

Independent Tests of Anti-Virus Software



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

TEST PERIOD: AUTUMN 2024

LAST REVISION: 28TH OCTOBER 2024

WWW.AV-COMPARATIVES.ORG

Contents

INTRODUCTION	3
TEST PROCEDURE	5
TESTED PRODUCTS	6
TEST RESULTS	8
CERTIFIED ADVANCED THREAT PROTECTION (ATP) ENTERPRISE PRODUCTS	10
TEST CASES EMPLOYED	11
ABOUT THIS TEST	13
COPYRIGHT AND DISCLAIMER	15

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 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.

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 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 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) Test¹, 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 Test² is only for enterprise products, and is more complex. The intended audience are IT security professionals in larger enterprises.

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

² <https://www.av-comparatives.org/the-difference-between-av-comparatives-epr-test-and-mitre-attck-engenuity/>

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

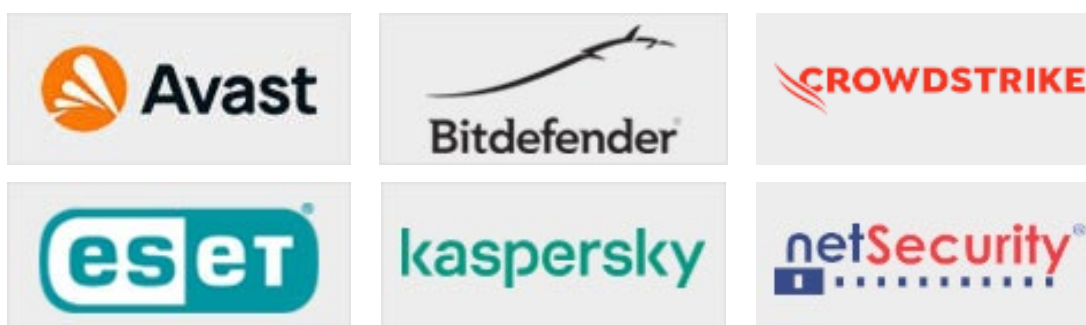
In the Advanced Threat Protection Test, we also include several different command-line stacks, CMD/PS commands, which can download malware from the network directly into RAM (staged) or base64 encoded calls. These methods completely avoid disk access, which is (usually) well-guarded by security products. 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 (ATP) Test, demonstrating their confidence in the protection capabilities of their products against targeted attacks. Participation in the AV-Comparatives ATP test reflects a commitment to transparency and the ability to prevent breaches and advanced attacks. While some vendors may choose not to undergo this evaluation yet for various reasons, we encourage analysts to take note of those who engage in this challenging and essential test.



Vendor	Product	Version
Avast	Ultimate Business Security	24.8
Bitdefender	GravityZone Business Security Premium	7.9
CrowdStrike	Falcon Pro	7.16
ESET	PROTECT Entry with ESET PROTECT Cloud	11.1
Kaspersky	Endpoint Security for Business – Select, with KSC	12.6
NetSecurity	ThreatResponder	3.5

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.

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):

Avast: default settings.

Bitdefender: "Sandbox Analyzer" (for Applications, Documents, Scripts, Archives and Emails) 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". "On Write Script File Visibility" and "Uploading of Unknown Detection-Related Executables" enabled. "On-demand Scans" and "Uploading of All Unknown Executables" disabled.

ESET: Under "Protections" all "Detection responses" were set to "Aggressive". "Detection of potentially unwanted programs" enabled.

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

NetSecurity: default settings.

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 test³:

	Test scenarios															FPs	Score
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Avast	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	13
Bitdefender	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	N	13
CrowdStrike	🛡️	✓	✓	✓	✓	🛡️	✓	✓	✓	🛡️	✓	✓	✓	✓	✗	N	14
ESET	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	14
Kaspersky	🛡️	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	13
NetSecurity	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	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/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) in our FP test, would not be certified.

