C2 channel.
- 3) This threat is introduced via Replication Trough Removable Media. A malicious CPL file has been created that runs shellcode via rundll32.exe to establish a Meterpreter C2 channel.
- 4) This threat is introduced via Valid Accounts. A malicious JavaScript has been crafted that runs shellcode via native APIs and DLL side-loading to open a Meterpreter C2 channel.
- 5) This threat is introduced via Spearphishing Attachment. A malicious .exe has been created that runs XOR-encrypted shellcode to open a Meterpreter C2 channel.
- 6) This threat is introduced via Spearphishing Attachment. A malicious JavaScript has been created that executes encrypted shellcode and opens a C2 channel to a commercial C2 framework.
- 7) This threat is introduced via Spearphishing Attachment. A malicious.exe has been created that can patch ETW and execute obfuscated/encrypted shellcode to establish a C2 channel to a commercial C2 framework.
- 8) This threat is introduced via Spearphishing Link. A malicious one-click MSI file was created to run obfuscated shellcode through a legitimate application and establish a C2 channel to a commercial C2 framework.
- 9) This threat is introduced via Spearphishing Link. A malicious masqueraded .exe has been created that runs shellcode via Native APIs to establish a C2 channel to a commercial C2 framework.

10) This threat is introduced via Spearphishing Link. A malicious Office document has been created that is capable of patching AMSI and running obfuscated shellcode to establish a C2 channel to a commercial C2 framework.

11) This threat is introduced via Spearphishing Attachment. A malicious .PIF shortcut file has been created that is capable of patching ETW, patching hooked user-mode APIs, and running obfuscated shellcode to establish an Empire C2 channel.

12) This threat is introduced via Spearphishing Attachment. A malicious masqueraded .exe file has been created that is capable of patching ETW, patching user mode APIs, and running encrypted shellcode to establish an Empire C2 channel.

13) This threat is introduced via Valid Accounts. A malicious PowerShell code has been created that first patches AMSI using hardware breakpoints and then executes encoded shellcode to establish an Empire C2 channel.

14) This threat is introduced via Trusted Relationship. A malicious .HTA file has been created that executes encrypted shellcode to establish an Empire C2 channel.

15) This threat is introduced via Spearphishing Attachment. A malicious obfuscated .JS file has been created that runs shellcode via powershell.exe to establish an Empire C2 channel.

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.

About this test

The Advanced Threat Protection Test for consumer products is an optional part of our Public Consumer Main-Test Series⁴. We commend the vendors who chose to participate, showcasing their dedication to producing reliable products. This particular test cannot be automated and requires manual intervention, which makes it resource-intensive to run. However, we offered vendors in the Consumer Main-Test Series an opportunity to join the Public Advanced Threat Protection Test of 2023 at no extra cost. This allows them to independently verify their product's claims through third-party testing.

In our test, some attack methods utilize legitimate system programs and techniques, making it relatively easy for a vendor to thwart these attacks by blocking the use of these legitimate processes. However, taking such an approach could lead to the product receiving lower ratings due to false positives, just as a security program might be penalized for indiscriminately blocking all unknown executable program files. In the same test, we do not permit the prevention of an attack by merely blacklisting servers, files, or emails originating from a specific domain name as a means of countering targeted attacks. Similarly, we do not endorse an approach that fails to distinguish between malicious and non-malicious processes, requiring administrators to whitelist those that should be allowed. It's worth noting that in enterprise environments, it is possible to restrict users' systems, preventing the execution of PowerShell scripts or macros. An ideal security product should be capable of distinguishing between malicious and non-malicious scripts and macros, thus enabling authorized users to work efficiently while maintaining robust security.

In our Consumer Main-Test Series, products are tested with their default settings. In the Enterprise 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.

We received requests from vendors seeking information about the attack methods for the test. Although we did not disclose specific attack method details upfront, post-test, we provided each participating vendor with sufficient data to demonstrate any missed test cases.

Our test is both highly challenging and a reflection of real-world scenarios. Penetration testers encounter genuine product capabilities daily in their work. Our comparative test strives to create a level playing field, enabling a fair assessment of the protection capabilities of various products against such attacks. This transparency benefits users by revealing the effectiveness of their protection, and it allows vendors, when necessary, to enhance their products in the future.

While this test is for consumer products, the attack techniques used are the same as those in our Enterprise ATP test. Hackers may be highly motivated to target the home computers of specific high-profile individuals, such as politicians or the very wealthy. Additionally, it's worth noting that targeted attacks on enterprises may commence by compromising the home computers of individuals like the CEO.

⁴ <https://www.av-comparatives.org/consumer/>

In the context of the test involving Windows User Accounts, none of the scenarios required administrator permissions on the target system. Therefore, from an attacker's perspective, it didn't matter whether the user was logged in with an Administrator Account (utilized for the Consumer Test) or a Standard User Account (utilized for the Enterprise Test).

In certain test cases, as specified in the testcase descriptions, Initial Access vectors, such as Trusted Relationships and Valid Accounts, were employed. This implies that the attacker already had the necessary user credentials to carry out the advanced attack. Numerous studies have shown that scenarios involving the use of stolen credentials for Initial Access are becoming increasingly prevalent in today's threat landscape.

In some cases, the test involved using redirected drives. Although the ATT&CK framework does not have a specific category for such instances, we categorize them as 'removable media,' as described in the documentation. However, in practice, the method of introducing malware into the system did not have a significant technical impact.

We've received positive feedback from security vendors regarding the thorough and realistic methodology of our annual security assessment. Notably, some vendors who were not included in this year's evaluation are actively improving their products to better defend against real-life targeted attacks, and they plan to participate in next year's assessment. We've also carefully considered feedback and suggestions from participating vendors and will make every effort to incorporate them into next year's assessment where applicable.

Copyright and Disclaimer

This publication is Copyright © 2023 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
(November 2023)