

# AMTSO Review Analysis Board Report

Testing organization: [NSS Labs](#)

Report title: [Endpoint Security – Socially Engineered Malware Protection Comparative Test Results](#)

Methodology version: 1.2

Date of publication: 08/09/2009 (8th September 2009)

Challenging organizations: [Sophos](#) (primary); [AVG](#) (supporting); [Panda Security](#) (supporting)

Reviewers: Simon Edwards, [Dennis Technology Labs](#) (chair); Dr. Leitold Ferenc, [CheckVir](#); Gabor Szappanos, [VirusBuster](#); Philipp Wolf, [Avira](#).

## Summary

The NSS Labs Endpoint Security test (see above) was challenged by Sophos, AVG and Panda Security. The following conclusions were reached by the Review Analysis Committee, which compared [AMTSO's Fundamental Principles of Testing](#) to the above report. Principle 9, which relates to active contact points, was evaluated through interviews with the testing organization and the challenging organizations.

The principles and results are as follows:

- ✓ 1. Testing must not endanger public.
- ✓ 2. Testing must be unbiased.
- ? 3. Testing should be reasonably open and transparent.
- ✓ 4. The effectiveness and performance of anti-malware products must be measured in a balanced way.
- ✓ 5. Testers must take reasonable care to validate whether test samples or test cases have been accurately classified as malicious, innocent or invalid.
- x 6. Testing methodology must be consistent with the testing purpose.
- x 7. The conclusions of the test must be based on the test results.
- ✓ 8. Test results should be statistically valid.
- ✓ 9. Vendors, testers and publishers must have an active contact point for testing-related correspondence.

The report was found to comply with the following principles: 1, 2, 4, 5, 8 and 9

The reviewers were unable to determine whether or not the report complied with the following principles: 3

The report was found not to comply with the following principles: 6 and 7

## Principle #1: Testing must not endanger public

- Was the test set of malicious samples inappropriately made available to the general public?  
No
- Was the test conducted in a controlled environment so that the tester was able contain the malware and its behavior?  
Yes
- What measures were taken to prevent the leakage of malware or the consequences of malware execution to the outside world, and do they seem appropriate and fit for purpose?  
Firewall
- Was new malware created as part of the testing process?  
No

**Conclusion: The report complies with this principle.**

## Principle #2: Testing must be unbiased

- Where did the funding for the conducting of the test come from? Could it represent an inducement to bias the results or conclusions?  
The reports are available to buy, hence the funding comes most likely from purchasers of the report.
- Could the sample selection process favor any vendor?  
According to the test description, no.
- Could the product configuration favor any vendor?  
The interaction between vendors and tester is not clear
- Did the tester use default configuration for all products? If so, was it declared as doing so?  
No, the vendors were offered an opportunity to affect the settings.
- Did the tester start off with a pre-determined target configuration and configure each product so that all tested products are configured to an equivalent functional level (or as near as possible)?  
Unknown
- Was the test environment for tested products consistent with the sample set, or were some samples selected irrespective of the test environment?  
Consistent
- Were products and product types mixed inappropriately in terms of (a) host environment (b) sample set?  
No
- Could the test methodology favor any vendor in any other way?  
Probably not

**Conclusion: The report complies with this principle, although there is an issue relating to Sophos' ability to change its product's settings (see Principle #9: Vendors, testers and publishers must have an active contact point for testing-related correspondence).**

### Principle #3: Testing should be reasonably open and transparent

- Which solutions were tested?  
As listed in the report.
- What versions of the products were used?  
Specified in the paper (though version numbers were not provided for all products).
- How were the solutions obtained and updated?  
Provided by the vendors.
- What were the sources of malicious and innocent samples or test cases?  
Extracting URLs from NSS Labs network of spam traps and honeypots
- How were the samples or test cases obtained and validated?  
NSS Labs honeypots fed samples into sandbox systems for verification. An array of anti-virus products were used to identify samples that were not identified using the sandboxes. They were not used to verify samples in isolation (which is good).
- How were the test samples or test cases selected?  
Unknown
- What product settings/configurations were used?  
Vendor-specific, in at least some cases
- What target configuration was used (if not default settings)?  
Internet Explorer 7 browser, no configuration options specified.
- When and under what conditions was the test conducted?  
July - August 2009, private test
- In what environment(s) was/were the test(s) conducted?  
Virtual client systems, Windows XP SP3
- How were the malicious and innocent samples or test cases applied?  
Downloaded from the original website
- Was malware presented to the tested software in a manner consistent with the way in which it would be presented “naturally”?  
Yes
- How was the response of the solutions measured?  
Unknown
- Was the test “apples to apples” (comparing products of similar type and functionality or range of functionalities) or “apples to oranges” (comparing products of significantly different type and/or functionality or range of functionalities)?  
Apples to apples
- If “apples to oranges”, how were the various solutions compared? Was it done in such a way that single products were not penalized for not being suites, for example?  
n/a
- How were the results calculated and interpreted?  
Results included stages of infection i.e. caught on download; subsequent execution.  
Products categorized into Recommend, Neutral and Caution

