Whole Product Dynamic Test

August-November 2010

Language: English
December 2010
Last revision: 10th December 2010

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 in visiting infected web pages, installing rogue/malicious software or open 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 behavior-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 ability of protection should be higher than in testing only parts of the product. We would recommend that all parts of a product would 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 project of AV-Comparatives and the University of Innsbruck, faculty of Computer Science and Quality Engineering. It is partially supported by the Austrian Government. Some details about the test process cannot be disclosed, as it could be easily misused by vendors to game the test systems.

Test Procedure

Testing hundreds of URL's a day with dozens of antivirus programs makes a total of thousands URL tests and only a high degree of automation makes this possible. This automation has been developed jointly with the Institute of Computer Science of the University of Innsbruck and AV-Comparatives.

Over the year we had to introduce several changes in the automated systems to circumvent and also prevent some AV vendors trying to "game" the system, as well as update/rewrite our tools due unannounced changes in the security products which made it harder to create automated systems. Due that, the start of our whole-product-dynamic test started with some delay. We kindly ask vendors to inform us in advance in case 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, each with its own external IP address. 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 not blocking any traffic) to ensure a stable internet connection. The tests are performed using a live internet connection. We took the necessary precautions (with special configured firewalls, etc.) not to harm others (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 Core2Duo	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. 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 suggesting trusting e.g. the malicious file/URL/behavior, we will consider it as a miss, as the ordinary user would. We consider a protection if 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 out-bound-firewall alert about a running malware process, which asks whether to block traffic form the users' workstation to the internet is too little, too late and not considered as protection by us.

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, like it would happen to the user in the real-world.

Testing Cycle for each malicious URL

First of all, there is researching. With our own crawler we are searching the web constantly for malicious sites. We are not focusing on zero-day malware/exploits (although it is possible that they are also present in the URL pool); we are looking for malicious websites that are currently out there and being a threat to the ordinary users. Before browsing to each new malicious URL/test-case we update the programs/signatures. New major product versions are installed once a month, that's 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. behavior-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 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 to mimic real-world conditions and due the way several different technologies work (like cloud scanners, reputation services, etc.), it is a matter of fact that such tests cannot be repeated or replicated like e.g. static detection rate tests. Anyway, we are trying 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 them with the additional proof/data they want. 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 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 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 increase the other protection features (like local real-time scan detection and heuristics, behavior-blockers, etc.) and not replace them completely, as e.g. offline cloud services mean the PC's may be 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, but on current and relevant malware that is currently out there and problematic to users. We are trying to include about 30-50% URLs pointing directly to malware. For example, if the user is tricked by social-engineering to 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. Those seem to be usually 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].

Comments

Most operating systems already include own firewalls, automatic updates and may even prompt 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 possibility to remove cookies.

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

Those are just some of the build-in protections, but despite all of them, systems can get infected anyway. The reason for this is 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 (i.e. executing unknown files / malware). We try to keep in mind the interests of the users and deliver good and easy-to-read test reports. We are continuously working on improving further our automated systems to deliver a better overview about product capabilities.

Tested products

The following products take part in the official Whole-Product-Dynamic main test-series[2]. We may test also other products which are not part of the main test-series, but only separately and for a limited time-period. In this type of test we usually included Internet Security Suites, although also other product versions would 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:

