Anti-Virus Comparative

Retrospective/Proactive test

(Heuristic detection and behavioural protection against new/unknown malicious software)

Language: English
March 2012
Last revision: 19th July 2012

www.av-comparatives.org

Contents

1. Introduction

This test report is the second part of the March 2012 test[1]. The report is delivered in late July due to the large amount of work required, deeper analysis, preparation and dynamic execution of the retrospective test-set. This year this test is performed only once, but includes also a behavioural protection element.

New in this test

There are two major changes in this test relative to our previous proactive tests. Firstly, because of the frequency of updates now provided by the vendors, the window between malware appearing and a signature being provided by the vendor is much shorter. Consequently we collected malware over a shorter period (~1 day), and the test scores are correspondingly higher than in earlier tests. Secondly, we have introduced a second (optional) element to the test: behavioural protection. In this, any malware samples not detected in the scan test are executed, and the results observed. A participating product has the opportunity to increase its overall score by blocking the malware on/after execution, using behavioural monitoring. The following vendors asked to be included in the new behavioural test: Avast, AVG, AVIRA, BitDefender, ESET, F-Secure, G DATA, GFI, Kaspersky, Panda and PC Tools. The results published in this report show results for all programs for the scan test, plus any additional protection by those products participating in the behavioural test. Although it was a lot of work, we received good feedback from various vendors, as they were able to find bugs and areas for improvement in the behavioural routines.

The products used the same updates and signatures they had on the 1st **March** 2012, and the same detection settings as used in March (see page 5 of this report) were used for the heuristic detection part. In the behavioural test we used the default settings. This test shows the proactive detection and protection capabilities that the products had at that time. We used 4,138 new malware variants that appeared around the 2nd March 2012. The following products were tested:

- AhnLab V3 Internet Security 8.0
- avast! Free Antivirus 7.0
- AVG Anti-Virus 2012
- AVIRA Antivirus Premium 2012
- BitDefender Anti-Virus Plus 2012
- BullGuard Antivirus 12
- eScan Anti-Virus 11.0
- ESET NOD32 Antivirus 5.0
- F-Secure Anti-Virus 2012
- Fortinet FortiClient Lite 4.3
- G DATA AntiVirus 2012
- GFI Vipre Antivirus 2012
- Kaspersky Anti-Virus 2012
- Microsoft Security Essentials 2.1
- Panda Cloud Antivirus 1.5.2
- PC Tools Spyware Doctor with AV 9.0
- Qihoo 360 Antivirus 3.0
- Tencent QQ PC Manager 5.3

At the beginning of the year, we gave the vendors the opportunity to opt out of this test. McAfee, Sophos, Trend Micro and Webroot decided not to take part in this type of test, as their products rely very heavily on the cloud.

[1] http://www.av-comparatives.org/images/docs/avc_fdt_201203_en.pdf

2. Description

Many new viruses and other types of malware appear every day, which is why it is important that antivirus products not only provide new updates, as frequently and as quickly as possible, but also that they are able to detect such threats in advance (preferably without having to execute them or contact the cloud) with generic/heuristic techniques; failing that, with behavioural protection measures. Even if nowadays most antivirus products provide daily, hourly or cloud updates, without proactive methods there is always a time-frame where the user is not reliably protected.

The data shows how good the proactive heuristic/generic detection capabilities of the scanners were in detecting new threats (sometimes also named as zero-hour threats by others) used in this test. By design and scope of the test, only the heuristic/generic detection capability and behavioural protection capabilities (on-execution) were tested (offline). Additional protection technologies (which are dependent on cloud-connectivity) are considered by AV-Comparatives in e.g. whole-product dynamic ("real-world") protection tests and other tests, but are outside the scope of retrospective tests.

This time we included in the retrospective test-set only new malware which has been seen in-the-field and prevalent in the few days after the last update in March. Additionally, we took care to include malware samples which belong to different clusters and that appeared in the field only after the freezing date. Due to the use of only one sample per malware variant and the shortened period (~1 day) of new samples, the detection rates are higher than in previous tests. We adapted the award system accordingly. Samples which were not detected by the heuristic/generic on-demand/on-access detection of the products were then executed in order to see if they would be blocked using behaviour-analysis features. As can be seen in the results, in at least half of the products the behaviour analyser (if even present) did not provide much additional protection. Good heuristic/generic detection remains one of the core components to protect against new malware. In several cases, we observed behaviour analysers only warning about detected threats without taking any action, or alerting to some dropped malware components or system changes without protecting against all malicious actions performed by the malware. If only some dropped files or system changes were detected/blocked, but not the main file that showed the behaviour, it was not counted as a block. As behaviour analysis only come into play after the malware is executed, a certain risk of getting compromised remains (even when the security product claims to have blocked/removed the threat). Therefore, it is preferable that malware gets detected before it gets executed, by e.g. the on-access scanner using heuristics (this is also one of the reasons for the different thresholds on the next page). Behaviour analyser/blockers should be considered as a complement to the other features inside a security product (multi-layer protection), and not as a replacement.

