Whole Product Dynamic “Real-World” Protection Test

March-June 2012

Language: English
July 2012
Last revision: 16th July 2012

www.av-comparatives.org

Content

Introduction 2
Content 2
Test Procedure 4
Lab-Setup 4
Hardware and Software 4
Settings 5
Preparation for every Testing Day 5
Testing Cycle for each malicious URL 5
Test Set 6
Tested products 7
Test Cases 8
Results 8
Summary Results (March-June) 9
Award levels reached in this test 12
Copyright and Disclaimer 13

Introduction

The threat posed by malicious software is growing day by day. Not only is the number of malware programs increasing, also the very nature of the threats is changing rapidly. The way in which harmful code gets onto computers is changing from simple file-based methods to distribution via the Internet. Malware is increasingly infecting PCs through e.g. users deceived into visiting infected web pages, installing rogue/malicious software or opening emails with malicious attachments.

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

In spite of these new technologies, it remains very important that the signature-based and heuristic detection abilities of antivirus programs continue to be tested. It is precisely because of the new threats that signature/heuristic detection methods are becoming ever more important too. The growing frequency of zero-day attacks means that there is an increasing risk of malware infection. If this is not intercepted by "conventional" or "non-conventional" methods, the computer will be compromised, and it is only by using an on-demand scan with signature and heuristic-based detection that the malware can be found (and hopefully removed). The additional protection technologies also offer no means of checking existing data stores for already-infected files, which can be found on the file servers of many companies. Those new security layers should be understood as an addition to good detection rates, not as replacement.

In this test all features of the product contribute protection, not only one part (like signatures/ heuristic file scanning). So the protection provided should be higher than in testing only parts of the product. We would recommend that all parts of a product should be high in detection, not only single components (e.g. URL blocking protects only while browsing the web, but not against malware introduced by other means or already present on the system).

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 Austrian Government.

Test Procedure

Testing dozens of antivirus products with 100 URLs each per day is a lot of work which cannot be done manually (as it would be thousands of websites to visit and in parallel), so it is necessary to use some sort of automation.

Lab-Setup

Every security program to be tested is installed on its own test computer. All computers are connected to the Internet (details below). The system is frozen, with the operating system and security program installed. The entire test is performed on real workstations. We do not use any kind of virtualization. Each workstation has its own internet connection with its own external IP. We have special agreements with several providers (failover clustering and no traffic blocking) to ensure a stable internet connection. The tests are performed using a live internet connection. We took the necessary precautions (with specially configured firewalls, etc.) not to harm other computers (i.e. not to cause outbreaks).

Hardware and Software

For this test we used identical workstations, a control and command server and network attached storage.

	Vendor	Type	CPU	RAM	Hard Disk
Workstations	Dell	Optiplex 755	Intel Core 2 Duo	4 GB	80 GB
Control Server	Dell	Optiplex 755	Intel Core 2 Duo	8 GB	2 x 500 GB
Storage	Eurostor	ES8700-Open-E	Dual Xenon	32 GB	140 TB Raid 6

The tests are performed under Windows XP SP3 with updates of 1st March 2012. Some further installed vulnerable software includes:

Vendor	Product	Version
Adobe	Flash Player ActiveX	10.1
Adobe	Flash Player Plug-In	10.0
Adobe	Acrobat Reader	8.0
Apple	QuickTime	7.1
Microsoft	Internet Explorer	7.0

Vendor	Product	Version
Microsoft	Office Professional	2003
Microsoft	.NET Framework	4.0
Mozilla	Firefox	9.0.1
Sun	Java	6.0.140
VideoLAN	VLC Media Player	1.1.11

Initially we planned to test this year with a fully updated/patched system, but we had to switch back using older/vulnerable/unpatched OS and software versions due to lack of enough exploits in-the-field to test against. This should remind users to keep their systems and applications always up-to-date, in order to minimize the risk of getting infected through exploits using unpatched software vulnerabilities.