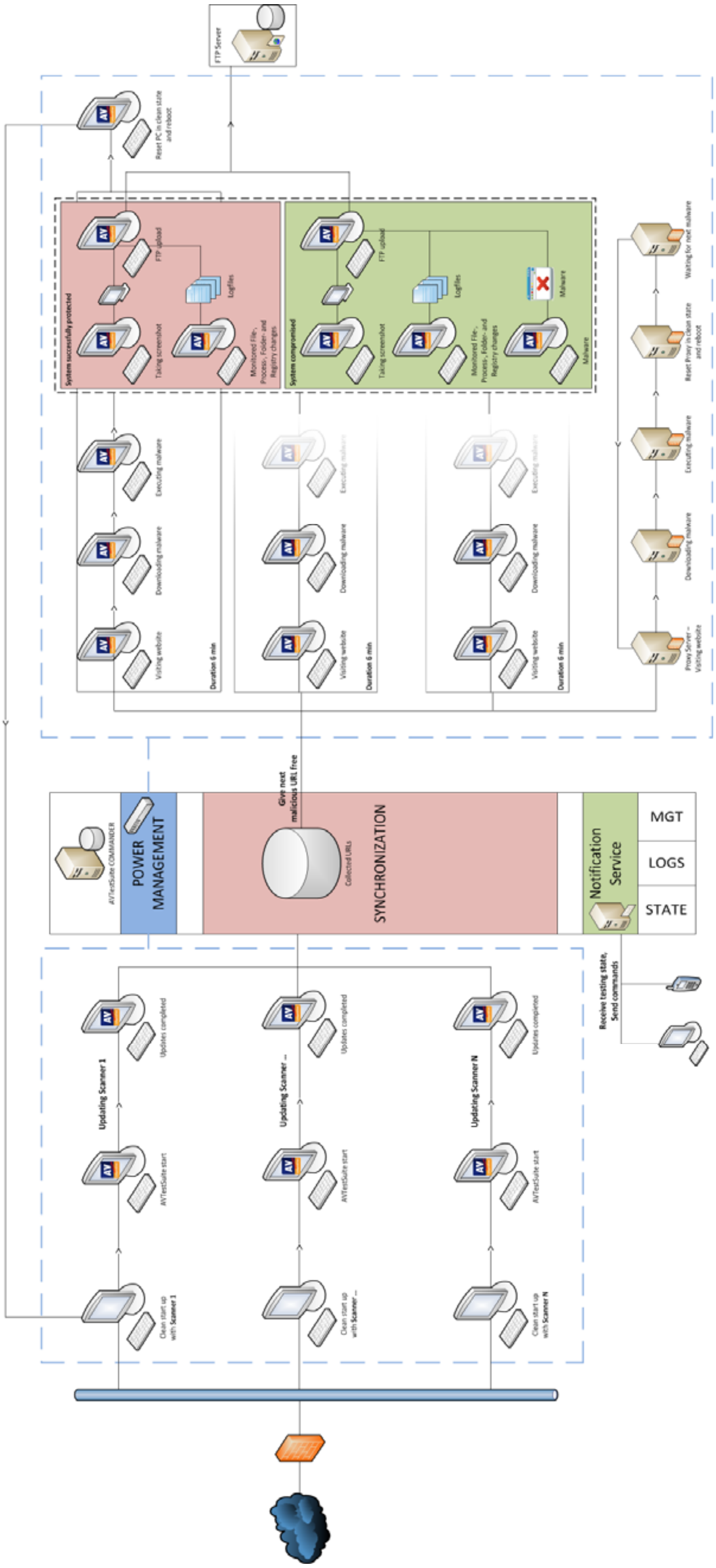
Vendor	Product	Version August	Version September	Version October	Version November
Avast	Internet Security	5.0	5.0	5.0	5.0
AVG	Internet Security	9.0	9.0	10.0	10.0
Avira	Premium Security Suite	10	10	10	10
BitDefender	Internet Security	2010	2011	2011	2011
ESET	Smart Security	4.2	4.2	4.2	4.2
F-Secure	Internet Security	2010	2011	2011	2011
G Data	Internet Security	2011	2011	2011	2011
Kaspersky	Internet Security	2011	2011	2011	2011
Kingsoft	Internet Security Plus	2010	2010	2011	2011
Norman	Security Suite Pro	8.0	8.0	8.0	8.0
Panda	Internet Security	2011	2011	2011	2011
PC Tools	Internet Security	2010	2011	2011	2011
Symantec	Norton Internet Security	2011	2011	2011	2011
Trend Micro	Titanium Internet Security	2010[3]	2011	2011	2011

Test Cases

Test period	Test-cases
10th to 26th August 2010	**304**
7th to 25th September 2010	**702**
7th to 21st October 2010	**454**
8th to 23rd November 2010	**508**
TOTAL	**1968**

[2] McAfee was not included due to miscommunication regarding their participation.
[3] Trend Micro Internet Security 2010.

