Whole Product Dynamic “Real-World” Protection Test

July-November 2016

Language: English
December 2016
Last revision: 12th December 2016

www.av-comparatives.org

Content

Introduction

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 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, reputation systems 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 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. 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.

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. Even with all the protection features available, the growing frequency of zero-day attacks means that some computers will inevitably become infected. As signatures can be updated, they provide the opportunity to recognize and remove malware which was initially missed by the security software. The newer, "non-conventional" protection technologies often 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 a replacement.

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.

The methodology of our 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

Test Procedure

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

Lab Setup

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

The tests were performed under Microsoft Windows 7 Home Premium SP1 64-Bit, with updates as of 1st July 2016. Some further installed software includes:

Vendor	Product	Version
Adobe	Flash Player ActiveX	20.0
Adobe	Flash Player Plug-In	20.0
Adobe	Acrobat Reader	11.0
Apple	QuickTime	7.79
Microsoft	Internet Explorer	11.0

Vendor	Product	Version
Microsoft	Office Home	2013
Microsoft	.NET Framework	4.5.2
Mozilla	Firefox	43.0.4
Oracle	Java	1.8.0.66
VideoLAN	VLC Media Player	2.1.5

The use of more up-to-date third-party software and an updated Microsoft Windows 7 SP1 64-Bit makes it much harder to find exploits in-the-field for the test. Users should always keep their systems and applications up-to-date, in order to minimize the risk of being infected through exploits which use unpatched software vulnerabilities.

Settings

We use every security suite with its default settings. Our 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 each 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

We aim to use visible and relevant malicious websites/malware that are currently out there, and present a risk to ordinary users. We usually try to include as many working drive-by exploits as we find –these are usually well covered by practically all major security products, which may be one reason why the scores look relatively high. The rest are URLs that point 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 products[1]. 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: http://www.av-comparatives.org/images/stories/test/statistics/somestats.pdf

Tested products

For this test, we normally use the Internet security suite, as any protection features that prevent the system from being compromised can be used. However, a vendor can choose to enter their basic antivirus product instead, if they prefer. The main versions of the products tested in each monthly test run are shown below:

Vendor	Product	Version July	Version August	Version September	Version October	Version November
Avast	Free Antivirus	12.1	12.2	12.3	12.3	12.3
AVG	Internet Security	2016	2016	2017	2017	2017
Avira	Antivirus Pro	15.0	15.0	15.0	15.0	15.0
Bitdefender	Internet Security	20.0	2016	2016	2017	2017
BullGuard	Internet Security	16.0	16.0	16.0	16.0	16.0
Emsisoft	Anti-Malware	11.9	11.10	11.10	11.10	12.0
eScan	Internet Security	14.0	14.0	14.0	14.0	14.0
ESET	Smart Security	9.0	9.0	9.0	9.0	10.0
F-Secure	Safe	2016	2016	2016	2016	2017
Fortinet	FortiClient (with FortiGate)[2]	5.4	5.4	5.4	5.4	5.4
Kaspersky Lab	Internet Security	2016	2017	2017	2017	2017
Lavasoft	Ad-Aware Pro Security	11.11	11.12	11.12	11.12	11.12
McAfee	Internet Security	18.0	18.0	18.0	18.0	19.0
Microsoft	Security Essentials	4.9	4.9	4.9	4.10	4.10
Quick Heal	Total Security	17.0	17.0	17.0	17.0	17.0
Sophos	Endpoint Security and Control	10.6	10.6	10.6	10.6	10.6
Tencent	PC Manager (English)	11.4	11.6	11.6	11.6	11.6
ThreatTrack	VIPRE Internet Security Pro	9.3	9.3	9.3	9.3	9.3
Trend Micro	Internet Security	10.0	10.0	10.0	10.0	11.0

