



Product Comparative Real-World Protection Test Focus on Exploit and In-The-Wild Malware

Tested Products

- Cylance / Cylance Protect
- Symantec / Symantec Endpoint Protection

February 2016

Language: English February 2016 Last revision: 17th February 2016

http://www.av-comparatives.org

http://www.mrg-effitas.com



Content

Introduction	3
General	3
Getting Products	3
Overview	4
Tested Products	4
Test Cases	4
List of Test cases	4
Test Period	4
Results	5
Detailed Results	6
Detailed Results Of The Exploit Protection Test	6
Detailed Results Of The In-The-Wild Malware Protection Test	7
Scoring / Calculation of Results	8
Scoring Of The Exploit Protection Results	8
Scoring Of The In-The-Wild Malware Protection Results	10
False positive test	10
Wrongly blocked files (while downloading/installing)	10
Test Procedure / Methodology	11
Exploit Test Setup	11
Analysis Of The Exploit Kits Used In The Exploit Test	13
Analysis Of The Exploits Used In The Exploit Test	13
In-Memory Malware In Exploit Tests	13
Real-World Protection Test Setup	14
Software Installed	14
Settings	14
Preparation For Every Testing Day	14
Testing Cycle For Each Malicious URL	15
Test Set For In-The-Wild Malware	16
About the test-labs	17
AV-Comparatives	17
MRG Effitas	18
Copyright and Disclaimer	19



Introduction

For this assessment, MRG Effitas and AV-Comparatives combined their strengths to conduct a joint test. The Real-World Protection Test was performed by AV-Comparatives, and the Exploit Test was performed by MRG Effitas.

After the test, a peer review was conducted.

General

Malicious software poses an ever-increasing threat, not only due to the number of malware programs increasing, but also due to the nature of the threats. Infection vectors are changing from simple file-based methods to distribution via the Internet. Malware is increasingly focussing on users, e.g. by deceiving them into visiting infected web pages, installing rogue/malicious software, cyber espionage, ransomware or opening emails with malicious attachments.

The scope of protection offered by antivirus programs like signatures and heuristics is extended by the inclusion of e.g. URL-blockers, content filtering, reputation systems, cloud based methodologies and user-friendly behaviour-blockers. If these features are perfectly coordinated with the signature-based and heuristic detection, the protection provided against threats increases.

However, we would recommend that all parts of a product should be as effective as possible. It should be borne in mind that not all malware enters computer systems via the Internet, and that e.g. a URL blocker is ineffective against malware introduced to a PC via a USB flash drive or over the local area network.

Getting Products

Both, AV-Comparatives and MRG Effitas tried to get a license for CylanceProtect. Unfortunately, this was nearly impossible. It was tried via two IT system houses (one in Italy, one in Austria). Both did not get any license, even if they asked for a regular sales. Fortunately, a third party granted access to the license of Cylance. This behaviour is seen by many of the newer products that claim to be next generation. It looks like they try to avoid getting tested in order to continue to attract users simple by unproven marketing claims.

The other tested vendor gave the license immediately for free.

The costs of the test have been covered by the testing labs, no vendor commissioned this specific test.



Overview

In this test, all protection features of the product can be used to prevent infection - not just signatures or heuristic file scanning. A suite can step in at any stage of the process – accessing the URL, downloading the file, formation of the file on the local hard drive, file access and file execution – to protect the PC. This means that the test achieves the most realistic way of determining how well the security product protects the PC. Because all of a suite's components can be used to protect the PC, it is possible for a product to score well in the test by having e.g. very good behavioural protection, but a weak URL blocker.

The test was split in two categories

- Protection against in the wild seen malware
- Protection against exploits

The test was run online. A detailed methodology can be found in the addendum.

