Independent Tests of
Anti-Virus Software

Advanced Threat Protection - Enterprise

Enhanced Real-World Test - Targeted Attacks

TEST PERIOD: SEPTEMBER-OCTOBER 2022

LAST REVISION: 10TH NOVEMBER 2022

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 specialised 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, PowerShell Empire, commercial frameworks, etc.).

To represent the targeted PCs, 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, 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.

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 and shared network drives.

It should be considered as a complement 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 testing different kinds of POC C2 malware, 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 EPR test).

Differences between our ATP Test and our EPR Test

Our ATP (Advanced Threat Protection) Test focusses on protection (as opposed to detection or information gathering). The stage at which the attack is blocked is not relevant, provided the system is ultimately protected. The ATP Test is run for both consumer and business products, and so is of interest to all users. Consequently, we have tried to make it easier to understand for non-expert users.

Our EPR (Endpoint Protection and Response) Test[2], on the other hand, does take into account which stage(s) an attack reaches before being detected and blocked. It also looks at any responses made, and considers total cost of ownership. The EPR Test is only for enterprise products, and is more complex. The intended audience are IT security professionals in larger enterprises.

[2] 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.

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 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.

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[3] against targeted attacks to take part in this public test.

Vendor	Product	Version
Acronis	Cyber Protect Cloud with Advanced Security pack	15.0
Avast	Ultimate Business Security	22.7 - 22.9
Bitdefender	GravityZone Business Security Premium	7.7
CrowdStrike	Falcon Pro	6.45
ESET	PROTECT Entry with ESET PROTECT Cloud	9.0
G Data	Endpoint Protection Business	15.3
Kaspersky	Endpoint Security for Business – Select, with KSC	11.10
Microsoft	Defender Antivirus for Business	4.18
VMware	Carbon Black Cloud Endpoint Standard	3.8

Most AV vendors did not participate with their respective EDR products, or disabled the EDR components of their participating products (see settings below). This may be explained by the following. The Enterprise ATP Test is an optional add-on to the Enterprise Main Test Series. We use the same product and configuration for all the tests within a series, and some EDR functions can have a negative impact on performance and false alarms.

[3] Information about additional third-party engines/signatures used by some of the products: **Acronis**, **G Data** and **VIPRE** use the Bitdefender engine (in addition to their own protection features). **VMware** uses the Avira engine (in addition to its own protection features). **G Data**'s OutbreakShield is based on **Cyren**.

Settings

In business environments, and with business products in general, it is usual for products to be configured by the system administrator, in accordance with vendor's guidelines, and so we invited all vendors to configure their respective products. Below we have listed relevant deviations from default settings (i.e. setting changes applied by the vendors):

Acronis: "Backup", "Vulnerability assessment", "Patch management", "Device control", "Data Loss Prevention" and "Data protection map" disabled. "Third-party scan engine" enabled.

Avast: "Scan for potentially unwanted programs (PUPs)" was enabled in "File Shield", "Web Shield" and "Mail Shield".

Bitdefender: "Sandbox Analyzer" (for Applications and Documents) enabled. "Analysis mode" set to "Monitoring". "Scan SSL" enabled for HTTP and RDP. "HyperDetect" and "Device Control" disabled. "Update ring" changed to "Fast ring". "Web Traffic Scan" and "Email Traffic Scan" enabled for Incoming emails (POP3). "Ransomware Mitigation" enabled. "Process memory Scan" for "On-Access scanning" enabled. All "AMSI Command-Line Scanner" settings enabled for "Fileless Attack Protection".

CrowdStrike: everything enabled and set to maximum, i.e. "Extra Aggressive". "Sensor Visibility" for "Firmware" disabled. Uploading of "Unknown Detection-Related Executables" and "Unknown Executables" disabled.

ESET: All "Real-Time & Machine Learning Protection" settings set to "Aggressive".

G Data: "BEAST Behavior Monitoring" set to "Halt program and move to quarantine". "G DATA WebProtection" add-on for Google Chrome installed and activated. "Malware Information Initiative" enabled.

Kaspersky: "Adaptive Anomaly Control" disabled; "Detect other software that can be used by criminals to damage your computer or personal data" enabled;

Microsoft: Google Chrome extension "Windows Defender Browser Protection" installed and enabled; "CloudExtendedTimeOut" set to 55; "PuaMode" enabled.