Test Cases

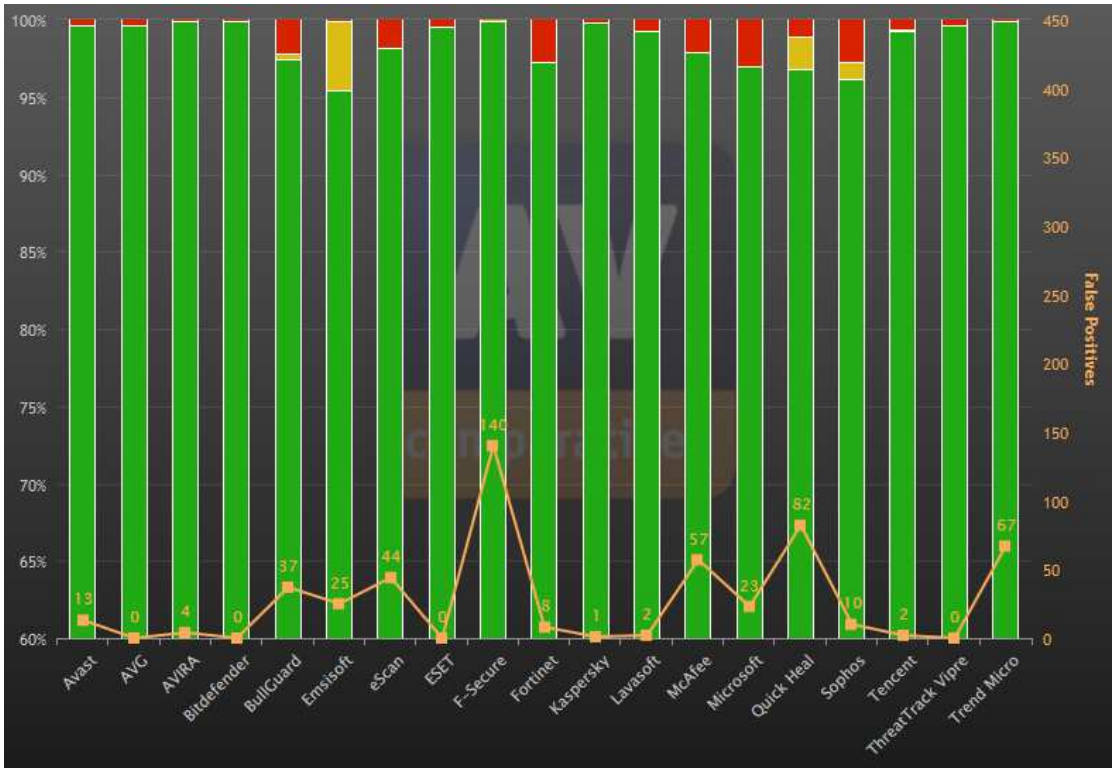
Test period	Test cases
1st to 26th July 2016	355
1st to 24th August 2016	346
1st to 25th September 2016	339
3rd to 24th October 2016	250
2nd to 25th November 2016	329
TOTAL	**1619**

[2] The cloud-based behaviour-analysis feature of Fortinet is only available to enterprises customers who also purchased a FortiGate.

Summary Results (July-November)

Test period: July – November 2016 (1619 Test cases)[3]

	Blocked	User dependent	Compromised	PROTECTION RATE [Blocked % + (User dependent %)/2][4]	Cluster[5]
F-Secure	1618	1	-	99.9%	1
Avira, Bitdefender, Trend Micro	1618	-	1	99.9%	1
Kaspersky Lab	1615	-	4	99.8%	1
ThreatTrack	1613	-	6	99.6%	1
Avast, AVG	1612	-	7	99.6%	1
ESET	1611	-	8	99.5%	1
Tencent	1608	1	10	99.4%	1
Lavasoft	1607	-	12	99.3%	1
eScan	1590	-	29	98.2%	2
McAfee	1585	-	34	97.9%	2
Quick Heal	1567	34	18	97.8%	2
Emsisoft	1545	73	1	97.7%	2
BullGuard	1577	7	35	97.6%	2
Fortinet	1576	-	43	97.3%	3
Microsoft	1570	-	49	97.0%	3
Sophos	1558	17	44	96.8%	3