Tested Products

For this test, we normally use the Endpoint Protection Suite, as any protection features that prevent the system from being compromised can be used. The main versions of the products are shown below:

Vendor	Product	Version
Symantec	Endpoint Protection	12.1 RU6 MP3
Cylance	CylanceProtect	1.2.1310.18

Test Cases

	Test cases
In-The-Wild Malware	110
Exploits	40
TOTAL	150

List of Test cases

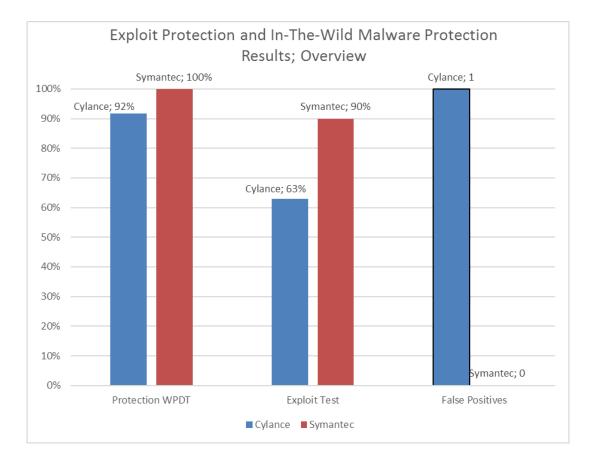
The test cases can be obtained from AV-Comparatives or MRG Effitas.

Test Period

29th of January 2016 – 9th of February 2016



Results



The following chart shows the results for the exploit protection and in-the-wild malware protection results.

In-The-Wild Malware Protection Test

	Protected	Missed	False Positives
Symantec	100%	0%	0
Cylance	92%	8%	1

Exploit Protection Test

	Protected	Missed	
Symantec	90%	10%	
Cylance	63%	37%	

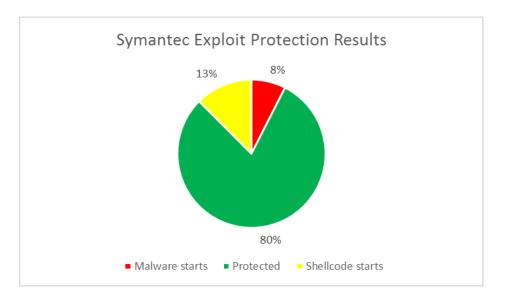
In this independent assessment Cylance clearly delivered inferior protection against In-the-Wild threats and exploits compared to Symantec.



Detailed Results

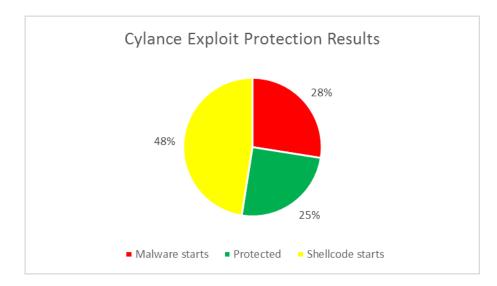
Detailed Results Of The Exploit Protection Test

The following chart shows the detailed results for the exploit protection results of Symantec Endpoint Protection.



Symantec fully protected the system in 80% of the cases. Shellcode was able to execute, but no additional harm was caused in 12.5% of the cases, and malware was able to start in 7.5% of the cases. The missed samples are in-the-wild exploit (malvertisement), Metasploit Macro code with in-memory Meterpreter and Sandworm Office exploit.

The following chart shows the detailed results for the exploit protection results of CylanceProtect.

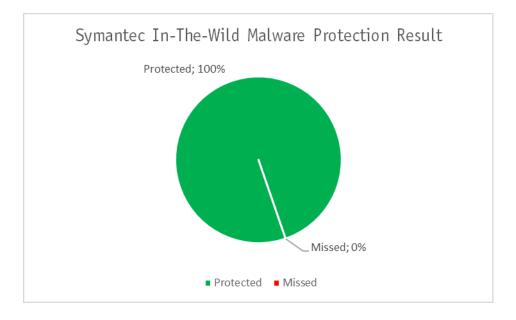


CylanceProtect fully protected the system in 25% of the cases. Shellcode was able to execute, but no additional harm was caused in 47.5% of the cases, and malware was able to start in 27.5% of the cases. Among the missed samples are Metasploit exploits with in-memory Meterpreter, Dridex financial malware, in-the-wild exploit (malvertisement) and Sandworm Office exploit.



