



AV-Comparatives
Quality Assurance

Based on the test-set of August 2007

Date: January 2008

Last revision of this report: 6th February 2008

Website: <http://www.av-comparatives.org>

1. Quality assurance based on the test-set of August 2007

We decided to do this quality assurance over the test-set of 2007 (the final published results) to see if the assumed presence of bad samples had any noticeable effect on the results. With the help (incl. scientific consultation and supervision) of the University of Innsbruck¹ and the Kompetenzzentrum.IT², we have reviewed in the last 6 months the whole test-sets, in order to improve our internal systems and developing new tools to prevent and minimise the occurrence of any similar faulty samples (non-working or grey samples) inclusions in future.

Due the exploding growth of new malware, all anti-virus vendors and testers which use large test-sets use nowadays many different automated tools (along with manual file analysis where the tools can not help or are not reliable) as help to be able to keep up with this samples flood. Beside the existing commercial utilities for such work, AV-Comparatives uses additionally own automated tools to check for functionality and maliciousness of the samples. Most of the removed files can now in fact be identified and sorted out automatically by our automated in-house tools/systems. Currently we are also working on a new sandbox for some types of malware.

We get samples from many various sources, including honeypots (of various sources and types), traps, crawlers, user submissions, infected customers in our region, infected websites and downloaders, independent groups/forums, online scan services, many various anti-malware vendors (which also get the samples from the field), etc. The largest amount of bad samples can be found in samples gathered thru honeypots/user submissions/online scan services. But also some few antivirus vendor collections contain quite a lot of obvious garbage (files labeled as damaged or unwanted not counted). If some vendors say they include/detect such files because they are often included in test-sets, they probably mean other testers or magazine testers, because nearly all of those files came first from them and were at no time included in the test-sets of AV-Comparatives. The truth is that any big collection contains also some amount of garbage, no matter how much efforts are done to avoid that, but it is also true that this little amount has no big effect on the results. Years ago the amount of garbage included in our tests was higher than in our recent test-sets, but especially since our team grow and some (new) automated tools were introduced, the quality compared to five years ago improved a lot.

After the tests, we usually give at least one-week time to AV vendors to peer-review their missed samples, in order to give them the possibility to report known bad samples. This way results can be crosschecked and be a bit more accurate before they get published.

Currently ~75% of our website visitors run Windows XP, ~19% run Windows Vista, ~4% run Windows NT/2000 and ~2% all the rest. That is why we focus on malware which may (still) pose a potential risk to modern 32-bit (and above) systems - due that, in February 2008 we will still test under Windows XP; starting from August 2008, we will probably perform the tests under Microsoft Windows Vista.

Note: this QA refers only to the August 2007 test-set. We will internally continue to monitor the quality and do our best to improve it further.

¹ supervised by Univ. Lect. Werner Wil

² <http://av.kompetenzzentrum.IT>

In the test-set of August 2007 we found nearly 0,4% non-working malware (which is relatively much). Anyway, below you see what impact they had on the published results:

Product	Before correction	Corrected percentage	Difference	Ranking	<u>AWARD</u>
avast!	95,24%	95,27%	+0,03%	unchanged	unchanged
AVG	97,75%	97,79%	+0,04%	unchanged	unchanged
AVIRA	99,45%	99,46%	+0,01%	unchanged	unchanged
BitDefender	97,51%	97,55%	+0,04%	unchanged	unchanged
Dr.Web	89,87%	89,94%	+0,07%	unchanged	unchanged
eScan	97,53%	97,57%	+0,04%	unchanged	unchanged
ESET NOD32	97,60%	97,64%	+0,04%	unchanged	unchanged
Fortinet	89,98%	90,06%	+0,08%	unchanged	unchanged
F-Prot	92,20%	92,40%	+0,20%	unchanged	unchanged
F-Secure	97,57%	97,62%	+0,05%	unchanged	unchanged
GDATA AVK	99,31%	99,33%	+0,02%	unchanged	unchanged
Kaspersky	98,46%	98,49%	+0,03%	unchanged	unchanged
McAfee	93,15%	93,19%	+0,04%	unchanged	unchanged
Microsoft	90,37%	90,43%	+0,06%	unchanged	unchanged
Norman	90,93%	91,00%	+0,07%	unchanged	unchanged
Symantec	98,80%	98,81%	+0,01%	unchanged	unchanged
TrustPort	99,64%	99,66%	+0,02%	unchanged	unchanged

As you can see, even in the worst case that nearly 0,4% were bad samples, the impact on the results was minimal, between ~0,01 and ~0,2% (note that AV-Comparatives predicted/stated that results can be ~0,1% different due bad samples and that users were instructed to look at the award and not at the percentages/decimals). **The bad samples included in the test-set of August 2007 had no impact on the rankings and the award given.** The reason why the impact was so minimal even if 0,4% were questionable files, is mainly that all products (with no exception) detect (on purpose, by mistake or coincidence) most of those questionable/corrupted malware samples and because results/samples are crosschecked before publishing (sometimes results need a significant correction, mainly due last-week submissions and additions which samples get validated by us when tests already started). Future test results may be more accurate, with even lower impact on the published results.