Settings

We use every security suite with its default (out-of-the-box) settings. Our whole-product dynamic protection test aims to simulate real-world conditions as experienced every day by users. If user interactions are required, we choose allow. If the system will be protected anyway, we count it as blocked even if there was first a user interaction. If the system gets 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 form 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. This ensures that even in the case the security product would not finish a bigger update during the day (products are being updated before each test case) or is not reachable, it would at least use the updates of the morning, as it would happen to the user in the real world.

Testing Cycle for each malicious URL

Before browsing to each new malicious URL/test-case we update the programs/signatures. New major product versions (i.e. the first digit of the build number is different) are installed once at the begin of the month, which is why in each monthly report we only give the product main version number. Our test software starts monitoring 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 reverted 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. This can either be while browsing to the website (e.g. protection through URL Blocker), while an exploit tries to run or while the file is being downloaded/created or while 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". Due to that, the yellow bars in the results graph can be interpreted either as protected or not protected (it's up to the user).

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 as reasonably possible to prove our findings and results. Vendors are invited to provide useful logs inside their products which can provide the additional data they want in case of disputes. Vendors were given after each testing month the possibility to dispute our conclusion about the compromised cases, so that we could recheck if there were maybe some problems in the automation or with our analysis of the results.

In the case of cloud products, we will only consider the results that the products had 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-downsides are often not disclosed/communicated to the users by the vendors. This is also a reason why products relying too much on cloud services (and not making use of local heuristics, behavior 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 and not replace them completely, as e.g. offline cloud services would mean the PCs being exposed to higher risks.

Test Set

We are focusing to include mainly current, visible and relevant malicious websites/malware that are currently out there and problematic to ordinary users. We are trying to include about 40-50% URLs pointing directly to malware (for example, if the user is tricked by social-engineering into follow links in spam mails or websites, or if the user is tricked into installing some Trojan or other rogue software). The rest are drive-by exploits - these usually are well covered by almost all major security products, which may be one reason why the scores look relatively high.

We use our own crawling system to search continuously for malicious sites and extract malicious URLs (including spammed malicious links). We also research 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 first exclusively to 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 among the tested products[1]. In fact, we consider even in our tests (with thousands of test-cases) products in the same protection cluster to be more or less equally good; as long as they do not wrongly block clean files/sites more than the industry average.

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

Comments

Most operating systems already include their own firewalls, automatic updates, and may even ask the user before downloading or executing files if they really want to do that, warning that downloading/executing files can be dangerous. Mail clients and web mails include spam filters too. Furthermore, most browsers include Pop-Up blockers, Phishing/URL-Filters and the ability to remove cookies. Those are just some of the build-in protection features, but despite all of them, systems can get infected anyway. The reason for this in most cases is the ordinary user, who may get tricked by social engineering into visiting malicious websites or installing malicious software. Users expect a security product not to ask them if they really want to execute a file etc. but expect that the security product will protect the system in any case without them having to think about it, and despite what they do (e.g. executing unknown files). We try to deliver good and easy-to-read test reports for end-users. We are continuously working on improving further our automated systems to deliver a better overview of product capabilities.

Tested products

The following products were tested in the official Whole-Product Dynamic "Real-World" Protection test series. In this type of test we usually include Internet Security Suites, although also other product versions fit (and are included/replaced on vendors request), because what is tested is the "protection" provided by the various products against a set of real-world threats.