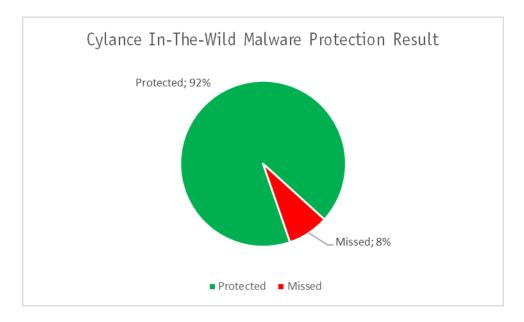
Detailed Results Of The In-The-Wild Malware Protection Test

The following chart shows the detailed results for the In-The-Wild Malware Protection results of Symantec Endpoint Protection.



Symantec Endpoint Protection blocked 100% of in-the-wild malware attacks with no false positives.

The following chart shows the detailed results for the In-The-Wild Malware Protection of CylanceProtect.



CylanceProtect blocked 92% of in-the-wild malware attacks with one false positives.



Scoring / Calculation of Results

Scoring Of The Exploit Protection Results

We defined the following stages, where the exploit can be prevented by the endpoint protection system:

1. Blocking the URL (infected URL, exploit kit URL, redirection URL, malware URL) by the URL database (local or cloud). For example, a typical result is the browser displaying a "site has been blocked" message by the endpoint protection.

2. Analysing and blocking the page containing a malicious HTML code, JavaScripts (redirects, iframes, obfuscated JavaScripts, etc.), or Flash files.

3. Blocking the exploit before the shellcode is executed.

4. Blocking the downloaded payload by analysing the malware before it is started. For example, the malware payload download (either the clear-text binary or the encrypted/encoded binary) can be seen in the proxy traffic, but no malware process starts.

5. The malware execution is blocked (no process create, load library).

6. There was a successful start by the dropped malware.

7. There was a successful start by the dropped malware, but after some time, all dropped malware was terminated and deleted ("malware starts, but blocked later").

The scoring of the results was calculated as the followings:

- If no malicious untrusted code was able to run on the endpoint, 5 points were given to the products. This can be achieved via blocking the exploit in step 1, 2 or 3.
- If malicious untrusted code run on the system (exploit shellcode, downloader code), but the final malware was not able start, 4 points were given to the product. This can be achieved via blocking the exploit in step 4 or 5. In a home environment, it does not really matter if exploit shellcode was able to run or not, but in an enterprise environment, investigating an incident can consume a lot of resources.
- If both the exploit shellcode (or downloader code) and the final malware was able to run, 0 points were given to the product.

We used this scoring for the following reasons:

- The scope of the test was exploit prevention and not the detection of malware running on the system.
- It is not possible to determine what kind of commands have been executed or what information exfiltrated by the malware. Data exfiltration cannot be undone or remediated.
- It cannot be determined if the malware exited because the endpoint protection system blocked it, or if malware quit because it detected monitor processes, virtualization, or quit because it did not find its target environment.



- Checking for malware remediation can be too time-consuming and remediation scoring very difficult in an enterprise environment. For example, in recent years we experienced several alerts that the endpoint protection system blocked a URL/page/exploit/malware, but still the malware was able to execute and run on the system. On other occasions, the malware code was deleted from the disk by the endpoint protection system, but the malware process was still running, or some parts of the malware were detected and killed, while others were not.
- In a complex enterprise environment multiple network and endpoint products protect the endpoints. If one network product alerts that malicious binary has been downloaded to the endpoint, administrators have to cross-check the alerts with the endpoint protection alerts, or do a full forensics investigation to be sure that no malware was running on the endpoint. This process can be time and resource consuming, which is why it is better to block the exploit before the shellcode starts.
- Usually the exploit shellcode is only a simple stage to download and execute a new piece of malware, but in targeted attacks, the exploit shellcode can be more complex. For example it can be used to leak information about the victim.

We believe that such zero-tolerance scoring helps enterprises to choose the best products, using simple metrics. Manually verifying the successful remediation of the malware in an enterprise environment is a very resource-intensive process and costs a lot of money. In our view, malware needs to be blocked before it has a chance to run, and no exploit shellcode should be able to run.