Diagrammatic Overview[4]

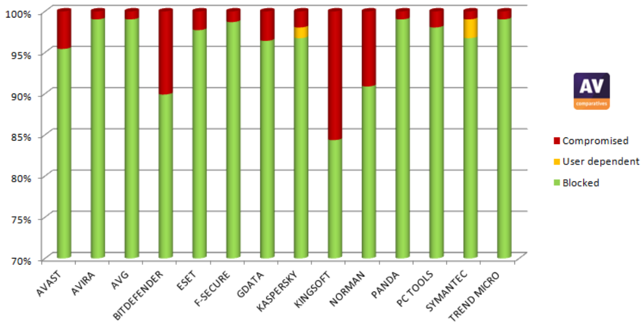


[4] As of August 2010. Further enhancements, changes or additions may have been implemented since then.

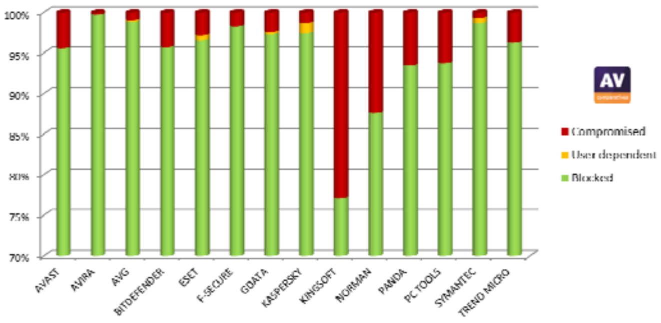
Results

Below you see an overview of the past single testing months. Percentages can be seen on the interactive graph on our website[5].

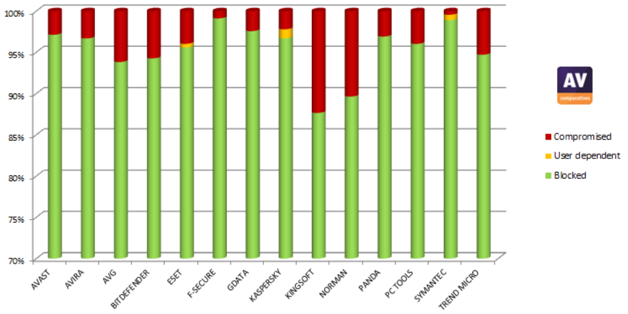
August 2010 – 304 test cases



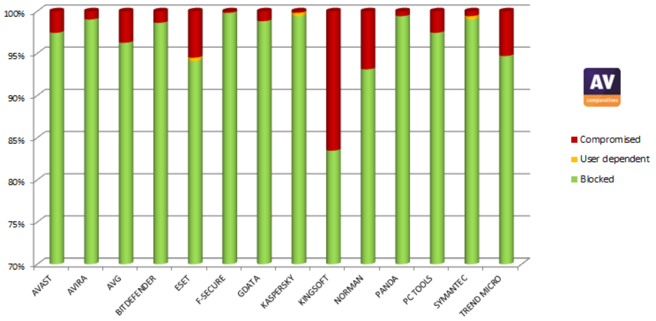
September 2010 – 702 test cases



October 2010 – 454 test cases



November 2010 – 508 test cases



We do not give in this report exact numbers for the single months on purpose, to avoid that little differences of 1-2 cases are misused to state that one product is better than the other on 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.