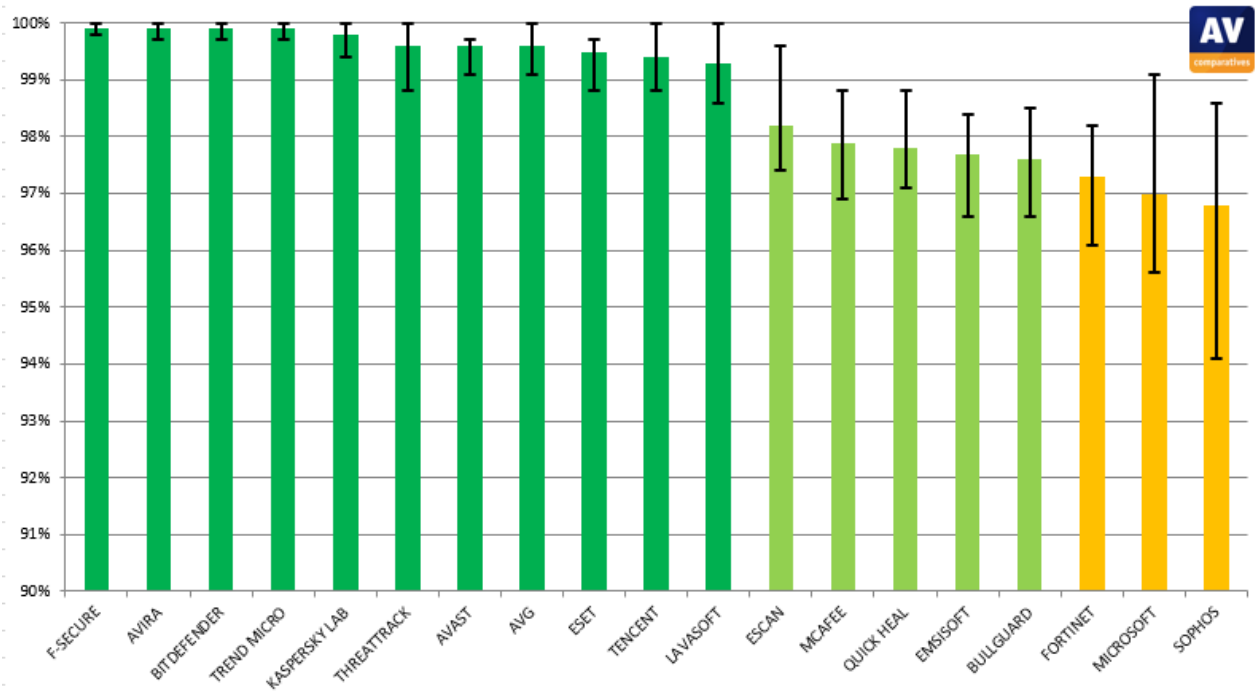


[3] Interested users who want to see the exact protection rates and FP rates for every month can see the monthly updated interactive charts on our website: http://chart.av-comparatives.org/chart1.php

[4] User-dependent cases are given half credit. For example, if a program blocks 80% by itself, and another 20% of cases are user-dependent, we give half credit for the 20%, i.e. 10%, so it gets 90% altogether.

[5] Hierarchical Clustering Method: defining clusters using average linkage between groups (Euclidian distance) based on the protection rate (see dendrogram on page 12).

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



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

a) Wrongly blocked domains (while browsing)

We used around one 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 not only risk causing a loss of trust in their warnings, but also possibly causing financial damage (besides the damage to 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 otherwise block just to 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 around two thousand different applications listed either as top downloads or as new/recommended downloads from various download portals. 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.

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

	Wrongly blocked clean domains/files (blocked / user-dependent[6])	**Wrongly blocked score[7]**
AVG, Bitdefender, ESET, ThreatTrack	0 / 0 (0)	0
Kaspersky Lab	1 / 0 (1)	1
Lavasoft, Tencent	2 / 0 (2)	2
Avira	4 / 0 (4)	4
Fortinet	8 / 0 (8)	8
Sophos	10 / 0 (10)	10
Avast	13 / 0 (13)	13
Emsisoft	2 / 23 (25)	12,5
Microsoft	23 / 0 (23)	23
	average (27)	***average 26***
BullGuard	36 / 1 (37)	36,5
eScan	44 / 0 (44)	44
McAfee	57 / 0 (57)	57
Trend Micro	67 / 0 (67)	67
Quick Heal	76 / 6 (82)	79
F-Secure	140 / 0 (140)	140

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 statistical methods and by looking at the average scores. The following products with above-average FPs have been downgraded: **BullGuard**, **eScan**, **McAfee**, **Trend Micro**, **Quick Heal** and **F-Secure**.

[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.

Prevalence of the FPs

According to some vendors, their own FPs are not seen at all in their user base (zero prevalence) or have a very low prevalence. Nevertheless, we want to give the best possible overview of prevalence data for the benefit of users of all our tested products. The table below shows the number of FPs for each product according to our amalgamated prevalence assessment, for which we used several sources of prevalence data.

Some products may block files based solely on their prevalence, i.e. if a vendor does not have any data for a particular file, their product may treat it as a threat. This of course helps to block many malicious files, but at the same time it can lead to higher false-alarm rates by blocking clean files which currently have zero or very low prevalence in the user base of the particular vendor.