Scoring Of The In-The-Wild Malware Protection Results

The scoring of the in-the-wild malware protection is straightforward, whenever the malware started on the test machine, 0 point were given to the product, and whenever the malware was blocked by any stage before it was executed, 1 point was given. If a pop-up was shifting the decision to the user, 0.5 point were given.

False positive test

The false-alarm test in the Whole-Product Dynamic "Real-World" Protection Test consists of two parts: wrongly blocked domains (while browsing) and wrongly blocked files (while downloading/installing).

It is necessary to test both scenarios because testing only one of the two above cases could penalize products that focus mainly on one type of protection method, either URL filtering or on-access/behaviour/reputation-based file protection. In this test, 100 non-malicious applications have been used.

Wrongly blocked files (while downloading/installing)

One hundred different applications listed either as top downloads or as new/recommended downloads from various download portals are used in the false positive test. The applications were downloaded from the original software developers' websites (instead of the download portal host), saved to disk and installed to see if they are blocked at any stage of this procedure. Additionally, we included a few clean files that were encountered and disputed over the past months of the Real-World Protection Test.

The duty of security products is to protect against malicious sites/files, not to censor or limit the access only to well-known popular applications and websites. If the user deliberately chooses a high security setting, which warns that it may block some legitimate sites or files, then this may be considered acceptable. However, we do not regard it to be acceptable as a default setting, where the user has not been warned. As the test is done at points in time and FPs on very popular software/websites are usually noticed and fixed within a few hours, it would be surprising to encounter FPs with very popular applications. Due to this, FP tests which are done e.g. only with very popular applications, or which use only the top 50 files from whitelisted/monitored download portals would be a waste of time and resources. Users do not care whether they are infected by malware that affects only them, just as they do not care if the FP count affects only them. While it is preferable that FPs do not affect many users, it should be the goal to avoid having any FPs and to protect against any malicious files, no matter how many users are affected or targeted. Prevalence of FPs based on user-base data is of interest for internal QA testing of AV vendors, but for the ordinary user it is important to know how accurately its product distinguishes between clean and malicious files.



Test Procedure / Methodology

Exploit Test Setup

Testing Cycle for Each Test Case

- One default install Windows 7 64 Service Pack 1 virtual machine (VirtualBox) endpoint was created. (Windows 7 64-bit was the most popular OS for the target audience.) The default HTTP/HTTPS proxy was configured to point to a proxy running on a different machine. SSL/TLS traffic was not intercepted on the proxy.
- 2) The security of the OS was weakened by the following actions:
 - a) Microsoft Defender was disabled
 - b) Internet Explorer SmartScreen was disabled
 - c) Vulnerable software was installed, see "Software Installed" for details.
 - d) Windows Update was disabled
- 3) From this point, different snapshots were created from the virtual machine, several with different endpoint protection products and one with none. This procedure ensured that the base system was exactly the same in all test systems.

The following endpoint security suites, with the following configuration, were defined for this test:

- a) No additional protection
 - this snapshot was used to infect the OS and to verify the exploit replay
- b) Product 1 installed (Cylance Protec)
- c) Product 2 installed (Symantec Endpoint Protection)

The endpoint systems were installed with default configuration, potentially unwanted software removal was enabled, and if it was an option during install, cloud/community participation was enabled.

- 4) The exploit sources can be divided into two categories. In-the-wild threats and Metasploit. VBscript based downloaders and Office macro documents were also in scope, as these threats are usually not included in other test scenarios.
- 5) The virtual machine was reverted to a clean state and traffic was replayed by the proxy server. The replay meant that the browser was used as before, but instead of the original webservers, the proxy server answered the requests based on the recorded traffic. When the "replayed exploit" was able to infect the OS, the exploit traffic was marked as a source for the tests. This method guarantees that exactly the same traffic will be seen by the endpoint protection systems, even if the original exploit kit goes down during the tests.
- 6) After new exploit traffic was approved, the endpoint protection systems were tested. Before the exploit site was tested, it was verified that the endpoint protection had been updated to the latest version with the latest signatures and that every cloud connection was working. If there was a need to restart the system, it was restarted. In the proxy setup, unmatched requests were allowed to pass through and SSL/TLS was not decrypted to ensure AV connectivity. VPN was used during the test on



the host machine. When user interaction was needed from the endpoint protection (e.g. site visit not recommended, etc.), the block/deny action was chosen. When user interaction was needed from Windows, we chose the run/allow options. No other processes were running on the system, except the Process Monitor/Process Explorer from SysInternals and Wireshark (both installed to non-default directories).