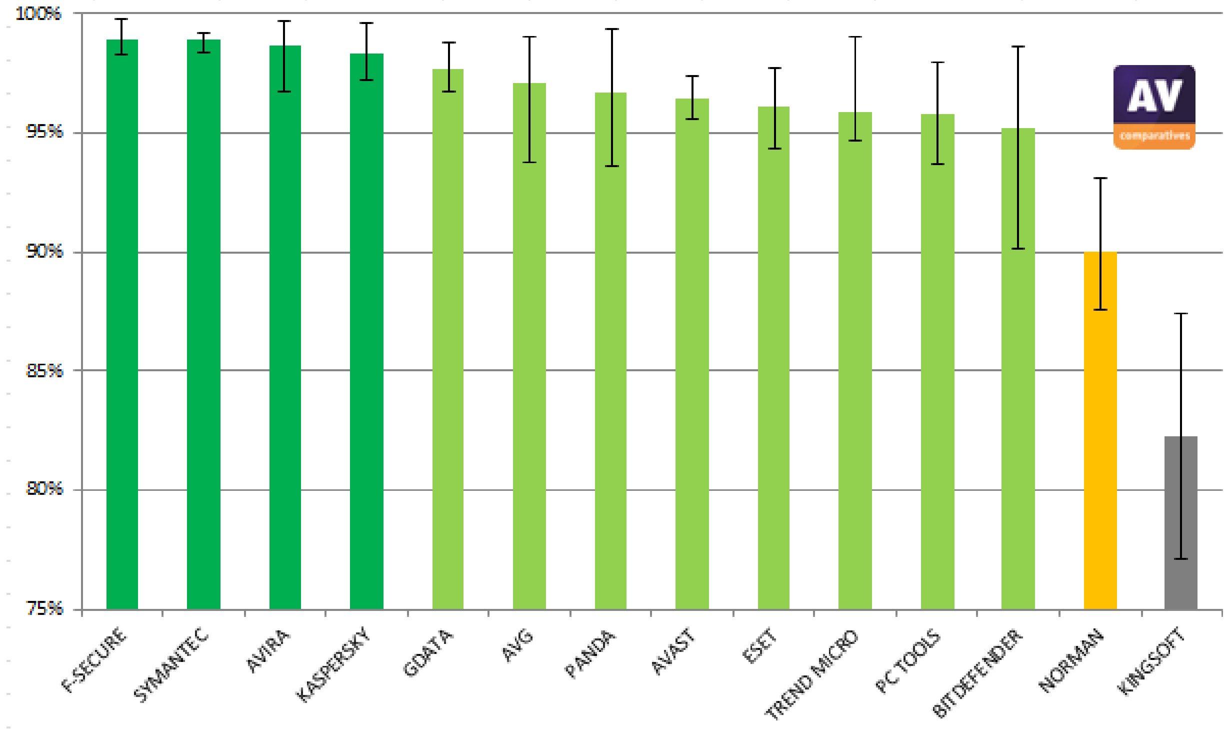
[5] http://www.av-comparatives.org/comparativesreviews/dynamic-tests

Summary Results (August-November)[6]

Test period: August – November 2010 (1968 Test cases)

	Blocked	User dependent	Compromised	PROTECTION RATE [Blocked % + (User dependent %)/2][7]	Cluster[8]
F-Secure	1946	-	22	98,9%	1
Symantec	1936	19	13	98,9%	1
AVIRA	1943	-	25	98,7%	1
Kaspersky	1925	19	24	98,3%	1
G DATA	1922	2	44	97,7%	2
AVG	1910	1	57	97,1%	2
Panda	1903	-	65	96,7%	2
Avast	1898	-	70	96,4%	2
ESET	1887	8	73	96,1%	2
Trend Micro	1888	-	80	95,9%	2
PC Tools	1886	-	82	95,8%	2
BitDefender	1874	-	94	95,2%	2
Norman	1771	-	197	90,0%	3
Kingsoft	1620	-	348	82,3%	4

The graph below shows the above protection rate over all samples, including the minimum and maximum protection rates for the single months.



[6] For detailed results of each month, please have a look at the graphical overview on our website.

[7] 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.

[8] Hierarchical Clustering Method: defining four clusters using average linkage between groups (Euclidian distance) on the protection rate.

Whole-Product False Alarm Test

The false alarm test in the Whole-Product-Dynamic test consists of two parts:

a) False Alarms on domains (while browsing)
b) False Alarms on 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/behavior/reputation-based file protection.

a) False Alarms on domains (while browsing)

For this False Alarm test we used domains listed in the Google Top1000[9] sites list of August 2010. We tested against those Top-Level-Domains twice, in September and in October. Non-malicious domains which were blocked at any time (either September or October) were counted as FPs (as they should never have been blocked). All below websites are among the most popular websites on the web (ranked on Alexa[10] between place ~300 and ~3000 worldwide)[11].

The domains below have been reported to the respective vendors for review and are now no longer blocked. We do not display the domains as we do not know if in future they may be still clean (and we also want to avoid making publicity for those domains).

By blocking the whole domains like in the cases below, the security products are causing potential financial damage (beside the damage on website reputation) to the domain owners, including loss of e.g. ads revenue. Due that, 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).