**Conclusion: The target system configurations are unknown, as are the methods used to measure the products' responses. For this reason the reviewers lacked the information required to find the test either in compliance or not in compliance with the AMTSO principles.**

#### **Principle #4: The effectiveness and performance of anti-malware products must be measured in a balanced way.**

- Were the products tested similar? If not, was the underlying functionality similar enough to justify inclusion in the same test, and was the methodology likely to measure that functionality accurately for all products?

Yes

- Were disproportionate ratings applied in the test?

No

**Conclusion: The report complies with this principle.**

#### **Principle #5: Testers must take reasonable care to validate whether test samples or test cases have been accurately classified as malicious, innocent or invalid.**

- Were the samples or sample sets used correctly classified [malicious, innocent, corrupted, and non-viable]?

Yes

- Could any problems with selection and validation lead to breaches of other principles (see principle 2, for example)?

Yes - the validation seems to be flawed - using sandboxes and anti-malware scanners is not the same as executing a threat on a clean system and monitoring for malicious behavior. Sandboxes are not 100 per cent accurate and scan results are well known to be a flawed verification method.

**Conclusion: The report complies with this principle, although there is a bias issue. Only samples that produced a positive output in one of the sandboxes were used. This ensured that all samples were active malware. It also meant that malicious samples that produced no output on the sandboxes were not used. This raises an important question about how important this bias is. The reviewers recommended to AMTSO that this principle should be discussed and the questions asked here in the template should maybe be rethought or reformulated to make it possible to find that a bias has been introduced but that perhaps it is a relatively insignificant one (or not, depending on the situation).**

## **Principle #6: Testing methodology must be consistent with the testing purpose.**

- Was the objective of the test clearly stated and defined?  
Yes
- Was the methodology clearly stated and defined?  
Yes
- Does the methodology align with the purpose or objective of the test?  
No. If you want to test "the protection of the products against socially-engineered malware", you should also test products against this situation. It was not taken into account how the URLs actually reached the endpoint system. This might happen through spam, for example. If a product uses a spam-filtering, the spam message might have never appeared on the system, therefore the user would be protected as well.
- Was the choice of product target consistent with the testing purpose?  
Yes
- Was the choice of configuration methodology consistent with the testing purpose?  
Yes
- Is the test objective internally consistent?  
Yes

**Conclusion: The report does not comply with this principle. The reviewers agreed that missing infection vectors (e.g. spam) can mislead the result. Nevertheless, they also thought that the test still did better than a lot of tests out there right now, since at least the malware was coming from the "real world" and also was executed afterwards in a dynamic test.**

## **Principle #7: The conclusions of the test must be based on the test results**

- Was the conclusion based only on the test results?  
Yes. No further data, from outside the test's remit, was referred to
- by the conclusion.
- Was the entire test results used for drawing the conclusion?  
Yes. The results were not cherry-picked or otherwise manipulated.
- Does the conclusion reflect the stated purpose?  
No. The report's Executive Summary states that test's purpose was to determine the protection of the products tested against socially-engineered malware only. Later in the report (Section 4 -product assessments) it says: "Products that earn a caution rating from NSS Labs should not be short-listed or renewed." This is clearly a conclusion that you can't make out of the detection for socially-engineered malware only, as the products have other layers of protection that the test did not evaluate.
- Does the interpretation of the results follow logically from the data as presented?  
No. As above, the conclusion is too general in its recommendations and
- condemnations, considering that only a portion of each product's
- functionality was tested.

**Conclusion: The report does not comply with this principle.**

## **Principle #8: Test results should be statistically valid**

- Are the statistical data available?  
Yes
- Do the data and stated results match?  
Yes
- Is the sample set sufficient and well represented to be statistically valid?  
Unknown
- If relevant, is the standard margin of error (or deviation) stated?  
Yes

**Conclusion: The report complies with this principle.**

## **Principle #9: Vendors, testers and publishers must have an active contact point for testing-related correspondence.**

- Was proper contact details of the tester made available?  
Yes
- Did the vendors, testers and publishers respond to correspondence in a reasonable timeframe?  
This issue was contested by the primary challenger (Sophos). Sophos claimed that NSS Labs responded to requests for information with a demand that Sophos buy a copy of the report. Over time NSS engaged with Sophos, providing information. Despite this Sophos issued its challenge to the report. While communicating with the reviewers, Sophos requested that its earlier challenge on this point be dropped.

**Conclusion: The report complies with this principle.**