Whole Product “Real-World” Dynamic Test

March-June 2011

Language: English
July 2011
Last revision: 19th August 2011

www.av-comparatives.org

Content

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 infected, 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 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. Some details of the test process cannot be disclosed, as the information could easily be misused by vendors to game the test systems.

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. This automation has been developed jointly by the Institute of Computer Science of the University of Innsbruck and AV-Comparatives.

Over the year we had to introduce several changes into the automated systems to prevent some AV vendors trying to "game" the system, as well as update/rewrite our tools due to unannounced changes in the security products, which made it harder to create automated systems. We kindly ask vendors to inform us in advance in the event of product changes which can affect automated testing systems.

Preparation for Test Series

Every antivirus program to be tested is installed on its own test computer (please note that the term "antivirus program" as used here may also mean a full Internet Security Suite). All computers are connected to the Internet (details below). The system is frozen, with the operating system and antivirus program installed.

Lab-Setup

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, an IBM Bladecenter and network attached storage (NAS).

	Vendor	Type	CPU	RAM	Hard Disk
Workstations	Fujitsu	E3521 E85+	Intel Core 2 Duo	4 GB	80 GB
BladeCenter	IBM	E Chassis	-	-	-
Blades	IBM	LS20	AMD Dual Opteron	8 GB	76 GB
NAS	QNAP	TS-859U-RP	Atom Dual Core	1 GB	16 TB Raid 6

The tests are performed under Windows XP SP3 with no further updates. Some further installed vulnerable software includes:

Vendor	Product	Version
Adobe	Flash Player ActiveX	10.1
Adobe	Flash Player Plug-In	10
Adobe	Acrobat Reader	8.0

Vendor	Product	Version
Microsoft	Internet Explorer	7
Microsoft	Office Professional	2003
Microsoft	.NET Framework	3.5
Sun	Java	6.0.140

Settings

We use every security suite with its default (out-of-the-box) settings. If user interactions are required, we will choose the default option. Our whole-product dynamic test aims to simulate real-world conditions as experienced every day by users. Therefore, if there is no predefined action, we will always use the same action where we consider the warning/message to be very clear and definitive. If the message leaves it up to the user, we will mark it as such, and if the message is very vague, misleading or even suggests trusting e.g. the malicious file/URL/behaviour, we will consider it to be a miss, as the ordinary user would. This year we will be stricter with required user decisions/interactions than last year. 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 antivirus software updates are downloaded and installed, and a new base image is made for that day. This ensures that even in the case the antivirus would not finish a bigger update during the day, it would at least use the updates of the morning, as 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 a 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 detection 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 "Malware Not Detected". 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 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 try to 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 proof/data they want in case of disputes. Vendors were given one to two weeks' time after each testing month 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 this is 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 etc.) can be risky, as in such cases the security provided by the products can decrease significantly. Cloud signatures/detection/reputation should be implemented in the products to complement the other protection features (like local real-time scan detection and heuristics, behaviour blockers, etc.) and not replace them completely, as e.g. offline cloud services would mean the PCs being exposed to higher risks.

Source of test cases

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 exclusively to AV-Comparatives. Although we have access to URLs shared between vendors and other public sources, we refrain from using these for the tests.

Test Set

We are not focusing on zero-day exploits/malware (although it is possible that they are also present in the URL pool), but mainly on current, visible and relevant malicious websites/malware that are currently out there and problematic to the ordinary users. We are trying to include about 30-50% 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/bigger part were exploits/drive-by downloads. These usually seem to be well covered by security products.

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.