³ 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
Avast	ON	PRE	-	-	PRE	ON	PRE	POST	PRE	ON	ON	PRE	ON	ON	ON
Bitdefender	PRE	PRE	PRE	ON	PRE	-	PRE	ON	PRE	ON	-	PRE	ON	PRE	ON
CrowdStrike	POST	ON	POST	ON	ON	POST	ON	ON	PRE	POST	ON	ON	ON	ON	-
ESET	PRE	ON	PRE	PRE	-	ON	ON	PRE	PRE	ON	ON	PRE	ON	PRE	ON
Kaspersky	POST	ON	-	POST	PRE	-	ON	POST	ON	POST	ON	ON	ON	PRE	ON
NetSecurity	ON	-	-	ON	ON	-	-	-	-	-	ON	ON	ON	PRE	ON

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

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

CrowdStrike: Most detections occurred on-execution. In three cases, there was no alert, but also no stable C2-session.

ESET: Detections occurred mostly pre- or on-execution.

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

NetSecurity: Detections occurred mostly on-execution. None of the five test-cases (6-10) using the Meterpreter C2 framework were blocked.

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.

- 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 screensaver SCR file has been created, which executes a payload to open a command-and-control channel via DNS. A commercial C2 framework was used.

2) This threat is introduced via Valid Accounts. A malicious JavaScript (JS) has been crafted that runs shellcode via native APIs and DLL side-loading to open a HTTP command and control channel. A commercial C2 framework was used.

3) This threat is introduced via Spearphishing Attachment. A malicious Visual Basic Script (VBS) has been created which executes an obfuscated payload to open a command-and-control channel via DNS. A commercial C2 framework was used.

4) This threat is introduced via Trusted Relationship. A malicious screensaver PE file has been created, which executes a payload to open a command-and-control channel via HTTP. A commercial C2 framework was used.

5) This threat is introduced via Replication Through Removable Media. A malicious LNK shortcut file has been created to execute an obfuscated payload via the CertUtil binary to open a command-and-control channel via DNS. A commercial C2 framework was used.

6) This threat is introduced via a Spearphishing Link. A malicious one-click MSI file has been created which uses a decoy file to run a legitimate application and also executes our payload via rundll32 to open a HTTP Meterpreter C2 channel.

7) This threat is introduced via Valid Accounts. A malicious JavaScript (JS) has been crafted that runs shellcode via native APIs and DLL side-loading to open a HTTP Meterpreter C2 channel.

8) This threat is introduced via Phishing: Spearphishing Attachment. A malicious CPL file has been created that runs shellcode via rundll32.exe to establish an HTTP Meterpreter C2 channel.

9) This threat is introduced via Trusted Relationship. A malicious DLL has been created that executes an XOR-encrypted shellcode via rundll32.exe to establish an HTTP Meterpreter C2 channel.

10) This threat is introduced via Replication Through Removable Media. 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 HTTP Meterpreter C2 channel.

11) This threat is introduced via a Spearphishing Link. A malicious one-click MSI file has been created which uses a decoy file to run a legitimate application and also executes our payload via rundll32 to open an HTTP Empire C2 channel.

12) This thread is introduced via Valid Accounts. A malicious Visual Basic Script (VBS) file has been created that executes shellcode to establish an HTTP Empire C2 channel.

13) This threat is introduced via Phishing: Spearphishing Attachment. A malicious HTA file has been created that executes encrypted shellcode to open an HTTP Empire C2 channel.

14) This threat is introduced via Trusted Relationship. A malicious DLL has been created that executes an AES encrypted shellcode via rundll32.exe to establish an HTTP Empire C2 channel.

15) This threat is introduced via Replication Trough Removeable Media. A malicious DLL has been created that executes an XOR-encrypted shellcode to establish an HTTP 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 enterprise products is an optional add-on to the Public Enterprise Main-Test Series. In other words, only enterprise products that are part of the Main-Test Series can participate in this add-on test. To gain a comprehensive understanding of the protection capabilities of any of the tested products, readers should also consider the results of the other tests in the Main-Test Series⁴.

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 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. 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.

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).

⁴ <https://www.av-comparatives.org/testmethod/business-security-tests-and-reviews/>

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, a number of vendors who were not included in this year's evaluation tell us that they are still actively improving their products to better defend against real-life targeted attacks, and they plan to participate in future assessments.

We've received positive feedback from security vendors regarding the thorough and realistic methodology of our annual security assessment. Several vendors not included in this year's evaluation have indicated that they are still working to improve their products to better defend against such real-life targeted attacks and intend to participate in future assessments. We look forward to seeing their progress reflected in upcoming tests when they participate publicly.

Copyright and Disclaimer

This publication is Copyright © 2024 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
(October 2024)