

The Curious Art of Anti-Malware Testing

David Harley FBCS CITP CISSP
ESET Director of Malware Intelligence



Perception is Everything

Common Assumptions:

- All comparative tests are sound
- All complaints are “vendor whining”

Complaints abound, but what is “good testing”?

Standard AV response:

If you have to ask, you’re not qualified

What does the Customer Want?

- 100% Protection
- Absolute Convenience
- No False Positives
- Preferably free of charge!

What does the Industry Want?

- To be assessed on what we can do, not what we don't do.
- Get feedback that allows us to improve products
- Good buying decisions = safer Internet = less pressure on the industry

What Functionality *can* be Tested?

- Malware –specific detection (one signature per instantiation). Nowadays, a little harder than it sounds.
- Generic signatures (several instantiations per signature: may catch unknown as well as unknown variants and sub-variants).
- Basic heuristics (proactive detection by static analysis}
- Advanced heuristics/behaviour analysis/emulation/sandboxing/behavior blocking (proactive detection by dynamic analysis)

What *Else* Can You Test?

Fairly “hard” (objective) data like:

- Scanning speed (clean machine)
- Scanning speed (infected machine)
- False Positives
- Processor Load
- Memory usage

Soft Data – Intangible Test-Targets

- Helpline support
- Usability
- Quality of Documentation
- Prettiness of the Interface
- Color of the box...

Unless you have unlimited time and resources,
Less is More™

So Who Can Run Comparative Tests?

In principle, *anyone*.

There are *no* mandatory requirements/certifications.

What Does the Customer *Really* Want?

- To be guided rather than misled.
- To be helped to make the best possible choice.
- But what's "best" for the customer, might not be best for the tester.
- That *doesn't* mean only saintly altruists and people with 20 years of research experience can produce a useful review.
- It *does* mean that you can't take competence or impartiality for granted.

Some Test Models

- Self-tests
- Sponsored/Independent Tests
- Independent Tests
- Multi-Scanner Submissions

Sponsored Test Red Lights

- Sponsored description of how good the product is. Actual testing is optional.
- No methodology description (that's *always* a red light!)
- Test with samples supplied by the vendor.

Dangerous Assumptions

- Vendor competence can be used as a substitute for real sample validation. (Assume no False Positives)
- Access to samples and sample processing priority is equal across all vendors.

The Glut Problem

- Tens of thousands of samples daily (or more):
- No-one has exactly (or near-exactly) the same sample set at any one time.
- Processing is prioritized according to a wide range of methodologies and criteria

The VirusTotal Fallacy

How To Be Independent, Impartial, Yet...Totally Useless...

- Sample set: found, presumed malicious objects (honeypots, honeynets, mailboxes)
- Methodology: files submitted to Virus Total (or a similar multi-scanner site)
- Validation: files submitted to Virus Total...

Pack Up Your Troubles

- In many cases, if you actually check the identifications, they're some variation on "suspicious".
- (You already knew that: that's why you submitted the file. 😊)

For instance...

- It's packed
- It's packed with a known "black" packer
- It's packed with a custom variant packer
- It's packed, but was already a small executable

These are all legitimate blocking criteria...

...but not exactly known malware detection.
More like blacklisting a whole class of
object.

Detection: Active versus Passive

- Passive scanning: check for signatures, generic signatures, passive heuristics.
- Active/Dynamic Scanning: analyse behaviour by observing code executed in a (hopefully) safe environment.

Static Testing

Scan a file/object without allowing it to execute.

- Very convenient testing practice.
- Can be done simply by running on-demand, even command-line scanner.
- Almost platform independent, if detection database is standard across platforms.

But...

- It can put products that use active/proactive techniques at a serious disadvantage.
- No execution, no behaviour to observe/analyse.
- In such a case, result doesn't reflect detection capability.

Dynamic Testing

- Better in principle, because a better reflection of threat landscape. But:
 - Harder to define
 - Harder to do
 - Expensive
 - Resource-Intensive (very difficult to do properly with large sample sets)
 - Complicated by vendors moving to In-the-Cloud technologies

AMTSO

- Anti-Malware Testing Standards Organization
<http://www.amtso.org>
- Target Membership:
 - Testing Organizations
 - Security Vendors
 - Academia
 - Reviewers and Publications

AMTSO Aims

Improve testing methodology across the board:

- Objectivity
- Quality
- Relevance

AMTSO Charter

- Providing a forum for discussions related to the testing of anti-malware and related products;
- Developing and publicizing objective standards and best practices for testing of anti-malware and related products;
- Promoting education and awareness of issues related to the testing of anti-malware and related products;
- Encouraging the provision of tools and resources to aid standards-based testing methodologies; and,
- Providing analysis and review of current and future testing of anti-malware and related products.

Deliverables So Far

[AMTSO Fundamental Principles of Testing](#)

[AMTSO Best Practices for Dynamic Testing](#)

[AMTSO Best Practices for Validation of Samples](#)

[AMTSO Best Practices for Testing In-the-Cloud Security Products](#)

[AMTSO Analysis of Reviews Process](#)

Issues Involved in the “Creation” of Samples for Testing

Guidelines for testing network based security products

<http://www.amtso.org/documents.html>

<http://www.amtso.org/related-resources.html>

AMTSO Fundamental Principles of Testing

1. Testing must not endanger the 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.
6. Testing methodology must be consistent with the testing purpose.
7. The conclusions of a 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.

AMTSO Fundamental Principles of Testing

Principle 1: Testing must not endanger the public.

AMTSO Fundamental Principles of Testing

Principle 2: Testing must be unbiased.

AMTSO Fundamental Principles of Testing

Principle 3: Testing should be reasonably open and transparent

AMTSO Fundamental Principles of Testing

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

AMTSO Fundamental Principles of Testing

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

AMTSO Fundamental Principles of Testing

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

AMTSO Fundamental Principles of Testing

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

AMTSO Fundamental Principles of Testing

Principle 8: Test results should be statistically valid

AMTSO Fundamental Principles of Testing

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

Quoting myself 😊

“AMTSO gives testers who already have a good working relationship with the industry a chance to maintain and build on those links, and that's good for testers, vendors, and consumers. But it's also a chance to break down the mistrust between the industry and testers that don't have such links, and that's even better.”

Next Steps

- 1) Make testers out of the mainstream more aware of their responsibilities to their audience.
- 2) Make it easier for audiences to distinguish between good and not-so-good tests and reviews
- 3) Build on co-operative relationships between the vendors