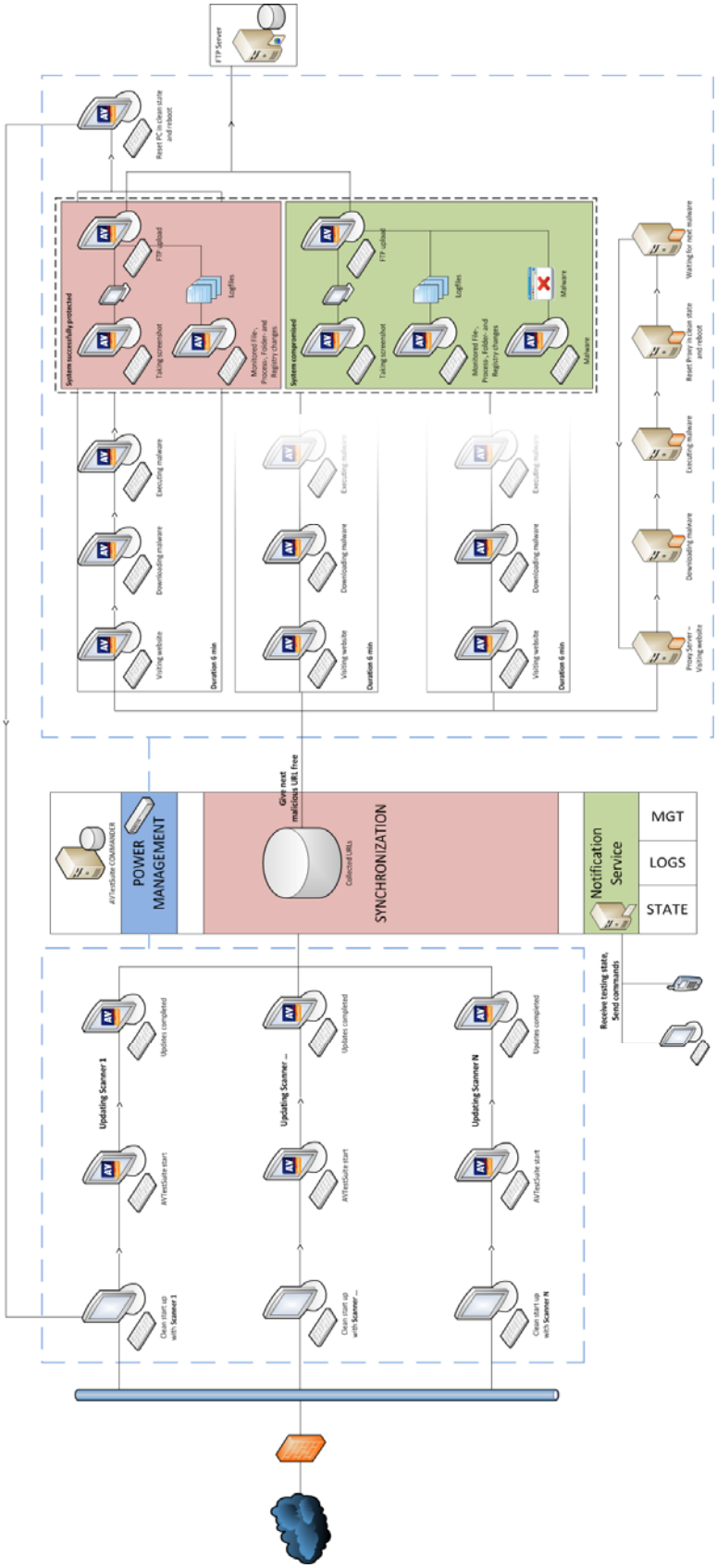
Tested products

The following products were tested in the official Whole-Product Dynamic main test series. In this type of test we usually include Internet Security Suites, although also other product versions[2] fit, 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
Avast	Internet Security	6.0	6.0	6.0	6.0
AVG	Internet Security	2011	2011	2011	2011
Avira	Premium Security Suite	10.0	10.0	10.0	10.0
Bitdefender	Internet Security	2011	2011	2011	2011
ESET	Smart Security	4.2	4.2	4.2	4.2
F-Secure	Internet Security	2011	2011	2011	2011
G DATA	Internet Security	2011	2011	2012	2012
K7	Total Security	11.1	11.1	11.1	11.1
Kaspersky	Internet Security	2011	2011	2011	2012
McAfee	Internet Security	2011	2011	2011	2011
Panda	Internet Security	2011	2011	2011	2012
PC Tools	Internet Security	2011	2011	2011	2011
Qihoo 360	Internet Security	2.0	2.0	2.0	2.0
Sophos	Endpoint Security	9.5	9.5	9.7	9.7
Symantec	Norton Internet Security	2011	2011	2011	2011
Trend Micro	Titanium Internet Security	2011	2011	2011	2011
Webroot	Internet Security Complete	7.0	7.0	7.0	7.0

[2] In the second half of 2011, we will include (at the request of the respective vendors) "Avast Antivirus Free", "Panda Cloud Antivirus" and "McAfee Total Protection" instead their respective Internet Security versions.