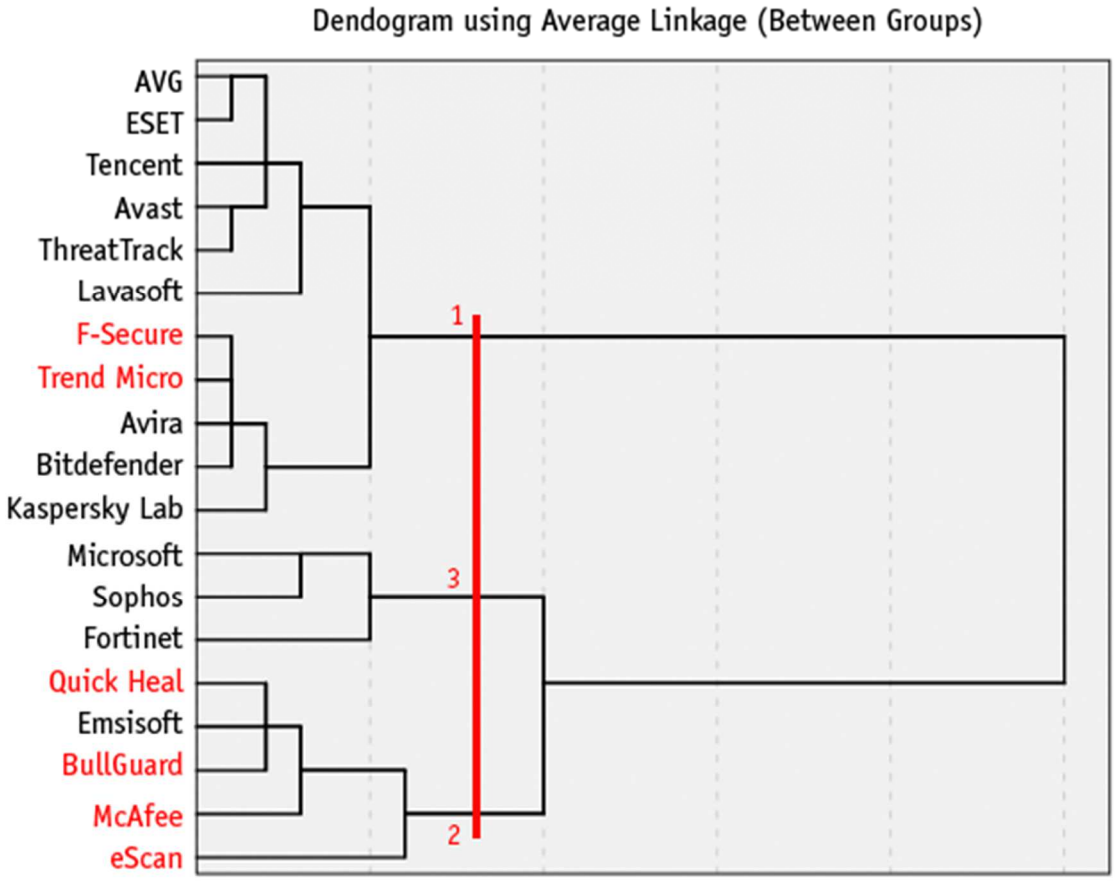
	Very low	Low	Medium	High
Avast	13	0	0	0
AVG	0	0	0	0
Avira	1	1	1	1
Bitdefender	0	0	0	0
BullGuard	17	10	7	3
Emsisoft	12	4	4	5
eScan	11	7	9	17
ESET	0	0	0	0
Fortinet	5	2	1	0
F-Secure	68	48	21	3
Kaspersky Lab	0	0	0	1
Lavasoft	1	1	0	0
McAfee	22	20	6	9
Microsoft	15	6	2	0
Quick Heal	30	28	18	6
Sophos	3	3	1	3
Tencent	1	0	1	0
ThreatTrack	0	0	0	0
Trend Micro	38	22	7	0

Key to prevalence ratings[8]
Very low: probably fewer than a hundred users
Low: probably several hundreds of users
Medium: probably several thousands of users
High: probably several tens of thousands of users

[8] These relate to our aggregated prevalence data, not to the data of the individual vendors.

Illustration of how awards were given

The dendrogram (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 3 groups)[9]. Products that had above-average FPs (wrongly blocked score) are marked in red (and downgraded according to the ranking system below).



Ranking system	**Protection score Cluster[10] 4**	**Protection score Cluster 3**	**Protection score Cluster 2**	**Protection score Cluster 1**
< ∅ FPs	Tested	Standard	Advanced	Advanced+
> ∅ FPs	Tested	Tested	Standard	Advanced

[9] As all products scored highly (over 95%) in this test, we have used three instead of four clusters.
[10] See protection score clusters on page 8.

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 this Whole-Product Dynamic "Real-World" Protection Test:

AWARD LEVELS	PRODUCTS
	Avira Bitdefender Kaspersky Lab ThreatTrack Avast AVG ESET Tencent Lavasoft
	F-Secure* Trend Micro* Emsisoft
	eScan* McAfee* Quick Heal* BullGuard* Fortinet Microsoft Sophos
	-

** downgraded by one rank due to the score of wrongly blocked sites/files (FPs); see page 13*

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.

Copyright and Disclaimer

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

AV-Comparatives (December 2016)