- 7) After navigating to the exploit site, the system was monitored to check for new processes, loaded DLLs or C&C traffic.
- 8) The process went back to step 5. until all exploit site test cases were reached.

The following hardware was dedicated to the virtual machine:

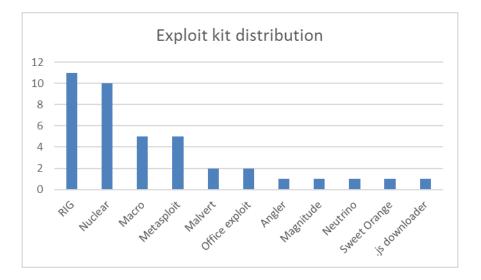
- 4 GB RAM memory
- 2 processors dedicated from AMD FX 8370E CPU
- 65 GB free space
- 1 network interface
- SSD drive

The VirtualBox host and guest system for the exploit test has been hardened in a way that common virtualization and sandbox detection techniques cannot detect the system as a test system.



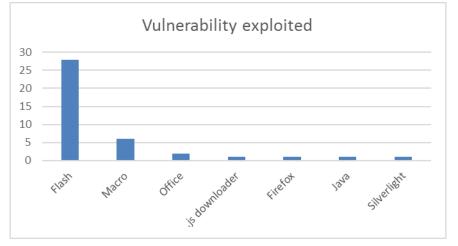
Analysis Of The Exploit Kits Used In The Exploit Test

The following graph displays the distribution of the detected exploit kits. The distribution is the result of choosing random malicious sites at the time of testing, in combination with Metasploit tests.



Analysis Of The Exploits Used In The Exploit Test

The following graph displays the distribution of the detected exploits.



In-Memory Malware In Exploit Tests

Not many comparative tests include in-memory (disk-less) malware. This is unfortunate, as many attacks include in-memory-malware, like Angler exploit kit or Metasploit. In-memory malware can bypass traditional AV protections, as there is no malware written to the hard disk, thus the malware is not checked at all. This threat can still be blocked before exploitation by URL block, analysing the JavaScript/HTML/SWF files, or even by blocking the exploit itself. Because the malware can be encrypted on the network level, it is not possible to detect the malware delivery by traditional methods. Also, the threat can be blocked after infection, when it starts to drop more malware onto the victim OS – which are traditional, persistent malware. In-memory malware is used by APT actors as well.



Real-World Protection Test Setup



The Whole-Product Dynamic "Real-World" Protection test is a joint project of AV-Comparatives and the University of Innsbruck's Faculty of Computer Science and Quality Engineering. It is partially funded by the Republic of Austria

Every potential test-case to be used in the test is run and analysed on a clean machine without antivirus software, to ensure that it is a suitable candidate. If the malware meets both of these criteria, the source URL is added to the list to be tested with security products. Any test cases which turn out not to be appropriate are excluded from the test set.

Every security program to be tested is installed on its own test computer. All computers are connected to the Internet. Each system is manually updated every day, and each product is updated before every single test case. Each test PC has its own external IP address. We make special arrangements with ISPs to ensure a stable Internet connection for each PC, and take the necessary precautions (with specially configured firewalls etc.) not to harm other computers (i.e. not to cause outbreaks).

Software Installed

The tests were performed under Microsoft Windows 7 Home Premium SP1 64-Bit with all updates till 11th of January 2016. Some further installed software includes and others not listed:

Vendor	Product	Version	Vendor	Product	Version
Adobe	Flash Player ActiveX	20.0.0.270	Microsoft	.NET Framework	4.5.2 (4.5.51209)
Adobe	Flash Player Plug In	20.0.0.267	Google	Chrome	47.0.2526.106
Adobe	Acrobat Reader	11.0.13	Microsoft	.NET Framework	4.5.1
Apple	QuickTime	7.79.80.05	Mozilla	Firefox	43.0.4
Microsoft	Internet Explorer	11.0.9600.18124	VideoLan	VLC Media Player	2.1.5