Main product versions used for the monthly test-runs:

Vendor	Product	Version March	Version April	Version May	Version June
AhnLab	V3 Internet Security	8.0	8.0	8.0	8.0
Avast	Free Antivirus	7.0	7.0	7.0	7.0
AVG	Internet Security	2012	2012	2012	2012
Avira	Internet Security	2012	2012	2012	2012
Bitdefender	Internet Security	2012	2012	2012	2012
BullGuard	Internet Security	12.0	12.0	12.0	12.0
eScan	Internet Security	11.0	11.0	11.0	11.0
ESET	Smart Security	5.0	5.0	5.0	5.0
F-Secure	Internet Security	2012	2012	2012	2012
Fortinet	FortiClient Lite	4.3.3	4.3.3	4.3.3	4.3.3
G DATA	Internet Security	2012	2012	2013	2013
GFI	Vipre Internet Security	2012	2012	2012	2012
Kaspersky	Internet Security	2012	2012	2012	2012
McAfee	Internet Security	2012	2012	2012	2012
Panda	Cloud Free Antivirus	1.5.2	1.5.2	1.5.2	1.5.2
PC Tools	Internet Security	2012	2012	2012	2012
Qihoo 360	Internet Security	3.0	3.0	3.0	3.0
Sophos	Endpoint Security	10.0	10.0	10.0	10.0
Tencent	PC Manager	6.6	6.6	6.6	6.6
Trend Micro	Titanium Internet Security	2012	2012	2012	2012
Webroot	SecureAnywhere Complete	2012	2012	2012	2012

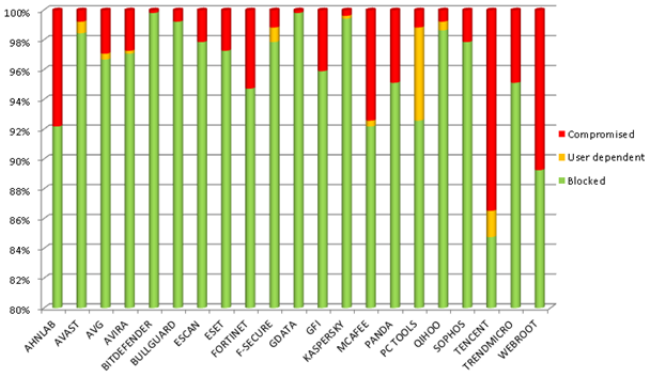
Test Cases

Test period	Test-cases
5th to 27th March 2012	512
2nd to 23rd April 2012	492
4th to 22nd May 2012	464
4th to 23rd June 2012	691
TOTAL	2159

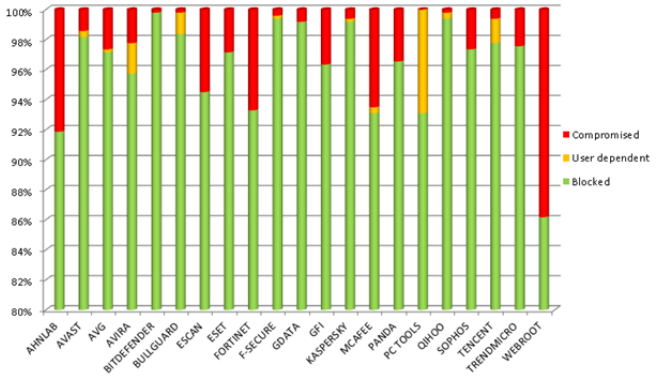
Results

Below you see an overview of the past single testing months. Percentages can be seen on the interactive graph on our website[2].