Diagrammatic Overview[3]



[3] As of August 2010. Some undisclosed enhancements/changes/additions have been implemented since then.

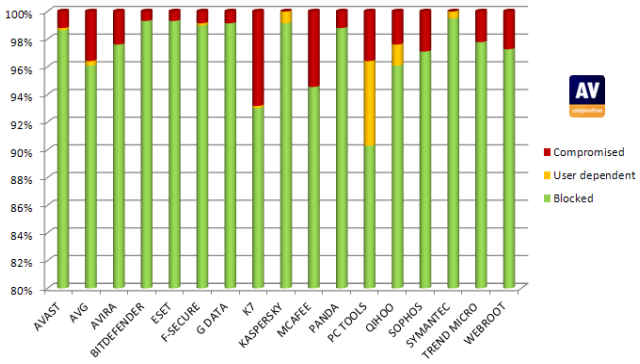
Test Cases

Test period	Test-cases
1st to 20th March 2011	587
4th to 26th April 2011	693
4th to 25th May 2011	622
6th to 23rd June 2011	578
TOTAL	2480

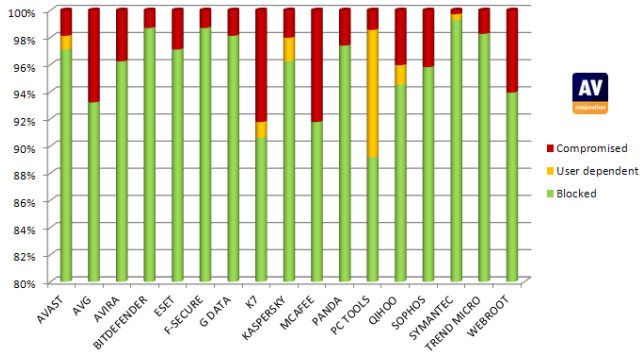
Results

Below you see an overview of the past single testing months. Percentages can be seen on the interactive graph on our website[4].

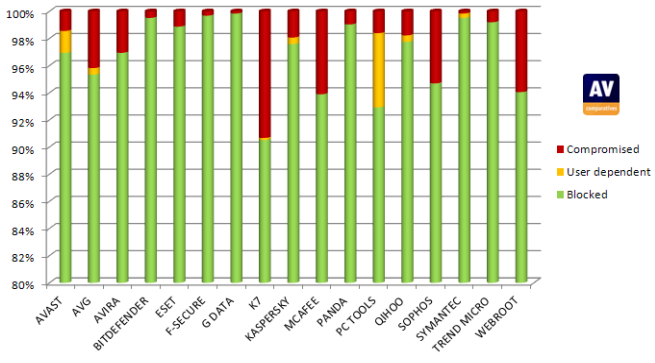
March 2011 – 587 test cases



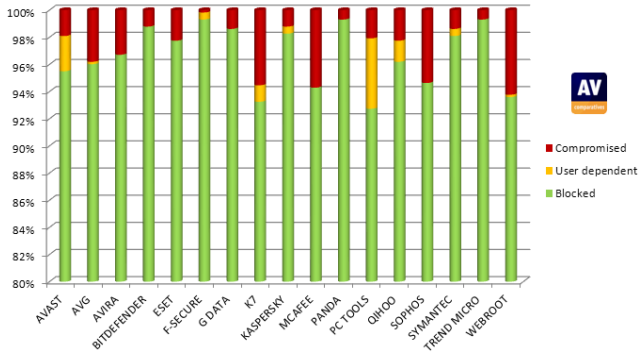
April 2011 – 693 test cases



May 2011 – 622 test cases



June 2011 – 578 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 summary, 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[5].