What about the cloud? Even in June (months later), many of the malware samples used were still not detected by certain products which rely heavily on the cloud. Consequently, we consider it a marketing excuse if retrospective tests - which test the proactive detection against new malware - are criticized for not being allowed to use cloud resources. This is especially true considering that in many corporate environments the cloud connection is disabled by the company policy, and the detection of new malware coming into the company often has to be provided (or is supposed to be provided) by other product features. Clouds are very (economically) convenient for security software vendors and allow the collection and processing of large amounts of data. However, in most cases (not all) they still rely on blacklisting known malware, i.e. if a file is completely new/unknown, the cloud will usually not be able to determine if it is good or malicious.

AV-Comparatives prefer to test with default settings. Almost all products run nowadays by default with highest protection settings or switch automatically to highest settings in the event of a detected infection. Due to this, in order to get comparable results for the heuristic detection part, we also set the few remaining products to highest settings (or leave them to default settings) in accordance with the respective vendor's wishes. In the behavioural protection part, we tested ALL products with DEFAULT settings. Below are notes about settings used (scan of all files etc. is always enabled) of some products:

F-Secure: asked to be tested and awarded based on their default settings (i.e. without using their advanced heuristics).
AVG, AVIRA: asked us not to enable/consider the informational warnings of packers as detections. Because of this, we did not count them as such.
Avast, AVIRA, Kaspersky: the heuristic detection test was done with heuristics set to high/advanced.

This time we distributed the awards by using the following award system / thresholds:

	Proactive Detection/Protection Rates			
Heuristic Detection	0-25%	25-50%	50-75%	75-100%
Heuristic + Behavioural Protection	0-30%	30-60%	60-90%	90-100%
Very few (or none)	tested	STANDARD	ADVANCED	ADVANCED+
Few FP	tested	STANDARD	ADVANCED	ADVANCED+
Many FP	tested	tested	STANDARD	ADVANCED
Very many FP	tested	tested	tested	STANDARD
Crazy many FP	tested	tested	tested	tested

NB: To qualify for a particular award level, a program needs to get EITHER the relevant score on heuristic detection alone, OR the relevant score for heuristic and behavioural protection, but not both. Thus a program that scores 85% on heuristic detection receives an Advanced+ award, even if it fails to improve its score at all in the behavioural test.

3. False alarm test

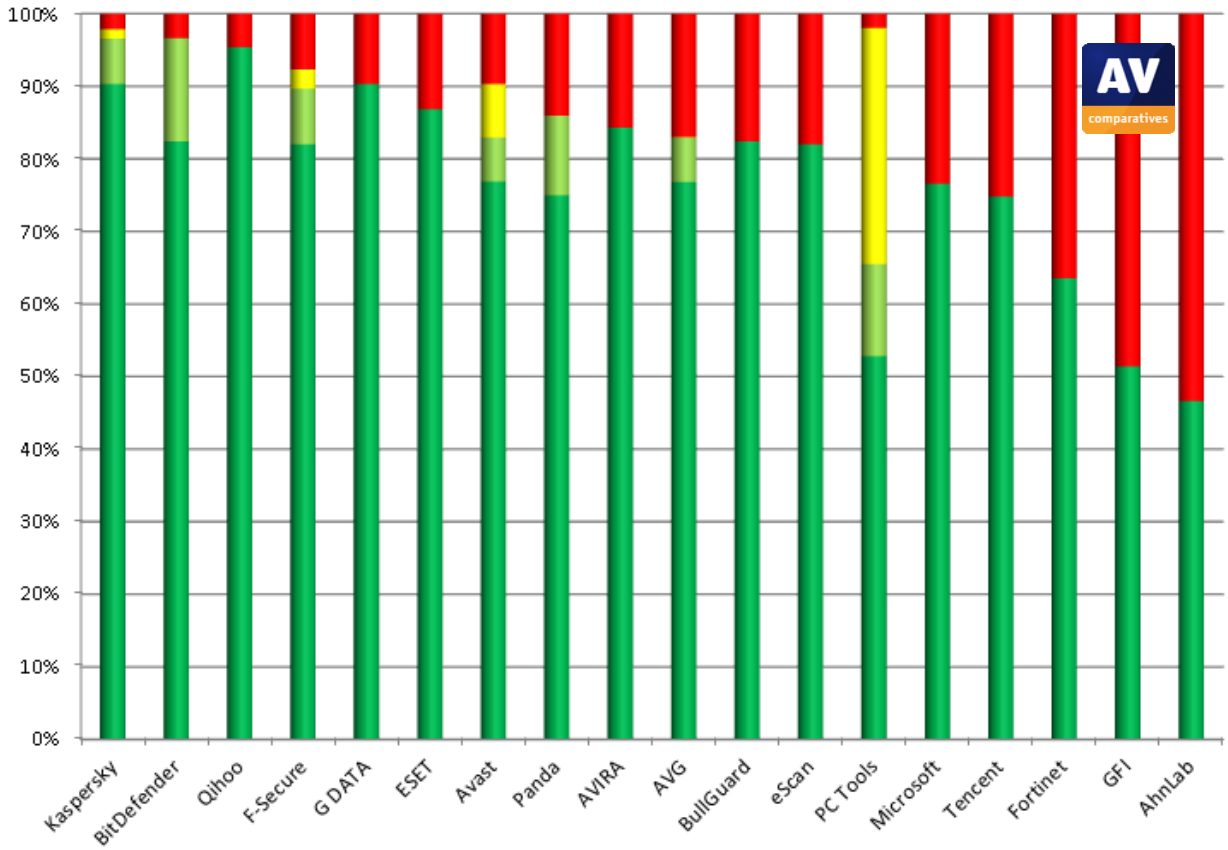
To better evaluate the quality of the detection capabilities, the false alarm rate has to be taken into account too. A false alarm (or false positive [FP])[2] is when an antivirus product flags an innocent file as infected although it is not. False alarms can sometimes cause as much trouble as real infections.
The false alarm test results (with active cloud connection) were already included in the March test report.