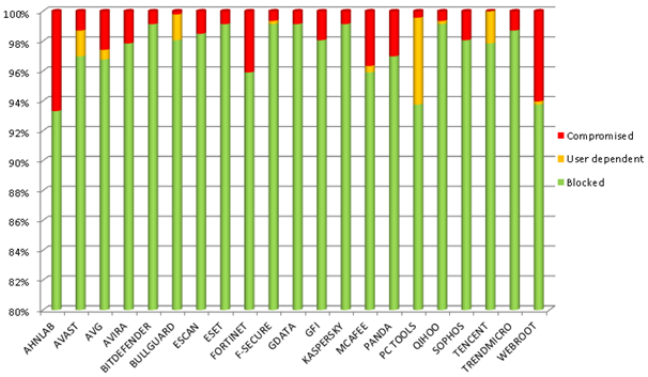
March 2012 – 512 test cases



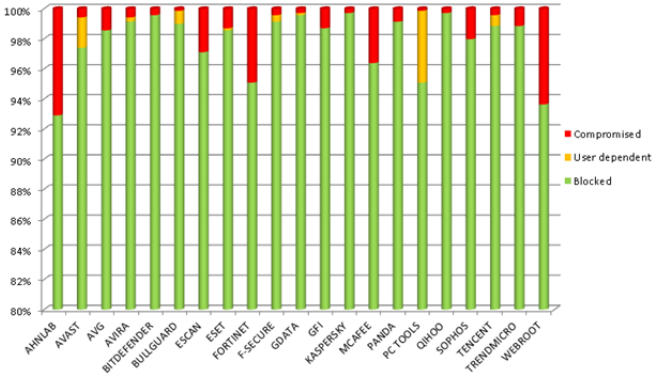
April 2012 – 492 test cases



May 2012 – 464 test cases



June 2012 – 691 test cases



We do not give in this report exact numbers for the single months on purpose, to avoid the little differences of few cases being misused to state that one product is better than the other in a given month and test-set size. We give the total numbers in the overall reports, where the size of the test-set is bigger, and more significant differences may be observed. Interested users who want to see the exact protection rates (without FP rates) every month can see the monthly updated interactive charts on our website[3].

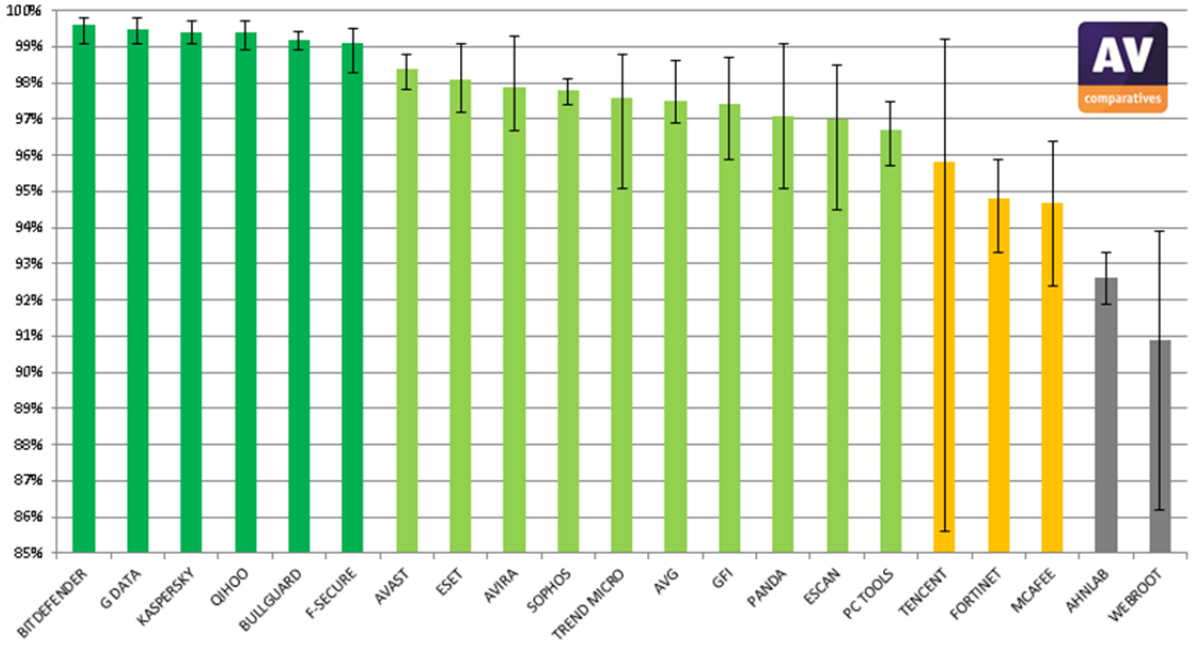
[2] http://www.av-comparatives.org/comparativesreviews/dynamic-tests
[3] http://chart.av-comparatives.org/chart2.php and http://chart.av-comparatives.org/chart3.php