Settings

<u>We use every security suite with its default settings</u>. The Whole-Product Dynamic Protection Test aims to simulate real-world conditions as experienced every day by users. If user interactions are shown, we choose "Allow" or equivalent. If the product protects the system anyway, we count the malware as blocked, even though we allow the program to run when the user is asked to make a decision. If the system is compromised, we count it as user-dependent. We consider "protection" to mean that the system is not compromised. This means that the malware is not running (or is removed/terminated) and there are no significant/malicious system changes. An outbound-firewall alert about a running malware process, which asks whether or not to block traffic from the users' workstation to the Internet, is too little, too late and not considered by us to be protection.

Preparation For Every Testing Day

Every morning, any available security software updates are downloaded and installed, and a new base image is made for that day. Before <u>each</u> test case is carried out, the products have some time to download and install newer updates which have just been released, as well as to load their protection modules (which in several cases takes some minutes). In the event that a major signature update for a



product is made available during the day, but fails to download/install before each test case starts, the product will at least have the signatures that were available at the start of the day. This replicates the situation of an ordinary user in the real world.

Testing Cycle For Each Malicious URL

Before browsing to each new malicious URL we update the programs/signatures (as described above). New major product versions (i.e. the first digit of the build number is different) are installed once at the beginning of the month, which is why in each monthly report we only give the main product version number. Our test software monitors the PC, so that any changes made by the malware will be recorded. Furthermore, the recognition algorithms check whether the antivirus program detects the malware. After each test case the machine is reset to its clean state.

Protection

Security products should protect the user's PC. It is not very important at which stage the protection takes place. It could be while browsing to the website (e.g. protection through URL Blocker), while an exploit tries to run, while the file is being downloaded/created or when the malware is executed (either by the exploit or by the user). After the malware is executed (if not blocked before), we wait several minutes for malicious actions and also to give e.g. behaviour-blockers time to react and remedy actions performed by the malware. If the malware is not detected and the system is indeed infected/compromised, the process goes to "System Compromised". If a user interaction is required and it is up to the user to decide if something is malicious, and in the case of the worst user decision the system gets compromised, we rate this as "user-dependent". Because of this, the yellow bars in the results graph can be interpreted either as protected or not protected (it's up to each individual user to decide what he/she would probably do in that situation).

Due to the dynamic nature of the test, i.e. mimicking real-world conditions, and because of the way several different technologies (such as cloud scanners, reputation services, etc.) work, it is a matter of fact that such tests cannot be repeated or replicated in the way that e.g. static detection rate tests can. Anyway, we log as much data as reasonably possible to support our findings and results. Vendors are invited to provide useful log functions in their products that can provide the additional data they want in the event of disputes. After each testing month, manufacturers are given the opportunity to dispute our conclusion about the compromised cases, so that we can recheck if there were maybe some problems in the automation or with our analysis of the results.

In the case of cloud products, we can only consider the results that the products achieved in our lab at the time of testing; sometimes the cloud services provided by the security vendors are down due to faults or maintenance downtime by the vendors, but these cloud-downtimes are often not disclosed to the users by the vendors. This is also a reason why products relying too heavily on cloud services (and not making use of local heuristics, behaviour blockers, etc.) can be risky, as in such cases the security provided by the products can decrease significantly. Cloud signatures/reputation should be implemented in the products to complement the other local/offline protection features, but not replace them completely, as e.g. offline cloud services would mean the PCs are being exposed to higher risks.



Test Set For In-The-Wild Malware

In this specific test, we used URLs that pointed directly to malware executables; this causes the malware file to be downloaded, thus replicating a scenario in which the user is tricked by social engineering into following links in spam mails or websites, or installing some Trojan or other malicious software.

We use our own crawling system to search continuously for malicious sites and extract malicious URLs (including spammed malicious links). We also search manually for malicious URLs. If our in-house crawler does not find enough valid malicious URLs on one day, we have contracted some external researchers to provide additional malicious URLs (initially for the exclusive use of AV-Comparatives) and look for additional (re)sources.