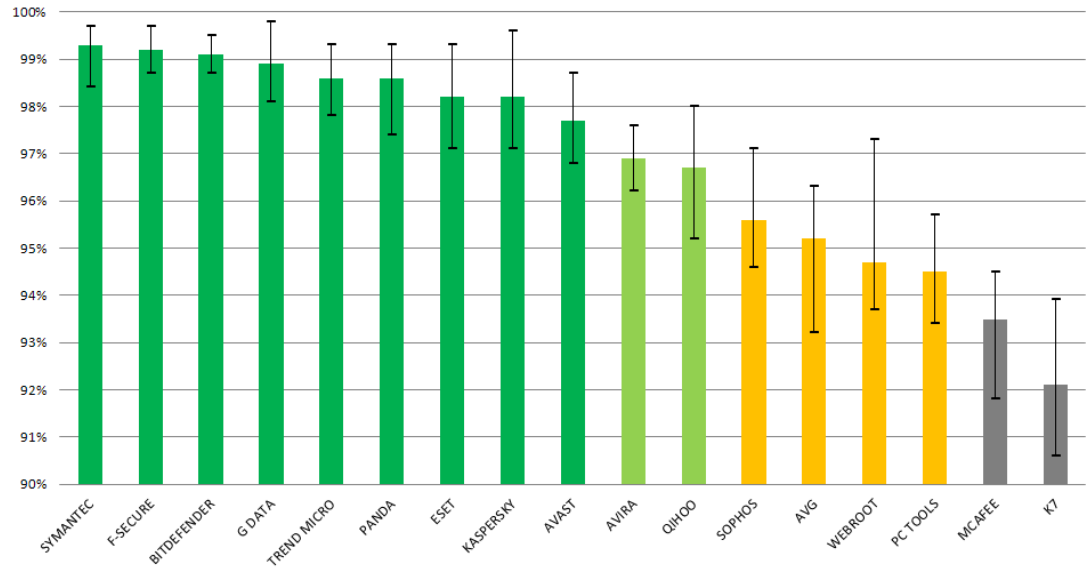
[4] http://www.av-comparatives.org/comparativesreviews/dynamic-tests
[5] http://chart.av-comparatives.org/chart2.php and http://chart.av-comparatives.org/chart3.php

Summary Results (March-June)

Test period: March – June 2011 (2480 Test cases)

	Blocked	User dependent	Compromised	PROTECTION RATE [Blocked % + (User dependent %)/2][6]	Cluster[7]
Symantec	2458	11	11	99,3%	1
F-Secure	2459	4	17	99,2%	1
Bitdefender	2457	-	23	99,1%	1
G DATA	2453	-	27	98,9%	1
Trend Micro	2446	-	34	98,6%	1
Panda	2445	-	35	98,6%	1
ESET	2436	-	44	98,2%	1
Kaspersky	2424	23	33	98,2%	1
Avast[8]	2407	33	40	97,7%	1
AVIRA	2402	-	78	96,9%	2
Qihoo	2383	31	66	96,7%	2
Sophos	2370	-	110	95,6%	3
AVG	2358	6	116	95,2%	3
Webroot	2348	1	131	94,7%	3
PC Tools	2262	165	53	94,5%	3
McAfee[9]	2320	-	160	93,5%	4
K7	2276	17	187	92,1%	4

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



[6] 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.

[7] Hierarchical Clustering Method: defining four clusters using average linkage between groups (Euclidian distance) on the protection rate. Statistically, products in same cluster don't significantly differ from each other.

[8] Avast! Free Antivirus would score the same as the Internet Security version.

[9] While "McAfee Internet Security" does include URL protection, it does not contain a proactive solution that actually blocks access to malicious sites. Rather it only warns users of potentially malicious sites through its browser traffic light system. However, McAfee does offer the additional SiteAdvisor Live component that can be purchased through the SiteAdvisor 'option' button. Alternatively, "McAfee Total Protection" includes McAfee SiteAdvisor Live by default. McAfee SiteAdvisor Live does have the capabilities to proactively block malicious URLs.

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

The false alarm test in the Whole-Product Dynamic 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. As announced in the previous report of 2010, we are now taking into account wrongly blocked domains/files when awarding the products.

a) Wrongly blocked domains (while browsing)

We used around two thousand randomly chosen popular domains listed in the Alexa[10] Top One Million sites[11]. Blocked non-malicious domains/URLs were counted as 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 over a hundred different applications listed either as top downloads or as new/recommended downloads from about 16 different popular download portals. The applications were downloaded from the websites (if original developer site was given, we used that source instead of the download portal host), saved to disk and installed to see if they get blocked at any stage of this procedure. Additionally, we included a few files whose status as malware had been disputed over the past months of the Dynamic Test. It is noteworthy that the status of some files we initially counted as missed malware[12] was disputed by several vendors, i.e. they regarded the files as clean/legitimate software. Also, the clean status of the very same files was disputed by other vendors[13] (and even by the same vendors who had initially insisted that they were clean) when we counted them as false positives (in one case, additional files were re-categorized by a vendor after the deadline/preview to come closer to the mean value). For the indisputable clean ones, some vendors tried to argue that they should not be counted as false positives, because although the files are clean, and their product blocks them (along with most other clean files from the same domains), only very few of their monitored users have downloaded those files or have

[10] http://www.alexa.com

[11] Currently (August 2011, http://www.domaintools.com/internet-statistics) over 130 million domains are active and about 150000 new Top-Level-Domains appear each day, which is far more than new unique malware appear each day. We used only domains/URLs from a pool of the top 1 million most popular websites (at time of testing).