We improved/changed the test-sets of 2008 also in other areas: The OtherOS category will no longer be included in the test-set. The old DOS virus samples have been completely removed already in 2006 from the test-sets. Files that were present in the test-set of August 2007 which do not work on Microsoft Windows NT/2000/XP/Vista have also been removed (such malware samples are anyway detected by practically all anti-virus products). Such viruses received in the last 6 months may be still present in the Windows category in February 2008, but will be removed completely from future test-sets. Some few files had the wrong executable extension, but that has now been fixed. Additionally we removed some files which could in our opinion be better classified as Adware, components, dialer, spyware, tools and other potentially unwanted programs. Results will be displayed with only one decimal instead of two, making readers clear that percentages should not be taken as absolute etc.

Only PE malware which can work under Windows NT/2000/XP/Vista and that (still) appeared or existed e.g. during last about 2 years (this period will be shortened further) will be in future test-sets. Older macro viruses will be removed during next months.

Some vendors say that they score low in large tests because what they miss is just old malware. To make readers clear, that this is just a defensive statement, here some numbers: In February 2004 we had 79000 samples in the test-set and most of them were already detected at that time, and what was missed at that time was added within that year by the vendors. In February 2007, we had 495000 samples in the test-set and in August 2007 we had 805000 samples, which means that nearly the half were samples received in the last 6 months.

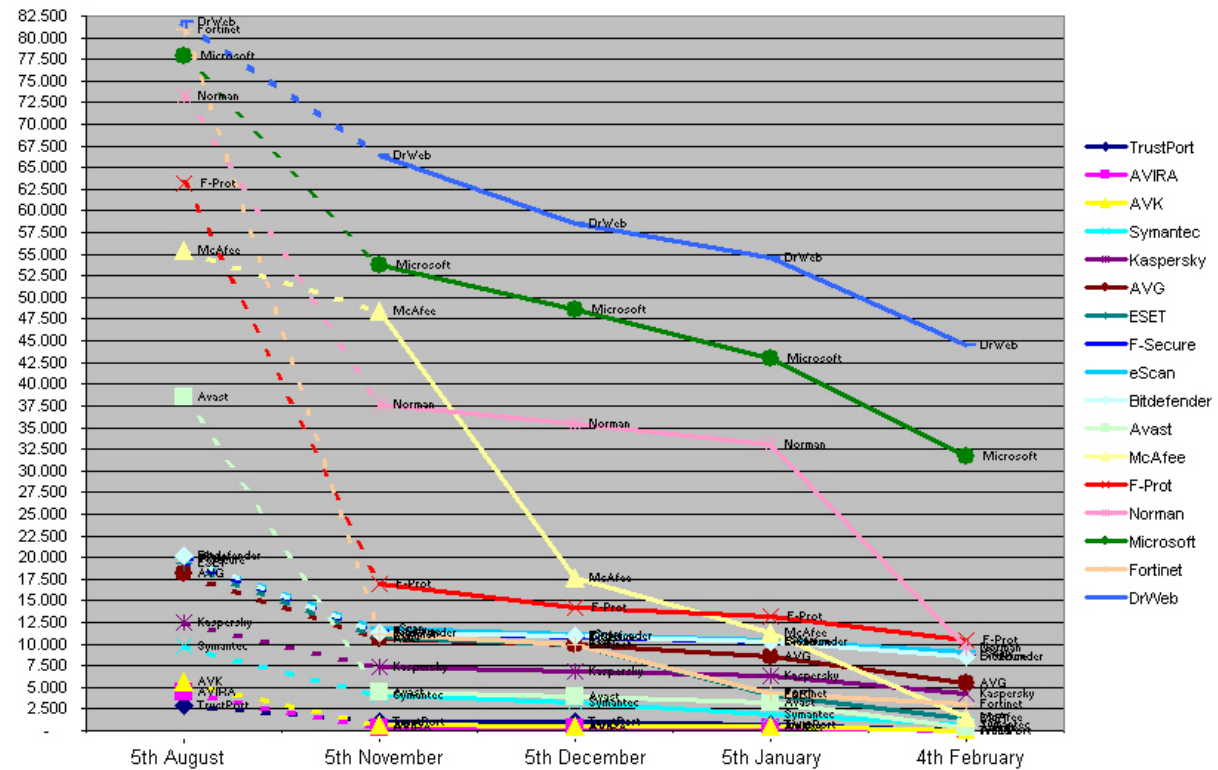
- February 2004: 79000 samples*
- August 2004: 103000 samples*
- February 2005: 150000 samples*
- August 2005: 178000 samples*
- February 2006: 243000 samples*
- August 2006: 321000 samples*
- February 2007: 495000 samples*
- August 2007: 805000 samples*
- February 2008: ~1,7 million samples*

* Please note that some old samples were removed by AV-Comparatives since the beginning twice at year, but nearly completely only during 2008.

When looking at the graphs of added samples, we see that most vendors add usually within 6-36 months what they missed, so what they miss in the test-set is usually mainly newer malware appeared during the last year.

Here also the graph of added missed samples from the August 2007 test-set in the last 6 months:

Missed samples



Also products which are (compared to others) slow at adding malware they missed in previous tests are theoretically able to achieve an ADVANCED+ rating if they are good at detecting newer/actual malware. The actual detection rates of various products will be released at the begin of March 2008 on our website.

2. Copyright and Disclaimer

This publication is Copyright (c) 2008 by AV-Comparatives. Any use of the results, etc. in whole or in part, is ONLY permitted after the explicit written agreement of Andreas Clementi, prior to any publication. AV-Comparatives 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. 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.

Andreas Clementi, AV-Comparatives (January 2008)