In this kind of testing, it is very important to use enough test cases. If an insufficient number of samples are used in comparative tests, differences in results may not indicate actual differences in protective capabilities among the tested products¹. Our tests use more test cases (samples) per product and month than any similar test performed by other testing labs. Because of the higher statistical significance this achieves, we consider all the products in each results cluster to be equally effective, assuming that they have a false-positives rate below the industry average.



¹ Read more in the following paper: <u>http://www.av-comparatives.org/images/stories/test/statistics/somestats.pdf</u>

About the test-labs

AV-Comparatives

AV-Comparatives is a vendor-independent organization offering systematic testing that checks whether security software such as PC/Mac-based antivirus products and mobile security solutions lives up to its promises. Using one of the largest sample collections worldwide, AV-Comparatives create real-world environments for accurate security tool testing offering freely accessible results to individuals, media and scientific institutions. Certification by AV-Comparatives provides an official seal of approval for software performance which is globally recognized. Currently, the Real-World Protection Test is the most comprehensive and complex test available when it comes to evaluating real-life protection capabilities of antivirus software. For this purpose AV-Comparatives runs one of the world largest IT security testing frameworks in a data centre located in Innsbruck.

Members of AV-Comparatives give frequently talks at the major IT security conferences like Virus Bulletin, AVAR, EICAR, IEEE Malware Conference, WATeR, AMTSO, BSides, Ninjacon.

The methodology of AV-Comparatives' Real-World Protection Test has received the following awards and certifications, including:

- Constantinus Award given by the Austrian government
- **Cluster Award** given by the Standortagentur Tirol Tyrolean government
- **eAward** given by report.at (Magazine for Computer Science) and the Office of the Federal Chancellor
- Innovationspreis IT "Best Of" given by Initiative Mittelstand Germany

AV-Comparatives' Management System is ISO 9001:2008 certified. The certification has been received from TÜV Austria for the management system with scope "Independent Tests of Anti-Virus Software".

AV-Comparatives is the first certified EICAR Trusted IT-Security Lab.

The data centre where AV-Comparatives runs the test equipment is ISO 27001:2013 certified.







MRG Effitas

MRG Effitas is a UK based, independent IT security research organisation which focuses on providing cutting edge efficacy assessment and assurance services, supply of malware samples to vendors and the latest news concerning new threats and other information in the field of IT security.

MRG Effitas' origin began when the "Malware Research Group" was formed in 2009 by Sveta Miladinov, an independent security researcher and consultant. In June 2009, Chris Pickard, joined, bringing expertise in process and methodology design, gained in the business process outsourcing market.

The Malware Research Group rapidly gained a reputation as being the leading efficacy assessor in the browser and online banking space and due to increasing demand for its services, was restructured in 2011 and became "MRG Effitas" with the parent company "Effitas".

Today, MRG Effitas has a team of analysts, researchers and associates across EMEA, USA and China, ensuring a truly global presence.

Since its inception, MRG Effitas has focused on providing ground-breaking testing processes, realistically modelling real world environments in order to generate the most accurate efficacy assessments possible.

MRG Effitas is recognised by several leading security vendors as being the leading testing and assessment organisation in the online banking, browser security and cloud security spaces and has become the partner of choice.

Members of MRG Effitas give frequently talks at the major IT security conferences like Botconf, DEF CON, WATeR (AMTSO), Hacktivity, Hacker Halted etc.

Our professionals hold the following certifications: CISSP, OSCP, OSCE, GPEN, SLAE, SPSE, CPTS, CHFI, MCP, OSWP.



Copyright and Disclaimer

This publication is Copyright © 2016 by AV-Comparatives ® / MRG Effitas ®. Any use of the results, etc. in whole or in part, is ONLY permitted with the explicit written agreement of the management board of AV-Comparatives / MRG Effitas, prior to any publication. AV-Comparatives / MRG Effitas and its testers cannot be held liable for any damage or loss which might occur as a 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 liability for the correctness of the test results cannot be taken by any representative of AV-Comparatives / MRG Effitas. 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 / MRG Effitas and the testing methodologies please visit our website.

AV-Comparatives / MRG Effitas (February 2016)