[12] Correctly disputed grey applications were removed after agreed cross-verification.

[13] Furthermore, some vendors try to put pressure on some testing labs with the aim of not getting unfavourable results published, or getting tested/rated as they would wish (to score better in favour of their product) by legal threats or defamation attempts by their marketing departments. As an independent lab, we will continue to publish reports for the users irrespective of the outcome, and rate the products the way we consider to be correct/impartial.

those files on their machines. In fact, other telemetry/prevalence data shows varying/higher numbers of downloads by users. As it would be easy to get a high protection score by blocking everything that is rare or unknown, wrongly blocked files/sites have to be taken into account too. The user-base argument cannot be independently verified, and any vendor could claim that false alarms or missed malware is not affecting their user base. 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. None of the products blocked extremely popular applications. 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 testing of AV vendors, but for the ordinary user it is important to know how accurately its product distinguishes between clean and malicious files. Although some vendors gave some of the FPs a very low prevalence rating among their user base, the real mean value (according to other clouds/telemetry data with different user-bases) of the highest reported current prevalence for the encountered file FPs are in the several thousands; the median value is 1000.

	Wrongly blocked clean domains/files (blocked / user-dependent[14])	**Wrongly blocked score[15]**
AVG	5 / - (5)	5
Bitdefender	7 / - (7)	7
ESET	7 / 1 (8)	7.5
Kaspersky	5 / 7 (12)	8.5
Qihoo, K7	10 / - (10)	10
G DATA, McAfee	12 / - (12)	12
Avast	12 / 6 (18)	15
AVIRA, Panda	16 / - (16)	16
F-Secure, Sophos, Trend Micro	17 / - (17)	17
	average (27)	*27*
Symantec	24 / 7 (31)	27.5
PC Tools	47 / 11 (58)	52.5
Webroot	207 / - (207)	207

To determine which products have to be downgraded in our award scheme due to the rate of wrongly blocked sites/files[16], we consulted and backed up our difficult decision by using a clustering method, and by looking at the average scores. The following products (with above average FPs) had to be downgraded: Symantec, PC Tools and Webroot.

[14] Although user dependent cases are extremely annoying (esp. on clean files) for the user, they were this time counted only as half for the "wrongly blocked rate" (like for the protection rate).

[15] Lower is better.

[16] Some vendors may consider relatively new files coming from websites that they e.g. do not yet have in their whitelist as being suspicious, esp. if their cloud has no or not much info about those files. This works well for blocking malicious files and increasing protection, but unfortunately it also leads to a higher degree of wrongly blocked clean files, until the popularity/reputation/information about the files improves.

Certification levels reached in this test

We use a ranking system (Tested, STANDARD, ADVANCED and ADVANCED+) to summarize the results (protection and "accuracy"). Overviews of levels reached in previous main tests can be found on our website[17]. The awards are decided and given by the testers based on the observed test results (after consulting statistical models).

The following certification levels are for the results reached in the Whole-Product Dynamic Test:

CERTIFICATION LEVELS	PRODUCTS (randomly ordered in each group)
	F-Secure BitDefender G DATA Trend Micro Panda ESET Kaspersky Avast
	Symantec* AVIRA Qihoo
	Sophos AVG
	Webroot* PC Tools* McAfee K7

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

Simplified[18] system to illustrate ranking model	**Protection score Cluster[19] 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 10 instead of our awards ranking which takes FPs in consideration.

[17] http://www.av-comparatives.org/comparativesreviews/main-tests/summary-reports

[18] We look mainly on the distance between the groups (clusters), but the mean is easier to illustrate to readers.

[19] See protection score clusters on page 10.

Copyright and Disclaimer

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

AV-Comparatives e.V. (August 2011)