From the tested vendors, the following vendors had FPs on the tested domains during the testing period:

F-Secure	**1 FP**	**http://www.[redacted].com**
G DATA	**1 FP**	**http://www.[redacted].com**
Panda	**1 FP**	**http://www.[redacted].com**
PCTOOLS	**4 FPs**	**http://www.[redacted].com** **http://www.[redacted].com** **http://www.[redacted].com** **http://www.[redacted].com**
Symantec	**1 FP**	**http://www.[redacted].pl**
Trend Micro	**4 FPs**	**http://www.[redacted].com** **http://www.[redacted].com** **http://www.[redacted].com** **http://www.[redacted].com**

[9] http://www.google.com/adplanner/static/top1000

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

[11] Currently (December 2010, http://www.domaintools.com/internet-statistics) about 125 million domains are active and about 100000 new Top-Level-Domains appear each day, which is far more than new unique malware appear each day.

Some few more websites were blocked by various products, but not counted as FPs here this time. Those cases were mainly websites or download portals currently still hosting/promoting also some adware or unlicensed software etc. Many products continue to block websites even when they are no longer malicious and have already been cleaned up for some time. This happens also with popular websites, but of course even more with less popular/prevalent websites, with the risk of turning the security products into a web censoring tool which goes too far in blocking websites (based on what the security vendor considers being a risk or potentially unwanted content for the user). It would be much better if the product were only to block the access to the malicious part/file instead of a whole website/domain which is not malicious by itself (e.g. not containing any drive-by/exploits etc.), unless the user wants and enables e.g. a parental control setting or similar. 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 block many unpopular and strange looking websites. A further option for future FP testing could be to use such URLs which are discarded as clean or down during malware testing.

At the moment the AV industry is discussing about what/when and under which circumstances a blocked website which is not or no longer malicious by itself can be considered as a "false alarm", as opinions are varying even among vendors. We will look at the outcome of that discussion and consider it if this makes sense also from a user perspective.

b) False Alarms on files (while downloading/installing)

For this False Alarm test we used software listed as Top 100 Freeware downloads in 2010 of a popular German download portal[12]. We may change the used download portals and clean site sources every time (and maybe also no longer disclose which portals/lists were used), in order to avoid that vendors focus on whitelisting/training mainly against those sites/lists/sources.

We tested only with 100 applications, as this test was done manually in order to install the programs completely and also use them afterwards to see if they get blocked. We may automate also this type of FP testing in order to get bigger test-sets in future.

None of the products had a false alarm on those very popular applications. There were some firewall alerts, but as we do not consider firewall alerts (for programs trying to access the internet) as protection in the dynamic tests, we are also not considering them as FPs. It would be surprising to encounter FPs on very popular software in the case of well-known and internationally used AV's, especially as the test is done at one point in time and FPs on very popular software are noticed and fixed within few hours. Probably it would make more sense to test against lower-prevalence software. We observed accidentally also some FPs on less popular software/websites, but have not included them this time as vendors do not see them as a big issue. If you think different, please let them (not us, as we know already!) know what you as a user think about FPs that happen to you on less popular files and websites.

As we do not yet have much experience about FP rates in Whole-Product-Dynamic testing, we are not considering the FPs in the awards this time, but we may give lower awards to products which will have FPs (or many user interactions) in future Whole-Product-Dynamic Tests.

[12] http://www.pcwelt.de

Certification levels reached in this test

We provide a 3-level-ranking-system (STANDARD, ADVANCED and ADVANCED+). Overviews of levels reached in previous main tests can be found on our website[13]. 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
AV comparatives ADVANCED+ ★★★ DYNAMIC PROTECTION TEST 2010	F-Secure Symantec AVIRA Kaspersky
AV comparatives ADVANCED ★★ DYNAMIC PROTECTION TEST 2010	G DATA AVG Panda Avast ESET PC Tools Trend Micro BitDefender
AV comparatives STANDARD ★ DYNAMIC PROTECTION TEST 2010	Norman
AV comparatives TESTED DYNAMIC PROTECTION TEST 2010	Kingsoft

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

Copyright and Disclaimer

This publication is Copyright © 2010 by AV-Comparatives e.V. ®. 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 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 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 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. (December 2010)

validEDGE