Summary Results (March-June)

Test period: March – June 2012 (2159 Test cases)

	Blocked	User dependent	Compromised	PROTECTION RATE [Blocked % + (User dependent %)/2][4]	Cluster[5]
BitDefender	2150	-	9	99,6%	1
G DATA	2147	1	11	99,5%	1
Kaspersky	2146	2	11	99,4%	1
Qihoo	2143	6	10	99,4%	1
BullGuard	2131	21	7	99,2%	1
F-Secure	2135	10	14	99,1%	1
Avast	2110	28	21	98,4%	2
ESET	2117	1	41	98,1%	2
AVIRA	2107	13	39	97,9%	2
Sophos	2112	-	47	97,8%	2
Trend Micro	2108	-	51	97,6%	2
AVG	2103	6	50	97,5%	2
GFI	2102	-	57	97,4%	2
Panda	2097	-	62	97,1%	2
eScan	2094	-	65	97,0%	2
PC Tools	2024	126	9	96,7%	2
Tencent	2052	32	75	95,8%	3
Fortinet	2046	-	113	94,8%	3
McAfee	2041	6	112	94,7%	3
AhnLab	1999	-	160	92,6%	4
Webroot	1963	1	195	90,9%	4

The graph below shows the above protection rate (all samples), including the minimum and maximum protection rates for the individual months.



[4] User-dependent cases were given a half credit. Example: if a program gets 80% blocked-rate by itself, plus another 20% user-dependent, we give credit for half the user-dependent one, so it gets 90% altogether.

[5] Hierarchical Clustering Method: defining four clusters using average linkage between groups (Euclidian distance) on the protection rate (see dendogram on page 11).

Whole-Product "False Alarm" Test (wrongly blocked domains/files)

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 which focus mainly on one type of protection method, either e.g. URL/reputation-filtering or e.g. on-access / behaviour / reputation-based file protection.

a) Wrongly blocked domains (while browsing)

We used around two thousand randomly chosen popular domains. Blocked non-malicious domains/URLs were counted as false positives (FPs). The wrongly blocked domains have been reported to the respective vendors for review and should now no longer be blocked.

By blocking whole domains, the security products are not only risking causing distrust in their warnings, but also eventually causing potential financial damage (beside the damage on website reputation) to the domain owners, including loss of e.g. advertisement revenue. Due to this, we strongly recommend vendors to block whole domains only in the case where the domain's sole purpose is to carry/deliver malicious code, and to otherwise block just the malicious pages (as long as they are indeed malicious). Products which tend to block URLs based e.g. on reputation may be more prone to this and score also higher in protection tests, as they may block many unpopular/new websites.

b) Wrongly blocked files (while downloading/installing)

We used about one hundred different applications listed either as top downloads or as new/recommended downloads from about a dozen different popular download portals. The applications were downloaded from the original developer websites of the software (instead of the download portal host), saved to disk and installed to see if they get blocked at any stage of this procedure.

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 on very popular applications. Due to this, FP tests which are done e.g. *only* on 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 which 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.

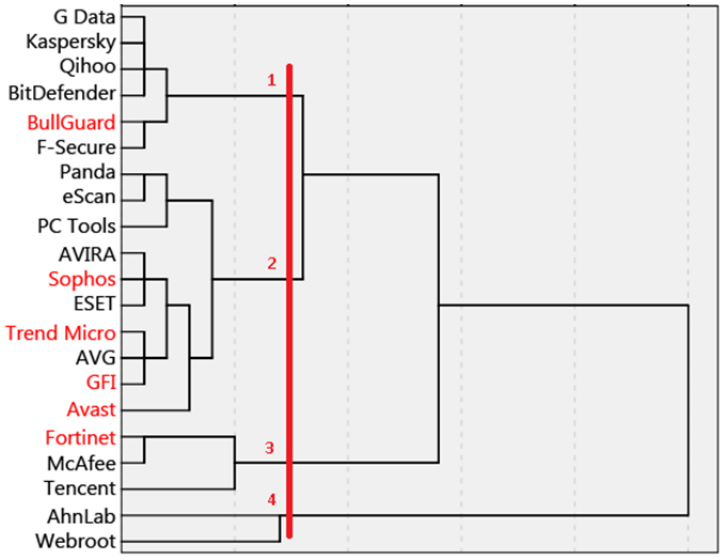
The below table shows the numbers of wrongly blocked domains/files:

	Wrongly blocked clean domains/files (blocked / user-dependent[6])	Wrongly blocked score[7]
AhnLab, AVG, BitDefender, eScan, G DATA, McAfee, Panda, Webroot	- / - (-)	-
F-Secure, Qihoo, Tencent,	1 / - (1)	1
Kaspersky	- / 3 (3)	1.5
ESET	2 / - (2)	2
PC Tools	4 / - (4)	4
AVIRA	5 / - (5)	5
	average (7)	*6*
Sophos, Fortinet	10 / - (10)	10
Avast	- / 28 (28)	14
Trend Micro	21 / - (21)	21
BullGuard	21 / 4 (25)	23
GFI	34 / - (34)	34

To determine which products have to be downgraded in our award scheme due to the rate of wrongly blocked sites/files, we backed up our decision by using a clustering method and by looking at the average scores. The following products with above average FPs have been downgraded: Avast, BullGuard, Fortinet, GFI, Sophos and Trend Micro.

Illustration of how awards were given

The dendogram (using average linkage between groups) shows the results of the hierarchical cluster analysis. It indicates at what level of similarity the clusters are joined. The red drafted line defines the level of similarity. Each intersection indicates a group (in this case 4 groups). Products which had above-average FPs are marked in red (and downgraded according to the ranking system on page 12).



[6] Although user dependent cases are extremely annoying (esp. on clean files) for the user, they were counted only as half for the "wrongly blocked rate" (like for the protection rate).
[7] Lower is better.

Award levels reached in this test

The awards are decided and given by the testers based on the observed test results (after consulting statistical models). The following awards are for the results reached in the Whole-Product Dynamic "Real-World" Protection Test:

AWARD LEVELS	PRODUCTS
	BitDefender G DATA Kaspersky Qihoo F-Secure
	BullGuard* ESET AVIRA AVG Panda eScan PC Tools
	Avast* Sophos* Trend Micro* GFI* Tencent McAfee
	Fortinet* AhnLab Webroot

** downgraded by one rank due to the score of wrongly blocked sites/files (FPs).*

Ranking system	**Protection score Cluster[8] 4**	**Protection score Cluster 3**	**Protection score Cluster 2**	**Protection score Cluster 1**
< ∅ FPs	Tested	Standard	Advanced	Advanced+
> ∅ FPs	Tested	Tested	Standard	Advanced

Expert users who do not care about wrongly blocked files/websites (false alarms) are free to rely on the protection rates on page 9 instead of our awards ranking which takes FPs in consideration.

[8] See protection score clusters on page 9.

Copyright and Disclaimer

This publication is Copyright © 2012 by AV-Comparatives e.V. ®. 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 e.V., prior to any publication. AV-Comparatives e.V. 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 e.V. 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. AV-Comparatives e.V. is a registered Austrian Non-Profit-Organization.

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

AV-Comparatives e.V. (July 2012)

Advertisement

Every second counts.
Who is attacking you? And how?

Even the best AV solution leaves you exposed
to zero-day and custom malware attacks.

Get real-time analysis.
No waiting for signature updates.

validEDGE
www.validedge.com

ValidEdge Malware Analysis Appliances
Free 30-day evaluation.

DETECT

ANALYZE

HEAL