VMware: policy set to "Advanced".

Please note that the results reached are valid only for the products tested with their respective settings. With other settings (or products) the scores could be worse or better.

Test Results

Below are the results for the 15 attacks used in this test[4]:

	Test scenarios																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	FPs	Score
Acronis	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗	N	8
Avast	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	N	10
Bitdefender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	🛡	N	14
CrowdStrike	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	N	11
ESET	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	N	14
G Data	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	N	12
Kaspersky	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	N	12
Microsoft	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗	✗	N	11
VMware	✓	✓	✗	✓	✗	✓	✗	✗	✓	✓	✗	✓	✓	✗	✗	N	8

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 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) in our FP test, would not be certified.

[4] Please note that the reached results are valid only for the products tested with their respective settings. With other settings (or products) the scores could be worse or better.

Observations on enterprise 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
Acronis	PRE	PRE	-	PRE	-	ON	-	ON	PRE	PRE	-	ON	-	-	-
Avast	POST	PRE	-	PRE	ON	ON	ON	ON	ON	ON	-	-	-	-	ON
Bitdefender	PRE	PRE	ON	PRE	ON	ON	ON	PRE	PRE	PRE	ON	PRE	PRE	-	POST
CrowdStrike	ON	ON	ON	ON	ON	ON	ON	ON	ON	POST	ON	-	-	-	-
ESET	POST	ON	PRE	PRE	ON	PRE	ON	POST	PRE	ON	PRE	-	ON	ON	ON
G Data	PRE	PRE	ON	PRE	POST	ON	-	PRE	PRE	ON	ON	PRE	ON	-	-
Kaspersky	PRE	ON	ON	ON	-	ON	-	POST	PRE	PRE	PRE	ON	-	ON	ON
Microsoft	PRE	PRE	PRE	PRE	ON	ON	PRE	-	ON	POST	PRE	-	PRE	-	-
VMware	PRE	ON	-	ON	-	ON	-	-	ON	PRE	-	ON	PRE	-	-

Acronis: Detections occurred either pre- or on-execution. In one case, the web console stated that the threat was successfully blocked, but in reality it was not, and the C2 channel was stable.

Avast: Most detections occurred on-execution.

Bitdefender: Detections occurred mostly pre- or on-execution, with one post-execution. In one case, there was no alert, but also no stable C2-session.

CrowdStrike: Most detections occurred 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 pre- or on-execution, with one post-execution.

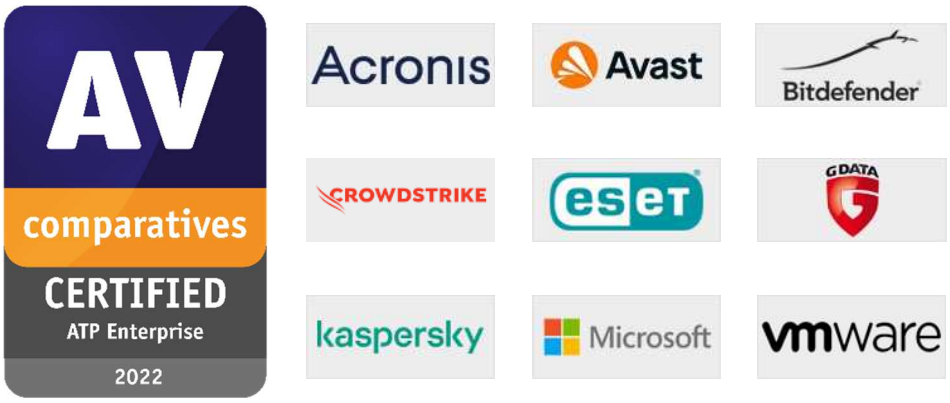
Microsoft: Detections occurred mostly pre- or on-execution, with one post-execution.

VMware: Detections occurred either pre- or on-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 covered by now.

Certified Advanced Threat Protection (ATP) Enterprise Products

AV-Comparatives' certification for Advanced Threat Protection is given to Approved Enterprise products which blocked at least 8 of the 15 attacks used in the Advanced Threat Protection Test, without blocking non-malicious operations. Business security programs are expected to deal with the kind of attacks used in this test, so detection of more than half of the test cases is required for certification.