Very few false alarms (0-3):	Microsoft, ESET
Few false alarms (4-15):	BitDefender, F-Secure, BullGuard, Kaspersky, Panda, eScan, G DATA, Avast, AVIRA
Many false alarms (over 15):	Tencent, PC Tools, Fortinet, AVG, AhnLab, GFI
Very many false alarms (over 100):	Qihoo

[2] All discovered false alarms were already reported to the vendors in March and are now already fixed. For details, please read the report available at http://www.av-comparatives.org/images/docs/avc_fps_201203_en.pdf

4. Test Results

The table below shows the proactive detection and protection capabilities of the various products. The awards given (see page 9 of this report) consider not only the detection/protection rates against new malware, but also the false alarm rates.



Key:

Dark green = detected on scan

Light green = blocked on/after execution

Yellow = user dependent

Red = not blocked

Some observations:

Behavioural detection was used successfully mainly only by Avast, AVG, BitDefender, F-Secure, Kaspersky, Panda and PC Tools.

BitDefender and Kaspersky scored very high and have even detected some few more samples, but failed to block or remove them (so they were counted as miss/not-blocked).

Qihoo detected a lot of malware using heuristics, but also had a high rate of false positives.

PC Tools was quite dependent on user decisions, i.e. gave many warnings.

5. Summary results

The results show the proactive (generic/heuristic) detection and protection capabilities of the security products against new malware. The percentages are rounded to the nearest whole number. Do not take the results as an absolute assessment of quality - they just give an idea of whom detected/blocked more and who less, in this specific test. To know how these antivirus products perform with actual signatures and cloud connection, please have a look at our File Detection Tests of March and September. To find out about real-life online protection rates provided by the various products, please have a look at our on-going Whole-Product Dynamic "Real-World" Protection tests.
Readers should look at the results and decide on the best product for them based on their individual needs. For example, laptop users who are worried about infection from e.g. infected flash drives whilst offline should pay particular attention to this Proactive test.

Below you can see the proactive heuristic detection and behavioural protection results over our set of new/prevalent malware appeared in-the-field within ~1 day in March (4,138 different samples):

	Heuristic Detection	Heuristic + Behavioural Protection Rate[3]	False Alarms
1. Kaspersky	90%	97%	few
2. BitDefender	82%	97%	few
3. Qihoo	95%		very many
4. F-Secure	82%	91%	few
5. G DATA	90%	90%	few
6. ESET	87%	87%	very few
7. Avast	77%	87%	few
8. Panda	75%	85%	few
9. AVIRA	84%	84%	few
10. AVG	77%	83%	many
11. BullGuard, eScan	82%		few
12. PC Tools	53%	82%	many
13. Microsoft	77%		very few
14. Tencent	75%		many
15. Fortinet	64%		many
16. GFI	51%	51%	many
17. AhnLab	47%		many

[3] User-dependent cases were given a half credit. Example: if a program blocks 80% of malware by itself, plus another 20% user-dependent, we give it 90% altogether, i.e. 80% + (20% x 0.5).

6. Awards reached in this test

The following awards are for the results reached in the proactive/retrospective test:

AWARDS	PRODUCTS
	Kaspersky BitDefender F-Secure G DATA ESET Avast Panda AVIRA BullGuard eScan Microsoft
	AVG* Tencent*
	Qihoo* PC Tools* Fortinet* GFI*
	AhnLab
NOT INCLUDED[4]	*McAfee* *Sophos* *Trend Micro* *Webroot*

*: these products got lower awards due to false alarms[5]

[4] As those products are included in our yearly public test-series, they are listed even if those vendors decided not to be included in retrospective tests as they rely heavily on cloud-connectivity.

[5] Considering that certain vendors did not take part, it makes sense to set and use fixed thresholds instead of using the cluster method (as by the non-inclusion of the low-scoring products clusters may be built "unfairly").

7. Copyright and Disclaimer

AV-Comparatives e.V. (July 2012)