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 Spearphishing Link. A malicious binary executes x86 shellcode to open a meterpreter C2 channel via http.

2) This threat is introduced via Spearphishing Link. A malicious JavaScript file executes x64 shellcode to open a meterpreter C2 channel via http.

3) This threat is introduced via Spearphishing Link. A malicious obfuscated JavaScript injects x64 shellcode into an Office process to open a meterpreter C2 channel via http.

4) This threat is introduced via Valid Accounts. A malicious HTA file opens a meterpreter C2 channel via http.

5) This threat is introduced via Valid Accounts. A malicious PowerShell command with some defense evasion capabilities opens a meterpreter C2 channel via http.

6) This threat is introduced via Valid Accounts. A malicious Batch file opens an Empire C2 channel via http using a non-standard port.

7) This threat is introduced via Trusted Relationship. A malicious, obfuscated binary with some defense evasion capabilities and file extension spoofing, opens a PowerShell Empire C2 channel via http using a non-standard port.

8) This threat is introduced via Trusted Relationship. A malicious CPL file executes a PowerShell payload, which opens an Empire C2 channel via http using a non-standard port.

9) This threat is introduced via Trusted Relationship. A malicious XSL file is executed via WMI, which opens an obfuscated Empire C2 channel via http using a non-standard port.

10) This threat is introduced via Spearphishing Attachment. A malicious binary with a spoofed file extension executes an Empire payload to open an Empire C2 channel via http using a non-standard port.

11) This threat is introduced via Spearphishing Attachment. A malicious JavaScript file with some defense evasion capabilities enables to execute malicious code via the control panel application, and opens a C2 channel to a commercial C2 framework via https.

12) This threat is introduced via Spearphishing Attachment. A malicious binary with some defense evasion capabilities opens a C2 channel to a commercial C2 framework via https.

13) This threat is introduced via Removable Media. A malicious DLL opens a C2 channel via https to a commercial C2 framework.

14) This threat is introduced via Removable Media. A malicious binary with advanced defense evasion capabilities opens a C2 channel via https to a commercial C2 framework.

15) This threat is introduced via Removable Media. A malicious office document injects into another user-space process and opens a C2 channel to a commercial C2 framework via https.

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 enterprise products is an optional add-on test to the Public Enterprise Main-Test Series, i.e. only enterprise products which are in the Main-Test Series can join this add-on test. 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[5] too.

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. We note that in enterprise environments, it is possible to lock down users' systems, e.g. to prevent the execution of PowerShell scripts or macros. The idea of a good security product is that it can distinguish between e.g. malicious and non-malicious scripts and macros, thus allowing authorised users to work efficiently whilst maintaining good security.

[5] https://www.av-comparatives.org/testmethod/business-security-tests-and-reviews/

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 had requests from vendors regarding the attack methods to be used in the test. We did not divulge specific details of the attack methods, but after the test, we provided each participating vendor with sufficient data to demonstrate the missed test cases.

The test is very challenging, but at the same time it also reflects realistic scenarios. 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.

As regards the Windows User Accounts used in the test, none of the test scenarios required administrator permissions on the targeted system. Hence, from an attacker's perspective it made no difference whether the user was logged on with an Administrator Account (as used for the Consumer Test) or Standard User Account (as used for the Enterprise Test).

In some of the attacks, as noted in the testcase descriptions, Initial Access vectors such as Trusted Relationships and Valid Accounts were used. That is to say, the attacker already had the necessary user credentials needed to proceed with the advanced attack. Various studies show that such scenarios (e.g. using stolen credentials for Initial Access) are very common nowadays.

In some cases, the test made use of redirected drives. As the ATT&CK framework does not have a specific category for such cases, we feel that they fall best into the category "removable media", which we have noted in the descriptions. Nevertheless, how the malware was introduced into the system made no technical difference in practice.

In general, we have received positive feedback on the thorough and realistic methodology of this test from security vendors. We are happy to say that a number of vendors not in this year's test are improving their products' protection against real-life targeted attacks, and aim to join the test next year. We have also considered some suggestions for improvement from some vendors, and will endeavour to include these in next year's test where appropriate.

Copyright and Disclaimer

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

AV-Comparatives
(November 2022)